Proceedings

First International Workshop on Decision Making and Recommendation Acceptance Issues in Recommender Systems (DEMRA 2011)

and

Second International Workshop on User Models for Motivational Systems: the affective and the rational routes to persuasion (UMMS 2011)

co-located with the:
19th User Modeling, Adaptation and Personalization Conference (UMAP 2011)

July 11, 2011, Girona, Spain

This volume is published and copyrighted by:

Francesco Ricci
Giovanni Semeraro
Marco de Gemmis
Pasquale Lops
Judith Masthoff
Floriana Grasso
Jaap Ham

ISSN 1613-0073

Table of Contents

Part I

First International Workshop on Decision Making and Recommendation Acceptance Issues in Recommender Systems (DEMRA 2011)

First Workshop on Decision Making and Recommendation Acceptance Issues in Recommender Systems (DEMRA 2011)

http://www.di.uniba.it/~swap/DM/index.html

co-located with the
19th User Modeling, Adaptation and Personalization Conference (UMAP 2011)

July 11, 2011, Girona, Spain

Preface

Recommender Systems (RSs) have proved to be a valuable kind of adaptive and intelligent systems for coping with the information overload problem. In recent years, the interest in RSs has dramatically increased:

- Many Internet sites and media companies (Amazon.com, YouTube, Netflix, Yahoo, Tripadvisor, Last.fm, IMDb) are developing and deploying RSs as part of the services they provide to their subscribers;
- At institutions of higher education around the world, undergraduate and graduate courses are dedicated entirely to RSs; tutorials on RSs are very popular at computer science conferences;
- There have been several special issues in academic journals covering research and developments in the RS field (AI Communications 2008; IEEE Intelligent Systems 2007; International Journal of Computer Science and Applications 2006; ACM Transactions on Computer-Human Interaction 2005; ACM Transactions on Information Systems 2004).

While a lot of discussion has been made on recommendation techniques and algorithms, few studies have stood from users' angles to consider their acceptance of recommendations.
Characterizing and evaluating the quality of user experience and users' subjective attitudes toward the acceptance of recommender technology is an important issue which merits attention from researchers and practitioners in both web technology and human factor fields.
Therefore, the main goal of the workshop is to stimulate the discussion around problems, challenges and research directions about the acceptance of recommender technology.

Some questions motivate this workshop:
1. What does influence and determine the acceptance of the suggestions computed by a RS?
2. How does the presentation of the computed recommendations can increase the acceptance of the suggestions and of the whole system?
3. How explanation techniques can contribute to establish trust?
4. Are there general rules or guidelines for system design that can be proved to be effective in influencing the user acceptance?
5. How the recommendations should be adapted to the context of the human computer interaction to increase their acceptance?

6. What Persuasion strategies could be more effective in increasing the recommendation take up?
7. What kinds of decision processes occur in users of recommender systems, and how RSs can support these processes?

In particular, the workshop will focus on the following aspects:

- **Presentation**: How the system presents and visualizes the computed recommendations is obviously a critical factor for the acceptance and helpfulness of the recommendations and the RS.
- **Explanation**: Presentation and explanation techniques are not easily separable. A good presentation technique is also capable of explaining recommendations but also in motivating the user to make further requests, including requests for explanations.
- **Trust**: Previous research indicates that transparency and the possibility of interaction with RSs increase user trust, defined as perceived confidence in a RS competence. Users may be more forgiving, and more confident in recommendations, if they understand why a bad recommendation has been made. In addition, the interface design of a RS may affect its credibility, in particular the importance of explanation interfaces in increasing user acceptance has been well recognized in a number of fields.
- **Persuasion**: Systems based on persuasion techniques can actively modify the user preferences and perceptions on the proposed items. Recommender systems may combine presentation and persuasion techniques to raise the expected utility of the suggested items.
- **Decision support**: A complementary perspective on recommender systems sees them as decision support systems that help users to make better choices. From this perspective, the focus is more on the various types of information that users require to make satisfactory decisions, including, for example, information that will enable them to justify their decisions to other people.

We would like to thank all the authors for their submissions, and our Program Committee for their precious work.

June 2011

Francesco Ricci
Giovanni Semeraro
Marco de Gemmis
Pasquale Lops

DEMRA 2011 Workshop Chairs

Organization

Francesco Ricci
Free University of Bozen-Bolzano
Faculty of Computer Science
Piazza Domenicani 3, I-39100 Bozen-Bolzano, Italy
Phone: 0471 016 971, fax: +39 0471 016 009
email: fricci@unibz.it

Giovanni Semeraro
Department of Computer Science
University of Bari Aldo Moro
Via E. Orabona 4, I-70126 Bari, Italy
Phone: +39 080 5442140 - Fax: +39 080 5443196
E-mail: semeraro@di.uniba.it

Marco de Gemmis
Department of Computer Science
University of Bari Aldo Moro
Via E. Orabona 4, I-70126 Bari, Italy
Phone: +39 080 5442276 - Fax: +39 080 5443196
E-mail: degemmis@di.uniba.it

Pasquale Lops
Department of Computer Science
University of Bari Aldo Moro
Via E. Orabona 4, I-70126 Bari, Italy
Phone: +39 080 5442276 - Fax: +39 080 5443196
E-mail: lops@di.uniba.it

Program Committee

- Pearl Pu, Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland
- Judith Masthoff, University of Aberdeen, UK
- Cosimo Palmisano, Ecce Customer, Turin, Italy
- Michele Gorgoglione, Technical University of Bari, Italy
- Paolo Massa, Fondazione Bruno Kessler, Trento, Italy
- Shlomo Berkovsky, CSIRO, Australia
- Jill Freyne, CSIRO, Australia
- Linas Baltrunas, Free University of Bozen-Bolzano, Italy
- Ivan Cantador, Universidad Autónoma de Madrid, Spain
- Antony Jameson, DFKI - German Research Center for Artificial Intelligence, Germany
- Alejandro Jaimes, Yahoo! Research, Barcelona, Spain
- Alexander Felfernig, Technische Universität Graz, Austria
- Dietmar Jannach, Technische Universität Dortmund, Germany
- Markus Zanker, University Klagenfurt, Austria

What Should Recommender Systems People Know About the Psychology of Choice and Decision Making?

Anthony Jameson
DFKI, the German Research Center for Artificial Intelligence

Abstract

The function of recommender systems, after all, is to help people make better choices. So you might expect work in this area to be based on a clear understanding of how people make choices and how these processes can be supported by recommender systems. But in fact we see only occasional attention to the psychology of choice and decision making in this area. One reason is that the most relevant knowledge is scattered around a number of areas of psychological research, including judgment and decision making, behavioral economics, social influence, habitual behavior, and learning.
This talk will give a sample of key concepts and results from these areas, showing how they suggest new research issues and design ideas for those who work on recommender systems.

Short bio

Anthony Jameson is a principal researcher at DFKI, the German Research Center for Artificial Intelligence. Some of his research since the early 1980s has concerned various forms of recommendation, including systems that conduct recommendation dialogs, employ decision-theoretic planning, exploit digital life logs, and/or make recommendations to groups. His is the author of the chapter Choices and Decisions of Computer Users in the forthcoming third edition of the Human-Computer Interaction Handbook and founding coeditor-in-chief (with John Riedl) of the ACM Transactions on Interactive Intelligent Systems.

Explanations in Proactive Recommender Systems in Automotive Scenarios

Roland Bader[1,2], Andreas Karitnig[3], Wolfgang Woerndl[2], and Gerhard Leitner[3]

[1] BMW Group Research and Technology, 80992 Munich, Germany
roland.bader@bmw.de
[2] Technische Universitaet Muenchen, 85748 Garching, Germany
woerndl@in.tum.de
[3] Alpen-Adria Universitaet Klagenfurt, 9020 Klagenfurt, Austria
Gerhard.Leitner@uni-klu.ac.at
andreas.karitnig@gmx.at

Abstract. Recommender techniques are commonly used to ease the selection and support the decision in the context of large quantities of items such as products, media or restaurants. Typically, recommender systems are used in contexts where users focus their full attention to the system. This is not the case in automotive scenarios, therefore we want to provide recommendations proactively to reduce driver distraction while searching for information. Our application scenario is a gas station recommender. Proactively delivered recommendations may will not be accepted, if the user does not understand why something was recommended to her. Therefore, our goal in this paper is to enhance transparency of proactively delivered recommendations by means of explanations. We focus on explaining items to convince the user of the relevance of the items and to enable an efficient item selection during driving. We describe a method based on knowledge- and utility-based recommender systems to extract explanations automatically. Our evaluation shows that explanations enable fast decision making for items with reduced information provided to the user.

1 Introduction

In recent years more and more information is digitally available. Due to the availability of Internet connections in many state-of-the-art cars, this information can be made accessible for drivers. As searching for information is not the primary task during driving, providing information as recommendations in a proactive manner seems to be a reasonable approach to reduce information overload and driver distraction [2]. As the user does not request recommendations by herself it is important to present the recommendations in a way that she quickly recognizes why this information is relevant for her.

The goal of this paper is to investigate the applicability of explanation techniques to make proactive recommendations comprehensible for drivers with limited amount of information. Explanations are already the focus of research in other areas of recommender systems, e.g. product recommendations ([9], [6]). To our knowledge there is no existing work on explanations for mobile proactive recommender systems. The challenge is to provide as little information as possible to make proactive decisions transparent without information overload. Our application scenario is a gas station recommender for driver, already presented in [1]. The contribution of this paper is first, an investigation what the requirements on explanations in our application scenario are, second, how short explanations for items can be generated out of the recommendation process described in [1], and third, an evaluation of generated explanations. Note that the scope of this paper is limited to an offline investigation to lay the groundwork for an infield study in a car.

The remainder of the paper is organized as follows. In Section 2 we describe fundamentals of explanations in recommender systems. Section 3 summarizes a preliminary study. In Section 4 we describe how explanations are generated out of the recommendation process and Section 5 includes a prototype evaluation of the presented method. Section 6 closes with conclusions and future work.

2 Fundamentals and Related Work

Recommender systems suggest items such as products or restaurants to an active user. Proactively delivered, recommendations should have high relevance, be non-intrusive and the system should have a long term memory [7]. We have already developed methods for proactivity in recommender systems in [2] and [1]. Based on this work we observed that proactively delivered recommendations lack user acceptance if the user does not know why something was recommended to her. Transparency and comprehensibility are two aspects a proactive system should fulfil to be accepted [5]. Our goal in this paper is to avoid loss of acceptance by providing explanations in our existing proactive recommender for gas stations.

An explanation is a set of arguments to describe a certain aspect, e.g. an item or a situation. An argument is a statement containing a piece of information related to the aspect which should be explained, e.g., "The gas station is inexpensive" or "Gas level is low". In an item explanation arguments can be for (positive) or against (negative) an item or neutral.

In [9] seven generalizable goals for explanations in recommender systems are provided. Which goals are accomplished by an explanation depends on the field of application. To give the user the chance to correct the system (*scrutability*) and to deliver *effective* recommendations is important for recommendation systems in general. For proactive recommender systems in a car, we think that especially *transparency* (Why was this recommended to me?), *persuasiveness* (Are the recommended items relevant for me?) and *efficiency* (Can I make a decision with little interaction?) are the most important reasons. If they are fulfilled *trust* and *satisfaction* can also be positively influenced.

The work described in [6] contains design principles for explanations in recommender systems. The principles are focused on categorizing alternative items and explain the categories. Due to limited amount of items represented in a proactive recommendation, we think that categorization can hardly be applied in our application domain. This applies to many explanation methods created for desktop systems, where the user can turn her attention fully to the interface. Hence, the challenge in proactive recommender systems is to convince the user quickly of the usefulness of the recommended items.

As we want to explain utility- and knowledge-based recommendations based on [2], a utility-based approach for explanations seems reasonable. The work in [4] presents a method based on the utility of a whole explanation to select and rank explanations. Instead of the utility of the whole explanation, [3] measures the performance of a single argument and combines arguments to structured explanations. We combine ideas from both works in our proposed method.

3 Preliminary Study

Before we implemented our methods for explanations in proactive recommender systems, we conducted a user survey to find out the main requirements for the generation of arguments in our application scenario of a gas station recommender.

The user survey was conducted on the basis of an online questionnaire. The subjects had to rate different kinds of arguments and structures on a 5 point Likert scale ranging from "very useful" to "not useful at all". We focused on aspects we found in [9], [6] and [3]. The most important question was what kind of arguments should be used for explaining items in our application domain. Arguments are build either on context-based (e.g. gas level, opening times) or preference-based (e.g. gas brand or price preference) criteria. Moreover, we wanted to know how many arguments to use and how to combine and structure them (independent vs. comparative to other items vs. comparative to an average). We also asked the respondents about the usefulness of other type of information like situation explanations, status information and reliability of item attributes and context data. The survey had 81 respondents who completed the questions. The group of participants consisted of 64 male and 17 female with an average age of 29 years.

The most important aspects influencing the decision for a certain gas station seem to be gas price, detour and gas level at the gas station. Following this pattern, arguments including detour, price and gas level have been rated mostly very good. Ratings for gas station context data, like opening times or a free soft drink, varied depending on the content of an argument. Arguments more related to the task of refilling, e.g. opening times, are rated better.

There is no clear subject's favourite for the structure of an explanation. Independent as well as comparative argumentation was rated equally. Two arguments seem to represent a good size for an explanation in the case of gas stations. Regarding the desired number of items in a gas station recommendation, which ranges from 3 to 5, two arguments seem to be reasonable to distinguish them. Arguments concerning situations leading to a recommendation were rated differently. Situations which are directly connected to the task and have an impact on the recommendation were rated best, e.g. "only gas stations along the route were recommended because you do not have much time" or "Just a few gas stations are available in this area". Status information as well as data reliability were not interesting for the subjects.

4 Our Approach for Explanations in Proactive Recommender Systems

Based on the results from the preliminary study, there are obviously two major aspects which should be explained to the user. First, we have to explain what has been the crucial situation for a recommendation. A low gas level is an obvious situation for a gas station recommendation, but there are some more situations which may lead to a recommendation: A rather good gas station along the route, e.g. very low priced, a deserted area with few gas stations or an important appointment which leads to a recommendation only with gas stations on the route. Without explanation a proactive recommendation in this situations may result in misunderstanding.

Second, it should be clear to the user why the recommended items are relevant for her based on her user profile. In this paper we focus on explanations for items. Our explanation method is designed for a small set of recommended items because many items overwhelm the user if they are provided proactively. There are two main goals we try to accomplish. First, we want to enable *efficiency* because item selection is no primary task while driving and much harder compared to situations where users can focus their attention to the system (e.g. parking). Second, the user should be *persuaded* that the items are relevant.

We use a ramping strategy like [8] to explain recommendations, i.e. explanations are distributed over several levels of detail. The lowest level (first phase) is provided automatically with the recommendations. Then gradually more and more information is accessible by the user manually. The elements in the first phase are short explanations for the situation and for the items. More detailed levels include a comparison of items, a list of all items or item details. The first phase is the most important one in the ramping strategy, as the user has to recognize quickly why the recommendation is relevant for her. The following description mainly comprises this phase.

The arguments for items in the first phase are structured independently, i.e. no comparative explanations are used. The preliminary study showed that it makes no difference for the user but an independent structure allows for shorter arguments. We use preference- as well as context-based arguments, starting with a positive argument in the first place and adding a second one if necessary. A maximum of 2 arguments are used for every item.

Information for arguments in an explanation can either be interpreted attribute values, e.g. gas level is low, or facts, e.g. gas level is 32 liter. An *interpretation* is a mapping from a specific value to a discrete interval. We used a generic nominal interval with *One, Very High, High, Medium, Low, Very Low, Null* to map values to a discrete value. Two kinds of values can be mapped. A *utility* *interpretation* maps the utility of an item, e.g. a gas level of 32 liter at a gas station can be mapped to *Null*, because most people do not refill at this level, therefore the utility is 0 on that decision dimension. Interpreting the attribute and context values leads to different results, e.g. a gas level of 32 liter is *Medium* if the tank has a capacity of 65 liters. This is called *attribute interpretation*.

4.1 Argument Assessment

Our argument generation method for items is based on a context-aware recommender system for gas stations presented in our previous work [1]. It uses *Multi-Criteria Decision Making Methods (MCDM)* to assess items I on multiple decision dimensions D by means of utility functions. For example, dimensions are price or detour. First, all item attributes and context (level 1) belonging together are aggregated to local scores $LS_{I,D}$ in the range $[0, 1]$ (level 2) on every dimension D. On level 3 all dimensions are aggregated to a global score GS_I. Users are able to set their preferences for the item dimensions explicitly which results in a weight w_D for every dimension D.

The argument assessment uses two additional scores. The *explanation score* $ES_{I,D}$ describes the explaining performance of an item dimension and the *information score* IS_D measures the amount of information in a dimension. The explanation score is calculated by multiplying the weight of a dimension w_D with the performance of the item I in that dimension: $ES_{I,D} = LS_{I,D} \cdot w_D$. This way, bad performing dimensions as well as aspects not important for the user are neglected. The score corresponds to the product of user interest in a dimension with the utility of an explanation for that dimension described in [4]. Instead of a whole explanation we measure the performance of the dimension directly. The problem of only using this score is that if every item performs well on a dimension and this dimension is important for the user, every item would be explained by the same information. This decreases the opportunity to make an effective decision as items are not distinguishable. Therefore the information score measures the amount of information in a dimension relative to an item set. It is calculated by $IS_D = \frac{R+I}{2}$. The value $R = max(x) - min(x)$ is the range of x in the set. The information can either be Shannon's entropy $I = -\sum_{i=1}^{n} p(x) log_n p(x)$ or simply $I = \frac{n-h}{n-1}$ where n is the number of items in the set and h is the frequency of the most frequent x in a set. Taking $x = LS_{I,D}$ is a good choice if local scores have a small value range, otherwise the utility interpretation of $LS_{I,D}$ performs better. The information score is low if either all x are similar (R is low) or same x appear frequently (I is low), e.g. all gas stations are average priced.

4.2 Explanation Process

Figure 1 shows the process to select arguments based on the scores we described in the previous section. It follows the framework for explanation generation described in [3] by dividing the process in the selection and organization of the explanation content and the transformation in a human understandable output.

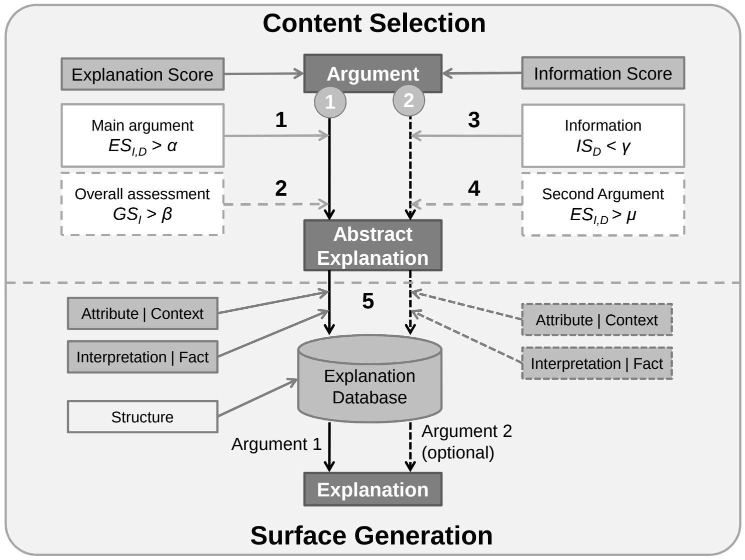

Fig. 1. Comparing scores to retrieve an explanation

In *content selection* our argumentation strategy selects arguments for every item I separately. A positive argument is selected first to help the user to instantly recognize why this item is relevant. For this, the best performing dimension D based on the explanation score $ES_{I,D}$ is compared to threshold α (1). Larger than α means the dimension is good enough for a first argument. The threshold α should be chosen so that the first argument is positive. If no dimension is larger α and thus no first argument can be selected, we look at the global score GS_I (2). If this score is larger β than the item is a good average, otherwise we suppose that the recommender could not find better alternatives. With a first argument we look at the information score of its dimension (3). A small information score (lower than γ) means that this dimension provides low information, therefore a second argument is selected by means of the explanation score: The explanation score $ES_{I,D}$ of the second argument must be larger μ to make sure the second argument is meaningful enough (4). Generally, $\mu < \alpha$ because the requirements on the second argument are lower. With the thresholds μ and γ the amount of information can be controlled.

The result of the content selection is an abstract explanation, which needs to be resolved to something the user understands. This is done in the *surface generation*. We map a key value pair, like $(gaslevel, low)$, to human understandable information, e.g. textual phrases or icons (5). Either facts or attribute interpretations can be used as values. Human understandable explanation information is uniquely stored in a database, e.g. in XML format. Also the structure of an explanation (icon, independent phrase, comparative phrase etc.) can be defined here.

5 Evaluation

To evaluate our generated explanations, we set up a user study with a desktop prototype. The prototype is a combination of a street map viewer and an explanation view. The map view is based on a street map from *OpenStreetMap.com* and is able to visualize a user's route, icons for recommended gas stations and detour routes for the gas stations. The displayed content depends on the current phase in the ramping strategy. The view for the first phase which is shown to the user automatically provides a list of maximum 3 gas station recommendations, 1 or 2 arguments for every gas station and a situation explanation. Due to shortness constraints of an explanation, negative arguments are avoided. From here, the subject can access the views for the second phase with item details and the third phase with a list of all gas stations prefiltered along the route.

We conducted a user interview with 20 participants with an average age of 29, 17 male and 3 female. For that, we created 6 different scenarios (2 short, 3 average and 1 long route). In every phase, the subjects were asked for missing and relevant information in the explanation as well as on the map. The *persuasiveness* was measured by asking the subjects for their satisfaction with a selection in the first phase and if they need more information. Looking at how often the subjects needed to switch to deeper phases with more information accounts for the *efficiency*. The explanations were all text-based. For example, a set of 3 gas stations could be explained by (1) very low priced (2) on the route (3) low priced, little detour. Acoustic and tactile modalities are out of scope of this survey. The recommendations were generated by the methods presented in [1] and every subject was asked to give her preference for gas price, detour, brand and preferred gas level at the gas station.

5.1 Results

The number of items provided by the recommender was rated as the right number by 14 subjects in average. The number of arguments was rated as too few by 7 subjects and exactly right by 8 subjects. Too few arguments have been criticized if two items could not be distinguished. Presenting the arguments either as facts or interpreted was rated differently. 11 subjects prefer facts, 9 interpretations. This may change in a real driving scenario, depending on which kind of argument imposes more cognitive effort.

Almost all information in the first phase was rated as useful by most of the subjects. In regular scenarios, most subjects could make a satisfying decision only with this information. Interestingly, the predicted gas level at the gas station was useless for most subjects, although it is an important decision dimension for most of the subjects. This may indicate that user's expectation plays also an important role: In our case, users only expect to get gas station recommendation if their gas level is low. The second phase only contained useful information and was selected if special details are needed, e.g. an ATM or a shop. In the beginning of the interview some subjects used the second phase to check the matching of interpreted values. The list of all items along the route was rarely selected and only if the recommendations do not corresponded to user expectations. In 70% of the cases the map played an important role for the decision process.

6 Conclusions and Future Work

We conclude that the explained strategy worked well offline. Most of the subjects were satisfied with the items based on the explanations provided in the first phase. Therefore we think that the amount of information was enough to convince the subjects of the relevance of the items. Further phases were rarely used and if needed than they were quickly accessible, therefore the selection could also be made efficiently. In this stage of the project it could not be derived if users prefer interpreted or specific information in an argument. Next, we investigate if the results are transferable to a driving scenario with real proactive recommendations. In our further research, we also will adjust the parameters based on the results of the study. Furthermore, we want to use Shannon's entropy on the whole prefiltered set of items to meet user expectations better. To further increase persuasiveness, we plan to integrate a dominance check like [6] over all arguments presented to the user to better distinguish items.

References

1. Bader, R., Neufeld, E., Woerndl, W., Prinz, V.: Context-aware POI recommendations in an automotive scenario using multi-criteria decision making methods. In: Workshop on Context-awareness in Retrieval and Recommendation. pp. 23–30. ACM Press, Palo Alto, CA (2011)
2. Bader, R., Woerndl, W., Prinz, V.: Situation Awareness for Proactive In-Car Recommendations of Points-Of-Interest (POI). In: Workshop on Context Aware Intelligent Assistance. Karlsruhe, Germany (2010)
3. Carenini, G., Moore, J.D.: Generating and evaluating evaluative arguments. Artificial Intelligence 170(11), 925–952 (Aug 2006)
4. Felfernig, A., Gula, B., Leitner, G., Maier, M., Melcher, R., Teppan, E.: Persuasion in Knowledge-Based Recommendation. In: 3rd International Conference on Persuasive Technology. pp. 71–82. Springer, Oulu, Finland (2008)
5. Myers, K., Yorke-smith, N.: Proactive Behavior of a Personal Assistive Agent. In: Workshop on Metareasoning in Agent-Based Systems. Honolulu, HI (2007)
6. Pu, P., Chen, L.: Trust building with explanation interfaces. In: 11th International conference on Intelligent User Interfaces. pp. 93–100. ACM Press, Sydney, Australia (2006)
7. Puerta Melguizo, M.C., Bogers, T., Boves, L., Deshpande, A., Bosch, A.V.D., Cardoso, J., Cordeiro, J., Filipe, J.: What a Proactive Recommendation System Needs: Relevance, Non-Intrusiveness, and a New Long-Term Memory. In: 9th International Conference on Enterprise Information Systems. vol. 6, pp. 86–91. Madeira, Portugal (Apr 2007)
8. Rhodes, B.J.: Just-In-Time Information Retrieval. Phd thesis, MIT Media Lab (2000)
9. Tintarev, N., Masthoff, J.: Designing and Evaluating Explanations for Recommender Systems, pp. 479 – 510 (2011)

Context-Aware Places of Interest Recommendations and Explanations

Linas Baltrunas, Bernd Ludwig, Stefan Peer, and Francesco Ricci

Free University of Bozen-Bolzano
Piazza Domenicani 3, 39100 Bolzano, Italy
lbaltrunas,bernd.ludwig,fricci@unibz.it

Abstract. Contextual knowledge has been traditionally used in Recommender Systems (RSs) to improve the recommendation accuracy of the core recommendation algorithm. Beyond this advantage, in this paper we argue that there is an additional benefit of context management; making more convincing recommendations because the system can use the contextual situation of the user to explain why an item has been recommended, i.e., the RS can pinpoint the relationships between the contextual situation and the recommended items to justify the suggestions. The results of a user study indicate that context management and this type of explanations increase the user satisfaction with the recommender system.

1 Introduction

Recommender Systems (RSs) are software tools and techniques providing suggestions for items to be of use to a user [8]. It is a matter of fact that more compelling and useful recommendations can be identified if the context of the user is known [1]. Here we adopt the definition of context provided by [5]: context is "any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves". For instance, in a travel recommender, the season and the duration of the travel are important contextual conditions that should be considered before suggesting a holiday.

For this reason, context-aware recommender systems (CARSs) have attracted a lot of attention, and in particular in the tourism domain [4, 3, 7]. But, in order to adapt the recommendations to the user's context one must first identify all the potential contextual factors that may influence the acceptance of a recommendation, e.g., distance to a target place of interest, motivations for the travel, etc. This knowledge can be obtained by referring to the vast consumer behavior literature, especially in tourism [9]. But this knowledge can only be used as a starting point. In a necessary second step the quantitative dependency of the user preferences (ratings for items) from each single contextual factor must be modeled. This dependency model can be built, in collaborative filtering RSs, by acquiring explicit ratings for some of the items to be recommended under several possible contextual conditions. So for instance, in our application domain, one should acquire the ratings of a museum (place of interest – POI) when the user is traveling with or without children or when she is alone.

In this paper, we briefly illustrate ReRex, a recommender system for places of interest (POIs), that exploits a context-aware rating prediction model to generate more useful recommendations and can explain the recommendations by referring to some selected factors describing the contextual situation of the target user [10]. We illustrate the evaluation methodology, based on the comparison of ReRex with a variant obtained by removing its context-awareness capability and recommendation explanations, showing that these two features of the system increase the user satisfaction with the recommender system.

2 Ratings in Context

Our working hypothesis is that a recommendation can be explained plausibly if at least the most important criteria that lead to the recommendation are communicated to the user. In our context-aware recommendation model, besides the user-item-matrix of ratings, the context, i.e., the set of conditions that hold when the recommendation is made, is of major importance for the recommendation.

Evidence that context matters for good recommendations is taken from a user study that we conducted. In this study subjects were asked to rate a selection of places of interest in Bolzano imagining that certain contextual conditions hold [2]. Table 1 lists some of the contextual factors that change the average ratings of particular categories of points of interest significantly (for lack of space only a selection of these categories is considered). For instance, "walking paths" are rated worse at "night time" or if the user is "far away" from that path. Note that in the table MCY, is the mean rating for items in that category when that contextual condition was considered, while MCN is the mean rating for the same selection of items when context was not considered.

This difference in the rating means is significant ($p < 0.001$: $***$; $0.001 \leq p < 0.01$: $**$; $0.01 \leq p < 0.05$: $*$). From this results we can conclude, for instance, that the rating prediction for a walking path should decrease if the user is far from it. Moreover, the distance to a walking path could be used as an argument for not suggesting that item even if based on other elements, e.g., the previous ratings of the user for similar items, it may seem a good recommendation. In contrast to this example, the mean rating of a walking path grows significantly if the user is with friends or she is in a lazy mood. Consequently, in that contextual conditions, the recommender could argue for its recommendation of a walking path by pointing out that since the user is with friends (or is in a lazy mood) then that particular walking path is a suitable activity.

The collected context-dependent ratings have been used to train a novel context-aware rating prediction model that extends and adapts the approach presented in [6]. We have introduced one model parameter for each contextual condition and item pair. To keep our approach tractable, we have modeled context as a set of independent contextual factors. The model then learns how the

Table 1. Effects of context on the mean rating for items. MCY is the mean of the ratings when that context is considered, while MCN is the mean of the ratings for the same items when context is not considered.

contextual condition	**factor**	***p*-value**	**MCN**	**MCY**	**Effect**
		Castle			
far away	distance	$***$	3.80	2.47	⇃
winter	season	$**$	3.81	2.63	⇃
		Museum			
sad	mood	$***$	2.79	1.64	⇃
activity/sport	travel-goal	$***$	2.64	1.33	⇃
active	mood	$***$	2.64	1.44	⇃
far away	distance	$**$	2.78	1.92	⇃
		Walking Path			
night time	day-time	$***$	3.78	1	⇃
far away	distance	$***$	3.86	2.38	⇃
cold	temperature	$***$	3.8	1.88	⇃
winter	season	$***$	3.91	2.33	⇃
with friends or colleagues	companion	$***$	3.85	4.83	⇑
crowded	crowdedness	$**$	3.88	2.75	⇃
working day	day-week	$**$	3.94	2.75	⇃
half day	time-available	$**$	4.01	1.6	⇃
more than a day	time-available	$**$	3.89	4.8	⇑
lazy	mood	$**$	4.03	4.71	⇑

ratings deviate from classical personalized predictions as effect of one selected contextual factor, for each possible value of the factor, i.e., contextual condition. This deviation is the *baseline* for that contextual condition and item combination. Broadly speaking, the system computes a rating prediction for a user-item pair and then adapts that prediction to the current contextual situation, i.e., a combinations of contextual conditions (values for contextual factors) using the learned context-dependent baselines.

More precisely, in our data set of context-aware ratings, a rating $r_{uic_1 \ldots c_k}$ indicates the evaluation of the user u for the item i made in the context $c_1, \ldots, c_k$, where $c_j = 0, 1, \ldots, z_j$, and $c_j = 0$ means that the j-th contextual factor is unknown, while the other index values refer to possible values for the j-th contextual factor. The tuples $(u, i, c_1, \ldots, c_k)$, for which rating the $r_{uic_1 \ldots c_k}$ is known, are stored in the data set $R = \{(u, i, c_1, \ldots, c_k) | r_{uic_1 \ldots c_k} \text{ is known}\}$. Note, that in our collected data set, only one contextual condition is known and all the others are unknown, hence in R there are ratings for which only one among the indices $c_1, \ldots, c_k$ is different from 0.

The proposed model computes a personalized context-dependent rating estimation using the following equation:

$$\hat{r}_{uic_1\ldots c_k} = \boldsymbol{v}_u \cdot \boldsymbol{q}_i + \bar{\imath} + b_u + \sum_{j=1}^{k} B_{ijc_j} \tag{1}$$

where $\boldsymbol{v}_u$ and $\boldsymbol{q}_i$ are d dimensional real valued vectors representing the user u and the item i. $\bar{\imath}$ is the mean of the item i ratings in the data set R, b_u is the baseline parameter for user u, and B_{ijc_j} are the parameters modeling the interaction of the contextual conditions and the items. The parameters $\boldsymbol{v}_u$, $\boldsymbol{q}_i$, b_u, and B_{ijc_j} are learned using stochastic gradient descent; this has been proved to be an efficient approach for similar learning problems [6].

In order to generate the explanation for a recommendation for item i in the contextual situation $c_1 \ldots c_k$ we identified $j = \arg\max_j B_{ijc_j}$, i.e., the factor that in the predictive model has the largest positive effect on the rating prediction for item i. Using one single factor in the generated explanation has the benefit of creating a simple, easy to grasp motivation, and to not overload the user. The implementation of a concrete recommender system, which is using this model, is discussed in the next section.

3 The ReRex Mobile Application

In a typical interaction with ReRex the user initially establishes the context of the visit. Using the system GUI the user can enable and/or set the values of important contextual factors. The user can switch on/off some of these factors, e.g., the "Temperature" or "Weather" (see Figure 1, left). When one of these factors is switched on the recommender system will take into account its current value in the recommendation generation process. The full set of contextual factors considered in ReRex, their values (contextual conditions), and whether they are automatically collected, using an external service, or manually entered by the user, is provided in the following:

- Distance to POI (automatic): far away, near by;
- Temperature (automatic): hot, warm, cold;
- Weather (automatic): sunny, cloudy, clear sky, rainy, snowing;
- Season (automatic): spring, summer, autumn, winter;
- Companion (manual): alone, friends, family, partner, children;
- Time day (automatic): morning, afternoon, night;
- Weekday (automatic): working day, weekend;
- Crowdedness (manual): crowded, not crowded, empty;
- Familiarity (manual): new to city, returning visitor, citizen of the city;
- Mood (manual): happy, sad, active, lazy;
- Budget (manual): budget traveler, price for quality, high spender;
- Travel length (manual): half day, one day, more than a day;
- Means of transport (manual): car, bicycle, pedestrian, public transport;

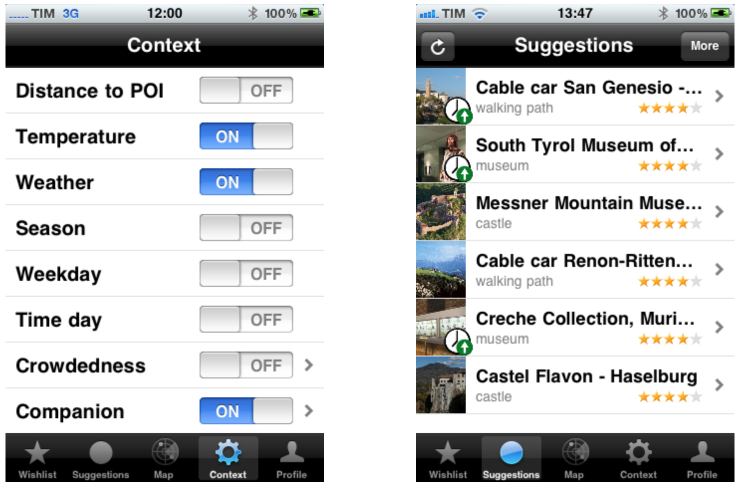

Fig. 1. ReRex context management (left); display for recommendations (right).

– Travel goal (manual): visiting friends, business, religion, health care, social event, education, cultural, scenic/landscape, hedonistic/fun, activity/sport.

After the user has entered the specification of the contextual situation (see Figure 1, left) the system can be requested to provide some recommendations. A short number of suggestions, namely six, are provided (see Figure 1, right). The recommendations are ordered according to their predicted rating. If the user is not happy with these suggestions she can request more recommendations. In the suggestion list the user can touch any of these suggestions to access a more detailed description of the POI (see Figure 2). It is worth noting that some of these suggestions are marked with an icon showing a small clock and a green arrow. This means that these recommendations are particularly suited for the current context of the request as it was previously acquired. For these recommendations (Figure 2) there is an explanation sentence like “This place is good to visit with family”. This refers to the contextual condition that was largely responsible for predicting an high ranking for this item. Note, that “with family” condition could even decrease the rank of some items, i.e., their relevance for the current context. However, some items become more attractive than others (this specific museum in our case) if the group is a family. The other items, i.e., those not marked with the clock icon, are suited as well for the current contextual situation. But we decided not to explain their relationship with the context to highlight and better differentiate those marked with the clock icon from the rest. This can be considered as a persuasive usage of the contextual information.

We have identified custom explanation messages for all the possible 54 contextual conditions listed previously. We note that even if more than one contextual condition holds in the current recommendation session, and all of them are actually used in the computation of the predicted score of each recommenda-

Fig. 2. ReRex screen for explaining recommendations.

tion, nevertheless the system exploits only one of them for the explanation. The contextual condition that is used in the explanation is the most influential one as estimated by the predictive model used by the recommender to predict the relevance (rating) of items in the current context. This design choice is motivated by a simplicity reason; we hypothesized that a single statement would be easily understood by the users and ultimately would produce the best effect on them. Naturally this issue, and more in general a better explanation functionality could be implemented in a future version of the system. In fact, as it will be illustrated in the next section, the quality of these canned explanations were not perceived by the users as strikingly good, indicating that better explanation messages could be generated.

Some additional functions have been implemented to enable the user to better exploit the system. The user can add a recommendation to her wish list, rate an item, show the position of an item on the map. We also note that ReRex recommendations are updated when a relevant contextual condition is changed either by the user manually or is automatically acquired.

4 Experimental Evaluation

In order to measure the effectiveness of this approach we developed two variants of our ReRex mobile recommender system. The first one is that described previously, the second variant is not context-aware, i.e., there is no possibility for the user to specify the current context, the UI screen shown in Figure 1 (left), has been removed, and no recommendation is marked with any icon, or explained to stress the appropriateness for the current contextual situation. The prediction model described in Equation 1 is simplified in this second variant, and the parameters B_{ijc_j} are not learned. This variant does not offer any explanation for the recommendation. Hence, comparing these two variants we could check if context management in the prediction model and the proposed explanation technique have a joint effect on user satisfaction compared to a system that does not exploit context at all.

To achieve this goal the test participants, 20 in total, tried out both variants of the system (within groups experimental model), in a random order, and executed, supported by each system, two similar but different tasks, related to travel planning. After the user completed the assigned task using one system, she was requested to fill out a usability questionnaire. These questions were extracted, and slightly adapted to the scope of our investigation, from the IBM Computer System Usability Questionnaire. Then finally the subjects were requested to compare the two systems. The full set of results of this evaluation are reported in detail elsewhere and are beyond the scope of this paper [2]. In summary, we can report that when the users were requested to directly compare the two variants, 85% of the users preferred the context-aware version, and 95% of the users considered the context-aware recommendations more appropriate. with respect to the explanation functionality, the subjects rated their agreements to the following two statements: (Q14) I am satisfied with the provided contextual explanations; and (Q15) I believe that the contextual explanations are useful. We observed a score of 1.05 for (Q14), and a higher score of 1.5 for (Q15) (scores range from -2, strongly disagree, to 2, strongly agree). This shows that the quality of the explanations is not yet optimal but the users clearly perceived the importance of such feature. Summarizing the evaluation results we observe that, even if this conclusion is supported by a limited number of testers, the context-aware recommendations were considered more effective than those produced by the non context-aware version. Moreover, the users largely agreed on the importance of explanations even if they complained about the quality of them. This indicates that the explanation is a very important component, it strongly influenced the system acceptance, but the user is particularly sensible to the quality of these explanation; and the formulation of these explanations can be surely improved.

5 Conclusions and Future Work

In this paper we have illustrated the importance of exploiting a traveler contextual conditions when recommending POIs. The proposed mobile application offers to the user context-aware recommendations that are justified and explained by referring explicitly to the contextual situation in which the user will experience them. We have shown that the proposed system can offer effective context-aware explanations that are generated by identifying the contextual conditions that show the largest influence on the predicted relevance score (rating) of the recommended items. In a live user study we have compared a context-aware version to a non context-aware one. We have shown that the user acceptance and satisfaction is larger for the context-aware version and that the users prefer this version compared to another, with a very similar user interface, which does not consider the request context and does not provide any explanations.

In a future work we want to better understand the individual role of personalization, contextualization, and explanations. In fact, in the study described in this paper we have compared a system offering contextualization of the recommendations and explanations with a variant that misses both features. We need to perform new experiments where the individual features are considered independently: for instance, comparing two context-aware systems: with and without explanations. A second issue was mentioned already in the paper and refers to the measured low user satisfaction for the generated explanations. We want to improve the quality of the explanations exploiting advanced natural language processing techniques to better adapt the explanation to the type of recommended item and using more information extracted from the predictive model.

References

1. G. Adomavicius and A. Tuzhilin. Context-aware recommender systems. In F. Ricci, L. Rokach, B. Shapira, and P. Kantor, editors, *Recommender Systems Handbook*, pages 217–256. Springer Verlag, 2011.
2. L. Baltrunas, B. Ludwig, S. Peer, and F. Ricci. Context relevance assessment and exploitation in mobile recommender systems. *(to appear) Personal and Ubiquitous Computing*, 2011.
3. V. Bellotti, J. Begole, E. Chi, N. Ducheneaut, J. Fang, E. Isaacs, T. King, M. Newman, K. Partridge, B. Price, P. Rasmussen, M. Roberts, D. Schiano, and A. Walendowski. Activity-based serendipitous recommendations with the magitti mobile leisure guide. In *Proceedings of the 2008 Conference on Human Factors in Computing Systems, CHI 2008*, pages 1157–1166. ACM Press, 2008.
4. F. Cena, L. Console, C. Gena, A. Goy, G. Levi, S. Modeo, and I. Torre. Integrating heterogeneous adaptation techniques to build a flexible and usable mobile tourist guide. *AI Communication*, 19(4):369–384, 2006.
5. A. Dey. Understanding and using context. *Personal and Ubiquitous Computing*, 5(1):4–7, 2001.
6. Y. Koren. Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge Discovery and Data mining*, KDD '09, pages 447–456, New York, NY, USA, 2009. ACM.
7. F. Ricci. Mobile recommender systems. *Journal of Information Technology and Tourism*, 12(3):205–231, 2011.
8. F. Ricci, L. Rokach, and B. Shapira. Introduction to recommender systems handbook. In F. Ricci, L. Rokach, B. Shapira, and P. Kantor, editors, *Recommender Systems Handbook*, pages 1–35. Springer Verlag, 2011.
9. J. Swarbrooke and S. Horner. *Consumer Behaviour in Tourism.* Butterworth-Heinemann, 2 edition, 2006.
10. N. Tintarev and J. Masthoff. Designing and evaluating explanations for recommender systems. In F. Ricci, L. Rokach, B. Shapira, and V. Kantor, editors, *Recommender Systems Handbook*, pages 479–510. Springer Verlag, 2011.

Group Decision Support for Requirements Negotiation

Alexander Felfernig, Christoph Zehentner, and Harald Grabner

Institute for Software Technology, Graz University of Technology,
Inffeldgasse 16b, A-8010 Graz, Austria
{alexander.felfernig,christoph.zehentner,harald.grabner}@ist.tugraz.at
http://www.ist.tugraz.at

Abstract. Requirements engineering is one of the most critical phases in software development processes. Requirements are verbalizing decision alternatives which are negotiated by stakeholders. In this paper we present the results of an empirical analysis of the effects of applying group recommendation technologies to requirements negotiation. This analysis has been conducted within the scope of software development projects at our university where development teams were supported with group recommendation technologies when deciding which requirements should be implemented. We summarize the results of this analysis and show how group recommendation can be applied to requirements negotiation.

1 Introduction

Requirements engineering (RE) is considered as one of the most critical phases in software projects and poorly implemented RE is a major risk for the failure of a project [8]. Requirements themselves are a verbalization of decision alternatives regarding the functionality and quality of the software [2]. Related individual as well as group decisions are extremely difficult due to the increasing size of requirement models as well as contradicting preferences of stakeholders [1]. In this paper we analyze the impact of applying group recommendation technologies [9] to improve the quality of decision processes in the context of *requirements negotiation* which is the process of resolving existing conflicts between requirements and deciding which requirements should be implemented. Typical functionalities of group recommender systems are the visualization of the preferences of other group members, recommendations for individual and group decisions, and recommendations for conflict resolutions in the case of inconsistent stakeholder preferences [9]. Our major motivation for applying group recommendation technologies is to improve the *usability* and the *quality of decision support* in requirements engineering environments (especially in the context of requirements negotiation – both are used as subjective measures in our evaluation).

Note that decision models based on rational thinking [11] are not applicable in most requirements negotiation scenarios since stakeholders do not exactly know their preferences beforehand [1]. Furthermore, preferences are not stable but rather change over time which is an important aspect to be taken into account by requirements negotiation environments [1].

For the purpose of supporting preference construction in requirements negotiation we have developed IntelliReq. Teams are allowed to configure the set of requirements that should be implemented. Note that our goal was to develop recommendation technologies which can be flexibly exploited in requirements negotiation; it is not our intention to replace existing requirements negotiation approaches (see, e.g., [3]) but to provide useful extensions.

This paper is organized as follows. In Section 2 we sketch the IntelliReq environment which supports group decision processes in requirements negotiation – for reasons of space limitations we omit screenshots. In Section 3 we present our hypotheses defined for the empirical evaluation of IntelliReq and discuss the corresponding study results. The paper is concluded with Section 4.

2 IntelliReq Environment

2.1 Application Scenario

IntelliReq is a group decision environment that supports computer science students at the Graz University of Technology in deciding on which requirements should be implemented within the scope of their software projects. Typically, a project team consists of 6–8 students who implement a software system with an average effort of about 8 man months. At the beginning of a project, students have to evaluate a set of requirements which have been defined by the course instructors and to figure out which requirements they will implement within the scope of their project (requirements negotiation phase). For example, the task could be the implementation of a tourist recommender application – the corresponding decision alternatives are depicted in Table 1. We will use this simple set of decision alternatives as a working example throughout the paper.

2.2 IntelliReq User Interface & Functionalities

With the goal of supporting the achievement of a common group decision, the IntelliReq user interface supports the following functionalities:

- Each stakeholder is enabled to define, adapt, and store his/her preferences (choices) regarding a given set of decision alternatives (see, e.g., Table 1).
- Each stakeholder can comment on, argue for, and discuss defined preferences.
- Each group can view and discuss recommendations for group decisions determined on the basis of already defined user preferences.
- Define and store a group decision; this is allowed for project managers.
- Each IntelliReq user can evaluate the application; this user feedback has been analyzed within the scope of an empirical study.

ID	Question	Decision Alternatives
1	which application domain?	20 destinations in Austria; 20 world-wide
2	which type of persistence management?	relational databases; XML; Java objects
3	which type of user interface?	text-based; Java Swing; Web application
4	which recommendation algorithms?	knowledge; collaborative & content-based
5	evaluation by whom?	by: students; other universities; instructors
6	type of user manual?	HTML-based; .pdf based
7	type of acceptance procedure?	live-demo; presentation with screenshots

Table 1. Example decisions to be taken by the project teams.

	with recommendation	without recommendation
preference view	version 1	version 3
no preference view	version 2	version 4

Table 2. The 4 *IntelliReq* versions. The variation points are: *group recommendation supported* (*yes/no*) and *preferences of other team members are visible* (*yes/no*).

3 Empirical Study

In order to evaluate the provided IntelliReq functionalities, we conducted an empirical study within the scope of the course *Object-oriented Analysis & Design* organized at the Graz University of Technology. The major focus of this study was to analyze the impact of group decision technologies on the dimensions *usability* of the system and *quality of decision support*.

3.1 Study Design

For the purpose of the empirical study we provided the IntelliReq environment in *four* versions. In order to analyze our hypotheses, we decided to implement a 2x2 study with the variation points *group recommendations available – recommendations are determined by majority voting* (*yes/no*) and *preferences of other users visible* (*yes/no*) – these versions are shown in Table 2. Both, group recommendations and preference visibility, are key functionalities provided by state of the art group recommendation environments [9, 13]. On the basis of this empirical study we wanted to investigate to which extent these functionalities are applicable within the scope of requirements negotiation.

N=293 participants (computer science students at the Graz University of Technology, 23.1% female and 76.9% male) selected their preferred requirements using the IntelliReq environment. The participants of the study were assigned to one of 56 different groups (the development teams) and defined (stored) 3733 individual preferences and 101 group decisions. For each development team the last stored group decision was interpreted as the final decision; after the published deadline no further adaptations of the taken decisions were possible. After a user had successfully articulated his/her requirements, he/she had the possibility to give feedback on the *usability* and the *decision support quality* of IntelliReq on a 10-point Likert scale.

3.2 Study Hypotheses

The empirical study is based on hypotheses derived from existing research in requirements engineering [1, 3], group recommender systems [9, 5], and decision & social psychology [4, 6, 12]. The list of hypotheses is shown in Table 3.

hypothesis	description
H1	*group recommendations* improve system usability
H2	*group recommendations* improve quality of decision support
H3	*group recommendations* trigger more discussions
H4	*preference visibility* deteriorates perceived usability
H5	*preference visibility* deteriorates perceived quality of decision support
H6	*preference visibility* triggers less preference adaptations
H7	*preference visibility* triggers a decision bias
H8	winning strategy: use *group recommendation* but no *preference visibility*
H9	unconsidered preferences: –usability and –quality of decision support

Table 3. Hypotheses used for evaluating the IntelliReq environment.

Group Recommendation (Hypotheses 1–3) Existing research in the field of recommender systems [9, 5] points out the potential of group recommendation technologies to significantly improve the quality of group decision processes. *First* we wanted to investigate the potential of group recommendation technologies to improve the quality of the dimensions *usability* and *decision support* in a requirements negotiation scenario. With *Hypothesis 1* we express the assumption that recommendation technologies can improve the overall system quality in terms of *usability*. *Hypothesis 2* expresses the assumption that recommendation technologies can help to improve the perceived *quality of decision support*. *Second* we wanted to know whether the availability of group recommendations has an influence on the frequency of applying discussion functionalities (*Hypothesis 3*) – the underlying assumption is that the availability of group recommendations intensifies discussions between group members. This phenomenon is well known and exploited by critiquing-based recommenders where the system proposes recommendations and the user can give feedback in terms of critiques [14]. Studies in social psychology show that frequent information interchange can improve the quality of group decisions [6, 12].

Visible User Preferences (Hypotheses 4–7) Existing research in the field of group-based recommendation points out the advantages of preference transparency in group decision making [9]. In contrast, literature in social psychology points out the fact that suboptimal outcomes of group decision processes are correlated with the visibility of individual preferences of other group members [12, 6]. The reason for groups not being able to take optimal decisions (hidden-profile identification problem) is explained by an insufficient discussion of unshared information which is triggered by the initial disclosure of individual preferences (focus shift from information interchange to preference comparison). *First* we wanted to investigate whether the group-wide visibility of individual preferences has an influence on the perceived usability and decision support quality (*Hypotheses 4* and *5*). *Second* we wanted to figure out whether the group-wide visibility of individual preferences has an influence on the frequency of preference adaptation (*Hypothesis 6*). One underlying assumption here is that persons follow the phenomenon of *social proof* [4], i.e., are doing or accepting things that others already did (accepted). The other underlying assumption is that persons tend to stick with their current decision due to the phenomenon of *consistency* [4], i.e., the effect that published personal opinions are changed less often. *Third*, a lower frequency of information exchange can lead to a different decision outcome [6]. With *Hypothesis 7* we wanted to investigate whether the group-wide visibility of preferences can lead to a decision bias (due to *social proof* [4]), i.e., whether preference visibility has an influence on the decision outcome.

Winning Strategy (Hypothesis 8) We wanted to provide an answer to the question which of the four different IntelliReq versions will be evaluated best regarding usability and quality of decision support. With *Hypothesis 8* we want to express the assumption that group recommendations improve system usability and decision support quality. In contrast, making preferences of other group members visible in the group decision process deteriorates the system evaluation. Consequently, *version* 2 (see Table 2) should be evaluated best.

Distance Matters (Hypothesis 9) Finally, we wanted to provide an answer to the question whether the distance of a users's preference to the final group decision has an impact on the overall system evaluation. With *Hypothesis 9* we express the assumption that users with a low number of considered requirements will not be satisfied with the system usability and the decision support quality.

Group recommendation heuristics The *majority* rule is a simple but very effective heuristic in group decision making [7]: each decision is taken conform to the majority of the votes of the team members. In addition to the majority rule, there exist a couple of heuristics which can be applied when generating recommendations for groups, for example, the *fairness* heuristic which guarantees that none of the group members will be disadvantaged. In the final part of our empirical study we will compare the *prediction quality* of different group recommendation heuristics in the context of our requirements negotiation scenario.[1]

3.3 Study Results

In order to identify statistically significant differences in the user quality feedback depending on the used IntelliReq version we conducted a series of two-sample t-tests. We will now discuss the results of our analysis.

Hypothesis H1 has to be rejected since the *usability* of IntelliReq versions with recommendation support (version 1 and version 2 in Table 2) is only better on the descriptive level (p=0.17, avg. 7.0, std.dev. 1.17) compared to versions without a recommendation support (avg. 6.42, std.dev. 2.47).

[1] Note that due to limited number of subjects (N=293) we were not able to compare the different recommendation heuristics w.r.t. the dimensions usability and quality of decision support. Such comparisons will be in the focus of future work.

Hypothesis H2 can be confirmed since we could detect a significant better evaluation of the IntelliReq *decision support* for recommendation-enhanced versions ($p<0.001$, avg. 7.07, std.dev. 2.03) compared to versions without a recommendation support (avg. 5.21, std.dev. 2.96).

Hypothesis H3 can be confirmed as well since the number of comments on individual preferences is significantly higher in versions which provided group recommendations ($p<0.0015$, avg. 7.96, std.dev. 5.90 vs. avg. 3.53, std.dev. 2.71). Thus we can interpret group recommendations as a stimulating elements for information interchange among group members which is a key factor for high-quality group decisions [12, 6].

Hypotheses H4 and *H5* can not be confirmed since users with no access to the preferences of other group members did not provide a significantly better rating for usability and quality of decision support. However, on the descriptive level the evaluation of versions *without* preference visibility for all group members is better (e.g., usability, avg. 7.0, std.dev. 2.08) compared to versions that make preferences visible (e.g., usability, avg. 6.46, std.dev. 2.09).

Hypothesis H6 can be confirmed since the number of adapted individual preferences is significantly *lower* in versions with access to the personal preferences of other group members ($p<0.001$). This can be explained by the fact that – due to preferences visible for other users – the current user inclines to be *consistent* [4] with his/her original requirements, i.e., the willingness to change articulated preferences decreases if preferences are accessible for other users [4].

Hypothesis H7 can be confirmed since users having access to the preferences of other group members articulate preferences which are more similar to the final group decision (avg. 0.28, std.dev. 0.09 vs. avg. 0.43, std.dev. 0.13). Being confronted with the preferences of other group members, persons base their decisions on the already known preferences and do not focus on a discussion of unshared information which is extremely important for finding optimal decisions [6]. There is a significant biasing effect due to the visibility of preferences ($p<0.001$). This effect can be explained by the phenomenon of social proof [4] which triggers group members to do things or accept things that other group members are doing (accepting).

Hypothesis H8 can not be confirmed. However, users with recommendation support and without insight into the preferences of other users provided the highest ranking for both, *usability* (avg. 7.62, std.dev. 1.84) and *quality of decision support* (avg. 7.11, std.dev. 2.06). Versions with recommendation support outperform versions without recommendation support in terms of *decision support quality* ($p<0.001$) and versions with recommendation support and without a view on the preferences of other users clearly outperform all other versions in terms of *usability* ($p<0.001$).

Hypothesis H9 can be confirmed since users with preferences having a higher distance from the final group decision rated the IntelliReq environment significantly worse in terms of *usability* ($p<0.05$). This result conforms to the *win-lose* situations discussed in [3] which typically turn into *lose-lose* situations. We could not detect a difference in the evaluation of the *quality of decision support.*

3.4 Comparison of Group Recommendation Heuristics

In our empirical study we applied the *majority voting* heuristic [7] for determining group recommendations. In addition to the majority heuristic we wanted to evaluate and compare different other group recommendation heuristics (see, e.g., [10]) w.r.t. to their applicability for our requirements negotiation scenario – for our comparison we used the following ones:

- RAND (randomized recommendation): a recommendation where each individual prediction has been generated randomly.
- LDM (least distance member): the preferences (selections) of the group member with the lowest distance to the preferences of all other group members is used as the group recommendation.
- FAIR (fairness): at least one preference of each group member is taken into account when generating the group recommendation.
- MP (most pleasure): for each question (see, e.g., Table 1) each possible answer is rated regarding its difficulty (in our case in terms of effort in man-months estimated by instructors). The alternative with the lowest overall difficulty is used as group recommendation.
- GBCF (group-based collaborative filtering): group decisions (of other groups) which are similar to the personal preferences of the members of the current group are used as group recommendation.
- MAJ (majority voting): decisions (preferences) supported by a majority of group members are integrated in the final group recommendation.
- MIN (minority voting): decisions (preferences) supported by a minority of group members are integrated in the final group recommendation.

On the basis of the data (individual preferences and taken group decisions) we compared these seven decision heuristics w.r.t. their *prediction quality* (see Table 4). This evaluation shows that (as expected) RAND and MIN should not be taken into account as serious heuristics for predicting user preferences. For our dataset, the MAJ heuristic outperforms all other decision heuristics in terms of the *average distance between predicted and actual group decision*.[2]

4 Conclusions

In this paper we have presented the results of an empirical study which investigated the impact of group recommendation technologies applied in the context of requirements negotiation. We introduced the IntelliReq decision support environment which is used at the Graz University of Technology for supporting group decision processes in small-sized software projects (6–8 team members). The major results of this experiment where that group recommendation technologies can improve the perceived usability and quality of decision support. It is not recommended to disclose the preferences of individual group members at the beginning of a decision process since the knowledge of the preferences of other group members can result in an insufficient discussion of unshared information.

[2] The IntelliReq dataset is available (anonymized): www.ist.tugraz.at/ase/intellireq.

heuristic	avg. dist. (all)	avg. dist. (rec.)	avg. dist. (no rec.)
RAND	0.55 (0.04)	0.55 (0.04)	0.55 (0.04)
LDM	0.31 (0.21)	0.26 (0.22)	0.35 (0.19)
FAIR	0.35 (0.23)	0.33 (0.24)	0.38 (0.22)
MP	0.47 (0.20)	0.47 (0.18)	0.46 (0.23)
GBCF	0.31 (0.19)	0.31 (0.21)	0.33 (0.17)
MAJ	0.27 (0.18)	0.22 (0.19)	0.32 (0.16)
MIN	0.80 (0.16)	0.81 (0.17)	0.79 (0.16)

Table 4. Average distances of recommended group decisions to the final group decision. Distances are measured in terms of the share of individual predictions different from the group decision (*rec.* = IntelliReq versions 1 and 2; *no rec.* = IntelliReq versions 3 and 4; *all* = all IntelliReq versions).

References

1. B. Alenljung and A. Persson. Decision-making activities in the requirements engineering decision processes: A case study. In *ISD 2005*, pages 707–718, 2005.
2. A. Aurum and C. Wohlin. The fundamental nature of requirements engineering activities as a decision-making process. *Information and Software Technology*, 45(14):945–954, 2003.
3. B. Boehm, P. Gruenbacher, and R. Briggs. Developing groupware for requirements negotiation: Lessons learned. *IEEE Software*, 18(3):46–55, 2001.
4. R. Cialdini. The science of persuasion. *Scientific American*, (284):76–81, 2001.
5. A. Felfernig, M. Mandl, M. Schubert, W. Maalej, and F. Ricci. Recommendation and decision technologies for requirements engineering. In *2nd Intl. Workshop on Recommendation Systems for Software Eng. (RSSE10)*, pages 11–15, 2010.
6. T. Greitemeyer and S. Schulz-Hardt. Preference-consistent evaluation of information in the hidden profile paradigm: Beyond group-level explanations for the dominance of shared information in group decisions. *Journal of Personality and Social Psychology*, 84(2):332–339, 2003.
7. R. Hastie and R. Kameda. The robust beauty of majority rules in group decisions. *Psychological Review*, 112(2):80–86, 2005.
8. H. Hofmann and F. Lehner. Requirements engineering as a success factor in software projects. *IEEE Software*, 18(4):58–66, 2001.
9. A. Jameson, S. Baldes, and T. Kleinbauer. Two methods for enhancing mutual awareness in a group recommender system. In *ACM Intl. Working Conference on Advanced Visual Interfaces*, pages 48–54, Gallipoli, Italy, 2004.
10. J. Masthoff. Group recommender systems: Combining individual models. In F. Ricci, L. Rokach, B. Shapira, and P. Kantor, editors, *Recommender Systems Handbook*, pages 677–702. Springer, 2011.
11. D. McFadden. Rationality for economists? *Journal of Risk and Uncertainty*, 19(1):73–105, 1999.
12. A. Mojzisch and S. Schulz-Hardt. Knowing other's preferences degrades the quality of group decisions. *Jrnl. of Personality and Social Psych.*, 98(5):794–808, 2010.
13. M. O'Connor, D. Cosley, J. Konstan, and J. Riedl. Polylens: A recommender system for groups of users. In *European Conference on Computer-Supported Cooperative Work*, pages 199–218, 2001.
14. P. Pu and L. Chen. User-involved preference elicitation for product search and recommender systems. *AI Magazine*, 29(4):93–103, 2008.

Exploring the Effects of Feed-forward and Feedback on Information Disclosure and User Experience in a Context-Aware Recommender System

Bart P. Knijnenburg[1], Alfred Kobsa[1], Simon Moritz[2], Martin A. Svensson[2],

[1] Department of Informatics, University of California, Irvine, USA
{Bart.K, Kobsa}@uci.edu
[2] Ericsson Research, Ericsson AB, Stockholm, Sweden
{simon.moritz, martin.a.svensson}@ericsson.com

Abstract. When disclosing information to a recommender system, users need to trade off its usefulness for receiving better recommendations with the privacy risks incurred through this disclosure. Our paper describes a series of studies that will investigate the use of feed-forward and feedback messages to inform users about the potential usefulness of their disclosure. We hypothesize that this approach will influence the user experience in several interesting ways.

Keywords: Recommender systems, privacy, information disclosure, context-aware recommenders, accuracy, user experience, satisfaction.

1 Introduction

Recommender systems for mobile applications need to provide immediate benefit to users, or else they may discontinue using them [1][2]. Many systems, however, give adequate recommendations after an extensive period of use only [3]. Context-aware recommender systems (CARS) use context data to overcome this new-user problem. Previous CARS have used location, system usage behavior, demographics, and implicit feedback [4][5]. Some users may feel uneasy providing such potentially privacy-sensitive information to the system [6][7]. Moreover, not all forms of context data are equally useful for the recommender [5]. From a privacy perspective, it is better to let users decide themselves whether or not they want to disclose some piece of information [8]. Research shows that a large majority of people is willing to trade off privacy for personal benefits [9]. However, users often have a hard time making an informed decision because they lack knowledge about their benefit from providing the information to the system and its consequences for their privacy [10][11].

Recent studies on users' election of privacy settings in an IM client [12] and a Facebook application [13] informed participants about the privacy decisions made by their friends and all other users, respectively. This "feed-forward" message facilitated social cues [14]; participants were slightly more likely to conform to the social norm in setting their privacy preferences. The current paper applies the idea of "feed-forward" to the field of recommender systems, and presents several extensions.

2 Other Types of Feed-forward and Feedback

While previous work has considered the impact of social cues only, we plan to investigate a variety of feed-forward messages that can help users make educated information disclosure decisions (Table 1). Wang and Benbasat [15] showed that providing feed-forward about the usefulness of the piece of information to be disclosed increased users' trust in the recommender system. Berendt and Teltzrow [16] suggest that providing such information might also increase the amount of disclosure. We propose a similar feed-forward message, which promises users that the recommendations will improve by a certain amount if they disclose a certain piece of information. The social cues and usefulness promises can be combined in a feed-forward message that tells users what percentage of other users received better recommendations after disclosing the information in question. The numbers in the feed-forward messages presumably affect the level of influence of the messages. As in previous work, they will not be based on real data but will rather be random within given ranges (Table 2).

Table 1. Different types of feed-forward messages to be investigated in our studies.

Type of feed-forward	Message to user
None	(no message)
Social	"XX% of our users gave us/allowed us to use…"
Usefulness	"The recommendations will be about XX% better when you give us/ allow us to use…"
Social usefulness (combined)	"XX% of our users received better recommendations when they gave us/allowed us…

Table 2. Different levels of influence that will be used in the feed-forward messages.

Level	Percentage
Low	A random number between 5% and 25%
Moderate	A random number between 40% and 60%
High	A random number between 75% and 95%

Whereas participants in previous studies chose their privacy settings once, we propose a system in which users can decide to change the amount and type of information they disclose. Users may base this decision on two pieces of feedback: the quality of the recommendations they receive, and a reflection of the information they are disclosing ('detailed profile inspection'). The effect of the quality of the recommendations on the amount of disclosure is unclear. In 'conversational' recommenders, where users incrementally disclose information, users tend to disclose more information if they see that this increases the recommendation quality [17]. This effect may however not occur in a system where most of the disclosure is at the beginning of the interaction.

For those types of disclosure that accumulate information over time, the user may initially not be aware of the exact extent of the disclosure. It is therefore assumed to be good privacy practice to allow users to inspect the 'profile' that the system has gathered over time [18]. Such detailed profile inspection may assist the user in deciding whether to change her information disclosure settings (see Table 3).

Table 3. Different levels of profile inspection (feedback).

Type of feedback	Implementation
Shallow	Shows the types of information being disclosed (e.g. “app usage”), but no specific information (e.g. the usage frequency)
Detailed	Shows the types of information being disclosed, as well as a detailed record of this information

3 Information Elicitation and the User Experience

In our proposed system, the amount of disclosure has a direct impact on the quality of the recommendations, and consequently on users’ satisfaction with the system. Information disclosure is thus a tradeoff between usefulness of disclosure and protection of privacy. Providing users with information can nudge users into over-protecting or under-protecting their privacy. If users are lured into over-protection, their satisfaction may decrease because the recommender may not have enough information to generate accurate recommendations. If users are lured into under-protection, they may later feel that their privacy was compromised.

Merely looking at users’ level of disclosure paints a one-sided picture; the complex nature of users’ interaction with the system warrants an integrative, user-centric approach. Based on Knijnenburg et al. [19], we hypothesize that several factors mediate the effect of feed-forward, feedback and disclosure on user experience: perceived privacy threat, perceived amount of control over the system, trust, and perceived quality of the recommendations. Table 4 and Fig. 1 show the hypothesized effects.

Table 4. Different levels of profile inspection (feedback).

Topic	Hypotheses
Feed-forward and feedback	The different types of feed-forward messages (H1) and levels of usefulness (H2) have a different impact on the initial amount of disclosure. The profile inspector (H3) and recommendation quality (H4) influence the change in disclosure. The profile inspector increases the perceived control over the privacy settings (H5), which increases the trust in the system (H6), which in turn causes a (negative) change in the level of disclosure (H7).
Privacy concerns and privacy threats	Users’ privacy concerns decrease the amount of initial disclosure (H8) and cause a (negative) change in disclosure (H9). The amount of initial disclosure (H10), change in disclosure (H11), and users’ privacy concerns (H12) influence the perceived privacy threat.
Recommendation quality and choice satisfaction	The amount of disclosure (H13) and change in disclosure (H14) influence the perceived recommendation quality, which in turn influences the satisfaction with the installed apps (H15)
System satisfaction and system use	The perceived privacy threat (H16), perceived recommendation quality (H17), and perceived control over the settings (H18) influence the system satisfaction, which is in turn related to the extent of system use (H19)

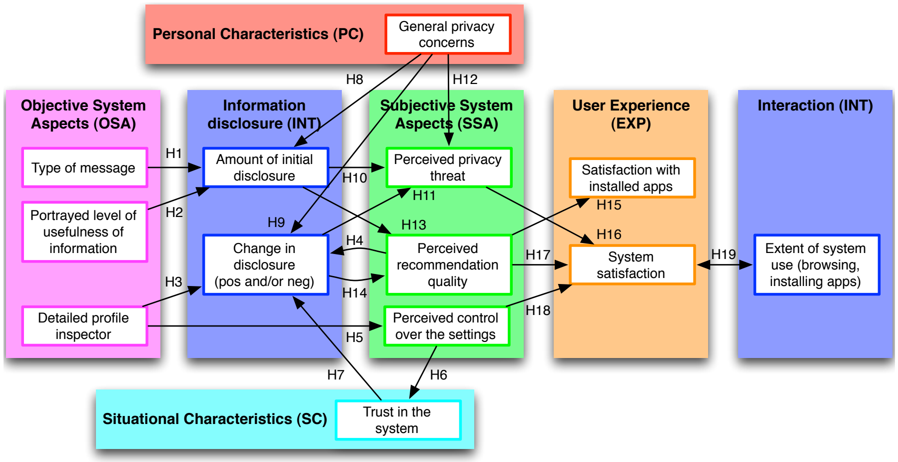

Fig. 1. A visual representation of the hypothesized effects in our studies.

4 Proposed Studies

We propose a series of studies that implement and test our feed-forward and feedback mechanisms in an app recommender system developed by Ericsson [3] with the working title "Applause" (Fig. 2). The system asks users to disclose their location (Fig. 2, screen 1), current app usage (e.g. app download, forwarding to friends, usage frequency, location, and time of day; screen 2), app browsing behavior in the system (screen 3), and demographics (e.g. age, income, occupation; screen 4). To guarantee that our findings are both comprehensive and statistically valid, we propose a variety of studies: qualitative user interviews, an online questionnaire, a highly controlled experiment with a system mockup, and a field test with real users of the real system.

4.1 Qualitative Study

The goal of the qualitative study is to get an in-depth insight into how users trade off the benefits of disclosing information with the threats that this poses to their privacy. 20-30 participants will be recruited, and given the opportunity to use the current version of Applause (without feed-forward and feedback) for at least a week.

Participants are asked to elaborate on their experience with the system. They are also asked about their phone usage, technological expertise, and privacy concerns. After that, they are shown different mockups of information disclosure screens (Fig. 2, screens 1-4). Screens will display different types of feed-forward messages, as well as different levels of influence. For each screen, users are asked if they would disclose the information or not, and to elaborate on their decision. Participants are also shown the different levels of profile inspection (screens 7-8), and asked for their comments. The goal of this study is to explore users' reactions to changes on each dimension.

Interview responses will be analyzed using grounded theory analysis [20], which models relationships between concepts (e.g. type of message and privacy concerns).

Models of each participant are compared to identify similarities and conflicts. The interviews are conducted in three batches, so that insights from the first analysis can influence the questions asked in the second batch of interviews. Finally, an integrated model is constructed, and interesting deviations from this model are highlighted.

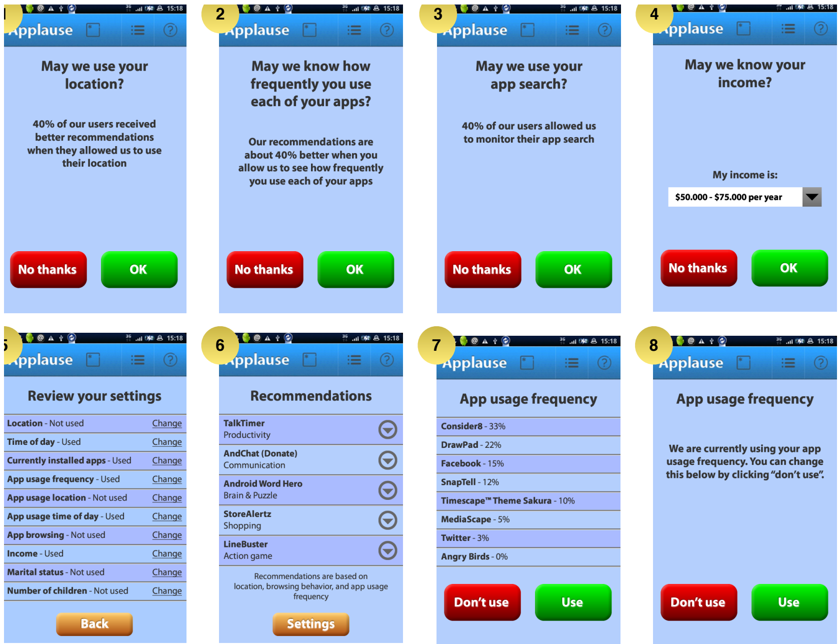

Fig. 2. Mockups of feed-forward and feedback conditions in proposed app recommender.

4.2 Online Survey

The online survey has the same goal as the quantitative study, but its results will be based on a larger sample (150-200 participants) and will have a quantitative character, allowing statistical validation of the results. This study also pre-tests the questionnaires that will be used in subsequent studies. Whereas participants in the quantitative study were asked to compare different types of feed-forward and feedback, the quantitative study presents each user with only one type of feed-forward message and one type of profile inspection. This results in 2x4 between-subjects conditions. The type of requested information and the level of influence are manipulated within subjects.

Participants are first asked several questions about their phone usage, technological expertise and privacy concerns. They are then randomly assigned to an experimental condition and shown mockups of information disclosure screens, with the feed-forward message and level of influence corresponding to this condition (Fig. 2, screens 1-4). For each screen, participants are asked to disclose this information or not. They are also shown one of the two profile inspectors (screens 7-8), and asked if they want to change their disclosure. Finally, they are asked several questions about the perceived privacy threat that this system poses, their perceived control over their profile, their satisfaction with a system that would use these features, and their intention to use that system and to recommend it to a friend.

Structural equation modeling will be used to extract relevant subjective concepts from the questionnaire responses and determine relationship between the experimental conditions and these concepts. The hypothesized effects that can be tested are a subset of the ones displayed in Fig. 1; specifically, due to the setup of the experiment we can only test H1-H3, H5-H12, H16, H18, and H19. As participants in this study are not interacting with the system, use can only be measured as an intention, and hypotheses related to the quality of the recommendations cannot be tested.

4.3 Fake Recommendation Experiment

Whereas the first two studies ask participants about their intended use of the system, the two experiments described in this and the next section consider actual system use. Research has shown that privacy attitudes and behaviors do not always align [21][22]. The fake recommendation experiment uses a semi-functional mockup of the recommender system that does not provide real recommendations (i.e., every participant receives the same recommendations), thereby controlling for the effects that would normally be mediated by the recommendation quality. Because the system is used only once, the different types of profile inspection will not be considered in this study. Type of feed-forward is again manipulated between subjects, and type of information and level of influence within subjects. The design of the study resembles [12]. The main difference to this work is that we test different types of messages.

100-150 participants are first asked several questions about their phone usage, technological expertise and privacy concerns. They then interact with the system in one cycle. The system first asks them to disclose information (Fig. 2, screens 1-4), where each screen shows a feed-forward message that corresponds to the randomly selected condition and the randomly selected level of influence. Then the system provides the (fake) recommendations (Fig. 2, screen 6). Finally, participants are asked about the perceived privacy threat posed by this system, the perceived quality of the recommendations, their perceived control over their own profile, their satisfaction with the system, and their intention to use the system if it would be available.

Structural equation modeling will be used to statistically test the relationships between the experimental conditions, the disclosure behavior, the subjective system aspects, and the user experience. The following hypotheses in Fig. 1 will be tested: H1, H2, H8, H10, H11, H13, H15, H16, H17 and H19. Note that participants use the system only once, so “extent of system use” can only be measured as an intention, and hypotheses related to changes in disclosure cannot be tested.

4.4 Field Experiment

The field experiment uses the fully operational app recommender. The study will sample 350 to 500 participants from existing users of the Applause system. Participants are shadowed over a period of time (in which they will be allowed to change their disclosure), and receive real recommendations based on their disclosure. The type of feed-forward message and the type of profile inspection are manipulated between subjects (leading to 2x4 conditions), and the type of requested information and the level of influence are manipulated within subjects.

Participants are first asked several questions about their phone usage, technological expertise and general privacy concerns. Consequently, they interact with the system repeatedly for a period of two weeks. Their initial interaction will be the same as in the fake recommendation experiment. However, after the first information elicitation screens (Fig. 2, screens 1-4), participants are asked to review their settings (screen 5) before moving on to the recommendations (screen 6). Participants are encouraged to revisit the recommendation screen throughout the study period. They will also be informed that they can return to the review screen to change their disclosure. When changing their disclosure, some participants are aided by a detailed profile inspector (screen 7), while others will only see a global profile inspector (screen 8).

The system logs participants' information disclosure and system usage (browsing recommendations, installing recommended apps). After two weeks, participants are asked several questions about the perceived privacy threat that this system poses, the perceived quality of the recommendations, and their satisfaction with the system and the apps they installed that were recommended by the system. Structural equation modeling will be used to evaluate all hypotheses in Fig. 1.

5 Conclusion and Future Work

Employing the user-centric framework for recommender system evaluation in [19], this paper applies (and extends) recent findings on information disclosure [12] to the field of recommender systems. Information disclosure is important for the proper operation of most recommender systems, and privacy issues are specifically salient in context-aware recommenders, where disclosure moves beyond the traditional elicitation of preferences. All proposed studies include "pretend" elements. Even the final study uses a "fake" feed-forward message (e.g., the expected usefulness of a certain piece of information that is not actually calculated). More research needs to be done to find 'real' metrics of information usefulness (e.g. the expected amount of change, or increase in accuracy, in the recommendations when providing the information).

References

1. Xiao, B., Benbasat, I.: E-commerce Product Recommendation Agents: Use, Characteristics, and Impact. MIS Quarterly. 31, 137--209 (2007).
2. Kobsa, A.: Privacy-Enhanced Web Personalization. In: Brusilovsky, P., Kobsa, A., and Nejdl, W. (eds.) The Adaptive Web: Methods and Strategies of Web Personalization. pp. 628--670. Springer, Heidelberg (2007). DOI: 10.1007/978-3-540-72079-9_21
3. Davidsson, C., Moritz, S.: Utilizing Implicit Feedback and Context to Recommend Mobile Applications from First Use. IUI 2011 Workshop on Context-awareness in Retrieval and Recommendation. pp. 19--22. ACM, New York (2011). DOI: 10.1145/1961634.1961639
4. Goy, A., Ardissono, L., Petrone, G.: Personalization in E-Commerce Applications. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web: Methods and Strategies of Web Personalization, 485-520. Springer, Heidelbg. (2007). 10.1007/978-3-540-72079-9_16

5. Adomavicius, G., Tuzhilin, A.: Context-Aware Recommender Systems. In: Ricci, F., Rokach, L., Shapira, B., and Kantor, P.B. (eds.) Recommender Systems Handbook. pp. 217--253. Springer, Heidelberg (2011). DOI: 10.1007/978-0-387-85820-3_7
6. Chellappa, R.K., Sin, R.G.: Personalization Versus Privacy: An Empirical Examination of the Online Consumer's Dilemma. Information Technology and. Management. 6, 181--202 (2005). DOI: 10.1007/s10799-005-5879-y
7. Teltzrow, M., Kobsa, A.: Impacts of User Privacy Preferences on Personalized Systems: a Comparative Study. In: Karat, C.-M., Blom, J., and Karat, J. (eds.) Designing Personalized User Experiences for eCommerce. pp. 315--332. Kluwer, Dordrecht (2004). DOI: 10.1007/1-4020-2148-8_17
8. Solove, D.: The Digital Person: Technology and Privacy in the Information Age. New York University Press, New York (2004).
9. Acquisti, A.: Privacy in Electronic Commerce and the Economics of Immediate Gratification. 5th ACM Conference on Electronic Commerce. pp. 21--29. ACM Press, New York (2004). DOI: 10.1145/988772.988777
10. Brodie, C., Karat, C.-M., Karat, J.: Creating an E-Commerce Environment Where Consumers are Willing to Share Personal Information. In: Karat, C.-M., Blom, J.O., and Karat, J. (eds.) Designing Personalized User Experience for eCommerce. pp. 185--206. Kluwer, Dordrecht (2004). DOI: 10.1007/1-4020-2148-8_11
11. Kobsa, A., Teltzrow, M.: Contextualized Communication of Privacy Practices and Personalization Benefits: Impacts on Users' Data Sharing Behavior. In: Martin, D. and Serjantov, A. (eds.) PET 2005., 329--343. Springer, Heidelberg (2005). DOI: 10.1007/11423409_21
12. Patil, S., Page, X., Kobsa, A.: With a Little Help from My Friends. In: ACM 2011 Conference on Computer Supported Cooperative Work. pp. 391--394. ACM Press, New York (2011). DOI: 10.1145/1958824.1958885
13. Besmer, A., Watson, J., Lipford, H.R.: The Impact of Social Navigation on Privacy Policy Configuration. In: 6th Symposium on Usable Privacy and Security. ACM Press, New York (2010). DOI: 10.1145/1837110.1837120
14. Dourish, P., Chalmers, M.: Running out of space: Models of Information Navigation. Short paper presented at HCI (1994).
15. Wang, W., Benbasat, I.: Recommendation Agents for Electronic Commerce: Effects of Explanation Facilities on Trusting Beliefs. Journal of Management Information Systems. 23, 217--246 (2007). DOI: 10.2753/MIS0742-1222230410
16. Berendt, B., Teltzrow, M.: Addressing Users' Privacy Concerns for Improving Personalization Quality. In: Mobasher, B., and Sarabjot, A.S. (eds.) LNCS, vol. 3169, pp. 69--88 (2005). DOI: 10.1007/11577935_4
17. Knijnenburg, B.P., Willemsen, M.C., Hirtbach, S.: Receiving Recommendations and Providing Feedback: The User-Experience of a Recommender System. In: Buccafurri, F. and Semeraro, G. (eds.) EC-Web 2010. LNBIP, vol. 61, pp. 207--216. Springer, Heidelberg (2010). DOI: 10.1007/978-3-642-15208-5_1
18. Kay, J.: Stereotypes, Student Models and Scrutability. In: Gauthier, G., Frasson, C., and VanLehn, K. (eds.) Intelligent Tutoring Systems. pp. 19--30. Springer, Heidelberg (2000).
19. Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H., Newell, C.: Explaining the user experience of recommender systems, http://db.tt/JG7079A.
20. Glaser, B.: The discovery of grounded theory: Strategies for qualitative research. Aldine Transaction, New Brunswick (2006).
21. Spiekermann, S., Grossklags, J.: E-privacy in 2nd Generation E-Commerce: Privacy Preferences versus actual Behavior. 3rd ACM Conference on Electronic Commerce. pp. 38--47. ACM Press, New York (2001). DOI: 10.1145/501158.501163
22. Tsai, J.Y., Egelman, S., Cranor, L.F., Acquisti, A.: The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study. Information Systems Research. 21, (2010). DOI: 10.1287/isre.1090.0260

Part II

Second International Workshop on User Models for Motivational Systems: the affective and the rational routes to persuasion (UMMS 2011)

User Models for Motivational Systems: The Affective and the Rational Routes to Persuasion

2nd international workshop
In conjunction with UMAP 2011
11 July 2011, Girona, Spain

Preface

Recent years have witnessed the growth of three parallel strands of research, all directing towards a more complex cognitive model of rational and extra-rational features, involving emotions, persuasion, motivation and argumentation.
On one side, Persuasive Technology is emerging as a very strong research field, interested in the use of interactive systems to influence human thought and behaviour. The international Persuasive conference is now well established at its 6th edition, and a series of other small events, like the Persuasive Technology Symposia (with AISB in 2008 and 2009), and workshops about persuasive technology at AmI2009 and Measuring Behavior 2010, confirm the importance of the field in the research landscape.
Parallel to this, Affective Computing is interested in the use, understanding and modelling of emotions and affect in computer systems. From the early 90s, which also saw two UM workshops (at UM03 and UM05), Affective Computing is now an established discipline, with an international conference (ACII), a professional society (HUMAINE) and, recently, a new journal (IEEE Trans. on Affective Computing).
Finally, Argument and Computation is also emerged in the past decade as a research strand interested in computational models of theories of argumentation and persuasion coming from Philosophy and Artificial Intelligence. Again, an increasing number of events dedicated to the topic, including two annual workshop series (Argumentation in MultiAgent Systems, now at its 8th edition, and Computational Models of Natural Argument, at its 11th edition) and a biennial international conference (COMMA), have recently been complemented by a new journal (Argument and Computation).

Following on from the workshop organised at UMAP 2010, this workshop intended to sit at the intersection between these three areas of research, and **focus on how adaptive and personalised systems can motivate people**, for instance to improve health, or to use sustainable resources, or to achieve goals or specific skills, by using persuasion and argumentation techniques and/or techniques involving the affective and emotional sphere.

The workshop's call focused on strategies, techniques and evaluation for motivational systems that tailor to cognitive and affective state of the individual. Suggested topics were:

- user models for persuasive motivational systems: Modeling receiver involvement, and position; Modeling personality and affective state for persuasion, Identifying relevant affective aspect in user modeling, Integrating affective and non-affective aspect in user models, Recognition and interpretation of the users' communicative intentions and affective states and updating of the user model, Investigating the relationship between recognized affective states and their impact on users' beliefs and motivation, Effect of cultural differences on persuasion;
- adaptive strategies for persuasion: Generating persuasive arguments; Ontologies for persuasion; Persuasive discourse processing: understanding what users say in terms of argumentation schemes; Computational models of argumentation tailored to a specific user; Rhetoric and affect: the role of emotions, personalities, etc. in models of persuasion and argumentation;
- motivation and affect: mutual interactions and synergies, peripheral routes of persuasion (humor, mood induction, enhancing source credibility)
- persuasive interfaces: ambient persuasion, use of embodied conversational agents, serious games
- applications and evaluations: in intelligent tutoring systems, health promotion, e-democracy, advertising, entertainment, coaching, decision support.
- ethical issues and evaluation of the impact of affective factors in motivation

We trust we managed to gather together an interesting set of papers on these topics, and we look forward to an interesting and stimulating event.

Floriana Grasso
Jaap Ham
Judith Masthoff

Programme Committee

Elisabeth Andre, University of Augsburg, Germany
Katie Atkinson, University of Liverpool, UK
Ruth Aylett, Heriot-Watt University, UK
Timothy Bickmore, Northeastern University, US
Nadja de Carolis, University of Bari, Italy
Peter De Vries, University of Twente, Netherlands
Susan Ferebee, University of Phoenix, US
Nancy Green, University of North Carolina Greensboro, US
Marco Guerini, ITC-IRST, Povo-Trento, Italy
Helmut Horacek, University of the Saarland, Saarbrücken, Germany
Irene Mazzotta, University of Bari, Italy
Cees Midden, Eindhoven University of Technology, Netherlands
Hien Nguyen, University of Aberdeen, UK
Nicole Novielli, University of Bari, Italy
Fabio Paglieri, ISTC-CNR, Rome, Italy
Helen Pain, University of Edinburgh, UK
Isabella Poggi, University Roma-Tre, Italy
Kaska Porayska-Pomsta, Institute of Education, University of London, UK
Chris Reed, University of Dundee, UK
Patrick Saint-Dizier, IRIT-CNRS, Toulouse, France
Oliviero Stock, ITC-IRST, Italy
Ielka van der Sluis, Trinity College, Dublin, Ireland
Julita Vassileva, University of Saskatchewan, Canada

Motivating People in Smart Environments

Berardina De Carolis and Irene Mazzotta

[1] Intelligent Interfaces Research Group
Dipartimento di Informatica, University of Bari
Via Orabona, 4 – 70125 Bari - Italy
{decarolis,mazzotta}@di.uniba.it

Abstract. In this paper we discuss the possibility to extend PORTIA, a persuasion system currently applied in human-agent dialogs, to support ambient persuasion. We have identified a fitness center as an appropriate smart environment in which ambient persuasion strategies can be applied. According to the Ubiquitous Computing vision, in the fitness center the user is surrounded by several connected devices that cooperate in the persuasion process, each of them with the most appropriate strategy, mode of persuasion, style of communication and ability of exploiting the kairos principle. To this aim we propose a multi-agent system able to support this distributed and intelligent approach to persuasion that allows to follow the user during the gradual change from the initial attitude to sustain of long term behaviours.

Keywords: Persuasion Systems, Ambient Intelligence, Multi-Agent systems.

1 Introduction

As stressed in Stock et al. [21] persuasion is a hot topic for intelligent interfaces since future interactive systems may have contextual goals to pursue which aim to induce and to convince the user to perform a specific action in the real world. It is feasible to imagine that persuasive technologies can be integrated into different aspects of daily life, and in this way they might have a greater persuasive power than traditional approaches to human-computer interaction. Under this perspective, the synergy between ambient intelligence and persuasion might be effective also because this solution, compared to traditional systems, could take the advantage to adapt the persuasion process, strategy and communication style to the context by using the kairos Principle [8].

In this paper, we present an approach to ambient persuasion [1] based on a combination of pervasive and distributed computation in which we aim at motivating people in the context of well-being. In particular we focus on how an intelligent environment may motivate the user to believe certain things, to behave in a certain way, or to abstain from performing certain actions, etc. This becomes important especially in certain kind of environments, such as those devoted to well-being, that intrinsically have this vocation. In fact, wellness is not limited to a single moment of people daily life – in which a person may consult a conversational agent or a web site in order to get advices or suggestions for improving life quality - but it is a continuous process along the temporal dimension and it is more central and peculiar in some environments than in others (i.e. fitness centers, food shops, homes, etc.). Moreover, the devices in these environments may cooperate in order to support people in achieving their goals.

Ambient Intelligence solutions may provide a great opportunity for achieving the aim of distributing and embedding persuasion and coaching strategies into the environments that the user attends, according to the Ubiquitous and Pervasive Computing vision [25], in order to apply persuasion methods and techniques usable *through several devices* and *in different usage contexts*.

Changing habits in the context of well-being is influenced by several -rational and emotional- factors depending on the context, that can be intended as: 'What the user is doing, Where is the user, With whom, When' [6]. Of course, attention should be paid to insure that arguments are relevant and strong to the user, especially in ambient intelligence context where it is essential to consider the conditions in which the message is communicated. Therefore, in our opinion, it is important not only to distribute the message through the existing devices in the environment and to adapt the persuasion strategy, the arguments and their expression to the user and the context, but it is also necessary that all the environments involved in the user's activities, task, etc, may communicate in order to cooperate to achieve the common goal of caring for the user.

To this aim, we propose a multi-agent architecture which includes different types of agents: (i) *Sensor Agents* –used in order to provide information about sensors parameters and context features (i.e. temperature, heart rate, humidity, presence of the user in a room, etc..); (ii) *Device agents* -typical of the environment- that manage the active devices in the environment (e.g. cardio fitness machines, public displays, mirrors, etc.) and convey to the user the training according to the context and the aim of the environment; (iii) *D-Me agents* [4], represent the users in the environment as a kind of digital image of the user; finally, (iv) the persuader agent that we call *Coach agent*, decides the most promising persuasion strategy to apply in a given context and communicate the action plan to Device agents.

In order to show how this architecture works, we will consider a fitness center as a suitable place to test the approach. In fact, a fitness center is equipped with enough technology for simulating a smart environment, the users are already confident with the technology during their workout and, moreover, most of them want to be constantly motivated in order to reach their goals concerning an healthier lifestyle [15].

The paper is structured as follows. In Section 2 we discuss the relation between ambient persuasion and wellbeing. Section 3 describes the proposed architecture of the system. Then, in Section 4, we illustrate a scenario example that is used to show the functioning of such a system. A final discussion and future work directions are reported in Section 5.

2 Ambient Persuasion and Wellbeing

Persuasion is a form of social influence and a ubiquitous part of contemporary life. It is a relatively new trend in the research community that shows a growing interest also into intelligent information technologies, and for better or for worse, persuasive technologies are already part of the everyday technological landscape (see examples in [8]). Coming from persuasion and technology, persuasive technologies are not exempt from ethical issue: they should be employed to change people's attitudes or behaviour without coercion or deception, acting therefore upon users' beliefs always in an atmosphere of free choice, where they are autonomous and able to change their mind. An application area in which persuasion can be used with great effectiveness is well-being, especially when its purpose is to persuade people to adopt a healthier diet, lifestyle, etc. In our opinion, wellness is a domain in which ambient persuasion technologies may increase its potential of alleviating the users' problem by helping them in triggering the decision to change their wrong habits and motivating them to achieving their goals. There are different examples of systems aimed to persuade in this application domain. Many of them are implemented as Embodied Conversational Agents that play a role aiming at inducing behavioural change in users, a role that traditionally was filled by coaches or therapists [2, 5, 12].

Currently, the most common persuasive systems used in fitness centers (at least in Italy) employ the feedback mechanism to show to users effects of the exercise (i.e. Polar Cardio or Cardio Fitness machines and so on) and are generally isolated without the ability to communicate and cooperate with other devices in order to achieve the common goal of taking care of the user. In addition, a fitness center has professionals responsible for this purpose, as personal trainers and wellness coaches: they have a very important role in helping the user to change their habits and find the motivations to work hard for achieving their goals. However, beside that they can be expensive or unavailable when users need them, many people feel shame and fear of being judged by their human coach: sometimes this can be a motivation for changing attitude, sometimes it may compromise the success of the coaching strategy, increasing the user's attitude at overcoming barriers -especially emotional- and decreasing self-esteem. Several coaching systems have been implemented on mobile devices (see for example, My Weight Loss Coach for Nintendo DS, Nokia Fitness Coach for Nokia phones and the so many sport trackers like Endomondo for the most popular mobile platforms, or CardioTrainer for Android, or Sports-Tracker for Nokia) aimed at monitoring, supporting and tracking users' progress and improving their energy balance. Again, in many of them the user has to input data about her workout, eating behaviour, etc. On the contrary, in other systems, as Nintendo Wii fit, My Body Coach by BigBen Interactive, or Your Shape: Fitness Evolved, the new edition of Ubisoft's training software for Microsoft console, the user is monitored and motivated during the exercise even though these are not integrated with other daily activities and situations of the user.

According to [10], when persuasion is used in ambient intelligence contexts it may take advantage of the distributed intelligence of the environment in order to improve the effectiveness of the persuasion process. For instance, since entities taking part of the persuasion process are multiple the system may use repetition for increasing compliance. Moreover, these multiple sources may have different roles in the process of persuading, motivating, sustaining the user and, therefore, may use different strategies.

Again, an intelligent environment is a social place and therefore, people may share personal experiences with others that have the same problems, goals, needs [19]. In this sense, perceived similarity through shared experience may have an effect on compliance [7].

Finally, the system should be perceived not has having a pure functional intelligence but has being an emotionally and socially intelligent actor that may monitor the user and intervene appropriately at the right moment.

In the light of these premises, we present an agent-based system that tries to apply the principles of ambient persuasion in a smart fitness center.

3 The proposed System

According to the ambient persuasion model proposed by Kaptein et al. [10], the first difference from traditional persuasive systems consists in the fact that the persuasion process can be distributed not only with respect of multiple sources but also according to the phases that constitute the gradual change from the initial attitude to sustain long term behaviour. In the application domain considered in this paper, the system provides a *first phase* in which the user should be persuaded to have the intention of adopting a certain behaviour, for instance a particular type of workout, and then, in a *subsequent phase* should be sustained using appropriate motivational cues, during the entire path of actuation of the suggested behaviour.

In order to generate the most appropriate persuasive message to the user, we extended PORTIA and used its reasoning and argumentation model. As far as the sustain phase it is necessary to reason on which motivational arguments have to be adopt for continuing to motivate the user according to the situation. To this aim, we started an empirical study aiming at exploiting the knowledge and rules that human personal trainer and fitness professionals use.

Before illustrating the architecture and the functioning of the system, let us introduce a brief overview of PORTIA.

3.1 An overview of PORTIA

PORTIA is a user-adapted persuasion system capable of simulating the persuasion process used by humans to convince someone to perform a given action. In this paper we provide a brief overview of the system. For a more detailed description of PORTIA, please refer to [14] It mainly focuses on two typical aspects of the human persuasion in order to produce effective persuasion attempt in different contexts: on one hand, the ability of reasoning on the potential strength of alternative persuasive strategies for a given user, in order to select the most appropriate one; on the other hand, the capability of combining rational and emotional modes of persuasion, according to the theory of *a-rational* persuasion [16]. The strategies represented in the model are the result of a combination of theoretical [22, 23, 18]) and empirical [13] background. The key points of the system are the separation between *reasoning* and *argumentation* phases in the persuasion process [24], and the use of *Belief Networks* to represent the uncertainty inherent in this form of practical reasoning [17].

PORTIA considers three knowledge bases: the User Model, the Persuasion and the Argumentation Knowledge Bases.

The ***User Model***. Understanding the presumed weight of user's goals is crucial to select the most promising persuasion strategy in a given context. User Model is employed to reason about the user's presumed characteristics. Rather than acquiring this knowledge through direct questions, PORTIA attempts to implicitly infer it, with some level of uncertainty, from information about user's personality traits and living habits. The User Model includes a *specific knowledge* and a *general knowledge* component. The former collects facts about the user (evidence).The second represents criteria to infer the user's goals and abilities under conditions of uncertainty in the form of *Elementary Belief Networks* (EBNs) that are belief networks with only one leaf node representing uncertain implications. In particular, user's rational and emotional goals can be inferred respectively from knowledge about user's habits and personality traits.

The ***Persuasion Knowledge Base*** is employed to model rational and emotional strategies. The Persuasion model is defined in term of goals and beliefs from the Persuader's perspective that may employ rational as well as emotional strategies (but also a mixture of them) to induce the user to perform a given action. Persuasion strategies are represented with EBNs too. In particular, emotions may be introduced in the persuasion process in two forms: by selecting an emotional goal or by activating, through arousal of user's emotion, an intermediate goal which is instrumental to the final one. The PORTIA's persuasion strategies are summarized in Table 1. For more details see [15].

Table 1.A summary of the Persuasion Strategies used by PORTIA

<table>
<tr><th colspan="2">PORTIA's Persuasion KB</th></tr>
<tr><td colspan="2"><u>General induction of intentions</u>
$[(VGoal\ U\ g_i) \wedge (AGoal\ U\ g_i) \wedge (Bel\ U\ Implies(a,gi)) \wedge (Bel\ U\ CanDo(U,a))\ ] \rightarrow ?\ (Int\ U\ Do(U\ a))$ [i]

It may be summarized as follow: "If User has the goal g (VGoal U g) and it is really relevant at this time (AGoal U g) and he believes that doing the action a implies achieving g in a more or less near future (Bel U Implies(a,gi)), and he believes that has the ability to do a (Bel U CanDo(U,a)), then probably user intends to do a (Int U Do(U a))" (from Miceli et al, 2006).</td></tr>
<tr><td>Rational induction of intention
$g_i \in$ {Rational goal set}
It focuses on rational goals like 'to be in good health', 'to have a good appearance', and so on.</td><td>Emotional induction of intention
$g_i \in$ {Emotional goal set}
It focuses on rational goals like 'to make friends', to be in good mood', and so on.</td></tr>
<tr><th colspan="2">Activation of goal strategy</th></tr>
<tr><td colspan="2">Activation through a belief or an emotion of an intermediate goal which is instrumental to the user's goal. It considers two possible applications: Rational Activation strategy or Emotional one.</td></tr>
<tr><th colspan="2">Induction of beliefs</th></tr>
<tr><td colspan="2">Argumentation about means-end implication. It represents the action-goal relation.</td></tr>
</table>

Appeal to Expert Opinion	Appeal to Popular Opinion	Appeal to Position to Know	Appeal to Friendly Personal Experience	Appeal to Examples	Others

The ***Argumentation Knowledge Base*** is employed to translate each strategy into an argument. Items to include in the argument correspond to the variables associated with nodes of EBNs, and the way these items are combined in the message (order in which to present them and relationships among the various parts) is represented into *Elementary Argumentation Plans* (EAPs) that are a coherent translation of EBNs. EAPs are built on two theoretical grounds: Walton's Argumentation Schemes [20] and Rhetorical Structure Theory (RST) [11]. In particular, EAPs represent the association between *rhetorical relations* (RRs) and argumentation scheme.

PORTIA considers two main modules: the Reasoning module (REASONER), and the Argumentation one (ARGUER).In the ***Reasoning module***, PORTIA exploits the information about the user (User Model KB), computes the degree of importance of the various -rational and emotional-goals on which focus the persuasion strategy, and evaluates the persuasiveness of different combination of strategies (Persuasion Strategies KB) and selects the most promising one with respect to the goal of inducing in the user the intention to do a certain action. For this purpose, PORTIA builds a complex Belief Network (BN) by dynamically chaining forward several EBNs. The BN is a representation of the user's mental state that enables to apply a "what-if" reasoning form for evaluating the persuasive power of the strategies, and to select the most promising one.

In the ***Argumentation module***, PORTIA has to construct the arguments to express the strategy selected in the previous step. To this aim, PORTIA explores the complex Belief Network and decides the items to mention, their presentation order and the rhetorical relations among them. Also, she has to decide whether to include an appeal to cognitive consistency (between the user's of goals and beliefs, and his behaviour) as a form of encouragement to adopt a more consistent behaviour. That is, PORTIA has to translate the complex Belief Network into a coherent discourse plan. The discourse plan is dynamically built by combining the elementary argumentation plans (Argumentation Plans KB) that represent the elementary beliefs networks included in the Belief Network. The discourse plan is then translated into a natural language message used as an attempt to persuade the user.

3.2 The System Architecture

In order to develop a system for ambient persuasion in the context described in the Introduction, we propose an extension of a multi-agent platform implemented in another project [4] which considers four types of agents:

i) *Sensor Agents* – they are used in order to provide information about sensors parameters and context features (i.e. temperature, heart rate, humidity, presence of the user in a room, etc..).
ii) *Device agents* – they control the active devices in the environment (e.g. cardio fitness machines, public displays, mirrors, etc.) and communicate with the user

by conveying the messages of the coach agent according to the display facilities typical of the controlled device.

iii) *D-Me agents* – they represent the users in the environment as a kind of digital alter-ego. In particular, a D-Me agent knows the user and monitors all his/her activities, when authorized, communicates the information required by the environment according to the privacy policies set by the user. Of course, the user can always decide which data to send to which environment and the level of detail of the information to be provided to the environment and the coach agent.

iv) *Coach agent*s – they decide the most promising strategy to apply in a given context in order to persuade/motivate/sustain the user involved in the workout and communicate the action plan to the Device agents or to the D-Me agents. Coach agents are specialised in persuasion strategies typical of the environment.

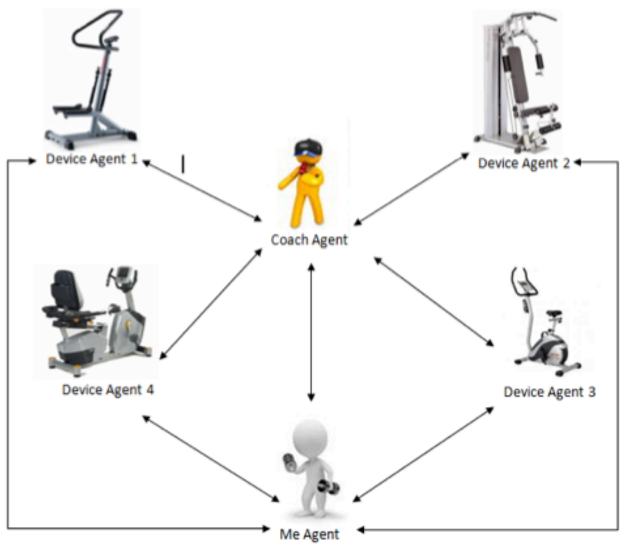

Fig. 1. A schema of a possible configuration of the multi agent platform.

It is worth noting that the architecture that we propose has not been conceived with the sole purpose of persuading the user, but it aims at implementing smart environments that aims at improving the quality of life of the user. In these environments all agents exchange data and information in order to provide services (recommendations, information, motivation) suitable for helping users in achieving their goals.

For instance, the D-Me agent, by monitoring the user behaviour, knows about his/her meals, and, through social networks, may know who are his/her friends, etc. Then, it may communicate this data to the Coach Agent that may adopt the optimal persuasion strategy and arguments accordingly. Again, suppose that the doctor recommended to the user to loose weight. When the user goes to the food shop with the intention to buy a sweet cream, the D-Me agent may communicate this information to the Coach Agent that may act to persuade him not to do it by adopting the most effective persuasion strategy and arguments.

For this reason, one might argue: why not delegating the entire ambient persuasion process to the D-Me agent? The idea is to build a platform environment-independent that may be applied in a fitness center as well as in a virtual home and, possibly, enabling interaction between different environments so as to support the user at different times of the day. According to this perspective, we believe that D-Me agent should not have specific knowledge of the environment or the technological devices because persuasion strategies and arguments used by a Personal-Coach in the fitness center are probably different from those applied by Personal-Butler in the smart home.

4. An example

The following example is a simulation of the system's behaviour in a typical scenario of the gym environment. Let us consider the following starting conditions.

> Robert is a *man below 40 years* who regularly *makes medical check-ups*. He is a *hypochondriac*, too. He is probably an extravert because he *feels comfortable around people*. The doctor suggested him to make some physical activity regularly. He decides to go to the gym but he is quite sceptical about this.

When Robert enters in the gym his D-Me agent has the permission to communicate to the other agents in the environment the anagraphical and physiometric data and other information about the user social network. Robert registers himself to the gym information system. Then, the system provides a personalized workout schema to Robert and, in order to persuade him to adopt the proposed workout, the Coach agent generates and communicates the following persuasion message, according to the BN in Figure 2:

> "*Hi Robert, I am your personal coach and this is your personalized workout schema. You should do it because I know that you care for your health and training has a lot of benefits on your health. In fact, the World Health Organization says that this is very important for health and it is a specialized agency of the United Nations that acts as a coordinating authority on international public health, it is an authoritative voice. In addition working out may be a great opportunity to know new friends. In fact, this is well known to all who attend gyms and there is no evidence against it. Come on! I'm sure you can do it if you wish".*

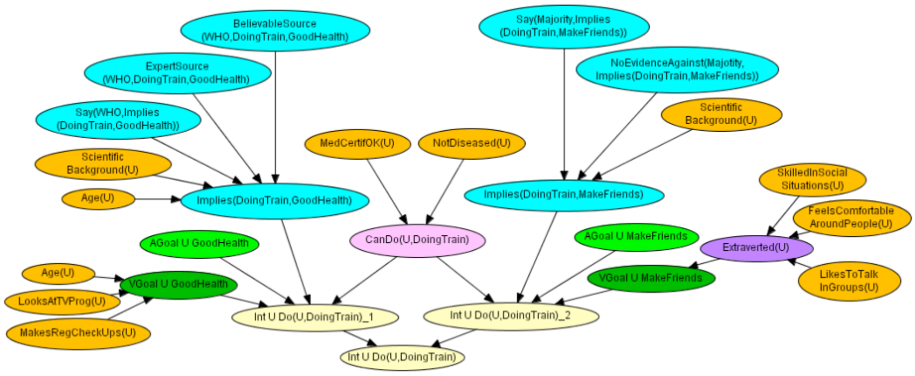

Fig. 2. The BN used by the Coach agent to simulate the effect of selected strategy on the user's mental state.

Figure 2 represents the result of the reasoning process of the persuasion model applied by the coach agent when tries to persuade the user to adopt the proposed workout. That is, the coach's representation of the user's mental state on which has been tested the effectiveness of the persuasion strategy. Reasoning component has propagated the user's evidence received by the D-Me agent into the EBN-KB and has inferred that, although '*to be in good health*' is the presumed most important goal to Robert, the associated rational persuasion strategy does not seem to induce in the user the desired level of intention to do the proposed workout schema and a mixed strategy could be more effective. Therefore, Coach agent selects the goal with the highest value among the emotional goals and infers that the two candidate goals on which the persuasion strategy focuses are the rational goal *to be in good health* and the emotional *goal to make friends*. Moreover, it selects the belief induction strategies through *appeal to expert opinion* and *appeal to popular opinion,* respectively, for the goals of being in good health and making friends.

In a dialog perspective, persuasion, rather then a predefined, integrated set of propositions, is seen as a sequence of moves in which two parties (Persuader and Receiver) are reasoning together on some argument [9]. While monologic persuasion is characterized by the three steps (planning, plan revision and surface realization), in the 'pure' persuasion dialogues the sequence of exchanges includes some typical phases, and forms of reasoning, by the Persuader (that is, making a proposal, observing, classifying and reasoning on the Receiver's reaction and replying to it). Therefore, a persuasion attempt may be criticized by the Receiver in several ways: by questioning the goal premises; by attacking them with counter-arguments alleging that one or more of them is false; by undercutting the inferential link between premises and conclusion with critical questions; by rebutting the practical reasoning inference with counter-arguments asserting that the conclusion is false or by putting forward a proposal arguing for a different action, and contending that the arguments for this opposed proposal are stronger. Persuader must be able to respond appropriately to all these situations.

For example, let's suppose that Robert is hypochondriac and he does not want to do the proposed workout schema because he is afraid of getting sick. The Coach agent classifies this reaction as an object on the user's capability and, after reasoning on the BN in order to select the most appropriate response, tries to reassure him because all the tests say he's fine:

> *"I know you're scared but try to stay calm because your medical certificate says you're fine!"*

Finally Robert goes to the bike and began his training.

Another strategy applied by the Coach agent is to motivate and sustain the user during the exercises. Let's suppose again that, despite the encouragement received, Robert starts going very slowly on the bike. Then the Device agent that monitor the user's exercises, communicates to the Coach agent the new situation of Robert. Then, the Coach tries again to reassure the user by conveying a following message on the display of the bike device:

> *"Robert, come on, do not worry, I'm tracking your heart and, at present, you are even below the threshold workout. Therefore, you can do more without fear of forcing your heart ... I monitor you, do not worry"*.

Robert starts to push a little more and retakes the right pace.

Then, finally, the user finishes the exercise. The Device agent transmit this information to the Coach that tells Robert that everything went well but that, the next time, he could do more.

5 Discussion

This contribution shows a preliminary work towards the development of a system for ambient persuasion in a fitness center based on a multiagent architecture. In the proposed architecture there are two agents that are central to the persuasion process. The D-Me agent, which manages the personal user profile, may transmit to the intelligent environment data about the user that may be important for adapting the persuasion strategy and the motivational messages. The Coach agent, which acts as a personal trainer, has the role of persuading the user to train and adopt a certain workout and also to sustain this behaviour during training. The Coach agent uses the reasoning and argumentation model of PORTIA for generating the persuasive message. While, for the generation of motivational messages to be provided in the sustain phase, we are collecting data from professionals expert in the fitness and wellness domain, such as personal trainers.

We are aware that in this domain the risk of producing a message that is not appropriate to the situation because of an inferential error on the user goal, personality traits, and so on may determine the selection of a wrong strategy or arguments and consequently may cause distrust in the user. In this case it is necessary to endow the environment with a formal model of trust [3], in order to give to the coach agent the capability of assessing the level of trust that the user has in the system behaviour and to reason on the cognitive factors involved on this project in order to recover the situation.

At moment, in order to test the effectiveness of the proposed system we used the knowledge of two personal trainers. To these people we proposed some scenarios, like the one in Section 4 of this paper, with the aim of collecting examples of motivational sentences, arguments to be used to motivate the users. Now we are conducting an experimental study that involves a greater number of experts in the fitness domain aiming at understanding:

- which are the features of the user relevant for adapting the motivational message;
- when to intervene with a motivational message;
- which are the strategies, at the reasoning and argumentation levels, most widely used according to the user features.

At present, the collected data give us some useful information for understanding how human personal trainers build in their mind the models of their clients and which are the features of the clients that influence their decision about how to motivate them.

For instance, the gender of the client seems to be important for choosing the arguments to use. Personality traits (mainly the levels of sociability and extraversion) influence the message style. While the cultural background, the age and the profession of the client influence the argumentation schema to be used to support some concepts and claims.

Moreover, from this initial analysis, seems clear that personal trainers initially classify their clients into stereotypes (*Lazy, Super, Model, Normal, Sociable*, …) that help in deciding how to motivate them initially. This capability is related to the level of experience of the trainer.

Then, in our future work we plan to analyze the collected data and build the initial knowledge of the coach agent relative to stereotypes and reasoning rules in order to generate motivational messages appropriate to the user and to the situation. Moreover, we intend to give to our coach the capability to learn from the user feedback in order to refine the rules driving the choice of the optimal strategy.

References

1. Aarts, E., Markopoulos, P., de Ruyter, B.: The persuasiveness of ambient intelligence. In M. Petkovic, W. Jonker (Eds.), Security, privacy, and trust in modern data management. (pp. 367-381) Berlin, Germany: Springer (2007)
2. Bickmore, T:"Relational Agents: Effecting Change through Human-Computer Relationships" PhD Thesis, Media Arts & Sciences, Massachusetts Institute of Technology (2003)
3. Castelfranchi, C., Falcone, R., Lorini, E.: A non-reductionist Approach to Trust. In J. Goldbeck (Ed.), Computing with Social Trust . Berlin, Springer, 45-72 (2008).
4. Cozzolongo, G., De Carolis, B., Pizzutilo, S. : A Personal Agent Supporting Ubiquitous Interaction. WOA 2004. Torino, Italia 2004: 55-61 (2004)
5. de Rosis, F., Novielli, N., Carofiglio, V., Cavalluzzi, A., De Carolis, B.: User Modeling And Adaptation In Health Promotion Dialogs With An Animated Character. International Journal of Biomedical Informatics, 514-531 (2006)
6. Dey, A. K., Abowd, G. D.: Towards a Better Understanding of Context and Context-Awareness. Workshop on 'The What, Who, Where, When, and How of Context-Awareness', as part of CHI 2000, The Hague, The Netherlands, (2000)
7. Festinger L (1954) A theory of social comparison processes. Hum Relat 7:117–140

8. Fogg, B.J.: Persuasive Technology: Using Computers to Change What we Think and Do. Morgan Kaufmann (2002)
9. Guerini M., Stock O., Zancanaro M., O'Keefe D.J., Mazzotta I., de Rosis F., Poggi I., Lim M. Y. & Aylett R. "Approaches to Verbal Persuasion in Intelligent User Interfaces". In P. Petta, R. Cowie and C. Pelachaud (eds.) The HUMAINE Handbook on Emotion-Oriented Systems Technologies. Springer, (2011).
10. Kaptein, M., Markopoulos, P., de Ruyter, B. & Aarts, E. (2009). Persuasion in Ambient Intelligence, Journal of Ambient Intelligence and Humanized Computing.
11. Mann, W. C., Matthiesen, C. M. and Thompson S. A.: Rhetorical structure theory and text analysis. Information Sciences Institute Research Report ISI/RR-89-242, 89-242, (1989).
12. Marsella, S. C., Johnson, W. L., LaBore, C. M.: Interactive pedagogical drama for health interventions (2003)
13. Mazzotta, I. and de Rosis, F.: Artifices for persuading to improve eating habits. AAAI Spring Symposium on "Argumentation for consumers of health care". Stanford, USA. Technical Report SS-06-01, 76-85 (2006).
14. Mazzotta, I., de Rosis, F., Carofiglio, V.: PORTIA: a user-adapted persuasion system in the healthy eating domain. IEEE Intelligent Systems, 22, 6, 42-51 (2007)
15. Mazzotta, I., Silvestri, V., and de Rosis, F.: Emotional And Non Emotional Persuasion Strength. Proceedings of AISB'08, Symposium on 'Persuasive Technology', 14-21 (2008)
16. Miceli, M., de Rosis, F., Poggi, I.: Emotional and non-emotional persuasion. Applied Artificial Intelligence: an International Journal, 20, 10, 849-880 (2006)
17. Pearl, J.: Probabilistic Reasoning in Expert Systems: Networks of Plausible Reasoning. San Mateo, CA: Morgan Kaufmann (Pubs.) (1988).
18. Petty, R. E., and Cacioppo, J. T.: The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), Advances in experimental social psychology, 19, 123-205. New York: Academic Press (1986).
19. Pinel EC, Long AE, Landau MJ, Alexander K, Pyszczynski T (2006) Seeing I to I: a pathway to interpersonal connectedness. J Pers Social Psychol 90:243–257
20. Reeves B, Nass C (1996) The media equation: how people treat computers, television, and new media like real people and places. Cambridge University Press, Cambridge
21. Stock O, Guerini M, Zancanaro M.: Interface design and persuasive intelligent user interfaces. chapter The foundations of interaction design. Lawrence Erlbaum, Hillsdale, NJ, (2006)
22. Walton, D. N.: Argumentation Schemes for Presumptive Reasoning. Mahwah, N. J.,Erlbaum (1996).
23. Walton, D. N.: The place of emotion in argument. The Pennsylvania State University Press (1992).
24. Walton, D.: What is reasoning? What is an argument? Journal of Philosophy, 87, 399-419 (1990).
25. Weiser, M.: Some computer science issues in ubiquitouscomputing. Commun. ACM, 36(7):75-84 (1993)

Arguing with Emotion

Martyn Lloyd-Kelly and Adam Wyner

University of Liverpool, Liverpool, L69 3BX, UK,
{mlk5060,azwyner}@liverpool.ac.uk,
http://www.csc.liv.ac.uk/

Abstract. Emotions are commonly thought to be beyond the pale of rational analysis, for they are subjective, may vary even with respect to the person experiencing the emotion, and may conflict with rational thought. In this paper, we develop the position that emotions can be the *objects* of argumentation, which we express by introducing emotion terms in *emotional argumentation schemes*. Thus, we can argue about whether or not, according to normative standards and available evidence, it is plausible that an individual had a particular emotion. This is particularly salient in legal cases, where decisions can depend on explicit arguments about emotional states.

Keywords: legal reasoning, emotional argumentation schemes

1 Introduction

Emotions are commonly thought to be beyond the pale of rational analysis. They are subjective; the same person in the same context may have different emotional responses to stimuli; a person's emotional response may conflict with rational thought. Emotions are also thought to only serve in an adjunct role in decision-making, by enhancing, moderating, or interfering with the persuasiveness of reasoning in an argument [19]. However, emotions can have a direct role where we normatively analyze and evaluate emotional appeals [4,11]. Emotions themselves can be viewed as *objects* of argumentation, not just adjuncts [13]. Thus, rather than filtering out or subordinating to rational argument, emotions can be *first class* citizens of argumentation. Developing this position, we introduce *emotional argumentation schemes*, where emotional terms are the components of the argument. This is particularly salient in legal cases, where reasoning about emotional states is a critical factor in reaching legal determinations.

In this paper, we briefly outline the legal context, computational analyses of emotions, and current research on emotions and argumentation. We then introduce our novel emotional argumentation schemes, where the key idea is that emotional terms can be central *components* in the schemes. These schemes model key parts of reasoning in the legal context and of the computational analysis of emotions. We use the schemes to model legal arguments that are relevant to legal cases. We close with some indications of future research.

2 Emotions in the Law

As emotions are a widespread, salient experience of our lives and in our social encounters, it is unremarkable that they are the subject of legal proceedings, where human experience and behaviour is reasoned about and regulated. Considering legal contexts bounds our discussion in three respects. First, there are explicit arguments about emotions, so we need only be concerned with *explicit statements about emotions rather than their psychological or physical reality*. Second, the legal context is *normative* and *truth determining*; judges and juries decide, relative to a normative model of human emotional responses. This means that though a party to a legal case may claim an emotional state as justification for an action, the courts may decide otherwise based on arguments about evidence, testimony, normative reasoning about emotional states, etc. Third, the arguments we consider are about emotional states *after the fact*, for we are not considering emotions engendered during the court proceedings. This means that scientific indicators of the embodiment of the emotion, e.g. MRI brain scans along with other physiological measurements, are not relevant to our discussion. While we acknowledge theories bearing on the embodiment of mind and emotion [10], we can only relate to the issues raised in terms of normative legal arguments about claims of a past emotional state rather than the real time indicators of emotion.

2.1 Various Forms

Emotions in law appear in a variety of ways. In common law, among the *causes of action* we find *intentional or negligent infliction of emotional distress* and *sexual harassment*, which have emotional referents. Over the course of litigation, there will be arguments as to whether distress was caused, the extent of distress, along with supporting evidence or expert testimony. In *hate crimes*, the emotional disposition of the perpetrator, whether the perpetrator felt hate towards the victim, may be subject to argument [7]. The difference between *murder* and *voluntary manslaughter* can hinge on the emotional state of mind of the perpetrator, e.g. *heat of passion*. Where emotionality is said to interfere with rationality, time may be a crucial factor, for the more time that passes between the incident that instigates the emotion and the action, the more the perpetrator is normatively taken to return to his "right mind", making the action more premeditated, and therefore more severely punishable. In arguing a case, lawyers make *rhetorical appeals* to a jury, attempting to elicit pity, fear, or sympathy in an effort to sway a decision on behalf of their client. *Jury instructions* are given by the judge to the jury about how the jury should reason with the evidence, law, and arguments in reaching its decision. For example, a jury might receive instructions to reason strictly about the facts of the case with respect to the law, leaving aside emotional appeals. In cases of particularly heinous crimes, the *degree of outrage to the sensibilities* is relevant in meting out punishment. Finally, in coming to a decision, the judges may seek any relevant mitigating factors which warrant *mercy* and counterbalance an otherwise harsh decision.

In all these uses, we can reason and argue about emotional content. For example, to counter an emotional conclusion, one might question whether certain actions, statements, or circumstances are consistent with a normative standard under which the claimed emotion obtains. Where such inconsistencies arise, one may counter-claim that the emotion did not normatively obtain, undermining the claimants argument. Alternatively, there may be procedural moves, as in where an emotional claim or emotional argument is ruled inadmissible in court. In these various ways, we reason explicitly about arguments with emotional content rather than simply ruling them out. As argued in [9], by making emotional arguments explicit and formal, we can present better, clearer, and fuller representations of legal case arguments and decision making. The question is, then, just how to represent emotions so as to be arguable?

2.2 Jury Instructions

One approach to modeling legal reasoning would be to model individual cases or a corpus of legal cases, e.g. as in legal case-based reasoning [1]. We take a different but related approach by modeling aspects of the reasoning found in jury instructions, e.g. the *Judicial Council of California Criminal Jury Instructions (2011)* [12], which are developed and maintained by criminal justice systems as instructions and standards for judges, juries, and litigants on how legal issues are to be decided, giving indicative cases. As such, in other words, jury instructions are intended to be distilled guidance about normative legal reasoning that takes the proceedings, evidence, and arguments of the case over time as input and produces a decision.

We consider, in particular, California Criminal Jury Instruction CALCRIM No. 511 *Excusable Homicide: Accident in the Heat of Passion*, which establishes the conditions under which a homicide is excusable on the grounds of extreme emotion and cites cases for various points of the conditions, e.g. Substantial Emotional Distress Defined in People v. Ewing (1999) 76 Cal.App.4th 199, 210 [90 Cal.Rptr.2d 177].

To ground our analysis, we provide the relevant extracts from the two pages of the jury instructions for CALCRIM No. 511. We index clauses I - VII for reference, and we have omitted clauses irrelevant to our discussion relating to *undue advantage*, *dangerous weapons*, *cruelty or unusualness of killing*, *intent to kill*, *great bodily injury*, or *criminal negligence*:

[I] CLAIM: The defendant is not guilty of (murder/ [or] manslaughter) if (he/she) killed someone by accident while acting in the heat of passion. Such a killing is excused, and therefore not unlawful, if, at the time of the killing:

- **1. The defendant acted in the heat of passion;**
- **2. The defendant was (suddenly provoked by <insert name of decedent>/ [or] suddenly drawn into combat by <insert name of decedent>);**

– **3 - 7 indicate other, non-emotional conditions.**

[II] A person acts in the heat of passion when he or she is provoked into doing a rash act under the influence of intense emotion that obscures his or her reasoning or judgment. The provocation must be sufficient to have caused a person of average disposition to act rashly and without due deliberation, that is, from passion rather than from judgment.

[III] Heat of passion does not require anger, rage, or any specific emotion. It can be any violent or intense emotion that causes a person to act without due deliberation and reflection.

[IV] In order for the killing to be excused on this basis, the defendant must have acted under the direct and immediate influence of provocation as I have defined it. While no specific type of provocation is required, slight or remote provocation is not sufficient. Sufficient provocation may occur over a short or long period of time.

[V] It is not enough that the defendant simply was provoked. The defendant is not allowed to set up (his/her) own standard of conduct. You must decide whether the defendant was provoked and whether the provocation was sufficient. In deciding whether the provocation was sufficient, consider whether a person of average disposition would have been provoked and how such a person would react in the same situation knowing the same facts.

[VI] The People have the burden of proving beyond a reasonable doubt that the killing was not excused. If the People have not met this burden, you must find the defendant not guilty of (murder/ [or] manslaughter).

The instructions also provide the duty of the trial court to give the instructions, related CALCRIM instructions, authorities (penal codes, case citations, secondary sources), and related issues (distinction between excusable, voluntary, and involuntary manslaughter).

[I1.] introduces the *heat of passion* element, which is clarified (somewhat) in [II] and [III] as a violent or intense emotion that interferes with rationality. In [II] and [V], the provocation must be *sufficient* to interfere in the rationality of a *person of average disposition*. [I2.] and [IV] highlight temporal dimensions: the provocation must be *sudden* (or combative) and be temporally close to the offending action; while the temporal extent of the overall provocation is underspecified, presumably the final "trigger" provocation is sudden. In [VI], the proof standard *beyond a reasonable doubt* is use to decide whether the killing was not excused; that is, if there is some reason that the killing was excused based on the conditions, then the jury should pass down this decision.

Having presented the elements of legal reasoning we model, we turn to outline computational models of emotions.

3 Analysis of Emotions

There has been substantial research on computational modeling of emotions in agents and in modeling the concerns of others. The Ortony, Clore, and Collins (OCC) model of the emotions [15] decomposes emotions according to whether they are reactions to the consequences of events pertaining to the goals of an agent, consequences of an agent's actions, and an agent's attitude towards certain objects. One of the key ideas of the model is that the same event/action/object (EAO) may elicit different emotional responses from different agents depending upon how it impacts upon their goals, standards, or attitudes (GSA). For example, suppose two agents (i and j) are held at gunpoint and threatened; agent i may feel fearful whereas agent j may feel angry. Furthermore, the emotional intensity of the emotion may vary according to the settings of several sorts of parameters. *Central* variables include desirability, praiseworthiness, and appealingness; they pertain to the intensity of emotions regarding events, actions, and objects respectively. *Global* variables, reality, proximity, unexpectedness, arousal, effect every emotion type: with *sense of reality*, the issue is whether the eliciting EAO actually occurred or was a hypothetical situation; *proximity* relates to how temporally close the EAO prompt is; *unexpectedness* bears on whether the agent was surprised or not with the EAO; and *arousal* expresses the degree to which the agent is attentive prior to and during the EAO. Finally, *local* variables are specific to one emotion type, for example, *likelihood* is associated with the emotion types *hope* and *fear*. Each variable has a value and weight that determines whether the emotion is triggered (the emotional threshold has been attained) and at what intensity. Emotions and their intensities also have rates of *decay* [18]. To determine whether a particular emotion holds or not of an agent, each of the values of the variables must be given, then input to calculate the values for *intensity*, *threshold*, and rate of *decay*.

[16,17] refine and formalize the OCC model in an agent specification language, introducing a logical language and its semantics. For our purposes, such a representation provides the *terms* that can be used in argumentation schemes to *justify* emotions. Models of agents emotional states can be modeled in knowledge bases. For example, *fear* occurs when an agent i with plan π believes that certain constituent parts of π, e.g. K, may not be achieved, resulting in a failure to execute the overall plan. This is formulated as: $fear_i\ (\pi, \neg K)$. Clearly, if any portion of the representation fails to hold, *fear* does not hold for that agents.

For our purposes, it is not only necessary to represent the emotions of individual agents, but also to be able to model the emotional representations of others, particularly the defendant and the abstract *person of average disposition* referred to in CALCRIM No. 511 since these are compared in giving a decision. [8] extends the OCC model to model and reason about the concerns of others (COO), including the emotions of other agents. Agents build and maintain databases of COOs and use them to reason deductively and abductively about the emotions of other agents in the environment. In [8], agents possess *interpretative* and *manifestative* personalities. The interpretive personality is used to generate an emotion from a certain situation by referring to the goals, standards, and preferences (GSP) of an agent. The manifestative personality is used to generate an action in accordance with the emotion generated. The two are used in conjunction in order to allow an agent to make an explanatory inference with respect to another agent. For an agent to model how another agent will behave it needs some understanding of both these personalities. In addition, [8] introduce the idea of satellite COO's, which are models that one agent has of another agent's models of others, e.g. *what I think you think of others (perhaps including me).* Such COOs may also used for *hypothetical reasoning* as in *how would I feel in such a situation?*, which could then be used to predict the behaviour of a stranger. In addition to the GSP of individual agents, we can have a system-wide GSP which sets a standard and can be consider to be the GSP of a the abstract *person of average disposition*.

4 Argumentation Schemes

Argumentation schemes describe normative, presumptive, defeasible reasoning patterns [21], that is, they describe patterns of how certain reasoning patterns do and should appear, how the conclusions are presumed to follow from the premises, and how the reasoning can be defeated in various ways. They cover a broad spectrum reasoning, including what is often referred to as *fallacious* argumentation, arguments which can be shown to be false in terms of reasoning or in light of additional facts or growth of information.

One example argument pattern is *Argument from Distress.*

Premise 1: Agent x is in distress (is suffering).
Premise 2: Agent y's bringing A will relieve or help to relieve this distress.
Conclusion: Agent y ought to bring about A.

There are various objections one might make about this argument: x is not in distress; even if y brings about A, it will not relieve this distress; it is not possible for y to bring about A; or, there are negative side effects to bringing about A that preclude bringing it about. If one agrees with one or more of these objections, then the presumptive conclusion does not hold, and the argument is defeated. The objections might, in a dialogue, be cast as questions such as *Is it the case that x is in distress?*, where the negative answer introduces the objection, while the positive answer upholds the presumptive conclusion.

In this scheme, the emotional term *distress* appears among the premises; that is, we do not have an argument for *distress*, where a statement such as *Agent x is in distress* is the conclusion of an argument which follows from some specified premises. While there are many other argumentation schemes that have emotional terms among their premises, e.g. Threat, Fear Appeal, Danger, Need for Help, and Distress [21], we know of no schemes for emotional conclusions, where the emotion statement is the conclusion of the argument rather than a premise; in other words, we have yet to presumptively argue *for* an emotion.

In a legal setting, as outlined in section 2.2, determining whether the emotion *normatively* and *plausibly* holds or not is crucial to the legal decision. Not only must the premises be supported with reports and evidence from the defendant and witnesses, but also a COO must be constructed for that emotion that represents the *person of average disposition.* The emotional models for both the defendant and the COO for the person of average disposition are compared. It may, in addition, be argued that the defendant and COO models must be relative (e.g. child, psychologically abnormal, unusual circumstance, etc), subclassing the person of average disposition relative to the defendant's class. It is also worth noting as an aside that argumentation schemes with such emotional terms among their premises may also be considered *rhetorical* schemes which are used to persuade others. For example, Argument from Distress might be used as an argument by a prosecuting attorney that the jury ought to make some particular decision in a case. As part of this, the attorney would construct a COO model of the individual bearing the distress. Alternatively, in an Argument from Fear Appeal, the jury members' own concerns might be offered as a reason for making a decision, thus requiring the prosecutor to model the jurors' hypothetical concerns.

Another important scheme in [21] for our purposes is the abductive *Backward Argumentation Scheme*, which allows reasoning from data to the most plausible hypothesis.

Premise 1: D is a set of data or supposed facts in a case.
Premise 2: Each one of a set of accounts A_1,..., A_n explains D.
Premise 3: A_i is the account that explains D most successfully.
Conclusion: A_i is the most plausible hypothesis in the case.

This is particularly useful in a legal setting where from known facts and several candidate theories, we reason to a plausible hypothesis, from which some legal decision will follow. Emotional conclusions may appear as parts of the accounts. For example, given as a fact that a perpetrator murdered a victim, the particular emotional context of the act may be significant in the legal judgment. If the best account for the murder includes a significant negative, shocking event which might (in the person of average disposition) induce emotional distress (even where this is not claimed by the defendant), this might be a mitigating factor in the judgment, deciding in favour of excusable homicide; alternatively, if no such abductive argument to an emotional state can be made, the absence of an emotion might be an aggravating factor. There are a range of objections one can raise for abductive arguments concerning the facts, the accounts for the facts, the success ranking, etc..

While argumentation schemes for emotions have not been discussed in the literature, the role of emotions in the *course* of arguing has been. In [19,20], fallacious arguments are conversational moves that, while appearing to contribute to the purpose of a conversation, interfere with it. In this view, emotional arguments have an adjunct status: "good" emotional arguments can be used to direct an agent towards a prudent course of action to achieve a desired goal, while "poor" emotional arguments can detract from it. Thus, normatively, one should only use good and avoid fallacious argument forms. While there are argumentation schemes with emotional content, the emphasis is on filtering "poor" arguments from the otherwise "rational" discussion rather than reasoning with them.

[14] integrates the OCC model into a decision-making model that uses an action formalism with the *Practical Reasoning Argumentation Scheme* [2], argumentation frameworks, and value-based argumentation [5]. In this analysis, emotions play an adjunct role of influencing an agent's decision-making with respect to what course of action to follow; emotions can increase or decrease the priority given to alternative value rankings, thereby influencing the choice of action.

5 Emotional Argumentation Schemes for CALCRIM No. 511

As outlined in section 4, emotions in the context of argumentation have been regarded as unargued for premises or as adjuncts in reasoning. However, as claimed in [4,11], emotions have a direct role in argumentation in terms of how we normatively analyze and evaluate emotional appeals. In [13] it is argued that emotions themselves should be viewed as *objects* of argumentation rather than serving only to enhance the persuasiveness of reasoning in an argument. Thus, rather than filtering out or subordinating to rational argument, emotions can be *first class* citizens of argumentation. In addition, we see the main advantage of introducing emotions as first class citizens of argumentation schemes is that we can then *argue about the emotions*, which is what occurs in legal contexts.

We introduce *emotional argumentation schemes*, where emotion terms are the conclusions of argumentation schemes and follow from premises which are given by the OCC. Thus, as with other defeasible arguments, we can argue for or against emotional arguments. These emotional conclusions may then serve as premises of other arguments such as *Argument from Distress* or as components of such premises as in the abductive argumentation schemes.

As we do not have the space in this paper to give analyses of all possible emotional argumentation schemes, we provide one illustrative example which represents the elements taken from the full analysis of the emotion *anger* in the OCC. We have been concerned to represent the key clauses of CALCRIM No. 511, particularly:

- Heat of passion.
- Sufficient provocation.
- Sudden provocation.
- Temporal proximity between provocation and offending action.
- Beyond reasonable doubt.

The objective of reasoning about these elements is to determine whether or not the defendant was irrational at the time of committing the offending action. Our strategy has been to identify sub-arguments which form a tree of justification, linking conclusions of one argument with premises of another till we conclude with the root of the whole argument. In these schemes, the root conclusion is

Agent x was irrational at the time of doing action a3, which is because Agent x was in the heat of a passion that interfered with rationality.

We relate the schemes here to the OCC in that we take into consideration the concepts and relationships the OCC uses to explain emotions. The OCC has formulae which calculate, from the values of several variables, the values of other variables, e.g. intensity; in addition, there are complex issues about *decay* rates. For our purposes, we do not provide a full analysis, including arguments and formulae, for all these elements. In addition, the OCC and related work analyse a spectrum of emotions in a range of degrees, while we are only interested in creating arguments relevant to CALCRIM No. 511. In the following, premises are introduced which would themselves require further argumentation and eventual grounding in some base model of the emotions (for related treatments of argumentation and semantic models see [22,3]).

The schemes we introduce below would be used in several different ways: *forward* or *backwards/abductive* inference; comparing the emotional states and actions of the defendant to those of the person of average disposition. The comparison may give rise to further schemes and objections, which we do not introduce here.

As we are providing defeasible argumentation schemes, used in context where knowledge is partial or perhaps inconsistent, there may be a variety of ways to defeat the arguments: one may object directly that some premise (or the conclusion) is false, giving the premises from which this objection presumptively follows; one may object that while a premise is not false, it is insufficiently supported, then providing an argument with the selected premise as conclusion, but the argument itself has a falsifiable premise; one may argue that the scheme is inapplicable in a particular circumstance; one might cite exceptions which hold, so the presumptive conclusion does not obviously follow. We leave implicit these various ways of arguing against the schemes. However, these various ways to attack the scheme represent the distinct ways that the arguments can be attacked, moving closer to the goal of making such reasoning explicit and formal [9]. In a legal setting, they could be used by legal professionals to analyse the emotional arguments.

Disapproval/Blameworthy Scheme
Premise 1a: Agent y performs action a1.
Premise 1b: Action a1 highly conflicts with the standards of Agent x.
Conclusion c1: Agent x highly disapproves of Agent y's highly blameworthy action a1.

Intense Displeasure Scheme
Premise 2a: Agent y performs action a2.
Premise 2b: Agent x intensely desires goal g.
Premise 2c: Action a2 results in not g.
Conclusion c2: Agent x is intensely displeased that not g holds.

Intense Anger Scheme
Premise 3a: Agent x highly disapproves of Agent y's highly blameworthy action a1.
Premise 3b: Agent x is intensely displeased that not g holds.
Premise 3c: The action a1 which Agent y performed is action a2 which results in not g.
Conclusion c3: Agent x was intensely angry at Agent y with respect to action a1.

Emotionally Overwhelmed Scheme
Premise 4a: Agent x was intensely angry at Agent y with respect to action a1.
Premise 4b: Agent x performs action a3, which is not equal to action a1.
Premise 4c: Action a1 happened in close temporal proximity to action a3.
Premise 4d: Action a1 was sudden and highly unexpected by Agent x.
Conclusion c4: Agent x was emotionally overwhelmed while doing action a3.

Irrationality Scheme
Premise 5a: Agent x was emotionally overwhelmed while doing action a3.
Premise 5b: Being emotionally overwhelmed precludes being rational.
Conclusion c5: Agent x was irrational at the time of doing action a3.

The schemes for **Disapproval/Blameworthy** and **Intense Displeasure** are used to argue for the conclusion of **Intense Anger Scheme**. The **Emotionally Overwhelmed Scheme** uses the intense anger conclusion along with temporal proximity and suddenness to conclude that the agent is emotionally overwhelmed. The **Irrationality Scheme** uses this conclusion along with a premise about the relationship between emotionality and rationality to conclude that the agent was irrational. This last conclusion is the target required (for our purposes) for *excusable homicide* – the killer was in the *heat of passion*, so not rationally in control of (or responsible for) his actions. Of course, a range of other conditions (not given) are required as well since the killing must also be accidental. Finally, for the burden of proof to be satisfied, there ought to be no reasonable means to defeat these arguments for irrational behaviour.

A fully spelled out range of argumentation schemes would be more extensive than these several schemes and include reasoning about the various elements of the OCC, the COO, the comparison between the defendant and a person of average disposition, auxiliary supporting evidence, and reported bodily states. Nonetheless, our analysis gives a clear indication of how emotional argumentation schemes can constructed, linked to further arguments, such as the relationship between emotionality and rationality, and elaborated further. In our view, a key advantage of presenting emotional argumentation schemes is not only the explicitness and clarity, but that we can introduce objections at key points which undermine the presumptive conclusions. Such objections are key in legal arguments and reaching judgments.

6 Future Work

We propose to continue to research into the many facets of emotions in legal reasoning so that they may be better understood and used in argumentation schemes and argumentation frameworks. Achieving this would facilitate a discussion of the relevant emotions present in the case by the judge, jury and lawyers rather than dismissing them as *ad hoc* arguments. One potentially useful approach is to use the argumentation schemes we have introduced along with argumentation schemes used to argue about stories and criminal evidence [6], where emotional states of participants may be important components. One current, generic problem with argumentation schemes of [21] is that other than the premise-claim structure, they are largely unconstrained; to make a computationally satisfactory theory, some well-formedness conditions would have to be introduced.

We have here *presumed* the OCC and COO accounts of the structures of the emotions rather than providing them explicitly either as formulae or as argumentation schemes. It remains to be developed how to account for intensity, decay, and the role of *moods* which alter the parameters. Nor have we provided argumentation schemes for the spectrum of emotions. Similarly, our schemes may need to be enriched with other aspects of reasoning about the emotions that are relevant in a legal context. This said, argumentation schemes along the lines such as we have provided do seem plausible as representations of emotional arguments in legal settings; they also provide an extensible and flexible structure for further development.

7 Acknowledgements

The second author was supported by the FP7-ICT-2009-4 Programme, IMPACT Project, Grant Agreement Number 247228. The views expressed are, however, those of the authors and should not be taken as representative of the project.

References

1. Ashley, K.: Modelling Legal Argument: Reasoning with Cases and Hypotheticals. Bradford Books/MIT Press, Cambridge, MA (1990)
2. Atkinson, K., Bench-Capon, T.: Action-based alternating transition systems for arguments about action. In: AAAI'07: Proceedings of the 22nd national conference on Artificial intelligence. pp. 24–29. AAAI Press (2007)
3. Atkinson, K., Bench-Capon, T., Cartwright, D., Wyner, A.: Semantic models for policy deliberation. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law (ICAIL 2011). Pittsburgh, PA, USA (2011), to appear
4. Ben-Ze'ev, A.: Emotions and argumentation. Informal Logic 17, 1–11 (1995)
5. Bench-Capon, T.J.M.: Persuasion in practical argument using value-based argumentation frameworks. Journal of Logic and Computation 13(3), 429–448 (2003)

6. Bex, F.: Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory. Springer, Dordrecht (2011)
7. Chakraborti, N., Garland, J.: Hate Crime: Impact, Causes and Responses. Sage (2009)
8. Elllott, C., Ortony, A.: Point of view: Reasoning about the concerns of others. In: Proceedings of the Fourteen Annual Conference of Cognitive Science. pp. 809–814. Cognitive Science Society, Bloomingto, Indiana (1992)
9. Kahan, D.: The anatomy of disgust in criminal law. Michigan Law Review, Vol. 96, No. 6, 1998. Michigan Law Review, Vol. 96, No. 6, 1998.(112), 1621–1657 (1998)
10. Lakoff, G., Johnson, M.: Philosophy In The Flesh: the Embodied Mind and its Challenge to Western Thought. Basic Books (1999)
11. Manolescu, B.: A normative pragmatic perspective on appealing to emotions in argumentation. Argumentation 20, 327–343 (2006)
12. Margulies, H.S.L. (ed.): Judicial Council of California Criminal Jury Instructions (2011). LexisNexis Matthew Bender (2010)
13. Micheli, R.: Emotions as objects of argumentative constructions. Argumentation 24, 1–17 (2010)
14. Nawwab, F.S., Bench-Capon, T., Dunne, P.: Exploring the role of emotions in rational decision making. In: Baroni, P., Cerutti, F., Giacomin, M., Simari, G. (eds.) Computational Models of Argument. Proceedings of COMMA 2010. pp. 367–378. No. 216 in Frontiers in Artificial Intelligence and Applications, IOS Press, Amsterdam (2010)
15. Ortony, A., Clore, G., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press (1988)
16. Steunebrink, B., Dastani, M., Meyer, J.J.: A logic of emotions for intelligent agents. In: Holte, R., Howe, A. (eds.) Proceedings of AAAI-07. pp. 142–147. AAAI Press., Vancouver, Canada (2007)
17. Steunebrink, B., Dastani, M., Meyer, J.J.: A formal model of emotions: Integrating qualitative and quantitative aspects. In: Mali, G., Spyropoulos, C., Fakotakis, N., Avouris, N. (eds.) Proceedings of the 18th European Conference on Artificial Intelligence (ECAI'08). pp. 256–260. IOS Press., Amsterdam (2008)
18. Velásquez, J.: When robots weep: Emotional memories and decision-making. In: Proceedings of American Association for Artificial Intelligence. pp. 70–75 (1998)
19. Walton, D.: The Place of Emotion in Argument. Pennsylvania State University Press (1992)
20. Walton, D.: A Pragmatic Theory of Fallacy. The University of Alabama Press (2003)
21. Walton, D., Reed, C., Macagno, F.: Argumentation Schemes. Cambridge University Press (2008)
22. Wyner, A., Bench-Capon, T., Atkinson, K.: Formalising argumentation about legal cases. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law (ICAIL 2011). Pittsburgh, PA, USA (2011), to appear

Adapting Engagement e-mails to Users' Characteristics

Claudia López[1] and Peter Brusilovsky[1]

University of Pittsburgh, Pittsburgh PA 15260, USA,
cal95, peterb @pitt.edu,
WWW home page: http://adapt2.sis.pitt.edu/wiki/

Abstract. Although some online communities have been able to produce high quality products and to engage thousands of users, community designers usually struggle to engage new users and increase the level of contribution of current users. Some researchers have explored approaches to persuade users to collaborate. An important strand of this research area is based on sending messages to the current users and manipulate the content of the message in order to evaluate their effectiveness. Mentioning the benefits of contributing has been tested, however the results of different studies have been contradictory. One of them have reported a positive effect in the contribution rate, but other one found that mentioning benefits has depressed the level of contribution. Our hypothesis is that the effectiveness of messages may depend on other users' variables and not in the content message only. To test our hypothesis, we performed a study to evaluate the effect of messages mentioning community and personal benefits in different users' cohorts. Levels of previous participation in the system and demographic data were tested in order to explain differences in the effectiveness of this engagement strategy.

Keywords: online community, engagement mechanisms, demographic data

1 Introduction

Several well-known online communities have demonstrated the potential of producing high quality products, enable people all around the world to share content or collaborate in geographically distributed teams. However, many other online community projects have failed in engaging enough users to achieve critical mass. Researchers have explored different ways to find out what motivates users to contribute, and how to increase their levels of contribution. Previous work has been mainly focused on using messages and manipulate the message content in order to encourage people to contribute to the community. Some studies have reported the effect of mentioning the benefits of contributing as a motivator, however the results in different studies has been contradictory. Mentioning the value of contributions has increased the level of contributions in one study, but it has decreased the contribution rate in another one.

We think that these contradictory results hint that the impact of a message may be affected by users' characteristics, not just by the message content itself. Users may have different motivations to collaborate, so different strategies that match with these diverse motivations may generate more effective results. These observations motivated us to explore adaptive engagement mechanisms in online communities. Our overall goal is to explore several ways of adaptation such as adapting to demographic data of users, user knolwedge, past levels of contribution, and the navigation patterns.

This paper reports our attempt to evaluate the effectiveness of adaptation to one aspect of user demography: user cultural background. Our initial hypothesis is that the effect of appealing to private vs. community benefits may be different for users with different cultural backgrounds. For example, given the popular belief that people from Asian countries are more community-oriented, they might get more motivated to work for community goals. In contrast, people from Western countries are more concerned with personal benefits and thus could be better motivated to do work for their own goals. This popular belief has been also supported by a multinational survey in [9]. We test this hypothesis by measuring the impact of mentioning community or personal benefits to users of different cultural background, i.e., graduate students from different home countries. Our results showed, however, that the community message was more effective in general, moreover the private benefits caused more contributions in users coming from Asian countries. The level of contribution, the academic program in which the user were enrolled and in some cases the gender also generated significant differences in the level of contribution after the message.

The rest of the paper is organized as follows: Section2 will describe general background about online communities, and related work on using benefits in the content of engagement messages. Section 3 will present the study design and the system that was used as testbed; Section 4 will detail the results of the study; Section 5 will include the discussion and future work and Section 6 will present the conclusions.

2 Related Work

2.1 Background: Online Communities

The term online community was first defined by Rheingold in 1994 [16] as cultural aggregations that emerge when enough people bump into each other often enough in cyberspace. Since then, the Web has enabled geographically distributed people to socially interact and create different kinds of communities. Discussion forums (e.g. BreastCancer Forum), Question and Answers sites (e.g. Yahoo Answers and Aardvark), sharing online social networks (e.g. Facebook, YouTube, Twitter and Flickr) and online community projects (e.g. Wikipedia, ClickWorkers and Open Source Software projects) are good examples of successful online communities that have been able to congregate thousands of active users. Collaboration among these (mainly volunteer) users has enabled fast world-wide information transfer of fun videos as well as breaking news, and produced high quality products such as a well-known encyclopedia and a secure operative system (i.e. Linux).

Along with these well-known online communities, many others starting online communities were never able to take off. Only 10.3 % of the Open Source projects that have been created in SourceForge have more than three members [15]. A third part of mailing lists get inactive over a four-month period [2]. Researchers argue that these successful examples have been possible because of intuitive and insightful design decisions, but we still lack of evidence-based, scientific guidance in building and maintaining online communities [10]. Several problems challenge the survival of online communities: 1) the cold start problem: there is few users that can create content, and there is little content to attract new users; and 2) managing the community: develop commitment, encourage contributions, reduce rate of user attrition, recruit and socialize newcomers, develop leaders, regulate behavior, manage coordination [10].

Several research groups have focused their efforts on finding out ways to maintain online communities alive longer. Several strands of work have been studied such as:

- how to socialize newcomers [4],
- how to encourage commitment to the community [17, 14],
- how to encourage more contributions [1, 11], and
- understanding people motives to be engaged in an online community [19].

One of the main strands of research has focused on how to encourage contributions. The main goal is to create the required amount of content (e.g. videos in Youtube, pages in Wikipedia, code in Open Source systems) to provide benefits to the whole online community, including casual visitors. Section 2.2 will details several findings related to encouraging contributions to online communities. Simply asking by contributions is the most popular strategy. Several different ways to do so has been reported:

- broadcasting an email asking for contribution [1] or a list of needed contribution [5],
- asking to specific people to do specific tasks [1, 4],
- emphasizing uniqueness [1, 12],
- asking people who is willing to contribute [5],
- providing social information and feedback [3, 13],
- assigning people to groups [1, 6] and
- setting goals [1, 6, 18] helps to increase the positive effect.
- reduce the effort required to know what needs to be done by by task routing (i.e. recommend possible tasks to users by matching users with tasks that they are more likely to want to do) [8, 5].

2.2 Using Benefits as Motivators in Engagement Messages

In 2004, Beenen et al. [1] reported an innovative study that used social psychology knowledge to create messages asking for more contributions in MovieLens, a movie recommender site. They run two experiments to test hypothesis borrowed from different psychological theories. The first experiment tested the effect of making salient user uniqueness and mentioning the benefits of collaborating in the community. The learned lessons are that sending a message asking for contribution boosts the number of contributions, at least during one week. Salience of uniqueness encouraged more contributions and the mention of benefit depressed ratings. The authors provided a discussion about why mention to benefits didn't work. They argue that reminding other reasons to contribute may undermine intrinsic motivations, for example user may like to rate because it is fun, but not to help others so mentioning that could have a negative effect. Other possible explanation is that the population was already committed, and the message undermined their commitment by contradicting their prior beliefs regarding who get the benefits of each contribution. An additional feasible reason was that the messages were too long, thus the effort required to understand the message about benefits may have drawn users attention away.

Another study in MovieLens [13] tested the effect of displaying the value of contributions as a GUI message. The lessons were that showing the value helped to increase the contributions. They also tested the effect of different kind of value: value to self, to the whole community, to a group of similar people, and to a group of different people. The message describing the value to groups was more effective than the one mentioning the value for the whole community. People also contributed more if the benefits are for similar people than for dissimilar people.

We believe that the reason to explain this contradictory results might be related to user's characteristics and its sensitivity to the kind of benefits that were mentioned in the messages, more that to the content itself.

3 The Study

Building upon current knowledge in the effectiveness of messages to encourage contributions, this study tested the effect of sending emails with different information to users with different cultural background and different levels of participation.

3.1 The System

We used CourseAgent system and its users as testbed of our studies. CourseAgent [7] is a community-based study planning system for graduate students of the School of Information (iSchool) at the University of Pittsburgh. CourseAgent allows students to plan their studies and rate courses that they have taken reflecting workload and relevance to personal career goals. CourseAgent serves as a communication platform and a source of knowledge about the suitability of iSchool courses to specific career goals.

Membership is restricted to the iSchool graduate students only. A new account is created for each new student who is enrolled in a graduate program at the iSchool. Recently, the system started to record when the students get their degree. So, there is partial knowledge about the student status. When we started the studies there were 1256 registered users. 123 users were already graduated according to the data in the system, 517 user had unknown student status and 616 were current students.

Out of 1256 registered users, 175 users (13,9%) have added at least one taken course to their study history. This is the most popular kind of contribution, others were done by fewer users. By the volume of contributions, the most successful feature is adding course evaluationsin in respect to a specific career goal. There were 1085 contributions of this kind. These numbers show that CourseAgent is a community that is in its early stages, and that is has not achieved a high number of contributions yet.

3.2 The Study Design

The study was designed to test the impact of messages appealing to community benefit versus messages appealing to a personal benefit to the behavior of students with different cultural background. The sample was a subset of current iSchool graduate students. The cultural background of students was modelled by their home country (represented as a part of student demographic data). The impact was measured by monitoring the changes in the database (such added course ratings) and tracking user actions through the system log mechanism. The latter allowed to observe the the level of previous and current users activity in the system even for the users who havent contributed any information that is stored the system database.

The experiment manipulated the kind of message and the cohorts that received each message. A user only received one message during the study, and the users activities before and after getting the message were tracked and analyzed. Cohorts were defined according an equally distributed users home country and the level of participation in the system before the message was sent.

The first execution of the study was run during Fall 2010, when the Spring term registration period begun. The message asked users to rate 3 courses in they have taken before Fall 2010, thus all the users who have started their programs in Fall 2010 were removed from the subjects sample. The second round of emails was sent after the end of the Fall 2010 semester (but before Spring 2011 registration is finished) to users who had started their programs in Fall 2010, so they were now able to rate courses they took in their first term. The messages that were sent in these two rounds are shown in Table 1.

The study was replicated in a slightly different form with newcomers. Students whose start term was Spring 2011 received a welcome email that mentioned community or personal benefit of contribution and asked to provide career goals and courses to be taken.

In total, e-mail messages were sent to 574 users. Six students received duplicated emails because they were students in the iSchool before, but started a new program in Fall 2010 or Spring 2011 so they were considered twice in the subject selection of different executions of the study. These users were removed from the analysis.

Table 1. Example of Community Benefit and Personal Benefit Messages

Community Benefit Message
CourseAgent enables the students to receive recommendations from other students, as well as advice from faculty, regarding their course of study, workload, and relevance of courses. The usefulness of CourseAgent recommendations for the student community increases as users provide more information including courses they have taken, their career goals, and their ratings of courses. We are trying to enhance the utility of CourseAgent before Spring registration starts. Please help your fellow students by adding and rating three courses you have taken and completed in the past by November 22th. Your contribution will empower the system to better recommend courses to all of the iSchool students just in time for their Spring registration.
Private Benefit Message
CourseAgent helps you to plan your course of study wiser by keeping track of your progress towards selected career goals and by offering advice from faculty and peer students about workload and relevance of courses. The usefulness of CourseAgent increases as you provide more information about courses taken, career goals, and your ratings of courses. We are trying to provide the best support for you before you start your Spring registration. To help us with that, please add and rate three courses you have taken and completed in the past by November 22th. Providing three course ratings by November 22th will help the system to present you a more complete picture of your progress (through the Career Scope tab) and better recommend you relevant courses just in time for your Spring registration.

The students who received these messages came from 30 different home countries to pursue their graduate degrees in the iSchool. Note that in our context, the home country is not just a country of birth, but a country where students lived and studied at least until finishing their high school. Moreover, with just a few exceptions, home country is also the country where iSchool graduate students received their undergraduate degree. As a result, in this context, student home country was used as reasonable indication of students cultural background. For this study, 6 groups of countries were defined considering their geographic and cultural similarities, and the number of iSchool students who came from those countries. The categories were defined as follows:

– Undefined: Students whose home country was not available at the moment of the study.
– United States: Students whose home country is United States.
– Asia: Students whose home country is China (PRC), Taiwan, Republic of Korea, Japan, or Thailand.
– India: Students whose home country is India.
– Middle East: Students whose home country is Islamic Republic of Iran, Turkey, Saudi Arabia, Kuwait, or Egypt.

– Others: Students whose home country is Mexico, Libyan Arab Jamahiriya, Trinidad y Tobago, Puerto Rico, Slovakia, Singapore, Nepal, Viet Nam, Canada, Chile, Russian Federation, Poland, Ukraine, Afghanistan, Uganda, Niger, Netherlands, Bangladesh, or Yugoslavia.

4 The Results

As a result of the study, 32 out of 568 message receivers used the system within one week after receiving the encouragement message (0.056%): 18 students who received the community benefit message and 14 who received the personal benefit message. Table 2 shows a detailed description of the results by country category. In our analysis of engagement we distinguished *contributions* (i.e. adding taken or planned courses and evaluation registration of courses) and *actions* that included both contributory actions and exploratory actions such as navigation through pages. Contributions add new information to the "community wisdom" and can measure the community-benefitial part of user engagement while the total volume of actions measures overall user engagement into working with the system. As the table shows, overall, the community message generated more actions in the system and more contributions.

Table 2. Number of Engaged Users and Level of Activity

		# Messages		# Engaged Users		# Actions		# Countributions	
		Message		Message		Message		Message	
	Total	Comm.	Pers.	Comm.	Pers.	Comm.	Pers.	Comm.	Pers.
Unknown	56	33	23	0	0	0	0	0	0
Asia	66	27	39	2	4	91	78	42	50
India	18	9	9	0	1	0	12	0	11
Middle East	11	6	5	1	1	8	8	3	7
Other	12	5	7	0	0	0	0	0	0
US	405	205	200	15	8	234	119	108	54
Total	568	285	283	18	14	333	217	153	122

The goal of the study was to test if the community benefit message could be more effective in people from Asian countries, and the personal benefit message more effective when sent to students from Western countries. Table 3 compares the numbers related to these two specific cohorts. To our surprize, bottom-level data showed the opposite effect - community benefit message engaged more users and produced more contributions among US students while personal benefit message engaged more Asian students and produced more contributions. However, a detailed analysis of the level of actions does not produce a clear picture. Asian users who received the community message executed more actions and contributed more to the system than Asian students who received the private message. US users provided a similar level of contribution and action when receiving the community benefit and the personal benefit message. A factorial logistic regression was run considering country category and kind of message as factors, and the fact of visiting the site within a week as the dependent variable. Although it seems that community message was able to engage more US and the personal benefit message engaged more Asian users, the predictor model using these factors didnt fit significantly better than the null model. However, the study results were still able to show significant differences in more specific cases that will be described below.

Table 3. Ratio of Engaged Users in US and Asia

% Engaged Users		
	US	Asia
Community Benefit Message	15/205 (0.073%)	2/27 (0.074%)
Personal Benefit Message	8/200 (0.04%)	4/39 (0.103%)
Mean Action Rate per Engaged User		
	US	Asia
Community Benefit Message	234/15 = 15.6	91/2 = 45.5
Personal Benefit Message	119/8 = 14.875	78/4 = 19.5
Mean Contribution Rate per Engaged User		
	US	Asia
Community Benefit Message	108/15 = 7.2	42/2 = 21
Personal Benefit Message	54/8 = 6.75	50/4 = 12.5
Mean Evaluation Rate per Engaged User		
	US	Asia
Community Benefit Message	31/15 = 2.07	8/2 = 4
Personal Benefit Message	16/8 = 2	26/4 = 6.5

Since the number of contributions and actions do not behave normally according to the normality tests, non-parametric tests were used to assess the significance of the difference of mean number of actions in different cohorts. All of the following reported results are based on non-parametric tests.

Table 4 illustrates the figures related to engaged users only. Asian students executed more actions in the system than US students for both kind of messages ($p<.049$), however the difference regarding number of contribution was not significant. Furthermore, users who have already contributed to the system provide significantly more contributions than newcomers (i.e. this includes current students who hasn't used the system before as well as new students - "No, but new" in the Table) ($p<.003$).

Table 5 shows the mean number of actions executed for users with different characteristics and the significance of mean differences considering the whole sample, not only engaged users. The number of actions executed for users who received the community benefit message was higher the number of actions done by those who received private value. The number of contributions was also higher,

Table 4. Users' Variables and Activity Level per Engaged User

Variable	Values	Action Mean	SD[1]	Sig.	Contribution Mean	SD	Sig.
Home Country	Asia	28.17	8.64	p<.049	15.33	6.048	p <.145
	US	15.35	2.452		7.04	1.576	
User has visited information before	No	20.15	3.840	p<.113	10.50	3.305	p <.003
	Yes	18.33	4.485		15.00	3.512	
	No, but new	10.22	1.234		2.22	.969	

however these differences were not significant. Since 53 out of 568 emails were sent to new students, and 9 of them were finally engaged in using the system (7 community message and 2 personal message). The mean actions of this sample is much higher than the other 2 cohorts: current student who haven't visited the system and those who have visited the system before (p<.001). This can be explained by the information needs of new students. New students usually require to get as much information as possible to make decisions, however most of them are recently arriving to the city so they do not have enough social contacts to get all the required information. The system offers them easy to access information about courses, and they spent most of the time looking for data in the system. However, they contribute less than current students. They do not have enough knowledge about courses to share, so their navigation pattern is more focused on browsing than contributing. Users who have contributed before to the system also contributed more after the message (p<.001). This can be related to the perception that the new time investment for contributing is low due to they have already invested time time in the system before. They just need to partially update their profiles in order to get the benefits. On the other side, newcomers have to invest more time in the system to achieve the same benefits.

Regarding the students who received the community benefit message, only the previous fact of lurking or contributing to the system were factors with a statistically significant differences in the level of activity. However, as the results have suggested before, the mean number of actions and contributions are higher than those computed when considering both messages. Table 6 shows these figures.

The analogous analysis for students who received the personal benefit message was executed, and the the fact of contributing to the system before is the only factor that is significant in this case. See Table 7 for a detailed description of the data.

5 Discussion and Future Work

Our hypothesis that community benefit message will be more effective with Asian students and the personal benefit will engage more US students was not confirmed. Unexpectedly, we found that the message appealing to the community

Table 5. Users' Variables After Receiving a Message

Variable	Values	Action Mean	SD	Sig.	Contribution Mean	SD	Sig.
Kind of Message	Community	1.17	.358	p <.468	.54	.198	p<.996
	Personal	.77	.250		.43	.149	
User has visited the system before	No	.89	.86	p<.001	.46	.142	p<.106
	Yes	.86	.518		.70	.422	
	No, but new	1.80	.589		.39	.202	
User has contributed information before	No	.77	.216	p<.005	.36	.122	p<.001
	Yes	2.12	.801		1.22	.453	
Home Country	Asia	2.56	1.253	p<.339	1.39	.745	p<.183
	India	.67	.667	all	.61	.611	all
	Middle East	1.45	.976	p<.249	.91	.667	p<.130
	US	.87	.223	US vs. Asia	.40	.119	US vs. Asia
Gender	Female	.79	.241	p < .128	.34	.120	p<.259
	Male	1.51	.486		.84	.292	

Table 6. Users' Variables and Activity Level After Receiving a Community Benefit Message

Variable	Values	Action Mean	SD	Sig.	Contribution Mean	SD	Sig.
User has visited the system before	No	1.08	.428	p<.000	.57	.240	p <.658
	Yes	.55	.552		.41	.414	
	No, but new	2.50	.892		.39	.269	
User has contributed information before	No	1.06	.360	p<.823	.43	.195	p <.000
	Yes	1.12	.801		1.14	.725	
Home Country	Asia	3.37	2.678	p<.917	1.56	1.518	p <.094
	US	1.14	.349		.53	.190	

Table 7. Users' Variables and Activity Level After Receiving the Personal Benefit Message

Variable	Values	Action Mean	SD	Sig.	Contribution Mean	SD	Sig.
User has contributed information before	No	.77	.216	p<.466	.36	.122	p <.000
	Yes	2.12	.801		1.22	.453	
Home Country	Asia	2.00	1.063	p<.663	1.38	.717	p <.093
	US	.60	.276		.27	.143	

benefit message engaged more US students than the personal one, however the personal message engaged more Asian students. Although these differences were not significant, the pattern is surprising and we plan to continue replicating the study to verify it. We did found, however, one significant difference related to the demography: Asian users executed significantly more actions and added more contributions in the system than the US students without regard of the kind of message they receive.

At the same time, we found a few important differences related not to user demograpy, but to their past experience and status in the system. Most importantly, users who have contributed to the system before contributed significantly more that the newcomers in the system. We think this is due to the fact that these users need to invest less time to improve their user profiles and get the benefits of the system. On the other side, newcomers can be discouraged by the fact that they to create their profile before getting personalized recommendations, so they quickly decide to stop contributing and start looking for useful information that can be obtained without a complete user profile.

Being a new student is also a significant factor of the number of actions to be executed in the system. Regarding the entire samples (not only engaged users), new students execute significantly more actions than the other cohorts. However, they do not contribute more than the others. We believe that this reflects an information seeking behavior. As new students they probably lack if information as well as social contacts within the iSchool, so the system offers them a way to explore information that they might need. However, they do not have enough information to share yet. We see this as an opportunity. We think that engaging new students might be easier that re-engaging those that have already decide not to use the system.

6 Conclusion

Online community designers usually struggle to encourage users to contribute enough content to make the site sustainable. One of the most common engagement mechanisms is to send messages to current users asking for contributions. Previous research has used the salience of benefits in the message as a motivator, however this has produced contradictory results in different studies. In this paper, we proposed that the difference could be explained by users' characteristics more than the message itself. We designed an experimental study to test the effectiveness of messages mentioning benefit and personal benefits of contributing in different cohorts. The subjects were assigned to different cohorts according to their home country and level of contributions in the past. We reported the results of the execution and replications of this study in an online community. Our original hypothesis that community benefit message would be more effective in Asian users, and the personal benefit message more effective in US users was not confirmed (in fact, the observed trend was rather opposite). Moreover, we were not able to register almost no reliable differences in actions and contributions when dividing students by demography. The only exception is the larger volume of actions pefrormed by Asian students. However, even this observation may not be considered reliable since the overal number of engaged Asian students was low.

At the same time, we discovered that the student status in the system (new or past user) and overal level of activity (active or passive users) appeare to be more reliable factors to predict student behavior. The fact of being a newcomer in the system, having contributed before to the system or being a new student are the most significant factors that predict the level of contribution that the messages generated.

While we are still interested to explore the value of demographic factors in personalizing engagement stragegies, we want to shift main focus of our work to adapting the engagement messages to the level of participation in the system. Another venue of research will evaluate the survival rates of the subjects of this study considering factors such as the kind of message they received and their navigation patterns. The ultimate goal is to propose adaptive engagement mechanisms as a way to increase the effectiveness of the engagement strategies.

References

1. Beenen, G., Ling, K., Wang, X., Chang, K., Frankowski, D., Resnick, P., Kraut, R.E.: Using social psychology to motivate contributions to online communities. In: Proceedings of the 2004 ACM conference on Computer supported cooperative work. pp. 212–221. CSCW '04, ACM, New York, NY, USA (2004)
2. Butler, B.: The dynamics of cyberspace: Examining and modeling online social structure, chap. When is a Group not a Group An Empirical Examination of Metaphors for Online Social Structure. Graduate School of Industrial Administration Pittsburgh: Carnegie-Mellon (1999)
3. Chen, Y., Harper, F.M., Konstan, J.A., Li, S.X.: Social comparisons and contributions to online communities: A field experiment on movielens. In: Computational Social Systems and the Internet (2007)
4. Choi, B., Alexander, K., Kraut, R.E., Levine, J.M.: Socialization tactics in wikipedia and their effects. In: Proceedings of the 2010 ACM conference on Computer supported cooperative work. pp. 107–116. CSCW '10, ACM, New York, NY, USA (2010)
5. Cosley, D., Frankowski, D., Terveen, L., Riedl, J.: Suggestbot: using intelligent task routing to help people find work in wikipedia. In: Proceedings of the 12th international conference on Intelligent user interfaces. pp. 32–41. IUI '07, ACM, New York, NY, USA (2007)
6. Drenner, S., Sen, S., Terveen, L.: Crafting the initial user experience to achieve community goals. In: Proceedings of the 2008 ACM conference on Recommender systems. pp. 187–194. RecSys '08, ACM, New York, NY, USA (2008)
7. Farzan, R., Brusilovsky, P.: Encouraging user participation in a course recommender system: An impact on user behavior. Comput. Hum. Behav. 27, 276–284 (January 2011), http://dx.doi.org/10.1016/j.chb.2010.08.005
8. Harper, F.M., Frankowski, D., Drenner, S., Ren, Y., Kiesler, S., Terveen, L., Kraut, R., Riedl, J.: Talk amongst yourselves: inviting users to participate in online conversations. In: Proceedings of the 12th international conference on Intelligent user interfaces. pp. 62–71. IUI '07, ACM, New York, NY, USA (2007)

9. Hofstede, G., McCrae, R.R.: Personality and culture revisited: Linking traits and dimensions of culture. CrossCultural Research 38, 52–88 (2010)
10. Kraut, R., Maher, M.L., Olson, J., Malone, T.W., Pirolli, P., Thomas, J.C.: Scientific foundations: A case for technology- mediated social- participation theory. Computer 43, 22–28 (November 2010)
11. Kraut, R.E., Resnick, P.: Evidence-based social design: Mining the social sciences to build online communities, chap. Encouraging Contribution to Online Communities. Cambridge, MA: MIT Press. (2011, to appear)
12. Ludford, P.J., Cosley, D., Frankowski, D., Terveen, L.: Think different: increasing online community participation using uniqueness and group dissimilarity. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 631–638. CHI '04, ACM, New York, NY, USA (2004)
13. Rashid, A.M., Ling, K., Tassone, R.D., Resnick, P., Kraut, R., Riedl, J.: Motivating participation by displaying the value of contribution. In: Proceedings of the SIGCHI conference on Human Factors in computing systems. pp. 955–958. CHI '06, ACM, New York, NY, USA (2006)
14. Ren, Y., Kraut, R., Kiesler, S., Resnick, P.: Evidence-based social design: Mining the social sciences to build online communities, chap. Encouraging commitment in Online Communities. Cambridge, MA: MIT Press. (2011, to appear)
15. Resnick, P., Janney, A., Buis, L.R., Richardson, C.R.: Starting an online community on demand: A case study of adding forums to a physical activity promotion program. Journal of Medical Internet Research (2010)
16. Rheingold, H.: A slice of life in my virtual community, pp. 57–80. MIT Press, Cambridge, MA, USA (1994)
17. Sassenberg, K.: Common bond and common identity groups on the internet: Attachment and normative behavior in on-topic and off-topic chats. Group Dynamics 6(1), 27 – 37 (2002)
18. Wash, R., MacKie-Mason, J.: Using a minimum threshold to motivate contributions to social computing (2009)
19. Wasko, M.M., Faraj, S.: Why should i share? examining social capital and knowledge contribution in electronic networks of practice. MIS Quarterly 29 (2005)

Discrediting moves in political debates

Isabella Poggi, Francesca D'Errico, Laura Vincze

DSE – Università Roma Tre
poggi@uniroma3.it; fderrico@uniroma3.it; laura.vincze@gmail.com

Abstract. The paper analyzes the move of discrediting the opponent as a means to persuasion in political debates. After analysis of a corpus of political debates, a typology of discrediting strategies is outlined, distinguished in terms of three criteria: the target – the feature of the opponent specifically attacked (dominance, competence, benevolence); the route through which it is attacked – topic, mode or directly the person; and the type of communicative act that conveys the attack (insult, criticism, correction…). The relevance of body signals in discrediting moves is highlighted.

Keywords: persuasion, fallacies, discrediting strategies, multimodality

1 Discrediting the opponent as a persuasive move

In persuasion an Agent A wants to convince an Agent B to pursue some goal GA by convincing B that GA is a subgoal to achieve a goal GB that B has [1]: the politician A may assure he will reduce taxes to convince B that voting A is a subgoal to his goal of paying lower taxes. But in persuading B, not only A's good reasons are of use: also the very person of A him/herself convinces B. In Aristotle's words, the audience is persuaded not only by *logos* and *pathos*, i.e., by rational argumentation and the appeal to emotions, but also by *ethos*, the character of the Persuader. In fact, we are not only persuaded by what people say, but what people *are*.

This is why the persuader not only has to induce positive evaluations of the goal s/he proposes to pursue, but also a positive evaluation of him/herself. Symmetrically to this, when argumentation takes place with an opponent C, like in a discussion or a debate, the persuader must also induce negative evaluation not only of the goals and arguments proposed by C, but also of the opponent C him/herself: in other words, convincing B to pursue the goal proposed by A may imply to discredit C.

This strategy has been called "*ad hominem* fallacy" in classical rhetoric and in argumentation theory. It is a "technique of argument used to attack someone's argument by raising questions about that person's character or personal situation" [2; p.140], and it "has the form: "My opponent here is a bad person, therefore you (the audience) should not accept his argument"" [3]. The Pragma-dialectic perspective [4], considers *ad hominem* as fallacious since it violates the "freedom rule", according to which participants in a discussion must be free to provide arguments without fearing of being attacked. In this attack, both Walton [2; 3] and van Eemeren [4] include issues of morality as well as expertise: [2] talks of "cutting down one's opponent by casting doubt on his [one's opponent's] expertise, intelligence, character, or good faith" (p.111), by portraying him "as stupid, unreliable, inconsistent, or biased" (p. 110). Walton [2] mainly speaks of "bad character for veracity, or bad moral character generally" (p. 140), but observes that "bad character" in political arguments does not necessarily imply a “moral” judgement; it may entail deficiency in some other qualities needed for the best candidate [3; p.115], like, for example, being a strong leader. Real debates are full of this sort of “arguments”, through which a participant may discredit the opponent. In this paper we define the notion of “discrediting move” during a debate, we analyze some cases of discrediting moves in Italian, Swiss and French political debates, and outline a first typology of them according to a model of social evaluation in terms of goals and beliefs.

2 Attacking the other’s face

We define discredit as the spoiling of another person’s image. According to a goal and belief view of mind and social action [5], a person’s *image* is the set of evaluative and non-evaluative beliefs that a person A conceives of person B. An *evaluation* is a belief about whether and how much some object, event, person have or give you the power to achieve some goal [6]. Persons are evaluated positively or negatively against several criteria (several goals) – ugly or handsome, selfish or altruistic, just or unjust, stupid or intelligent, honest or unethical – and to have a positive image (to be evaluated well against a number of criteria) becomes a permanent goal for people, since the image others have of you determines the type of relationships others want to entertain with you. Further, there are two kinds of negative evaluations: one of inadequacy, if you lack the power necessary for some goals; and one of noxiousness, if you are endowed with power, but a negative power that risks of thwarting someone’s goals. So an elector may not vote for leader C because, despite his honesty and moral integrity, he is not very smart in his political strategy (evaluation of lack of power); or else he may not vote for A because, though being very smart, he is not honest or abuses of his power (negative evaluation of noxiousness). But these two kinds of evaluation are both necessary for trust: to trust a person, I must assume 1. that s/he has a benevolent attitude toward me – s/he is willing to act for my good, she does not want to hurt me, and in her attempts to persuade me, she is not trying to cheat – and 2. that s/he is a competent person, one who has the necessary skills and knowledge to plan and to predict outcomes of actions, who has updated knowledge about the topics to decide upon, and so forth.

Now, while looking competent and benevolent may be sufficient conditions in everyday persuasion, in political discourse the orator, besides exhibiting benevolence and competence, must also show dominance. If a candidate tries to persuade me to vote for him, I will check not only his benevolence (towards my category of electors) nor only his competence in politics or economy, but also take into account how strong and effective he looks in carrying out his goals. Thus, the image a politician should project – at least for some types of electors – is also one of a dominant person: he must be totally devoid of features of lack of power.

From this it stems our hypothesis that, to lower the likeliness for electors to vote one’s opponent, one should attack the opponent not only on the image of benevolence and competence, but also on that of dominance.

Of course an aspect linked to dominance is how charismatic the opponent is; in the classical definition by Weber the charismatic leader “is treated as endowed with supernatural, superhuman, or at least specifically exceptional powers and qualities" [7]; so charisma is a more complex construct which includes also the so called dimension "emotional identification" [8] of the followers with the leader. Moreover, recently Williams and colleagues [9] emphasized the role, in the emerging of charismatic leader, of contextual variables, like the perception of economic or political crisis, or the status of the candidate as the "incumbent" or the "challenger" within an electoral contest.

3. How to discredit others

When A wants to discredit C s/he casts doubts on C's characteristics, or definitely expresses negative evaluations about them. The negative evaluations may be both ones of noxiousness or of lack of power, according to what is the image of C that A wants to convey to the addressee(s). Discredit may be expressed verbally, or simply by body signals, or by a combination of the two. Moreover, it can be expressed in an indirect way, that is, not through the meanings of explicit signals, but through the inferences that may be implied by them. To find out cases of "casting discredit" on the opponent in political debates, we ran an exploratory study.

3.1 Method

To find examples of discrediting moves we analyzed fourteen video-recorded political debates, among which one in the presidential campaign of 2007 in France and the others during Italian election campaigns in 2008 and 2011, for a total of 150 minutes of debate. After an overview of these debates, 46 fragments were selected in which a debater discredits another. The fragments were transcribed, analyzed and classified by two expert independent coders. For some fragments only the verbal communication was transcribed, while for those in which the discrediting move was mainly performed through body signals, communication in all relevant modalities was annotated. An annotation scheme was constructed to the purpose according to the principles of the "musical score of multimodal communication" [18].
Compared to other schemes like Allwood et al. [10], Kendon, [11] McNeill [12], Kipp [13] and Ekman & Friesen [14], our scheme, beside describing each signal in terms of its physical features (say, handshape, location, orientation and movement of a gesture) aims at classifying it through attributing it a specific meaning. Speech and its parallel body signals are described in terms of their parameters (for gestures, handshape, location, orientation and movement, and the expressivity parameters of temporal extent, spatial extent, fluidity, power and repetition; for gaze, direction of the eyes, eyebrows and eyelids position and movements; for mouth, position of chin and lip corners...). Then, based on the assumption that any signal, by definition, conveys a meaning that can be translated in words, and beside its literal meaning, may imply a further (indirect) meaning, to be understood by the Addressee through inference, each verbal or body signal is attributed a literal and possibly an indirect meaning, expressed through a verbal paraphrase. Based on these meanings, a typology of

discrediting moves was built up. Let us provide an example of annotation (Table 1). In this fragment Marco Travaglio, a left wing journalist, is talking of the numerous indictments of the right wing premier Berlusconi, and Elisabetta Casellati, an under-secretary of Berlusconi's government, trying to demonstrate that the leader of her party is not alone in having pending indictments in many trials, alludes to trials (for defamation) in which Travaglio has been condemned.

Table 1. An annotation scheme of discrediting acts

1. Time stamp Sender	2. Speech	3. Meaning	4. Body signals	5. Meaning	6. Indirect meaning	7. Discrediting Strategy
1. Travaglio 7.35	*(I miei processi) non riguardano*	(My own trials) do not concern	**Prosody:** *raising intonation*	I am going to explain precisely		
2. Travaglio 7.38	*prostituzione minorile,*	child prostitution,	**Prosody***:* *Stress on* ***ri*** *and* ***na*** **Gaze:** *Eyes open wide and eyebrows raised* **Gesture:** *Right thumb up*	I am scanning words → I emphasize these words Number one of a list	I am explaining very clearly → you are stupid Berlusconi's indictments are more serious than mine I count them →They are numerous	Negative evaluation: stupid
3. Travaglio 8.00	*riguardano degli articoli scritti su un giornale*	they concern some articles written on a newspaper	**Prosody:** *singsong intonation* **Gesture:** *Right hand palm to Hearer, thumb and index in precision pick, moves to right as if writing*	I remind this as a poem to learn by heart I iconically depict what "written" means	You should learn this once for all → you are like a pupil I explain very clearly → you are stupid	Negative evaluation: stupid

(1) Travaglio replies: *"…Facciamo una puntata sui miei processi, che non riguardano […] prostituzione mino**rì**le, corruzione di testi**mò**ne, concussione della que**stu**ra, frode fiscale per centinaia di milioni di **è**uro […] riguardano degli articoli scritti sul giornale che non sono piaciuti a qual**cù**no, soprattutto perché ho criticato qual**cù**no".*

"Let us have a talk show about my trials, that do not concern child prostitution, witness corruption, police bribery, tax fiddle for hundreds of millions euros […]; they concern some articles written on a newspaper that someone did not like, mainly because I have criticized someone".

Travaglio at time 7.35 (Col.1) says *"che non riguardano"* ([my trials] that do not concern) (col.2 -3), with a raising intonation (4) meaning his sentence is not finished, and he is going to explain more (5). Then – 7.38 – he says *"prostituzione minorile"* (child prostitution) by stressing the tonic vowel of the adjective "*mino**rì**le*". This is the first item of a list of four (Berlusconi's) indictments, and to stress it Travaglio, in correspondence with this item (and later with each of the next three items), opens eyes wide and raises his eyebrows, while counting on fingers (here raising his thumb, col. 4) to convey he is making a list (5). The indirect meaning conveyed (col.6) is that Berlusconi's indictments are more and more serious than his own. After finishing the list of Berlusconi's misdeeds, Travaglio says: *riguardano degli articoli scritti su un giornale* ([my trials] concern some articles written on a newspaper), and while uttering this he moves his hand, palm to Interlocutor, with joint thumb and index, rightward (col.4), iconically depicting the action of writing (5). But the very fact of using a very clear representation, even, an iconic one that might be addressed to small children, indirectly implies his interlocutor (Casellati) is stupid (col.6). The multimodal communication (speech and other signals) analyzed in columns 2 – 6 is finally classified in col.7 in terms of the typology of discrediting moves described in Section 4. In Table 1, the multimodal communication of both lines 1-2 and line 3 is classified (col. 7, lines 2 and 3) as a discrediting move that points at the opponent's lack of competence.

3.2 Results. Types of discrediting moves

A qualitative analysis of our data allowed us to distinguish various types of discrediting moves in terms of various criteria. A first criterion is whether the Sender directly attacks the other person, or more indirectly attacks the person through criticizing what s/he did or said, or the way s/he is presently behaving. Another criterion is the target of the attack, i.e., the characteristic of the person that is subject to negative evaluation. A third criterion is the type of speech act (or, considering nonverbal signals, the type of communicative act) specifically used to discredit – an insult, a trick question, an ironic statement, an allusion, an insinuation – which might be an open list. In what follows we illustrate some types of discrediting moves in terms of the former two criteria.

3.2.1 The person, the topic, the mode. The final goal of any discrediting move is to spoil the image of the opponent; yet, this can be done in a direct way by attacking the person herself (Person), but also indirectly. For instance, you can express a very negative evaluation of what that person is saying (Topic), either by denying that what she says is true, or providing correction, clarification, or more precise information. Or finally, you can attack the way in which the opponent is conducting her argumentation during the debate (Mode).

A. The person. Insulting is a clear way to convey discredit by directly attacking a person. Being generally caused by a personal or moral attack, the insult tends to be considered an emotional reaction of anger, a "loss of control" (which conveys a negative evaluation of the insulter). Of course, the insult is a hard blow, since the insulter has the goal not only to communicate to others how bad or dangerous that person is, but also to communicate to the person him/herself one's intention of offending him, that is, to publicly spoil his/her image [15]. But this can be a strategic move to persuade by damaging the opponent's image, hence, as Schopenhauer puts it, a last weapon in a contest. Here is an example.

(2) Alessandro Sallusti, a right wing journalist, is discussing with Massimo D'Alema, a left wing leader, about the right wing minister Claudio Scajola, who has received an expensive house paid by a building company to corrupt him. Sallusti compares Scajola to D'Alema, saying that he too is a "privileged" politician because he lived for a long time in a popular apartment paying a very cheap rent. D'Alema tries to demonstrate that the comparison with the minister under investigation is wrong, and when Sallusti goes on provoking he replies: *Ma vada a farsi fottere. Lei è un bugiardo e un mascalzone* (Get screwed! You are a liar and a scoundrel!) with very high vocal intensity, tilting head forward and staring at the opponent with eyebrows raised.

A more indirect way to attack a person is to attack someone close to him/her (according to the principle that "bad company corrupts"). In an example from our corpus, Travaglio, who is politically close to the party of a former judge, Antonio Di Pietro, is debating with Daniele Capezzone, one of the spokesmen of Berlusconi. They both skip the contents of the debate and start accusing each other by trying to demonstrate that the reciprocal political referents are immoral or unfair; in this case Travaglio accuses Capezzone of being at the service of a politician who takes advantage of his legislative power to escape from court trials.

(3) Travaglio, to Capezzone: *Stai al servizio di uno che si fa le leggi per farsi assolvere*!
(You are at the service of one who makes laws in order to be discharged)

In this sentence, Travaglio emphasizes the *relation* between Capezzone and Berlusconi: the former is "at the service" of the latter: a slave of a cheater. Such discrediting communicative act entails a triple attack: (1) Capezzone has a relation of slave-master with his leader, (2) Capezzone is a slave, and (3) his master is a cheater.

B. The topic. A way to indirectly attack the person is to judge his or her action. Here is a case of this strategy.

(4) Giuliano Pisapia and Letizia Moratti are running for being elected mayor of Milan. A few days before the vote Letizia Moratti launches an accusation to Pisapia of having

been charged, 20 years before, of stealing a car, and then discharged under an amnesty. She does that at the very end of the debate, when Pisapia has no time to reply.

Moratti with her sentence casts a very negative evaluation over Pisapia through mentioning a past action of him.

C. The mode. A somewhat indirect way to discredit the opponent is to highlight his blameworthy behavior during the debate.

(5) In example (4), Moratti's accusation to Pisapia is a calumny, because Pisapia in that case actually had given up the amnesty, he had chosen to be put on trial and had been acquitted. Thus Pisapia, while Moratti is still speaking, just before the debate is over, only has the time to say: *"E' calunnia, questa"* (this is calumny).

Here Pisapia retorts the negative evaluation over the opponent by remarking Moratti's unfair behaviour during the debate. In other less serious cases the Speaker remarks the opponent's unfair floor management, in order to describe him as more generally unfair. During a harsh debate between Marco Travaglio and Daniele Capezzone, the discussion is not focused on what they are talking about but on the way they discuss: Travaglio in fact by teasing the opponent wants to demonstrate that being a right wing politician means not to let the other express his opinion, by continuously interrupting and overlapping on his discourse:

(6) Travaglio, addressing Capezzone, says: *Ora mi metto a fare come te, guarda, mi iscrivo al partito liberale e ti parlo sopra*; ("Now I start doing like you, I join the liberal party and speak over you"). And when Lilli Gruber, the Moderator, tells him that this way nobody can understand, he adds: "lo so che non capisce niente nessuno! Per questo loro fanno cosi" (I do know that no one can understand anything: this is why they do so!"). While saying he will join the liberal party, Travaglio performs an *asymmetrical* (then simulated) *smile*, he *puts his right arm on his hip* and *sways bust* as if provoking and defying the opponent, in an amused pose that unmasks his ironic intent.

3.2.2. Ends and means of discrediting moves: attacked features and communicative acts. Attacks to the person during a political debate can be distinguished in terms of two criteria: the *target feature*, i.e., the characteristic of the opponent subject to negative evaluation, and the *speech act* (or if not only verbal, the *communicative act*) through which the negative feature is highlighted; in fact, the same target may be pointed out by different verbal or nonverbal acts. Here we only incidentally take note of the specific communicative acts performed (written in small capitals), while we mainly focus on the types of target features.

The feature that is subject to attack may concern all three aspects of the opponent's image: Competence, Dominance, and Benevolence, with the former two being subject to negative evaluations of lack of power, and the third to ones of noxiousness. On the *Competence* side, one may cast doubts on the opponent being *ignorant* or *stupid*, on the *Dominance* side, concerning his/her being *helpless*, *ridiculous* or *inconsequential*, and on the *Benevolence* side, concerning his being *immoral*, *dishonest* or *cheating*. As shown in the following examples, various communicative acts may be exploited to point at these features.

A.1. Competence – Ignorant

(7) Travaglio is criticizing the premier Silvio Berlusconi, and Elisabetta Casellati, an under-secretary of his government, to cast doubts on what Travaglio is saying, reminds that he has been condemned twice for defamation. He then replies: *"Se la sottosegretaria conoscesse la differenza che c'è fra il reato di opinione e lo scrivere il falso…"* ("if the under-secretary knew the difference between a thought crime and writing the false…").

Here Travaglio implies that Casellati does not even know the difference between thought crime and defamation, thus performing an INSINUATION about her ignorance on legal issues.

One more example. During a debate before the president election in 2007, Ségolène Royal aims at showing that Nicholas Sarkozy does not have precise and updated knowledge concerning nuclear energy, and to do so adopts, in both verbal and body behavior, a didactic attitude while talking to him.

(8) Nicolas Sarkozy: *Vous confirmez l'EPR?*
Ségolène Royal: *Non. Je suspends l'EPR dès que je suis élue.*
S: *C'est-à-dire vous suspendez les nouvelles centrales et vous prolongez les vieilles.*
R: *Mais l'EPR n'est pas une nouvelle, n'est pas une centrale* (she *frowns*, as if meaning « I am annoyed by the silly things you say »).
S: *Si. Bien sûr.*
R: *Vous mélangez tout. L'EPR c'est un prototype* (she *frowns*, and *tilts her head back*, expressing superiority and thus remarking her correction)
S: *Non madame.*
R: *L'EPR est un prototype de quelle génération?* (she *leans across the table* towards Sarkozy and *points at him with her index finger*, as a teacher asking the pupil a question)
S: *Ce n'est pas un prototype. C'est la quatrième génération.*
R: *Non, c'est la troisième génération.*

Nicolas Sarkozy: *Do you confirm the EPR*[1]*?*
Ségolène Royal: *No, I intend to suspend the EPR as soon as I am elected.*
S : *That is to say that you suspend the new (nuclear) centrals and you prolong the old ones.*
R: *But the EPR is not a new, not a nuclear plant* (she *frowns*, as if meaning « I am annoyed by the silly things you say »).
S: *Yes, of course.*
R: *You are mixing everything up. The EPR is a prototype* (she *frowns*, and *tilts her head back*, expressing superiority and thus remarking her correction)
S: *No, madame.*
R: *The EPR is a prototype of which generation?* (she *leans across the table* towards Sarkozy and *points at him with her index finger*, as a teacher asking the pupil a question)
S: *It is not a prototype, it's the fourth generation.*
R: *No, it is the third generation.*

[1] *EPR* (European Pressurised Reactor)

In this example Royal highlights Sarkozy's ignorance about nuclear plants, first by a speech act of CORRECTION (« EPR is not a new, not a nuclear plant »), then REMARKING what his error is (« You are mixing everything up »). Then she puts a TRICK QUESTION (« EPR is a prototype of which generation ? ») displaying a body behavior typical of a teacher with her pupil, to unmask Sarkozy's ignorance, and finally REMARKS his wrong answer (« No, it's third generation »).

A.2. Competence – Stupid. Sometimes, a Speaker implies that the opponent is not so smart as to understand some subtle but important differences. A typical way to imply an image of stupidity is a *didactic attitude*, that is well exemplified by example (1) above, in which, to reply to Casellati, Travaglio says: "Let us have a talk show about my trials, that do not concern child prostitution, witness corruption, police bribery, tax fiddle for hundreds of millions euros […]; they concern some articles written on a newspaper that someone did not like, mainly because I have criticized someone".

He says so while using a singsong intonation, and stressing the last tonic syllable of each item of the list, both with a higher pitch and by raising his eyebrows and opening his eyes wide, much like when talking to small children or teaching pupils. While listing Berlusconi's pending indictments, he numbers them with gestures (one, two, three). Finally when saying that his own trials concerned "articles written on the newspapers", with his right hand, thumb and index touching, palm to interlocutor, he draws circles in the air moving from left to right: the iconic gesture for "writing".

In this passage, by his words, Travaglio attacks Casellati on the content of what she said – the trials undergone by him as concerning only his "thought crimes", not the much more serious misdeeds charged to Berlusconi. But by the very way he says so – his recurrent stress and recurrent intonation, his iconic gesture, all concurring to the general form of a very clear and didactic explanation – he implies that Casellati needs such an explanation since she cannot see the difference between Travaglio's and Berlusconi's trials. In other words, he is treating her as a stupid person, thus discrediting her image in an indirect way (I am didactic → you are stupid) and only through his body behavior.

B.1. Dominance – Helplessness. A negative image that a Speaker in a debate may cast on the opponent regarding aspects of dominance is an image of helplessness.

(9) La Russa often takes the floor by interrupting Di Pietro, his opponent, also despite the intervention of the Moderator, Bianca Berlinguer. When La Russa interrupts once more, Berlinguer says: *Però adesso lo faccia finire*. (But now let him finish), and La Russa says *Ma sìì…* (but suuure….), with *raised eyebrows* and *closed eyelids*. Both the way he says *sìì (suuure)* and his gaze expression convey haughtiness, and indulgence, thus implying that Di Pietro is a poor thing who cannot intimidate anybody.

(10) In another passage in which Berlinguer defends Di Pietro from his interruptions, La Russa says: *Ma povero, poverino* (Oh poor, poor thing!), with a voice quality typical of one who pulls a long face of disappointment.

(11) The Moderator is interviewing Margherita Hack, a famous old Italian scholar in astrophysics, who is talking against Berlusconi and the "ad personam" laws that he

made to save himself from trials. Roberto Formigoni, a politician on Berlusconi's side, while looking at her, shows an *asymmetrical smile*, with *left lip corner raised*, and *oblique eyebrows*, expressing ironic compassion.

La Russa and Formigoni discredit the other's image of dominance implying his or her helplessness, and do so by showing ironic compassion, in (9) by prosody and gaze, in (10) by words and voice, in (11) by gaze and smile.

B.2. Dominance – Ridicule. In some cases, finally, a dramatic lack of power is attributed to the opponent through ridicule. *Ridiculization* is the act of remarking a negative evaluation of lack of power in someone who, unlike one who deserves compassion, has a pretence of superiority; the contrast between pretended superiority ad actual inferiority results in a violated expectation that is, though, not threatening [16; 17]. Thus the person made fun of is abased, not even being credited with the power of being feared. Here is a case of ridiculization.

(12) Travaglio has previously ironically called Berlusconi's spokesmen *trumpeters*, and later has reported that, as Berlusconi himself declared, he had "payed" 45.000 euros to a girl to save her from prostitution. Casellati has shown disappointment by his using a "rude" language, and Travaglio has replied by reminding her of some very dirty jokes publicly told by Berlusconi. Later, while talking of Berlusconi's justification for his donation of 45.000 euros to a supposed prostitute, he says:
Prendiamo atto che il Presidente del Consiglio è il redentore di queste ragazze e quindi le pagava per toglierle dalla strada, per toglierle dal marciapiede, per fargli aprire […] un centro estetico con una macchina costosissima per la depilazione, ***mi scuso per la parola depilazione****, ma l'ha usata lui.*

"We take note that the Prime Minister is the redeemer of these girls and therefore he used to pay them to save her from prostitution, to let them open […] a beauty center with a very expensive machine for hair removal, *I apologize for the term "hair removal"*, but it was him who used it".

Travaglio's IRONIC APOLOGY for his using the term "hair removal" remarks how ridicule (and hypocritical) Casellati's prudishness is, since she was shocked for a mild ironic word ("trumpeters") previously used by Travaglio, while her chief violates all the rules of linguistic politeness.

B.3. Dominance – Inconsequential. An even worse image of lack of power is one of an inconsequential person, one who only elicits indifference. A typical move to imply a such image in the opponent is to *diminish the other*, for instance by addressing her not by her institutional allocution, *"sottosegretaria"* (under-secretary) but as a simple woman, *"signora"* ("madam"); or by pretending not to remember the opponent's name and mispronouncing it ostentatiously. Another move is to disregard the opponent's individuality, like Travaglio does when he refers to his opponent Capezzone as yet another spokesmen of Berlusconi, by saying: *"Un altro replicante. Li sfornano a raffica, li sfornano a raffica"* ("One more replicant. They churn them out at full blast"). Or finally, you can ostentatiously ignore the other.

C.1. Benevolence – Immoral. In Aristotle's speculation, benevolence is "rendering a service to one in need, not in return for something or benefit for the subject, but in order to benefit the other person.". Extending this definition to political area and persuasion, a benevolent politician is one full of civil values, taking care of "the other" as opposed to selfish interests. In this sense a discrediting move focused on a benevolence target will tend to show how that politician doesn't care about the citizens' interests by reminding for example his immoral or unfair past, or his malevolent wrongs certified by judicial decisions or by public stigmatization.

(13) The left wing politician Massimo D'Alema is talking of the right wing minister Scajola, who had to resign for corruption: he received a house facing Colosseum as a gift from a building company. To counter attack, Sallusti, the director of a pro-government newspaper, remarks that also a left wing politician like D'Alema is not flawless from a moral point of view, since he lived for long time in a popular apartment paying a very cheap rent.

Sallusti: *L'onorevole d'Alema credo che possa darci lezione a noi e al paese, lo dico sinceramente, su tanti temi ma non sulla casa. Il moralismo del Presidente D'Alema sulla casa […] è inaccettabile. Lei non si era accorto che pagava un decimo del valore di mercato. Tant'è vero che se n'è andato, presidente. Sulle case lei non può… Da un punto di vista etico-morale. Gli operai pagavano tre volte di quello che pagava lei. Lei è un privilegiato. Lei si era accorto che pagava poco rispetto al mercato?* […]

D'Alema: *Io non pagavo troppo poco. Io pagavo quello che era previsto dalla legge.*

Sallusti: *E allora perché se n'è andato, scusi? Da un punto di vista etico-morale lei ha approfittato della sua posizione!*

Sallusti: *MP D'Alema I think can lecture us and the Country, I say this sincerely, on a lot of topics, but not on houses. President D'Alema's moralism on houses […] is unacceptable. You did not realize you were paying one tenth of the commercial value. In fact you left, president. Concerning houses you cannot… From the ethical-moral point of view. Workers used to pay three times as much as what you paid. You are a privileged person. Had you realized that you paid a low price with respect to the commercial value?* […]

D'Alema: *I did not pay too low a price, I paid the price stated by the law.*

Sallusti: *Then, I beg your pardon, why did you leave? From an ethical-moral point of view you took advantage of your position!*

Sallusti's head movements play a complementary role to words while accusing D'Alema. He reinforces his message by repeated *nods of emphasis* (and by a slow speech rhythm), but at the same time, since he is probably aware this is a serious attack to an influential politician, he performs some unexpected signals of submission, like a *head canting* and *head down looking downward*. Finally while saying "you can lecture... but not on houses" he *shakes his head* reinforcing what is creeping into words. Sallusti's move is a case of *"tu quoque"* [4]: a fallacy (very frequently used in our corpus) through which, to weaken the impact of moral criticism or accusation, one retorts the accusation of immorality on the critic.

C.2. Benevolence – Dishonest. In the same debate D'Alema provides an example of attack to the benevolence side of the opponent, casting an image of (at least intellectual) dishonesty over Sallusti.

(14) D'Alema: *Io capisco che la pagano per venire qui e fare il difensore d'ufficio del governo. Lo capisco, lo capisco, capisco che si deve guadagnare lo stipendio a proposito di etica ma dicendo mascalzonate non si guadagna lo stipendio Lei è pagato dal giornale della famiglia Berlusconi, le daranno un premio. Io capisco che deve guadagnarsi il pane ma questo modo è vergognoso.*

D'Alema: I can understand that you are paid to come here to be the public defender of the government. I understand it, I understand it, I understand that you have to earn your salary concerning ethics, but by saying knaveries one does not earn one's salary. You are paid by the newspaper of Berlusconi family, you will be rewarded. I understand that you have to earn your bread, but this way is shameful.

C.3. Benevolence – Cheater. A last very important negative evaluation concerning the benevolence side is an accusation of cheating, that often debaters launch to each other. Only two examples.

(15) Casellati: [Travaglio] *racconta sempre, ha un'attitudine a raccontare sempre delle cose che non corrispondono a verità.*
Casellati: "[Travaglio] always tells stories, he has an attitude of always telling things not corresponding to truth".

Casellati uses very polite and euphemistic words (*"tell things not corresponding to truth"*, instead of uttering the word *lies* or definitely calling Travaglio a *liar*), but by uttering the word *sempre* (always) twice she implies a steady attitude of her opponent to tell the false.

(16) Travaglio (to Capezzone): *Tu hai un padrone che ogni sera ti manda in televisione a raccontare balle!*
Travaglio. You have a boss who every night sends you on TV to tell lies!

Here the accusation of cheating is not only to the actual opponent, but also to his boss.

4 Quantitative analysis of discrediting strategies

Once we distinguished various discrediting strategies, we can now see their quantitative distribution. We first calculated the inter-judge agreement between two independent judges in classifying discredit cases in terms of their *route* (mode, topic and person): K Cohen is 0.63 ($p<0.000$). Table2 shows their distribution in the debate. Overall, attacks are most often directly to the person (67%), sometimes to the topic (26%) and only in a few cases to the mode (7%).

	n.	**%**
Person	31	67
Topic	12	26
Mode	3	7
TOT.	46	100

Table 2. Discrediting moves attacking Person, Topic, Mode

A chi-square test aimed at differentiating the most frequently exploited route by rightwing vs. leftwing politicians [$\chi^2(46)=9.616$; $p<0.05$] reveals that those from the right tend to discredit more by a person route (87,5%) and less by topic (12,5%), while the leftists adopt a more complex pattern of discrediting moves, referring mainly to the person (56,50%) and the topic (33.50%) but sometimes also to the mode (10%) (Fig. 1).

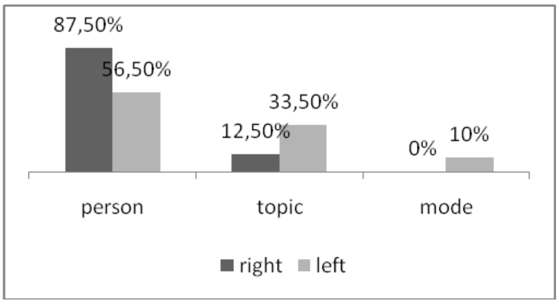

Figure 1 Political orientation* route

For the *target* feature, K Cohen was of 0.82 ($p<0.000$). Fig. 2. shows the distribution across target features.

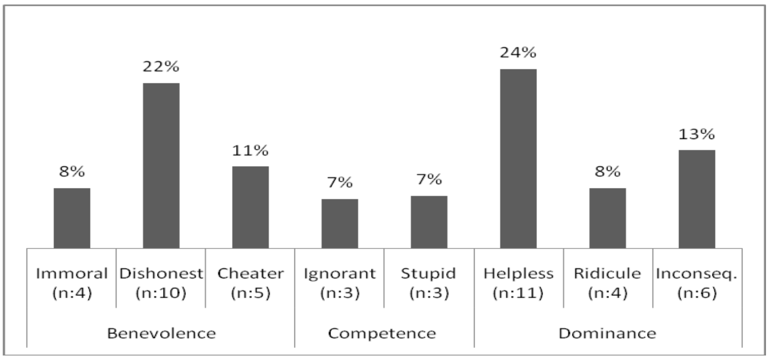

Figure 2 Discrediting moves across target features

Negative evaluations concerning the opponent's dominance features are generally the most frequent (45%), followed by benevolence (41%), while competence is more rarely addressed (14%). But let us describe results taking into account the controlling variables *political orientation* and *gender* of the politicians analyzed.

As to political orientation the videos analyzed are not balanced (30 discredits from the left vs 16 from the right). Rightwing politicians seem to choose more aggressive moves targeted on the benevolence side than the leftist ones (60% vs 32%); on the other hand, the prevailing strategy of the leftists is to discredit the opponent's dominance, for example casting ridicule on him/her or relativizing his/her power (55% vs 33%) (Figure 3).

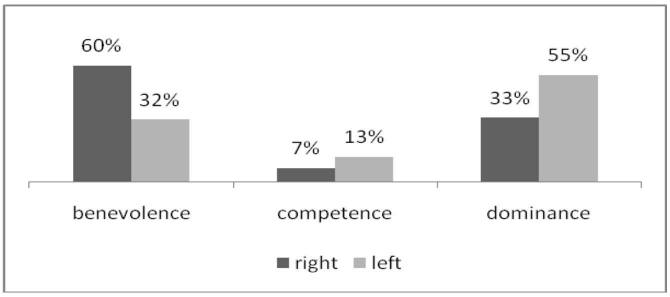

Figure 3 Political orientation*target feature

By analyzing in detail the types of discredits (Table 5), we can see that a historical theme of Italian Left parties – the so-called "moral issue", according to which the leftist should be more sensitive to ethical issues – is not confirmed at all by our data. It seems that it is more typical of right-wing politicians to highlight the other's "dishonest" and "cheater" behaviors (27% for both).

On the other hand, the left-wing politicians' discrediting moves tend to characterize the opponent mainly as ridicule (16%), helpless (19%) and inconsequential (17%): leftists do not seem to value the arguments of right-wing politicians so much, thus focusing on one type of discredit that neutralizes the opponent's dominance rather than attacking his/her image from an ethical or competencies point of view.

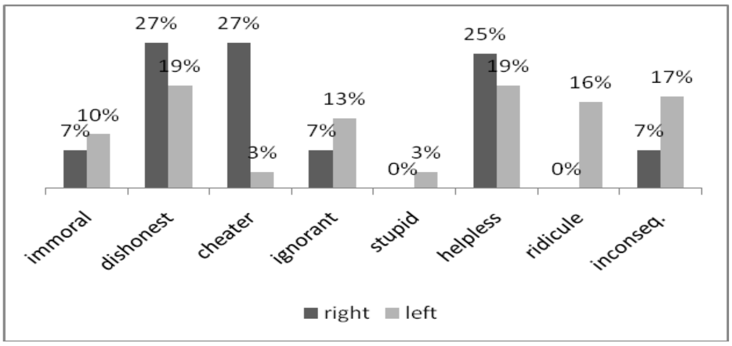

Figure 4. Political orientation*target feature in detail

Looking at the gender variable, discrediting moves based on dominance rather than benevolence seem to be a male strategy (55% vs 31%), while women look more "moralistic": they focus more on discrediting the other's benevolence (61%). In particular, as shown in Figure 5, a chi-square analysis [$\chi^2(46)=17,22$; $p<0.015$] highlights the significant differences in the strategies used by women, who typically focus on ethical negative evaluations like "cheater"(38%) and "dishonest" (23%).

Discrediting trough the dimension of dominance, in particular the prevalence of "helplessness"(24%) and "ridicule"(15%) seems to be more a male strategy.

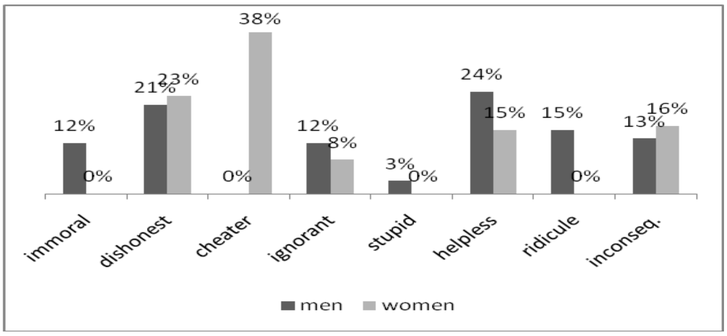

Figure 5. Gender*target features

5 Conclusion

In this work we have analyzed a move often used to persuade the audience during debates: discrediting the opponent. We have outlined a typology of discrediting strategies, distinguished in terms of three criteria: the target feature – the aspect of the opponent that is specifically attacked (dominance, competence, benevolence); the route through which it is attacked – topic, mode or directly the person; and the type of communicative act that conveys the attack (insult, criticism, correction…). From our exploratory study it resulted that in political debates the person is most often directly discredited, and mainly its features of strength and power are an object of evaluation, even more than morality. Yet, the choice of the route and the feature differ from male to females, and from right to left politicians. Many intriguing questions remain open, if only, the role of irony in discrediting, that we did not tackle here. Again, we might wonder what is the relation between frequence of a discrediting move and its seriousness: for a politician is it worse to be marked as stupid or evil, as to dishonesty or impotence? Further, are there particular words, specific syntactic or argumentative structures, or particular (combinations of) body signals that are typically used in discrediting the opponent? An overview of debates in different countries would also allow assess if there are cultural differences in the preferred discrediting strategies.

Acknowledgments. Research supported by SSPNet Seventh Framework Program, European Network of Excellence SSPNet (Social Signal Processing Network), Grant Agreement N.231287.

References

1. Poggi, I. The Goals of Persuasion, Pragmatics and Cognition 13, 298-335 (2005)

2. Walton, D., N., Types of dialogue, dialectical shifts and fallacies, In: Frans H. van Eemeren, Rob Grootendorst, J. Anthony Blair and Charles A. Willard (eds). Argumentation Illuminated, Amsterdam, SICSAT, p. 133-147 (1992)
3. Walton, D.,N., Ad Hominem Arguments. [Studies in Rhetoric and Communication], University of Alabama Press (1998)
4. Eemeren, F.H. van, Grootendorst, R. Argumentation, communication, and fallacies: A pragma-dialectical perspective. Hillsdale, NJ: Lawrence Erlbaum Associates (1992).
5. Conte, R., Castelfranchi, C. Cognitive and social action. University College, London (1995)
6. Miceli, M., Castelfranchi, C., The role of evaluation in cognition and social interaction, In: Dautenhahn, K. (ed.), Human cognition and agent technology, Amsterdam, John Benjamins (1998).
7. Weber, M. In G. Roth & C. Wittich (Eds.), Economy and society. New York, NY: Bedminister , Vol.1-3, (1968).
8. Williams, E.A, Pillai, R., Lowe, K.B., Jung, D. Herst, D. Crisis, charisma, values, and voting behavior in the 2004 Presidential election, The Leadership Quarterly, 20 (2), 70-86, (2009), Pages ISSN 1048-9843, DOI: 10.1016/j.leaqua.2009.01.002.
9. Rosenberg, A., Hirshberg, J. Charisma perception from text and speech. Speech Communication 51 640–655 (2009).
10. Allwood J., Cerrato, L., Jokinen, K., Navarretta, C. and Paggio, P.. The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. Language Resources and Evaluation, 41(3):273–287, (2007).
11. Kendon, A. (2004 a). Gesture. Visible action as utterance. Cambridge: Cambridge UniversityPress.
12. McNeill, D. Hand and Mind. Chicago: University of Chicago Press (1992).
13. Kipp, M. From Human gesture to synthetic action. In C. Pelachaud e I. Poggi (Eds.), Multimodal Communication and Context in Embodied Agents. Proceedings of the Workshop W7 at the 5th International Conference on Autonomous Agents, Montreal,Canada pp. 9-14, (2001).
14. Ekman P., & Friesen W.V. Facial Action Coding System. Palo Alto, CA: Consulting Psychologists Press (1978)..
15. Castelfranchi, C., Parisi, D. Linguaggio, conoscenze e scopi, Bologna, Il Mulino (1980)
16. Castelfranchi, C., Che figura. Emozioni e immagine sociale, Bologna, Il Mulino (1988).
17. Poggi, I., Irony, humour ad ridicule. Power, image and judicial rhetoric in an Italian political trial In: Vion R., Giacomi A. et Vargas C. (éds). La corporalité du langage : Multimodalité, discours et écriture, Hommage à Claire Maury-Rouan. Aix en Provence : Presses Universitaires de Provence (2011)
18. Poggi,, I., Mind, Hands, Face and Body. A goal and belief view of multimodal communication. Weidler Buchverlag.(2007).

Towards an embodied view of flow

Pablo Romero & Eduardo H. Calvillo-Gámez

[1]IIMAS, UNAM, México, DF
[2]Div. Nuevas Tec. De la Inf. Universidad Politécnica de SLP
pablor@unam.mx, eduardo.calvillo@gmail.com

Abstract. Flow is a psychological construct that has been used to describe states of optimal experience and intrinsic motivation. We claim that adaptions of flow made from psychological into computing studies have been done without a careful consideration of the original concept and that frequently they are the product of conceptual misunderstandings. We propose a view of flow for computing studies based on notions of phenomenology and embodied interaction and analyse the major characteristics of this concept from this embodied view.

Keywords: Flow, Human-Computer Interaction, Embodied Interaction.

1 Introduction

It can be argued that the most direct motivation to pursue any activity is the enjoyment we obtain as a result of doing it. If we know that doing a particular activity would produce an optimal positive experience, then we could consider this knowledge as a motivation to pursue the said activity. This is the case of flow. The concept of flow has been used to describe psychological states of optimal experience that are characterised by a deep concentration in the task at hand and have been associated with intrinsic motivation, skills promotion and academic excellence [1]. In computing, a number of studies of video gaming, e-shopping, web marketing and e-learning, among others areas, have reported their environments as conducive to flow and promoting positive attitudes and outcomes for users [2-4]. However, there has not been a consensus on a uniform way to conceptualise, model, operationalise and measure flow in those studies [2]. Models of flow, for example, do not agree on which characteristics of flow to include, and how they can be defined, categorised or related among them.

According to Finneran and Zhang [2], the discrepancies of those models indicate underlying problems in the conceptualisation of flow. Rather than trying to evaluate, refine, integrate or create new models of flow, what we propose in this paper is to go back and revise the way in which this concept has been adapted to the computing area. The paper has four sections. The second section briefly describes the way the concept of flow and its characteristics have been understood and adapted in computing studies and highlights inconsistencies and possible misunderstandings. The third section proposes a view of flow that addresses those inconsistencies and misunderstandings and revisits some of the characteristics of flow in terms of that view. Finally, the fourth section presents some conclusions.

2 Conceptual inconsistencies and gaps in studies of flow

Computing studies of flow have defined it in a number of different ways: as engagement and immersion in an activity [4], as absorption in a virtual space and the fading away of the physical world [5], as a playful and exploratory experience [6], and as an experience which is undertaken for its own sake [7], among other definitions. Although Csikszentmihalyi has warned against reifying flow [8], the ambiguity of the concept of flow has created a situation in which research in the area might be studying altogether different phenomena. Identifying at least a central characteristic that could be used to better model and operationalise flow would be particularly useful for future research in the area.

Computing studies of flow have adopted a set of nine characteristics associated with this construct: a balance between challenges and skills, clear goals, immediate feedback, intense concentration, merging of action and awareness, loss of self-consciousness, a sense of control, time distortion and experiencing the activity as intrinsically rewarding [7, 9-11]. Models of flow have incorporated those characteristics and tried to establish causality links among them. However there is no agreement as to what characteristics to take into account or what their dependencies are [2, 4].

Also, frequently it is unclear whether different studies understand the characteristics in the same way. Two examples that are relevant to our main discussion have to do with the definition of the challenges-skills balance and of intense concentration. The challenges-skills balance has been understood as the match between the person's skill and the challenges associated with the task. However there is disagreement on whether it refers to the potential challenge of learning and mastering the use of a digital system or of a task related with some aspect of reality other than the digital system per se [2]. Additionally, psychology research has suggested that challenges and skills might be multimodal in the sense that they are associated with the cognitive, physical and emotional parts of the person [12]. However studies and models of flow in computing have not taken this into account.

The other characteristic, intense concentration, has been defined as a narrow attention that focuses entirely in the interaction with the digital system, to the degree that users screen out irrelevant thoughts and perceptions and loose awareness of everyday life [9,13]. However this understanding is at odds with characterisations of this feature in psychology studies of flow where frequently it has been described as the opposite, an attention characterised by an expansive type of awareness [14].

Finally, studies of flow in computing have focused mainly on desktop interaction, forgetting about movement interaction, a relatively recent but promising area that includes research in tangible user interfaces, ubiquitous computing and product design [15]. The following section presents a view of flow that addresses this as well as the other issues mentioned above.

3 A view based on phenomenology and embodiment

The main reference of our approach is the embodied interaction framework of Dorish [16]. From this framework, and in general from its foundations, embodiment and phenomenology, we take four points as central for our embodied view of flow. The first is attention; a central issue for phenomenology is to be able to turn one's attention to the lived experience instead of being just inattentively immersed in it. The second is the importance of the context, the world in which people think, act and live. The third are the notions of present-at-hand and ready-to-hand; whether when performing a task, the user is directly concerned with the digital system or with any other aspect of reality. Finally the fourth is the importance of the body; within the phenomenological tradition, Maurice Merlau-Ponty [17] gives special importance to the body as the entity that makes the act of experiencing possible and to the bodily skills and knowledge that enable us to act on and experience the world. The relevance of these points for our view is described below.

Captive, effortful and effortless attention
While not attempting to provide a precise definition, we would like to highlight the fact that computing studies of flow have largely ignored a central characteristic of this construct: effortless attention. Flow has been defined as a state of deep concentration that is perceived as effortless. People perceive this experience as their attention being effortlessly carried by a current, hence the analogy with flow [18]. Under ordinary circumstances, subjective attentional effort in a task is proportional to the demands of the task, until there comes a point in which no increase in effort is possible (see figure 1a) [19]. In contrast to this effortful attention scenario, there are occasions in which paradoxically, at some point in the execution of the task, one is concentrated so thoroughly in the activity that suddenly attention seems effortless. At these moments, increased demands can be met with a sustained level of efficacy but without an increase in the perceived attentional effort (although the real level of attention is high. See figure 1b).

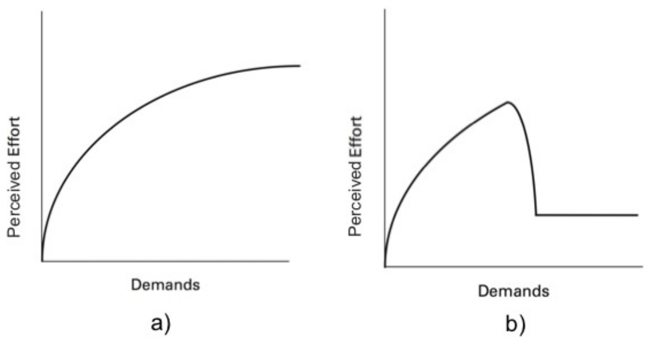

Fig. 1. Effort vs. demands in a) effortful and b) effortless attention. From Bruya (2010).

Effortful attention has been defined as people focusing and maintaining attention on specific stimuli intentionally [20]. Most of the time, however, attention is captured by external stimuli: smells, noises and images of the external world, for example. People pay attention to these stimuli without spending any effort. This captive type of attention has a passive quality to it; the person is not in control of which stimuli are attended to. In contrast, effortful attention has a more active quality to it; some effort is required to keep such focus. The fact that effortless attention occurs in contexts that should demand an effortful type of attention makes it difficult to imagine that recreational, non-goal oriented online activities like browsing aimlessly on the web [21] could promote states of flow.

The expansive type of awareness reported in the psychological flow literature could also be strongly associated with the effortless nature of attention in flow episodes. Being able to register an unusual amount of context detail means that the detail does not act as a distractor anymore. Ordinarily, attention is effortful because the tendency to attend to constant external distractors must be overridden in order to keep an intentional focus on the chosen stimuli. In flow episodes, what is usually considered as external distractors can instead be regarded as part of the activity; they do not distract anymore but are included in the experience.

In practical terms, studies and models of flow should discriminate between captive, effortful and effortless attention.

The importance of context

The view of flow we propose is in line with the emphasis of the embodied interaction framework of taking technology to the world of people. This stance contrasts with other approaches such as virtual reality, where people are the visitors in the world of computers. Our embodied view of flow instead suggests that it is digital applications that are drawn into the world of the user. We believe this is a more appropriate view because it is more in tune with the concept of effortless attention and the expansive awareness it can bring; and also because it is more suited to movement interaction research. In practical terms this view would lead to question the widely held assumption that when in flow, users are so absorbed in the task that they loose awareness of everyday life [9].

In flow with or through the system

Another important point of our embodied view of flow is related to whether the task promoting the flow experience is directly concerned with the digital system or with any other aspect of reality. In terms of an example, the task might have to do with, say, learning to use a graphics editing application or with using such an application to retouch a photograph. In the first case users will probably be concerned with analysing and reflecting about the system, while in the second they will want to achieve a task that ultimately is not about the system but that will be accomplished through the use of the graphics editing application. This differentiation is known as *present-at-hand* or *ready-to-hand* in the phenomenological terminology of Heideger [22]. When *present-at-hand*, the digital system becomes the focus of the users' attention, and depending of the task, they will explore it, learn it or analyse it, for example. When *ready-to-hand*, the system becomes a tool and, if a good quality tool, it disappears from the users' immediate concerns.

Studies of flow often are not clear about this difference; as a result they have interpreted the characteristics of flow in different ways (the challenges for example) and therefore operationalised it in dissimilar forms. According to our view of flow, studies should make clear whether the task is of a *present-at-hand* or a *ready-to-hand* nature.

We are not the first ones to notice this difference in the focus of the task; the Person-Artifact-Task (PAT) model of flow antecedents had already gone some way in clearing this confusion [23]. However in this model the task and the artifact are considered as alternatives for the users' focus of attention. This is not strictly speaking correct as users will always be engaged on a task, what is important is to clarify whether the task is of a present-at-hand or a ready-to-hand nature.

A wholesome view of interaction

In our view of flow, the body plays a central role. However the importance of the body does not lie on itself as a separate element, but "in the harmonious focusing of physical and psychic energy" [14]. Of course not all activities require full-body engagement, but the body has a critical role for any type of perception and action [17], even for using computers. Paraphrasing Bayliss [24], human-computer interaction has always consisted of embodied action, traditionally of small movements of the hands on the keyboard and mouse but embodied action nevertheless. Also, taking the body into account has a clear benefit for flow studies in movement interaction. In practical terms, studies and models of flow should take into account that the challenges and skills are multimodal composites, with physical, emotional and cognitive components.

4 Conclusion and Future Work

We have presented a view of flow that attempts to characterise this state at a conceptual level. This view is based on notions of embodied interaction and addresses some of the conceptual inconsistencies and misunderstandings in the area of computing studies of flow. The view stresses the importance of four main points: effortless attention, the importance of the context where interaction takes place, whether the task is directly concerned with the digital system or with any other aspect of reality, and the body and its role in the interaction with the system. Further work comprises developing models of flow and eventually user models that could be used to predict the probability of users reaching flow as result of using a particular application.

5 References

1. Nakamura, J. and M. Csikszentmihalyi, The concept of flow, in Handbook of positive psychology, C.R. Snyder and S.J. Lopez, Editors. 2002, Oxford University Press: Oxford, UK. p. 89-105.
2. Finneran, C.M. and P. Zhang, Flow in Computer-Mediated Environments: Promises and Challenges. Communications of the Association for Information Systems, 2005. 15(4): p. 82–101.

3. Voiskounsky, A.E., Flow Experience in Cyberspace: Current Studies and Perspectives, in Psychological Aspects of Cyberspace: Theory, Research, Applications, A. Barak, Editor 2008, Cambridge University Press: New York. p. 70-101.
4. Hoffman, D.L. and T.P. Novak, Flow Online: Lessons Learned and Future Prospects. Journal of Interactive Marketing, 2009. 23(1): p. 23-34.
5. Chen, V.H.-H., et al., Communicative behaviors and flow experience in tabletop gaming, in Proceedings of the International Conference on Advances in Computer Enterntainment Technology2009, ACM: Athens, Greece. p. 281-286.
6. Webster, J., L.K. Trevino, and L. Ryan, The dimensionality and correlates of flow in human-computer interactions. Computers in Human Behavior, 1993. 9(4): p. 411-426
7. Cowley, B., et al., Toward an understanding of flow in video games. Computers in Entertainment, 2008. 6(2): p. 1-27.
8. Csikszentmihalyi, M., A response to the Kimiecik& Stein and Jackson papers. Journal of Applied Sport Psychology, 1992. 4: p. 181-183.
9. Sweetser, P. and P. Wyeth, GameFlow: a model for evaluating player enjoyment in games. Computers in Entertainment, 2005. 3(3): p. 3-3.
10. Pace, S., A grounded theory of the flow experiences of web users. International Journal of Human-Computer Studies, 2004. 60(3): p. 327-363.
11. Voiskounsky, A.E., O.V. Mitina, and A.A. Avetisova, Playing online games: Flow experience. Psychnology Journal, 2004. 2(3): p. 259–281.
12. G.D.Ellis, J.E. Voelkl, and C. Morris, Measurement and Analysis Issues with Explanation of Variance in Daily Experiences Using the Flow Model. Journal of Leisure Research, 1994. 26(4): p. 337-356.
13. Hoffman, T.P. and D.L. Novak, Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations. Journal of Marketing, 1996. 60: p. 50-68.
14. Jackson, S.A. and M. Csikszentmihalyi, Flow in Sports1999, Champaign, IL.: Human Kinetics.
15. Larssen, A.T., et al., Introduction to the special issue on movement-based interaction. Personal Ubiquitous Computing, 2007. 11(8): p. 607-608.
16. Dourish, P., Where the Action Is: The Foundations of Embodied Interaction2001: The MIT Press.
17. Merleau-Ponty, M., Phenomenology of Perception1962, London: Routledge.
18. Csikszentmihalyi, M. and J. Nakamura, Effortless Attention in Everyday Life: A Systematic Phenomenology, in Effortless Attention, B. Bruya, Editor 2010, The MIT Press: Boston.
19. Bruya, B., Introduction: Toward a Theory of Attention That Includes Effortless Attention, in Effortless Attention: A New Perspective in the Cognitive Science of Attention and Action, B. Bruya, Editor 2010, MIT Press.
20. Schmeichel, B.J. and R.F. Baumeister, Effortful attention control, in Effortless attention: A new perspective in the cognitive science of attention and action, B. Bruya, Editor 2010, MIT Press: Cambridge, MA. p. 29-50.
21. Novak, T.P., D.L. Hoffman, and A. Duhachek, The Influence of Goal-Directed and Experiential Activities on Online Flow Experiences. Journal of Consumer Psychology, 2003. 13(1-2): p. 3-16.
22. Cerbone, D.R., Understanding Phenomenology2006, Cheham: Acumen.
23. Finneran, C.M. and P. Zhang, A Person-Artefact-Task (PAT) Model of Flow Antecedents in Computer-Mediated Environments. International Journal of Human-Computer Studies, 2003. 59(4): p. 475-496.
24. Bayliss, P. Notes towards a sense of embodied gameplay. in 2007 Digital Games Research Association Conference. 2007. Tokyo.

Addressing the New User Problem with a Personality Based User Similarity Measure

Marko Tkalčič, Matevž Kunaver, Andrej Košir, and Jurij Tasič

University of Ljubljana Faculty of electrical engineering,
Tržaška 25, 1000 Ljubljana, Sovenia
{marko.tkalcic,matevz.kunaver,andrej.kosir,jurij.tasic}@fe.uni-lj.si
http://ldos.fe.uni-lj.si

Abstract. The new user problem is a recurring problem in memory based collaborative recommender systems (MBCR). It occurs when a new user is added to the system and there are not enough information to make a good selection of the user's neighbours. As a consequence, the recommended items have poor correlation with the user's interests. We addressed the new user problem by observing the user similarity measure (USM). In this paper we present two novelties that address the new user problem : (i) the usage of a personality based USM to alleviate the new user problem and (ii) a method for establishing the boundary of the cold start period. We succesfully used a personality based USM that yielded significantly better recommender performance in the period where the new user problem occurs. Furthermore we presented a new methodology for assessing the boundary of the period where the new user problem occurs.

Keywords: memory based collaborative recommender system, new user problem, personality based user similarity measure

1 Introduction

The new user problem is an important issue in memory based collaborative recommender systems [Adomavicius and Tuzhilin, 2005]. It occurs when a new user joins the system and there are no (or there are too few) overlapping ratings to calculate good estimates of user similarities with rating-based user similarity measures (USM). We will denote this initial period as the cold start period (CSP). The consequences of being in the CSP are bad rating predictions for unseen items and thus poor quality of the recommender system. Usually, the new user problem (NUP) has been addressed by introducing content-based approaches which resulted in hybrid systems [Adomavicius and Tuzhilin, 2005, Ahn, 2008]. Once the system has enough overlapping items it is not in the CSP and rating based USM can be used.

We introduced a personality-based USM using the five factor model (FFM) in Tkalčič et al. [2009]. The same approach was later used by Hu and Pu [2010] for the NUP in a music recommender system. In this paper we present (i) the results of the proposed USM in a CF recommender system for images and (ii) a methodology for assessing the boundary of the cold start period.

The proposed approach to use a personality-based USM in the NUP allows us to calculate user similarities immediately, without waiting for the user to rate several items. The underlying assumption for choosing personality as the basis for the proposed user similarity measure is that people with similar personalities have similar tastes for products. In psychology, personality is described as a set of factors that account for the majority of between-user variance in emotive, interpersonal, experiential and attitudinal styles [John and Srivastava, 1999].

The second novelty is a statistical method for determining at which point the new user problem stops occurring. A review of literature showed that authors either (i) did not set limits for the CSP [Schein et al., 2002] or (ii) provided limits without further argumentation, e.g. Massa and Bhattacharjee [2004] defined *cold start users as the users who have expressed less than 5 ratings*. We propose to determine the boundary of the CSP with a statistical approach, as the number of ratings where the recommender's performance stops being significantly lower than the performance with higher number of ratings given by the user.

2 The new user problem

The new user problem in collaborative filtering recommenders is described as the period from the moment when a user joins the system to the moment when there are enough ratings to yield a stable list of neighbours (i.e. users with similar preferences). We rewrote this description from various sources [Adomavicius and Tuzhilin, 2005, Schein et al., 2002, Ahn, 2008]. To the best of the authors' knowledge no formal definition of the new user problem period is available.

In this section we define the boundary of the CSP. Let us have a user u joining the system. The user starts using the system and gives ratings $r(u,h)$ to items $h \in H$ where $H \subset \{h_1 \dots h_J\}$, a set of J items. At any given moment the user has given n ratings to n different items which yields the set

$$R_u^n = \{r_1^u \dots r_n^u\} \tag{1}$$

The boundary of the new user problem period (the CSP) for the selected user is the number of ratings N_u^{CS} after which the system starts to yield stable sets of users. The consequence of a stable set of users is a stable confusion matrix of recommended items. We define that the confusion matrix is stable if a sequence of F-measure values, has statistically equivalent means at different n.

We choose the F measure as a scalar measure of the confusion matrix. We denote the F measure when n ratings have been used to calculate neighbours as F^n. We define the CSP boundary as the point N where the means of F values of the sets

$$R_u^{NJ} = \{F^N \dots F^J\} \qquad R_u^{(N-1)J} = \{F^{(N-1)} \dots F^J\} \tag{2}$$

are significantly different.

In Tkalčič et al. [2009] we presented a user similarity measure that takes two vectors $\boldsymbol{b_i} = (b_{i1} \dots b_{i5})$ and $\boldsymbol{b_j} = (b_{j1} \dots b_{j5})$ containing the personality values of two users u_i and u_j and yields the scalar similarity value

$$d_W(\boldsymbol{b_i}, \boldsymbol{b_j}) = \sqrt{\sum_{l=1}^{5} w_l (b_{il} - b_{jl})^2} \tag{3}$$

We use the proposed user similarity measure to calculate similarities in the CSP.

3 Experiment

The goal of the experiment was to (i) assess the CSP boundary and to (ii) see whether the personality-based USM performs better in the CSP. We conducted two experiments with the CF recommender system: (i) one with the personality based USM and (ii) one with the rating based USM.

We used the LDOS-PerAff-1 [Tkalčič et al., 2010] dataset which contained all data necessary to carry out our experiments. The dataset provided the usage history (i.e. the log of users' interactions) of 52 users consuming 70 content items (a subset of images from the IAPS[1] dataset) and giving explicit ratings to each item. The users' task was to assess the images for their computer's wallpaper. The users' personalities vectors $\boldsymbol{b}$ were assessed using the IPIP50[2] questionnaire. We used the personality based USM as defined in Eq. 3 to calculate the distances between the users. We calculated the predicted ratings based on the neighbours' ratings using the adjusted Pearson's coefficient as defined in Kunaver et al. [2007]. We then compared the predicted ratings with the ground truth ratings which yielded the confusion matrix.

We calculated the rating based USM $d(u_i, u_j)$ between two arbitrary users u_i and u_j based on their respective ratings $e(u, h)$ of the overlapping items h_m, where m is the index of the items that both the users have rated

$$d(u_i, u_j) = \sqrt{\sum_{m} (e(u_i, h_m) - e(u_j, h_m))^2} \tag{4}$$

The dataset used in our experiments had a full ratings-items table without missing values with I users and K ratings. To simulate the CSP we determined a usage history path in the form of a random sequence of ratings, for each user separately. We iterated through cold start stages s from one (the user has given only one rating) to K (the user has rated all items) for each user separately. At each stage $1 \leq s \leq K$ we performed the recommender procedure and calculated the confusion matrix for the observed user u at the observed stage s. We chose the F measure as the performance measure of the recommender system. The

[1] `http://csea.phhp.ufl.edu/media.html`
[2] `http://ipip.ori.org/New_IPIP-50-item-scale.htm`

experimental procedure thus yielded a table of F values at different stages $s \in \{1..K\}$ and for each user $u \in U$.

3.1 Evaluation methodology

We compared the performance of the rating based USM and the personality based USM by testing the hypothesis $H_0 : \mu_R = \mu_{B5}$ at different cold start stages using the t-test. The value μ_R represents the mean F values using the rating based USM and μ_{B5} represents the mean F values using the personality based USM.

We determined the position of the CSP boundary by testing the hypothesis $H_0 : \mu_s = \mu_{s-K}$ where μ_s represents the mean F value at stage s and μ_{s-K} represents the mean F value from stages $s+1$ to K, where K is the last observed cold start stage.

4 Results

When seeking for the CSP boundary we calculated the p values which are shown in Fig. 1. On the dataset used we observed that $p < 0.05$ occurs when the cold start stage is $s < 6$.

We analyzed the CSP by graphing the quality rate of the recommender(F) versus the number of ratings used. At each cold start stage s the F measures for each user were calculated.

The results of the t test showed that, on the dataset used, the personality based USM yields a significantly higher mean of F values than the rating based USM when the number of ratings taken in account for the calculation of the neighbours is lower than 50 (see Fig. 2). When the number of ratings is higher than 50 the means of F values for both similarity measures are not significantly different at $\alpha = 0.05$.

5 Discussion and conclusion

Experimental results showed that the personality based USM performs significantly better than the rating based USM in cold start conditions. A positive outcome is also the fact that the personality USM is statistically equivalent to the rating based USM which makes it a good candidate for a complete replacement of the rating based USM.

However, the results presented here were verified only on the specific dataset and we don't have any ground to conclude that the presented approach is useful also in other domains. We do speculate that hedonistic-content domains would benefit from the presented approach but this should be verified as future work.

The main drawback of the personality based USM is the difficulty of acquisition of end users' personality parameters. There are two main obstacles in this: (i) it is annoying for the end user to fill in questionnaires and (ii) the acquisition

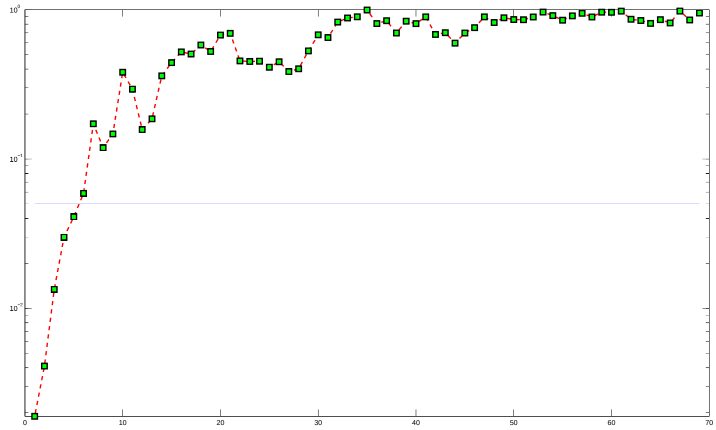

Fig. 1. p values of the t test for the CSP boundary. On the dataset used the CSP occurs when $s < 6$.

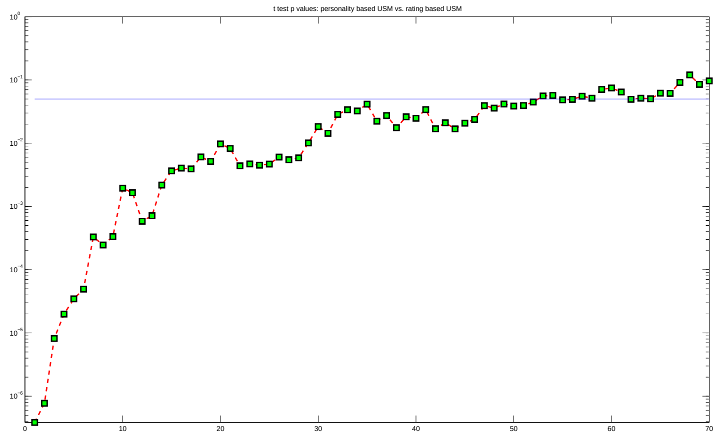

Fig. 2. p values of the t test of the comparison of the personality based USM and rating based USM.

of personality data raises ethical and privacy issues that need to be addressed first. The progress beyond the state of the art here is the knowledge that personality does account for between-users variance in entertainment applications.

In the lack of existing methodologies for assessing the boundaries of the new user problem we chose a statistical approach. We acknowledge that further investigations should be conducted to determine how to test for the CSP boundary and that these investigations might conclude that a different approach is more suitable.

We provided a methodology for the assessment of the new user boundary. The results presented should not be taken for granted and several repetitions of the procedure should be carried out on different datasets.

In this paper we have evaluated a personality based USM under cold start conditions. The results showed that the personality based USM performed significantly better than the rating based USM. Furthermore we described a methodology for the assessment of the CSP border. Both novelties are important in the field of memory based collaborative filtering recommender systems and should be further explored.

References

G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

H.J. Ahn. A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. *Information Sciences*, 178(1):37–51, 2008.

R. Hu and P. Pu. Using Personality Information in Collaborative Filtering for New Users. *Recommender Systems and the Social Web*, page 17, 2010.

Oliver P. John and Sanjay Srivastava. The big five trait taxonomy: History, measurement, and theoretical perspectives. In Lawrence A. Pervin and Oliver P. John, editors, *Handbook of Personality: Theory and Research*, pages 102–138. Guilford Press, New York, second edition, 1999. URL `http://www.uoregon.edu/\~{}sanjay/pubs/bigfive.pdf`.

Matevž Kunaver, Tomaž Požrl, Matevž Pogačnik, and Jurij Tasič. Optimisation of combined collaborative recommended systems. *International Journal of Electronic Communications*, 61:433–443, 2007.

P. Massa and B. Bhattacharjee. Using Trust in Recommender Systems: An Experimental Analysis. In *Trust management: second international conference, iTrust 2004, Oxford, UK, March 29-April 1, 2004: proceedings*, page 221. Springer-Verlag New York Inc, 2004.

A.I. Schein, A. Popescul, L.H. Ungar, and D.M. Pennock. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 253–260. ACM, 2002.

Marko Tkalčič, Matevž Kunaver, Jurij Tasič, and Andrej Košir. Personality based user similarity measure for a collaborative recommender system. In C. Peter, E. Crane, L. Axelrod, H. Agius, S. Afzal, and M. Balaam, editors, *Proceedings of the 5th Workshop on Emotion in Human-Computer Interaction - Real world challenges*, pages 30–37. Fraunhofer Verlag, September 2009.

Marko Tkalčič, Jurij Tasič, and Andrej Košir. The LDOS-PerAff-1 Corpus of Face Video Clips with Affective and Personality Metadata. In Michael Kipp, editor, *Proceedings of the LREC 2010 Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*, 2010.

Impact of implicit and explicit affective labeling on a recommender system's performance

Marko Tkalčič, Ante Odić, Andrej Košir, and Jurij Tasič

University of Ljubljana Faculty of electrical engineering,
Tržaška 25, 1000 Ljubljana, Sovenia
{marko.tkalcic,ante.odic,andrej.kosir,jurij.tasic}@fe.uni-lj.si
http://ldos.fe.uni-lj.si

Abstract. Affective labeling of multimedia content can be useful in recommender systems. In this paper we compare the effect of implicit and explicit affective labeling in an image recommender system. The implicit affective labeling method is based on an emotion detection technique that takes as input the video sequences of the users' facial expressions. It extracts Gabor low level features from the video frames and employs a kNN machine learning technique to generate affective labels in the valence-arousal-dominance space. We performed a comparative study of the performance of a content-based recommender (CBR) system for images that uses three types of metadata to model the users and the items: (i) generic metadata, (ii) explicitly acquired affective labels and (iii) implicitly acquired affective labels with the proposed methodology. The results showed that the CBR performs best when explicit labels are used. However, implicitly acquired labels yield a significantly better performance of the CBR than generic metadata while being an unobtrusive feedback tool.

Keywords: content-based recommender system, affective labeling, emotion detection, facial expressions, affective user modeling

1 Introduction

Recently, investigations, that evaluate the use of *affective metadata* (AM) in content-based recommender (CBR) systems, were carried out [Arapakis et al., 2009, Tkalčič et al., 2010a] and showed an increase of the accuracy of recommended items. This improvement of CBR systems that use affective metadata over systems that use *generic metadata* (GM), like the genre, represents the motivation for the work presented in this paper. Such systems require that the content items are labeled with affective metadata which can be done in two ways: (i) explicitly (i.e. asking the user to give an explicit affective label for the observed item) or (ii) implicitly (i.e. automatically detecting the user's emotive response).

1.1 Problem statement and proposed solution

Each of the two approaches for affective labeling, explicit and implicit, has its pros and cons. The explicit approach provides unambiguous labels but Pantic and Vinciarelli [2009] argue that the truthfulness of such labels is questionable as users can be driven by different motives (egoistic labeling, reputation-driven labeling and asocial labeling). Another drawback of the explicit labeling approach is the intrusiveness of the process. On the other hand implicit affective labeling is completely unobtrusive and harder to be cheated by the user. Unfortunately the accuracy of the algorithms that detect affective responses might be too low and thus yield ambiguous/inaccurate labels.

Given the advantages of implicit labeling over explicit there is a need to assess the impact of the low emotion detection accuracy on the performance of recommender systems.

In this paper we compare the performance of a CBR system using explicit affective labeling vs. the proposed implicit affective labeling. The baseline results of the CBR with explicit affective labeling are those published in Tkalčič et al. [2010a]. The comparative results of the implicit affective labeling are obtained using the same CBR procedure as in Tkalčič et al. [2010a], the same user interaction dataset [Tkalčič et al., 2010c] but with affective labels acquired implicitly.

1.2 Related work

As anticipated by Pantic and Vinciarelli [2009], affective labels are supposed to be useful in content retrieval applications. Work related to this paper is divided in (i) the acquisition of affective labels and (ii) the usage of affective labels.

The acquisition of explicit affective labels is usually performed through an application with a graphical user interface (GUI) where users consume the multimedia content and provide appropriate labels. An example of such an application is the one developed by Eckhardt and Picard [2009].

On the other hand, the acquisition of implicit affective labels is usually reduced to the problem of non-intrusive emotion detection. Various modalities are used, such as video of users' faces, voice or physiological sensors (heartbeat, galvanic skin response etc.) [Picard and Daily, 2005]. A good overview of such methods is given in Zeng et al. [2009]. In our work we use implicit affective labeling from videos of users' faces. Generally, the approach taken in related work in automatic detection of emotions from video clips of users' faces is composed of three stages: (i) pre-processing, (ii) low level features extraction and (iii) classification. Related work differ mostly in the last two stages. Bartlett et al. [2006], Wang and Guan [2008], Zhi and Ruan [2008] used Gabor wavelets based features for emotion detection. Beside these, which are mostly used, Zhi and Ruan [2008] report the usage of other facial features in related work: active appearance models (AAM), action units, various facial points and motion units, Haar based features and textures. Various classification schemes were used successfully in video emotion detection. Bartlett et al. [2006] employed both the Support Vector Machine (SVM) and AdaBoost classifiers. Zhi and Ruan [2008] used the k-nearest neighbours (k-NN) algorithm. Before using the classifier they performed a dimensionality reduction step using the locality preserving projection (LPP) technique. In their work, Wang and Guan [2008] compared four classifiers: the Gaussian Mixture Model (GMM), the k-NN, neural networks (NN) and Fisher's Linear Discriminant Analysis (FLDA). The latter turned out to yield the best performance. The survey Zeng et al. [2009] reports the use of other classifiers like the C4.5, Bayes Net and rule based classifiers. Joho et al. [2009] used an emotion detection techique that uses video sequences of users' face expressions to provide affective labels for video content.

Another approach is to extract affective labels directly from the content itself, without observing the users. Hanjalic and Xu [2005] used low level features extracted from the audio track of video clips to identify moments in video sequences that induce high arousal in viewers.

In contrast to emotion detection techniques the usage of affective labels for information retrieval has only recently started to gain attention. Chen et al. [2008] developed the EmoPlayer which has a similar user interface to the tool developed by Eckhardt and Picard [2009] but with a reversed functionality: it assists users to find specific scenes in a video sequence. Soleymani et al. [2009] built a collaborative filtering system that retrieves video clips based on affective queries. Similarly, but for music content, Shan et al. [2009] have developed a system that performs emotion based queries. Arapakis et al. [2009] built a complete video recommender system that detects the users' affective state and provides recommended content. Kierkels and Pun [2009] used physiological sensors (ECG and EEG) to implicitly detect the emotive responses of users. Based on implicit affective labels they observed an increase of content retrieval accuracy compared to explicit affective labels. Tkalčič et al. [2010a] have shown that the usage of affective labels significantly improves the performance of a recommender system over generic labels.

2 Affective modeling in CBR systems

2.1 Emotions during multimedia items consumption

In a multimedia consumption scenario a user is watching multimedia content. During the consumption of multimedia content (images in our case), the emotive state of a user is continuously changing between different emotive states $\epsilon_j \in E$, as different visual stimuli $h_i \in H$ induce these emotions (see Fig. 1). The facial expressions of the user are being continuously monitored by a video camera for the purpose of the automatic detection of the emotion expressions.

The detected emotion expressions of the users, along with the ratings given to the content items, can be used in two ways: (i) to model the multimedia content item (e.g. *the multimedia item* h_i *is funny - it induces laughter in most of the viewers*) and (ii) to model individual users (e.g. *the user* u *likes images that induce fear*).

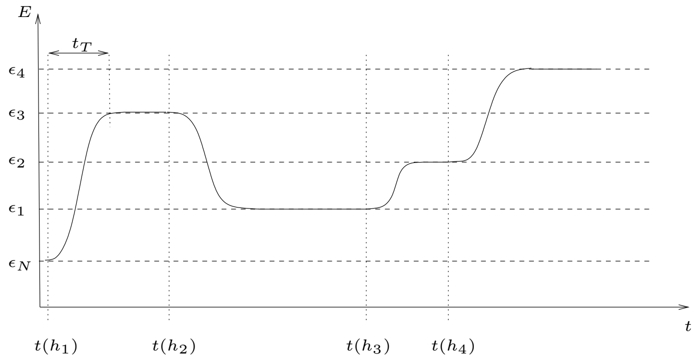

Fig. 1: The user's emotional state ϵ is continuously changing as the the time sequence of the visual stimuli $h_i \in H$ induce different emotions.

2.2 Affective modeling in a CBR system

Item modeling with affective metadata We use the valence-arousal-dominance (VAD) emotive space for describing the users' emotive reactions to images. In the VAD space each emotive state is described by three parameters, namely valence, arousal and dominance. A single user $u \in U$ consumes one or more content items (images) $h \in H$. As a consequence of the image h being a visual stimulus, the user u experiences an emotive response which we denote as $er(u, h) = (v, a, d)$ where v, a and d are scalar values that represent the valence, arousal and dominance dimensions of the emotive response er. The set of users that have watched a single item h are denoted with U_h. The emotive responses of all users U_h, that have watched the item h form the set $ER_h = \{er(u, h) : u \in U_h\}$. We model the image h with the item profile that is composed of the first two statistical moments of the VAD values from the emotive responses ER_h which yields the six tuple

$$\mathcal{V} = (\bar{v}, \sigma_v, \bar{a}, \sigma_a, \bar{d}, \sigma_d) \tag{1}$$

where $\bar{v}$, $\bar{a}$ and $\bar{d}$ represent the average VAD values and σ_v, σ_a and σ_d represent the standard deviations of the VAD values for the observed content item h. An example of the affective item profile is shown in Tab. 1.

User modeling with affective metadata The preferences of the user are modeled based on the explicit ratings that she/he has given to the consumed items. The observed user u rates each viewed item either as relevant or non-relevant. A machine learning (ML) algorithm is trained to separate relevant from non-relevant items using the affective metadata in the item profiles as features and the binary ratings (relevant/non-relevant) as classes. The user profile $up(u)$ of the observed user u is thus an ML algorithm dependent data structure. Fig. 2 shows an example of a user profile when the tree classifier C4.5 is being used.

Metadata field	Value
$\bar{v}$	3.12
σ_v	1.13
$\bar{a}$	4.76
σ_a	0.34
$\bar{d}$	6.28
σ_d	1.31

Table 1: Example of an affective item profile $\mathcal{V}$ (first two statistical moments of the induced emotion values v, a and d) .

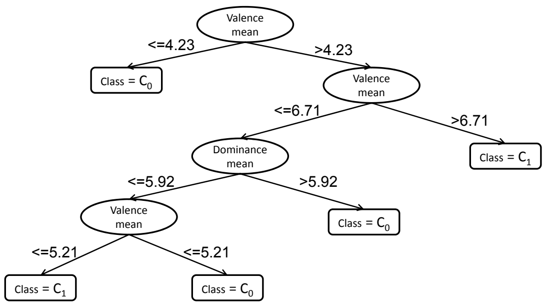

Fig. 2: Example of a user profile when the C4.5 tree classifier is used for inferring the user's preferences. The labels C_0 and C_1 represent the relevant and non-relevant classes, respectively.

3 Experiment

We used our implementation of an emotion detection algorithm (see Tkalčič et al. [2010b]) for implicit affective labeling and we compared the performance of the CBR system that uses explicit vs. implicit affective labels.

3.1 Overview of the emotion detection algorithm for implicit affective labeling

The emotion detection procedure used to give affective labels to the content images involved three stages: (i) pre-processing, (ii) low level feature extraction and (iii) emotion detection. We formalized the procedure with the mappings

$$I \rightarrow \Psi \rightarrow E \tag{2}$$

where I represents the frame from the video stream, Ψ represents the low level features corresponding to the frame I and E represents the emotion corresponding to the frame I.

In the pre-processing stage we extracted and registered the faces from the video frames to allow precise low level feature extraction. We used the eye tracker developed by Valenti et al. [2009] to extract the locations of the eyes. The detection of emotions from frames in a video stream was performed by comparing the current video frame I_t of the user's face to a neutral face expression. As the LDOS-PerAff-1 database is an ongoing video stream of users consuming different images we averaged all the frames to get the neutral frame. This method is applicable when we have a non supervised video stream of a user with different face expressions.

The low level features used in the proposed method were drawn from the images filtered by a Gabor filter bank. We used a bank of Gabor filters of 6 different orientation and 4 different spatial sub-bands which yielded a total of 24 Gabor filtered images per frame. The final feature vector had the total length of 240 elements.

The emotion detection was done by a k-NN algorithm after performing dimensionality reduction using the principal component analysis (PCA).

Each frame from the LDOS-PerAff-1 dataset was labeled with a six tuple of the induced emotion $\mathcal{V}$. The six tuple was composed of scalar values representing the first two statistical moments in the VAD space. However, for our purposes we opted for a coarser set of emotional classes $\epsilon \in E$. We divided the whole VAD space into 8 subspaces by thresholding each of the three first statistical moments $\bar{v}$, $\bar{a}$ and $\bar{d}$. We thus gained 8 rough classes. Among these, only 6 classes actually contained at least one item so we reduced the emotion detection problem to a classification into 6 distinct classes problem as shown in Tab. 2.

				centroid values		
class E	$\bar{v}$	$\bar{a}$	$\bar{d}$	v	a	d
ϵ_1	$\bar{v} > 0$	$\bar{a} < 0$	$\bar{d} < 0$	0.5	−0.5	−0.5
ϵ_2	$\bar{v} < 0$	$\bar{a} > 0$	$\bar{d} < 0$	−0.5	0.5	−0.5
ϵ_3	$\bar{v} > 0$	$\bar{a} > 0$	$\bar{d} < 0$	0.5	0.5	−0.5
ϵ_4	$\bar{v} < 0$	$\bar{a} < 0$	$\bar{d} > 0$	−0.5	−0.5	0.5
ϵ_5	$\bar{v} > 0$	$\bar{a} < 0$	$\bar{d} > 0$	0.5	−0.5	0.5
ϵ_6	$\bar{v} > 0$	$\bar{a} > 0$	$\bar{d} > 0$	0.5	0.5	0.5

Table 2: Division of the continuous VAD space into six distinct classes $E = \{\epsilon_1 \dots \epsilon_6\}$ with the respective centroid values.

3.2 Overview of the CBR procedure

Our scenario consisted in showing end users a set of still color images while observing their facial expressions with a camera. These videos were used for implicit affective labeling. The users were also asked to give explicit binary ratings to the images. They were instructed to select images for their computer wallpapers. The task of the recommender system was to select the relevant items for each user as accurate as possible. This task falls in the category *find all good items* for the recommender systems' tasks taxonomy proposed by Herlocker et al. [2004].

Figure 3 shows the overview of the CBR experimental setup. After we collected the ratings and calculated the affective labels for the item profiles, we trained the user profiles with four different machine learning algorithms: the SVM, NaiveBayes, AdaBoost and C4.5. We split the dataset in the train and test sets using the ten-fold cross validation technique. We then performed ten training/classifying iterations which yielded the confusion matrices that we used to assess the performance of the CBR system.

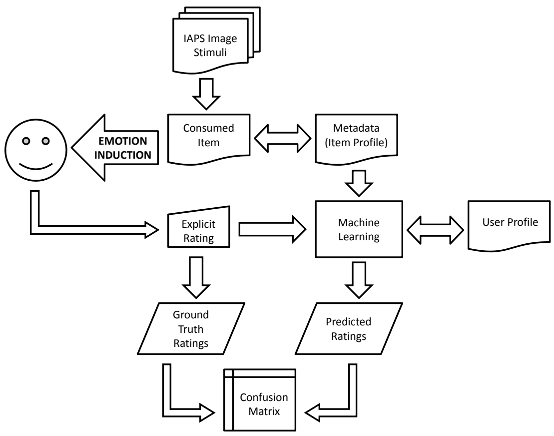

Fig. 3: Overview of the CBR experiment.

The set of images $h \in H$ that the users were consuming, had a twofold meaning: (i) they were used as content items and (ii) they were used as emotion induction stimuli for the affective labeling algorithm. We used a subset of 70 images from the IAPS dataset Lang et al. [2005]. The IAPS dataset of images is annotated with the mean and standard deviations of the emotion responses in the VAD space which was useful as the ground truth in the affective labeling part of the experiment.

The affective labeling algorithm described in Sec. 3.1 yielded rough classes in the VAD space. In order to build the affective item profiles we used the classes' centroid values (see Tab. 2) in the calculation of the first two statistical moments. We applied the procedure from Sec. 2.2.

We had 52 users taking part in our experiment (mean = 18.3 years, 15 males).

3.3 Affective CBR system evaluation methodology

The results of the CBR system were the confusion matrices of the classification procedure that mapped the images H into one of the two possible classes: relevant or non-relevant class. From the confusion matrices we calculated the recall, precision and F measure as defined in Herlocker et al. [2004].

We also compared the performances of the CBR system with three types of metadata: (i) generic metadata (genre and watching time as done by Tkalčič et al. [2010a]), (ii) affective metadata given explicitly and (iii) affective metadata acquired implicitly with the proposed emotion detection algorithm. For that purpose we transferred the statistical testing of the confusion matrices into the testing for the equivalence of two estimated discrete probability distributions [Lehman and Romano, 2005]. To test the equivalence of the underlying distributions we used the Pearson χ^2 test. In case of significant differences we used the scalar measures precision, recall and F measure to see which approach was significantly better.

4 Results

We compared the performance of the classification of items into relevant or non relevant through the confusion matrices in the following way: (i) Explicitly acquired affective metadata vs Implicitly acquired metadata, (ii) explicitly acquired metadata vs. generic metadata and (iii) implicitly acquired metadata vs. generic metadata. In all three cases the p value was $p < 0.01$. Table 3 shows the scalar measures precision, recall and F measures for all three approaches.

metadata/labeling method	classifier	P	R	F
implicit affective labeling	AdaBoost	0.61	0.57	0.59
	C4.5	0.58	0.50	0.53
	NaiveBayes	0.56	0.62	0.59
	SVM	0.64	0.47	0.54
explicit affective labeling	AdaBoost	0.64	0.56	0.60
	C4.5	0.62	0.54	0.58
	NaiveBayes	0.56	0.59	0.58
	SVM	0.68	0.54	0.60
generic metadata	AdaBoost	0.57	0.41	0.48
	C4.5	0.60	0.45	0.51
	NaiveBayes	0.58	0.57	0.58
	SVM	0.61	0.55	0.58

Table 3: The scalar measures P, R, F for the CBR system

5 Discussion

As we already reported in Tkalčič et al. [2010b], the application of the emotion detection algorithm on spontaneous face expression videos has a low performance. We identified three main reasons for that: (i) weak supervision in learning, (ii) non-optimal video acquisition and (iii) non-extreme facial expressions.

In supervised learning techniques there is ground truth reference data to which we compare our model. In the induced emotion experiment the ground truth data is weak because we did not verify whether the emotive response of the user equals to the predicted induced emotive response.

Second, the acquisition of video of users' expressions in real applications takes place in less controlled environments. The users change their position during the session. This results in head orientation changes, size of the face changes and changes of camera focus. All these changes require a precise face tracker that allows for fine face registration. Further difficulties are brought by various face occlusions and changing lighting conditions (e.g. a light can be turned on or off, the position of the curtains can be changed etc.) which confuse the face tracker. It is important that the face registration is done in a precisely manner to allow the detection of changes in the same areas of the face.

The third reason why the accuracy drops is the fact that face expressions in spontaneous videos are less extreme than in posed videos. As a consequence the changes on the faces are less visible and are hidden in the overall noise of the face changes. The dynamics of face expressions depend on the emotion amplitude as well as on the subjects' individual differences.

The comparison of the performance of the CBR with explicit vs. implicit affective labeling shows significant differences regardless of the ML technique employed to predict the ratings. The explicit labeling yields superior CBR performance than the implicit labeling. However, another comparison, that between the implicitly acquired affective labels and generic metadata (genre and watching time) shows that the CBR with implicit affective labels is significantly better than the CBR with generic metadata only. Although not as good as explicit labeling, the presented implicit labeling technique brings additional value to the CBR system used.

The usage of affective labels is not present in state-of-the-art commercial recommender systems, to the best of the authors' knowledge. The presented approach allows to upgrade an existing CBR system by adding the unobtrusive video acquisition of users' emotive responses. The results showed that the inclusion of affective metadata, although acquired with a not-so-perfect emotion detection algorithm, significantly improves the quality of the selection of recommended items. In other words, although there is a lot of noise in the affective labels acquired with the proposed method, these labels still describe more variance in users' preferences than the generic metadata used in state-of-the-art recommender systems.

5.1 Pending issues and future work

The usage of affective labels in recommender systems has not reached a production level yet. There are several open issues that need to be addressed in the future.

The presented work was verified on a sample of 52 users of a narrow age and social segment and on 70 images as content items. The sample size is not big but it is in line with sample sizes used in related work [Arapakis et al., 2009, Joho et al., 2009, Kierkels and Pun, 2009]. Although we correctly used the statistical tests and verified the conditions before applying the tests a repetition of the experiment on a larger sample of users and content items would increase the strength of the results reported.

Another aspect of the sample size issue is the impact of the size on the ML techniques used. The sample size in the emotion detection algorithm (the kNN classifier) is not problematic. It is, however, questionable the sample size used in the CBR. In the ten fold cross validation scheme we used 63 items for training the model and seven for testing. Although it appears that this is small, a comparison with other recommender system reveals that this is a common issue, and is usually referred as the sparsity problem. It occurs when, even if there are lots of users and lots of items, each user usually rated only few items and there are few data to build the models upon [Adomavicius and Tuzhilin, 2005].

The presented work also lacks a further user satisfaction study. Besides just aiming at the prediction of user ratings for unseen items research should also focus on the users' satisfaction with the list of recommended items.

But the most important thing to do in the future is to improve the emotion detection algorithms used for implicit affective labeling. In the ideal case, the perfect emotion detection algorithm would yield CBR performance that is identical to the CBR performance with explicit labeling.

The acquisition of video of users raises also privacy issues that need to be addressed before such a system can go in production.

Last, but not least, we believe that implicit affective labeling should be complemented with context modeling to provide better predictions of users' preferences. In fact, emotional responses of users and their tendencies to seek one kind of emotion over another, is tightly connected with the context where the items are consumed. Several investigations started to explore the influence of various contextual parameters, like being alone or being in company, on the users' preferences [Adomavicius et al., 2005, Odić et al., 2010]. We will include this information in our future affective user models.

6 Conclusion

We performed a comparative study of a CBR system for images that uses three types of metadata: (i) explicit affective labels, (ii) implicit affective labels and (iii) generic metadata. Although the results showed that the explicit labels yielded better recommendations than implicit labels, the proposed approach significantly improves the CBR performance over generic metadata. Because the approach is unobtrusive it is feasible to upgrade existing CBR systems with the proposed solution. The presented implicit labeling technique takes as input video sequences of users' facial expressions and yields affective labels in the VAD emotive space. We used Gabor filtering based low level features, PCA for dimensionality reduction and the kNN classifier for affective labeling.

Acknowledgement

This work was partially funded by the European Commission within the FP6 IST grant number FP6-27312 and partially by the Slovenian Research Agency ARRS. All statements in this work reflect the personal ideas and opinions of the authors and not necessarily the opinions of the EC or ARRS.

References

G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin. Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems (TOIS)*, 23(1):103–145, 2005.

I. Arapakis, Y. Moshfeghi, H. Joho, R. Ren, D. Hannah, J.M. Jose, and L. Gardens. Integrating facial expressions into user profiling for the improvement of a multimodal recommender system. In *Proc. IEEE Int'l Conf. Multimedia & Expo*, pages 1440–1443, 2009.

M.S. Bartlett, G.C. Littlewort, M.G. Frank, C. Lainscsek, I. Fasel, and J.R. Movellan. Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia*, 1(6):22–35, 2006.

Ling Chen, Gen-Cai Chen, Cheng-Zhe Xu, Jack March, and Steve Benford. Emoplayer: A media player for video clips with affective annotations. *Interacting with Computers*, 20(1):17–28, January 2008. doi: http://dx.doi.org/10.1016/j.intcom.2007.06.003. URL `http://dx.doi.org/10.1016/j.intcom.2007.06.003`.

Micah Eckhardt and Rosalind Picard. A more effective way to label affective expressions. *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–2, September 2009. doi: 10.1109/ACII.2009.5349528. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5349528`.

Alan Hanjalic and Li-Qun Xu. Affective video content representation and modeling. *IEEE Transactions on Multimedia*, 7(1):143–154, February 2005.

J.L. Herlocker, J.A. Konstan, L.G. Terveen, and J.T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1):53, January 2004.

H. Joho, J.M. Jose, R. Valenti, and N. Sebe. Exploiting facial expressions for affective video summarisation. In *Proceeding of the ACM International Conference on Image and Video Retrieval*, pages 1–8. ACM, 2009.

J.J.M. Kierkels and T. Pun. Simultaneous exploitation of explicit and implicit tags in affect-based multimedia retrieval. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–6. IEEE, 2009.

P.J. Lang, M.M. Bradley., and B.N. Cuthbert. International affective picture system (iaps): Affective ratings of pictures and instruction manual. technical report a-6. Technical report, University of Florida, Gainesville, FL, 2005.

E. L. Lehman and J.P. Romano. *Testing Statistical Hypotheses*. Springer Science + Business Inc., 2005.

Ante Odić, Matevž Kunaver, Jurij Tasič, and Andrej Košir. Open issues with contextual information in existing recommender system databases. *Proceedings of the IEEE ERK 2010*, A:217–220, September 2010.

M. Pantic and A. Vinciarelli. Implicit Human-Centered Tagging. *IEEE Signal Processing Magazine*, 26(6):173–180, 2009.

Rosalind Picard and Shaundra Briant Daily. Evaluating affective interactions: Alternatives to asking what users feel. In *CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches*, Portland, OR, April 2005.

Man-Kwan Shan, Fang-Fei Kuo, Meng-Fen Chiang, and Suh-Yin Lee. Emotion-based music recommendation by affinity discovery from film music. *Expert Syst. Appl.*, 36(4):7666–7674, 2009. ISSN 0957-4174. doi: http://dx.doi.org/10.1016/j.eswa.2008.09.042.

Mohammad Soleymani, Jeremy Davis, and Thierry Pun. A collaborative personalized affective video retrieval system. *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–2, September 2009. doi: 10.1109/ACII.2009.5349526. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5349526`.

Marko Tkalčič, Urban Burnik, and Andrej Košir. Using affective parameters in a content-based recommender system. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, 20(4), 2010a.

Marko Tkalčič, Ante Odić, Andrej Košir, and Jurij Tasič. Comparison of an emotion detection technique on posed and spontaneous datasets. *Proceedings of the IEEE ERK 2010*, 2010b.

Marko Tkalčič, Jurij Tasič, and Andrej Košir. The LDOS-PerAff-1 Corpus of Face Video Clips with Affective and Personality Metadata. In Michael Kipp, editor, *Proceedings of the LREC 2010 Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*, 2010c.

R. Valenti, Z. Yucel, and T. Gevers. Robustifying eye center localization by head pose cues. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009. URL `http://www.science.uva.nl/research/publications/2009/ValentiCVPR2009`.

Yongjin Wang and Ling Guan. Recognizing human emotional state from audiovisual signals. *IEEE Transactionson multimedia*, 10(5):936–946, 2008.

Zhihong Zeng, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(1):39–58, Jan. 2009. ISSN 0162-8828. doi: 10.1109/TPAMI.2008.52.

R. Zhi and Q. Ruan. Facial expression recognition based on two-dimensional discriminant locality preserving projections. *Neurocomputing*, 71(7-9):1730–1734, 2008.