Application of Inclusive User Modelling Web Service

Pradipta Biswas[1], Sanat Sarangi[2], Pat Langdon[1] and Subrat Kar[2]

[1]University of Cambridge, Cambridge, UK
[2]Indian Institute of Technology, Delhi, India
[1]Contact author: pb400@cam.ac.uk

Abstract: This paper presents an application of a user modelling web service in personalizing interfaces for users with age-related and physical impairment. The user model stores user profile in standardized format independent of application and device. It allows users to carry the profile with them irrespective of device and application. The user modelling web service is used to adapt interfaces of a sensor network based geo-visualization system and an agriculture advisory system for UK and Indian rural population. A pilot study found that all users preferred the adapted interface and their performance was also improved with adaptation.

1. Introduction

User model can be defined as a machine-readable representation of user characteristics of a system. We have developed a user model that considers users with physical, age-related or contextual impairment and can be used to personalize electronic interfaces to facilitate human machine interaction. We have identified a set of human factors that can affect human computer interaction and formulated models [2, 3] to relate those factors to interface parameters. We have developed inclusive user model, which can adjust font size, font colour, inter-element spacing (like line spacing, button spacing and so on) based on age, gender, visual acuity, type of colour blindness, presence of hand tremor and spasm of users. The model is more detailed than GOMS model [6], easier to use than Cognitive Architecture based models [1,9], and covers a wider range of users than existing generic user models [8] for disabled users, a detailed literature survey can be found in a different paper [2] . The user profile is created using a web form and the profile is stored in cloud. The sign-up page can be accessed at www-edc.eng.cam.ac.uk/~pb400/CambUM/UMSignUp.htm Once created, this profile is accessible to the user irrespective of application and device. We have worked with different development teams to integrate this user model into their applications. So far, the inclusive user modelling system has found applications in a wide variety of systems including a digital TV framework for elderly users (EU GUIDE system), an electronic agricultural advisory system, weather monitoring system and an emergency warning system. In parallel we conducted user trials to validate the user model. Users preferred and performed better with the adaptive system than the non-adaptive version.

2. User Modelling Framework

We have developed the Inclusive User Model and used it to develop a user modelling web service that can automatically adjust font size, colour contrast, line and button spacing of interfaces based on visual acuity, type of colour blindness, grip strength, active range of motion of wrist and static tremor of users. The user modelling system

- follows a standardized user profile format specified by a EU cluster [7] and published by International Telecommunication Union [5].
- does not propose to change content of an interface rather specifies layout parameters, so it is easily integrated to different applications.
- can automatically convert interface parameters (like font size or button spacing) for multiple devices (e.g. TV, computer, laptop, mobile phone and so on) by assuming a viewing distance for different devices and taking the screen resolution as input parameter.
- has investigated details of visual, auditory and motor functions of humans and is developed through extensive user trials to relate human factors to interface parameters [3].

The personalization framework (Fig 1) takes input about users' functional parameters (like visual acuity, colour blindness, short term memory capacity, first language and dexterity level) and subjective requirements. It also takes input from a sensor network [10] in runtime to detect the context of application. The Wisekar (Wireless Sensor Knowledge Archive) system is an Internet of Things (IoT) based repository for archival of sensor-derived information.

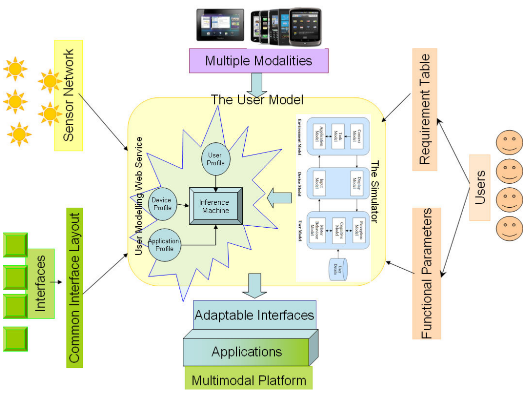


Figure 1. Personalization framework

These input parameters are fed into the Inclusive User Model that consists of perception, cognition and motor-behaviour models. The Inclusive User Model [2, 3] can predict how a person with visual acuity v and contrast sensitivity s will perceive an interface or a person with grip strength g and range of motion of wrist (ROMW) w will use a pointing device. Our user survey [4] generated a range of values of human factors for end users and we used this data in Monte Carlo simulation to predict a set of rules relating users' range of abilities with interface parameters. The rule based system along with the user, device and application profiles are stored in a cloud based server. The client application can access the web service using a plug-in.

The framework is integrated to applications using a simple Javascript program. The client application reads data from the user model and sensor network and changes the font size, font colour, line spacing, default zooming level and so on by either selecting an appropriate pre-defined stylesheet or changing parameters for each individual webpage or standalone application. A personalised weather monitoring system can be found at http://wisekar.iitd.ernet.in/wisekar_mm_full/index.php/main while a personalised electronic agriculture system can be found at http://e-vivasaya.rtbi.in/aas_cambridge/login.php . Different renderings can be generated with usernames user-1, user-2, user-3, user-4 and so on. In each case the password is same as the username.

3. User Trial

The following user trial reports a controlled experiment on the weather monitoring application. It compared users' objective performance and subjective preference for an adaptive and a non-adaptive version of the weather monitoring system. We purposefully used two different devices for signing up and using the application to highlight the notion of transporting user profile across multiple devices.

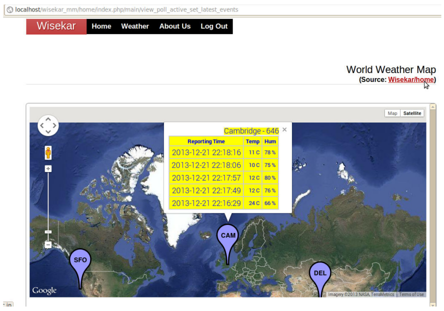
We collected data from 7 users (6 male, 1 female, average age 54.87 years) with moderate visual impairment, mainly age related short-sightedness and cataract.

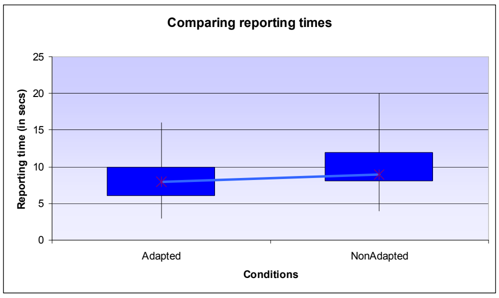
We have used a Windows 7 HP computer with 54 cm × 33 cm monitor having 1920 x 1080 pixels resolution to record users' performance with the weather monitoring system. We used a standard Logitech mouse for pointing. Users signed up using a HP Tx2 laptop with 30 cm × 20 cm screen and 1280 × 800 pixels resolution.

The participants were initially registered with the user modelling system using the Laptop. After that participants were briefed about the weather monitoring system. The task was to report temperature and humidity of cities on a specific date. Each participant was instructed to report temperature and humidity six times for each of adapted and non-adapted conditions. The order of adapted and non-adapted conditions was altered randomly to eliminate order effect.

During the sign up stage we found that different users preferred different font sizes ranging from 14 points to 18 points. We also noticed that one user was Protanomalous colour blind and he read 45 instead of 42 in the plate 16 of Ishihara colour blindness test. During use of the weather monitoring system, we measured the time interval between pressing the left mouse button on the bubble with the city name and reporting of the required temperature and humidity data. In total we analysed 84 tasks (42 for adapted and 42 for non-adapted). We found that users took significantly less time in adapted condition (average 8.25 secs, std dev 3.1 secs) than non-adapted condition (average 9.75 secs, std dev 3.63 secs). All participants were already familiar with mouse and also practiced the system before the actual trial. So we assumed that each pointing task was independent to each other. Under this assumption, the difference is significant in a two-tailed paired t-test with $p < 0.05$ and with an effect size (Cohen's d) of 0.44 (Fig 2). Without this assumption, the difference is also significant in Wilcoxon Signed-Rank test ($Z = -2.1$, $p < 0.05$). We conducted a subjective questionnaire to understand users' subjective preference. All users noticed bigger font and preferred it. One user was colour-blind and he preferred the change in colour contrast too.

The user study shows that users prefer different font sizes and colour contrast even for a simple system. The study also confirms that even for a simple text searching task users performed and preferred an adaptive system that can automatically adjusts font size, line spacing and colour contrast. The user modeling system successfully converted users' preference across two different devices having different screen resolutions. Future studies will collect data from more users and will use more complicated tasks than the present study.





a. Example of Adapted Wisekar b. Comparing weather reporting times

Figure2. Application of Inclusive User Model

4. Conclusions

This paper presents an application of a user modeling system that is used to store a user profile online and uses it to adapt user interfaces across different applications running on different devices. The user model follows standardized format to store the profile so that it can be easily integrated to multiple applications developed by different development teams. Our user study confirms that systems adapted by the user modeling system are preferred by users and it also statistically significantly reduces task completion times.

References

[1] Anderson J. R. and Lebiere C. "The Atomic Components of Thought." Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1998.

[2] Biswas P., Robinson P. and Langdon P., Designing inclusive interfaces through user modelling and simulation, International Journal of Human-Computer Interaction 28(1), Taylor & Francis, Print ISSN: 1044-7318, 2012

[3] Biswas P. and Langdon P., Developing multimodal adaptation algorithm for mobility impaired users by evaluating their hand strength, International Journal of Human-Computer Interaction 28(9), , International Journal of Human-Computer Interaction 28(9), Taylor & Francis, 2012

[4] Biswas P. and Langdon P., An HCI Survey on elderly users of India, Inclusive Designing - Joining Usability, Accessibility, and Inclusion, Springer 2014

[5] ITU Technical Report of the Focus Group on Smart Cable Television, Availabe at: http://www.itu.int/dms_pub/itu-t/opb/fg/T-FG-SMART-2013-PDF-E.pdf Accessed on 25/01/2013

[6] John B. E. and Kieras D. "The GOMS Family of User Interface Analysis Techniques: Comparison And Contrast." ACM Transactions On Computer Human Interaction 3 (1996): 320-351.

[7] Kaklanis N. et al,, An interoperable and inclusive user modelling concept for simulation and adaptation, International Conference on User Modeling (UMAP) 2012

[8] Kobsa A., Generic User Modelling Systems, The Adaptive Web: Mathods and Strategies of Web Prsonalization, Springer 2012

[9] Newell A. "Unified Theories of Cognition." Cambridge, MA, USA: Harvard University Press, 1990.

[10] Sarangi S. and Kar S., Wireless Sensor Knowledge Archive, Proceedings of the International Conference on Electronics, Computing and Communication Technologies (IEEE CONECCT 2013), Bangalore, Jan 17-19.

Users Ranking in Online Social Networks to Support POI Selection in Small Groups

Antonio Caso[1] and Silvia Rossi[2] *

[1] Dipartimento di Fisica, University of Naples "Federico II", Napoli - Italy
antonio.caso@unina.it
[2] Dipartimento di Ingegneria Elettrica e Tecnologie dell'Informazione
University of Naples "Federico II", Napoli - Italy
silvia.rossi@unina.it

Abstract. The process of content personalization in planning of a city tour requires taking into account the presence of small groups of users. In this context, social network analysis may play an important role both in evaluating users' common interests, and in determining users' relationships. In particular, cohesion, dominance and mediation can be helpful in the design of automatic processes to help users in reaching a consensus. This paper provides initial insights towards this goal and cues to derive simple models of user dominance through intra-group ranking.

Keywords: small groups, activity networks, decision support systems.

1 Social Interaction and Decision Making

The long-term goal of our research is to provide tourist users with recommender systems and decision support applications. An example of such applications is a city tour planner. When tourists visit a city, they are usually aggregated in small groups and stay in the city only for few days. Hence, it is necessary to choose certain Points of Interest (POI) that maximize the group satisfaction, taking into account that the members' preferences can be different.

Group recommendation approaches rely either on building a single user profile, resulting from the combination of users' profiles, or on merging the recommendation lists of individual users, at runtime, using different group decision strategies. Many of these techniques do not consider the social relationships among group members [1], while the design and implementation of Group Recommendation Systems, and, more generally, of Decision Support Systems, should take into account the type of control in the group decision-making process [2]. For example, there may be cases where the participants follow a democratic process in order to find a possible solution, and cases where the group is supported by a human leader or mediator. Following to this idea, the work of [1] starts to evaluate the group members' weights, in terms of their importance or influence, for TV recommendations.

* The research leading to these results has received funding from the Italian Ministry of University and Research and EU under the PON OR.C.HE.S.T.R.A. project.

In this work we provide initial insight towards the evaluation of small group relationships (leadership and mediation) with the aim to use such analysis to automatically generate a consensus function. In particular, we take concepts from research on users connectivity in order to identify key users in Online Social Networks (OSNs). OSNs interaction analysis can provide a viable way to obtain information about the social relationships and pattern of activities among the group of visitors, without bothering the users with questions.

1.1 A Popularity Ranking in Online Social Networks

The analysis of relationships through OSNs relies on several centrality measures, as formalized in [3]. However, the basic definitions of these measures are only designed for binary networks, with symmetric relationships of equal value between all directly connected users, while, in reality, an individual has relationships of varying qualities [4].

In our work, to compute the users' centrality, we took cues from the PageRank algorithm [5]. The idea behind this choice is to use a simple, but effective, algorithm to evaluate the rank of a person, interpreted as an index of popularity in a small group of friends. According to [6] concepts of popularity and leadership are highly correlated in close groups. A similar approach was used in [7], where the authors defined a centrality measure for undirected graphs. Instead, in this work, we present another variant that use directed and weighted graphs. In our opinion, the degree of activity of a person and the directionality of specific communication activities provides meaningful information on the social relationships. Our ranking is defined as follows:

$$R(x) = \frac{1-d}{|F|} + d\sum_{i \in F} \frac{w_{i,x}}{w_i} R(i) \tag{1}$$

where, $|F|$ is the total number of Friends in the group F and d (with $0 \leq d \leq 1$) is a dampening factor (set to 0.85). According to Eq. 1, the user x inherits a proportion of popularity from other i group's members. This is calculated considering both the i-th friend's popularity and the weight of the communication activity of the i-th friend towards the user x ($w_{i,x}$), normalized with respect to the total communication activity of the i-th friend with the group (w_i). Such weights are calculated considering some of the communication activities between two users on the OSN *Facebook.com*, collecting a combination of data arising from [8]. In particular, we use the number of posts and links published on the wall of the user x by the user i; the number of comments and likes from the user i on the posts published by the user x; the number of tags of user x inserted by i. Finally, the same computation is done with the comments, likes and tags from the user i on the photos of the user x. Note that the friend contribution is normalized with respect to its global activity with the group (as in PageRank). However, PageRank assumed only one link between pages x and i (hence, i equally contributes to the centrality of all the pages it points to), while, here, we represent the weight of the directed connections from i to x determining the level of one-sided communication.

% Sim	Leader	Average	Mediator	Lead & Med	Lead No Med	LNM Low STD
Average	61 ± 17	59 ± 11	63 ± 13	53 ± 15	73 ± 10	75 ± 10

Table 1: Cumulative results in the pilot study.

2 A Pilot Study

We conducted a pilot study with real users planning a trip in the city of Naples in order to gather useful information on social network relationships vs. face-to-face interactions. In our first pilot study we evaluated the behavior of 14 groups composed, in the average, of 3.4 people. 46 users took part in the experimentation (26 male and 20 female). The average age was 27.3 with a graduate education. In half of the groups there was a mediator, which was identified as the person to whom we asked to create the group and to help the group in performing the experiment. The leader of each group was identified as the member with the highest score according to Eq. 1.

Each person was asked to register on a specific web site using the credentials of *Facebook.com*. Once registered, it was asked to imagine to plan a one-day visit to the city of Naples (Italy) and to select from a checklist of ten items only three activities (places to visit) for the day. After that, it was asked to select two places to eat (from a check list of eight). Since we do not want the user to be involved in strategic reasoning, we did not ask the user to express ratings and preferences among the selected choices. The group was, then, asked to discuss, face-to-face, in order to obtain a shared and unique decision for the group.

Table 1 summarizes the cumulative data of all groups involved in the experiment. For each group, the following data are calculated: the similarity percentage between the choices of the leader and the group final decision (Leader); the similarity percentage between the choices of the mediator, if applicable, and the group final decision (Mediator); the similarity percentage between of the choices of each users and the group final decision (Average).

From the amount of analyzed interactions, with a very high standard deviation, we can conclude that the groups' behaviors on the OSN were very different and with a good value of cohesion (Average = 59%). Considering the aggregated data, the average similarity value of the leader choices (Leader) is on average 61%, which is comparable with the Average similarity, and the Mediator similarity (63%) with the final decision of the group. Apart from the aggregated data, that shows similar results on the average, what is interesting, from our point of view, is to compare the behavior of groups with a mediator with groups without this specific role. We observed that in the case of a member of the group acting as mediator the similarity of the group decision w.r.t. the leader was on average 53% (Lead & Med); instead, in the second case, the similarity with the leader was, on average, equal to 73% (Lead No Med). The p-value, calculated on these two sets, is 0.0058, which means that such difference is not due to chance. Finally, we analyzed the standard deviation of the leadership values (evaluated according to Eq. 1) for the set without a mediator and subdivided it in two sets (with low and high standard deviation). The groups with low standard deviation (LNM Low STD), which is interpreted as a measure of cohesion and similarity in the behavior of the group members, showed a similarity of the leader choices with the group final decision of 75%. Hence, we can infer that, in case there is not a mediator, the leader got a much more important role in the consensus making, especially in close groups where the popularity index, we evaluated, better identifies a possible leader.

3 Conclusions and Future Works

In this paper, we started to analyze the users' interactions in a social network in order to gather useful information to help groups in decision-making. In particular, we were interested in the role of cohesion, dominance and mediation for reaching a consensus. We showed that it is possible to derive a simple model of user dominance, through intra-group ranking, and such a role is fundamental in the absence of a mediator.

In this pilot study we used a small number of alternatives for planning only a single day in a delimited neighbourhood of a city. The scalability of our results, increasing the number of choices with more complex real settings, have to be deeply analyzed, including also the possibility to express an explicit ranking on the selected choices. Moreover, we limited our groups to people (mainly friends) without any hierarchical relationships between them, while also social intra-group roles have to be taken into account.

References

1. Gartrell, M., Xing, X., Lv, Q., Beach, A., Han, R., Mishra, S., Seada, K.: Enhancing group recommendation by incorporating social relationship interactions. In: Proc. of the 16th ACM International Conference on Supporting Group Work. (2010) 97–106
2. Jelassi, M.T., Foroughi, A.: Negotiation support systems: an overview of design issues and existing software. Decision Support Systems **5**(2) (1989) 167 – 181
3. Freeman, L.C.: Centrality in social networks conceptual clarification. Social Networks **1**(3) (1978) 215 – 239
4. Banks, L., Wu, S.: All friends are not created equal: An interaction intensity based approach to privacy in online social networks. In: International Conference on Computational Science and Engineering, 2009. Volume 4. (2009) 970–974
5. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems **30**(17) (1998) 107 – 117 Proceedings of the Seventh International World Wide Web Conference.
6. Theodorson, G.A.: The relationship between leadership and popularity roles in small groups. American Sociological Review **22**(1) (1957) pp. 58–67
7. Heidemann, J., Klier, M., Probst, F.: Identifying key users in online social networks: A pagerank based approach. In: Proc. of the Int. Conf. on Information Systems, ICIS, Association for Information Systems (2010)
8. Gilbert, E., Karahalios, K.: Predicting tie strength with social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '09, New York, NY, USA, ACM (2009) 211–220

Visualising Uncertainty for Open Learner Model Users

Carrie Demmans Epp[1,2], Susan Bull[1] and Matthew D. Johnson[1]

[1] Electronic, Electrical and Computer Engineering, University of Birmingham, UK
[2] Department of Computer Science, University of Toronto, Canada

carrie@taglab.ca, s.bull@bham.ac.uk, m.d.johnson.1@bham.ac.uk

Abstract. There is widening use of open learner models (OLM) to support learning and promote metacognitive behaviours, but learner model visualisations do not typically include information about uncertainty. We consider findings from the field of information visualisation and apply these to OLMs. Examples are given for how uncertainty visualisation might be usefully achieved.

Keywords: Open learner model, uncertainty, information visualisation.

1 Introduction

Open learner models (OLM) are learner models which are inspectable or can be directly interacted with in some way, by students or others [5]. There are now widening deployments of OLMs, especially at university level (e.g. [3,4,9,10,15,17]). The visualisations need to be understandable by users so that they can benefit from the purpose of the visualisations. These purposes are often to support or encourage metacognitive activities such as reflection, progress monitoring, planning and taking responsibility for learning [6], and studies have indicated improvements in learning with OLMs (e.g. [12,13,15]). Examples of OLM visualisations are given in Figure 1.

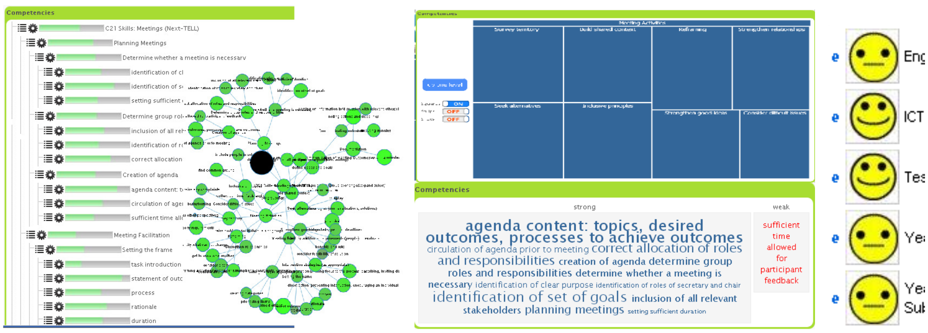


Fig. 1. OLM visualisations: skill meters, competency network, treemap, word cloud, smilies [4]

Uncertainty has long been a recognised problem in user and learner modelling (see [11]), with continuing interest in modelling techniques to overcome uncertainty in the modelling process (see examples in [8]), or methods to allow users to update or contribute information to their models [5]. Nevertheless, some level of uncertainty may still be present. This has implications for personalisation in adaptive learning systems, but is also crucial in OLMs: if users access a visualisation of their knowledge, etc., to prompt metacognitive activities, how can uncertainty be incorporated into the visualisation to enable them to take appropriate decisions according to what is shown?

2 Uncertainty Visualisation

The field of information visualisation aims to communicate complex information in a way that enables people to more easily understand the data, and make appropriate inferences from it [7]. Ways of presenting information on the quality of data people reason over has received growing attention across a range of disciplines [2]. However, people can still have difficulty understanding visual representations of uncertainty even if trained in their use [16]. In education, learning analytics dashboards [18] and OLMs [5] are increasingly used, but instructors often have minimal training in how to interpret visualisations, and if uncertainty is involved, this can be even harder. Moreover, as one of the primary aims of OLMs is to encourage metacognitive behaviours in learners, failure to understand visualisations of their learner models and the uncertainty therein, can negatively impact users' metacognitive processes and, consequently, their learning. We therefore propose some generic methods to visualise uncertainty in OLMs, with visual variables that can be processed pre-attentively [19] or selectively [1], such as position, closure, opacity or grain, while avoiding visual complexity that may impede pre-attentive processing [14]. We illustrate with the Next-TELL OLM, as it has multiple visualisations which are all used by students [4].

As shown in Figure 1, one of the visualisations uses skill meters. However, while quite easy to interpret, skill meters typically provide no information about the uncertainty of data. To avoid learners taking this as indisputable data, we propose indicating uncertainty using, for example, the skill meter fill (grain or opacity); or more precise uncertainty information represented similarly to error bars, as in Figure 2. For discrete skill meter-like visualisations (used in the Next-TELL OLM for users to input self, peer and teacher assessments), opacity could be incorporated to reflect uncertainty. The Next-TELL competency network uses node size and shade to indicate level of competency of elements within a domain structure. To avoid difficulty processing the information if additional features were included within the nodes, we propose grain or dashed outlines (closure) to map uncertainty in the information indicated by a node. In systems where uncertainty in relationships between nodes are modelled, manipulating the style of the connector lines is an option. The treemap uses only size to indicate competency strength. Therefore change in shade, grain or opacity could be used to indicate uncertainty. However, in the Next-TELL context, care must be taken to ensure consistency between visualisations (the competency network uses shading to show strength of competencies). Smilies are also available in the Next-TELL OLM, but adding other 'face features' would increase the complexity of the visualisation. We therefore propose opacity. Unlike common uses of word clouds to show word frequency (e.g. in a document or discussion), the Next-TELL word cloud indicates strength of competencies or understanding. In cases of uncertainty, the arrangement of (part of) a word cloud could be made 'messier', to reflect this.

Reasons for including uncertainty in OLM visualisations not only apply to supporting decision-making relating to the next stage of learning, but can also help focus user attention onto exploring their agreement with the learner model data (e.g. by viewing the evidence for the model which is also available in the Next-TELL OLM, or to suggest changes to the OLM to improve its accuracy, such as in [12]).

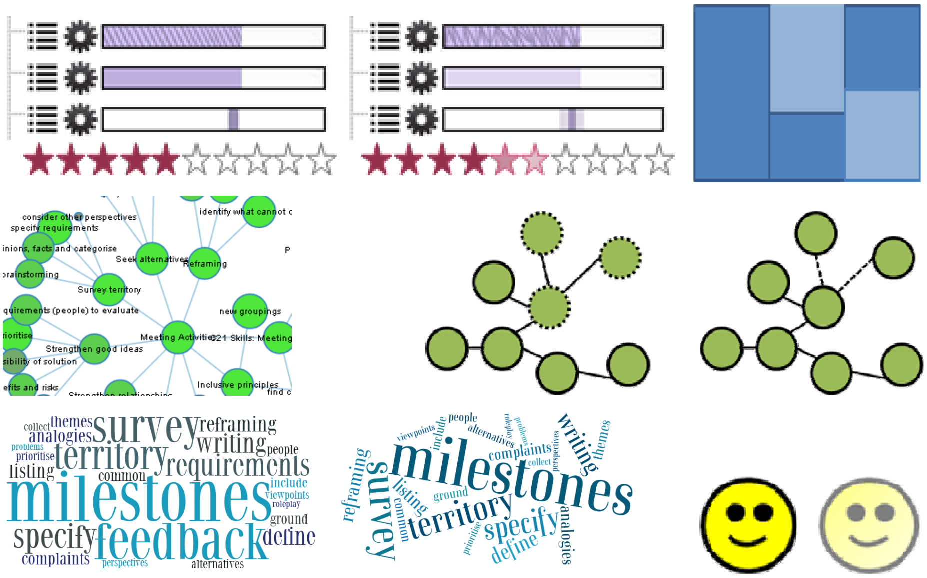


Fig. 2. Uncertainty visualisations (left: low uncertainty; right: high uncertainty; shade: treemap)

4 Summary

There is increasing use of OLMs, but few consider uncertainty in model data. This paper has highlighted potential methods to indicate uncertainty in various OLM visualisations, based on principles of uncertainty visualisation and the knowledge that OLM users are typically not trained in visualisation interpretation. In many cases opacity is a solution, as long as there are no other variables that may result in opacity making the visualisation over-complex for processing. Other solutions include grain and closure. We recommend such approaches be considered by OLM designers.

Acknowledgements

The Next-TELL project is supported by the European Community (EC) under the Information Society Technology priority of the 7th Framework Programme for R&D under contract no 258114 NEXT-TELL. This document does not represent the opinion of the EC and the EC is not responsible for any use that might be made of its content. Demmans Epp held a W. Garfield Weston Fellowship and a Walter C. Sumner Memorial Fellowship. The work performed was made possible via the support of the GRAND research network.

References

1. Bertin, J. (1983). Semiology of graphics, University of Wisconsin Press, Madison.
2. Brodlie, K., Allendes Osorio, R., & Lopes, A. (2012). A Review of Uncertainty in Data Visualization. In J. Dill, R. Earnshaw, D. Kasik, J. Vince, & P. C. Wong (Eds.), Expanding the Frontiers of Visual Analytics and Visualization, Springer, London, 81-109.
3. Bull, S. Jackson, T. & Lancaster, M. (2010). Students' Interest in their Misconceptions in First Year Electrical Circuits and Mathematics Courses, IJEEE 47(3), 307-318.
4. Bull, S., Johnson, M.D., Alotaibi, M., Byrne, W. & Cierniak, G. (2013). Visualising Multiple Data Sources in an Independent Open Learner Model, in H.C. Lane & K. Yacef (eds), Artificial Intelligence in Education, Springer-Verlag, Berlin Heidelberg, 199-208.
5. Bull, S. & Kay, J. (2010). Open Learner Models, in R. Nkambou, J. Bourdeau & R. Mizoguchi (eds), Advances in Intelligent Tutoring Systems, Springer, Berlin, 301-322.
6. Bull, S. & Kay, J. (2013). Open Learner Models as Drivers for Metacognitive Processes, in R. Azevedo & V. Aleven (eds), International Handbook of Metacognition and Learning Technologies, Springer, New York, 349-365.
7. Chen, C. (2010). Information visualization. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 387–403.
8. Chrysafiadi, K. & Virvou, M. (2013). Student Modeling Approaches: A Literature review for the Last Decade, Expert Systems with Applications 40, 4715-4729.
9. Demmans Epp, C. & McCalla, G. (2011). ProTutor: Historic Open Learner Models for Pronunciation Tutoring. In G. Biswas, S. Bull., J. Kay & A. Mitrovic (eds.), Artificial Intelligence in Education, Springer-Verlag, Berlin Heidelberg, 441-443.
10. Hsiao, I-H, Bakalov, F. & Brusilovsky, P. (2013). Progressor: Social navigation Support through Open Social Student Modeling, New Review of Hypermedia & Multimedia 19(2).
11. Jameson, A. (1995). Numerical Uncertainty Management in User and Student Modeling, User Modeling and User-Adapted Interaction 5(3-4), 193-251.
12. Kerly, A. & Bull, S. (2008). Children's Interactions with Inspectable and Negotiated Learner Models, in B.P. Woolf, E. Aimeur, R. Nkambou & S. Lajoie (eds), Intelligent Tutoring Systems, Springer-Verlag, Berlin Heidelberg, 132-141.
13. Long, Y. & Aleven, V. (2013). Supporting Students' Self-Regulated Learning with an Open Learner Model in a Linear Equation Tutor, in H.C. Lane, K. Yacef, J. Mostow & P. Pavlik (eds), Artificial Intelligence in Education, Springer, Berlin Heidelberg, 219-228.
14. MacEachren, A. M., Roth, R. E., O'Brien, J., Li, B., Swingley, D., & Gahegan, M. (2012). Visual Semiotics & Uncertainty Visualization: An Empirical Study, IEEE Transactions on Visualization and Computer Graphics 18(12), 2496–2505.
15. Mitrovic, A., Martin, B. (2007). Evaluating the Effect of Open Student Models on Self-Assessment, Int. J. of AI in Education 17(2), 121-144.
16. Newman, G.E., & Scholl, B.J. (2012). Bar Graphs Depicting Averages are Perceptually Misinterpreted: The within-the-bar bias. Psychonomic Bulletin & Review, 19(4), 601–607.
17. Perez-Marin, D. & Pascual-Nieto, I. (2010). Showing Automatically Generated Students' Conceptual Models to Students and Teachers, Int. J. of AI in Education 20(1), 47-72.
18. Verbert, K., Duval, E., Klerkx, J., Govaerts, S., Santos, J.L. (2013). Learning Analytics Dashboard Applications, American Behavioral Scientist, 57(10), 1500-1509.
19. Ware, C. (2004). Information visualization: Perception for Design. San Francisco, CA: Morgan Kaufman.

Balancing Exploration – Exploitation in Image Retrieval

Dorota Głowacka and Sayantan Hore

HIIT, Department of Computer Science,
University of Helsinki, Finland

Abstract. In recent years there has been an increased interest in developing exploration–exploitation algorithms for image search. However, little research has been done as to what type of image search such techniques might be most beneficial. We present an interactive image retrieval system that combines Reinforcement Learning with an interface designed to allow users to actively engage in directing the search. Reinforcement Learning is used to model the user interests by allowing the system to trade off between exploration (unseen types of image) and exploitation (images the system thinks are relevant). A task-based user study indicates that for certain types of searches a traditional exploitation-based system is more than adequate, while for others a more complex system trading off exploration and exploitation is more beneficial.

Image retrieval techniques operating on meta-data, such as textual annotations, have become the industry standard. However, with the explosive growth of image collections, tagging new images quickly is not always possible. Secondly, there are many instances where image search by query is problematic, e.g. finding an illustration for an article about "youth". A solution to such a problem is content-based image retrieval (CBIR) [6]. Early experiments show that CBIR can be improved through relevance feedback by involving the user in the search loop [1]. However, relevance feedback can lead to a context trap, where users specify the context so strictly that they can only exploit a limited area of information space. Combining *exploration/exploitation* strategies with relevance feedback is a popular attempt at avoiding the context trap [2, 3, 7]. However, few studies have been done showing the advantages (and disadvantages) of exploratory image retrieval systems. We report preliminary studies showing under what conditions exploratory image search might be most beneficial and where exploratory search may actually hinder the search results. For this purpose, we built a query-less image search system incorporating state of the art reinforcement learning (RL) techniques to allow the system to efficiently balance between exploration and exploitation.

System Overview. The system assists users in finding images in a database of unannotated images without query typing. The RL methods and interactive interface allow users to direct the search according to their interests. The interface and an example search are presented in Figure 1. The search starts with a display of a collage of images. To ensure that the initial set is a good representation of the entire image space, we cluster all the images in k clusters, where k is the number of displayed images and then we sample an image from each cluster. Our pre-user study shows that this technique provides a good starting point for the search. When the mouse hovers over an image,

(a) Iteration 1 (b) Iteration 2

(c) Iteration 3 (d) Iteration 4

Fig. 1. The first four iterations of an example search for “City by night”.

a slide bar appears at the bottom allowing the user to rate that particular image. The feedback ranges from -1 (no interest to the user) to 1 (highly relevant). Users can score as many images as they like. Images not rated by the user are assumed to have score of 0. Each image can be displayed at most once throughout the entire search session. We illustrate the interface and interaction design through a walkthrough exemple. The user wants to find an image to illustrate an article about “city by night”. Initially (Figure 1a), the user is presented with a collage of images uniformly selected from the database and marks the fifth image in the second row and the second image in the third row as highly relevant. The user moves to the next iteration by pressing the “Next” button at the top of the page. In the second iteration (Figure 1b), more images related to “night” are presented and the user selects four images. In iterations 3 and 4 (Figures 1c and 1d), more relevant images are presented and the user can further narrow down his search.

To help the user to explore the image space, we use Gaussian Process bandits with Self-Organizing Maps (GP-SOM), with dependencies across arms, which in our system translates into similarities between images. The algorithm uses function f that makes predictions with regards to the relevance of all the images to the user’s interests. When selecting the next set of images to display, the system might select images with the highest estimated relevance score but since the estimate of f may be inaccurate, this exploitative choice might be suboptimal. Alternatively, the system might exploratively select an image for which the user feedback improves the accuracy of f, enabling better future image selections. A detailed description of the algorithm and the similarity measure between the images can be found in [5].

Experiments. We conducted a set of user studies to evaluate the impact of exploration on three types of searches [1]: (1) *Target search* - looking for a particular image; (2) *Category search* - looking for any image from a given category, e.g. image of a cat; (3)

Open search - browsing a collection of images without knowing what the target may look like. The study included three conditions: 1) our Gaussian Process system (GP), 2) a version of our system that uses only exploitation (EXPLOIT), and 3) a system that presents random images at each iteration (RAND). In EXPLOIT, the exploration level was set to 0, which means that the system can only present images similar to the ones marked as relevant. The same interface was used in all settings. We used the MIRFLICKR-25000 dataset [4] consisting of 25000 images from the social photography site Flickr and commonly used in assessment of image retrieval and annotation tasks. We recruited 20 post-graduate students to run the experiments. Each participant was asked to perform three tasks for all three types of searches, i.e. each participant performed 9 searchers. We counterbalanced between the tasks and the systems for each subject so that each task was performed the same number of times with each system. The participants were asked to finish the task when they find the target image (in target search) or when they feel they found the ideal image in category and open searches. In all the tasks, the search was limited to 25 iterations. In target search, participants were presented with an image and a short description of that image and then asked to look for that image. In category and open searches, no example images were provided and participants were only given a short description of what to look for, e.g. red rose or illustration for an article about gardening.

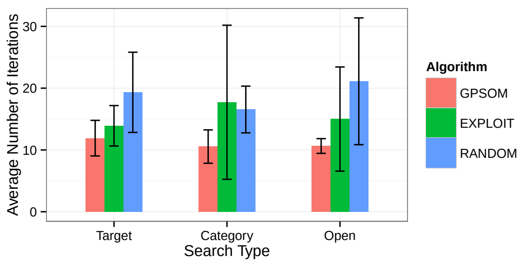
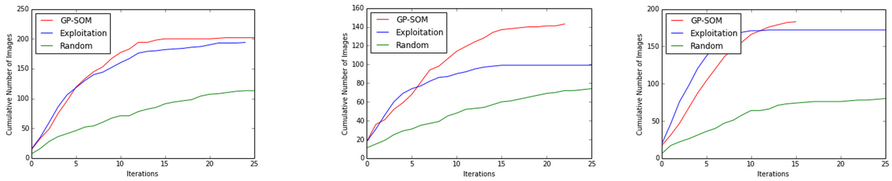


Fig. 2. Average number of iterations along with 95% confidence intervals.

We measured the average number of iterations to complete each task (Figure 2), which is a standard performance measure to evaluate CBIR systems [1]. GP-SOM outperforms EXPLOIT and RAND in all search types indicating that adding exploration to image search provides better support for user needs. There is little difference between GP-SOM and EXPLOIT in target search, indicating that when users have a specific image in mind from the onset, adding exploration makes little improvement. GP-SOM is more suitable for searches that are more exploratory in nature, such as category or open search, where the user first wants to browse the dataset before deciding what image they really want. We also counted the cumulative number of images that received positive feedback over search sessions in order to assess users' engagement in the search process (Figure 3). In target search, GP-SOM and EXPLOIT behave in a similar way. In category and open searches, GP-SOM displays relevant images throughout the search, while EXPLOIT stops providing relevant images after about 10 iterations, which indicates that users "get stuck" in a very limited area of the image space.



(a) Target Search (b) Category Search (c) Open Search

Fig. 3. Cumulative number of images marked as positive by user over iterations.

To summarize, GP-SOM exposes users to a higher number of relevant images in searches that are more vague in nature compared to EXPLOIT which narrows down the image space available to the user from the onset, which makes it more suited for target search. The results have significant implications for design of image retrieval system, where different strategies should be applied depending on the type of search, e.g. if we know that users will always make short searches then an EXPLOIT-type system will do a good job. However, if users have to perform longer open-ended searches (e.g. browsing a database of missing people), then a system based on exploration-exploitation might be more appropriate. In the future, we plan to run extensive user studies to get a better understanding of the relationship between search type and various levels of exploration.

Acknowledgements. The project was supported by The Finnish Funding Agency for Innovation (under projects Re:Know and D2I) and by the Academy of Finland (under the Finnish Centre of Excellence in Computational Inference).

References

1. I. Cox, M. Miller, T. Minka, T. Papathomas, and P. Yianilos. The bayesian image retrieval system, pichunter: theory, implementation, and psychophysical experiments. *Image Processing*, 9(1):20–37, 2000.
2. D. Głowacka, T. Ruotsalo, K. Konyushkova, K. Athukorala, S. Kaski, and G. Jacucci. Directing exploratory search: Reinforcement learning from user interactions with keywords. In *Proc. of IUI*, 2013.
3. D. Głowacka and J. Shawe-Taylor. Content-based image retrieval with multinomial relevance feedback. In *Proc. of ACML*, pages 111–125, 2010.
4. M. Huiskes and M. Lew. The mir flickr retrieval evaluation. In *Proc. of MIR*, 2008.
5. K. Konyushkova and D. Głowacka. Content-based image retrieval with hierarchical gaussian process bandits with self-organizing maps. In *Proc. of ESANN*, 2013.
6. H. Kosch and P. Maier. Content-based image retrieval systems-reviewing and benchmarking. *JDIM*, 8(1):54–64, 2010.
7. N. Suditu and F. Fleuret. Iterative relevance feedback with adaptive exploration/exploitation trade-off. In *Proc. of CIKM*, 2012.

Supporting the Fast Prototyping of Personalised Narratives for Tangible Interaction in Cultural Heritage

Robin Goldberg[1] and Daniela Petrelli[2]

[1]University of Stuttgart, Stuttgart, Germany
[2]Sheffield Hallam University, Sheffield, United Kingdom
D.Petrelli@shu.ac.uk

Abstract. Our research combines personalisation and tangible interaction to support visitors of cultural heritage sites. It is a complex setting of technology, people, objects and content all affecting each other and changing the final rendering and ultimately the visitor's experience. We present a completely open architecture for the fast prototyping of such personalised experiences tuned to the physical context. We used it to test personalisation applied to different types of tangible interaction, i.e. a soundscape controlled via physical movements and the a book like device or a haptic device that leads to personal destinations.

Keywords: Personalisation; Tangible Interaction; Fast Prototyping

1 An Open Framework to Model Content and Context

To date, personalisation applied to cultural heritage has not experimented with the possibilities offered by tangible interaction [1]. This new area of research needs to develop tools that allow to explore the many facets personalisation can take when applied in a tangible and/or embodied context, e.g. when objects and spaces react to people's presence, movement and gesture. Design uses fast prototyping as a technique to give form to an idea, show its strengths and weaknesses and identify new directions [2]. In computing most of the time prototypes are created to put forward a solution to a problem (experimental prototypes) or as steps toward the final product (evolutionary prototypes); but computational prototypes can be as exploratory as those developed in design that are informal, offers alternatives, are unstructured and messy, are used to communicate and then are thrown away [3]. The power of computational exploratory prototypes comes at an early stage in the project life cycle and lays in supporting the widest possible exploration of alternative solutions at the minimal cost. As part of meSch [6], we have developed such a framework for the fast prototyping of personalised interactive narratives, i.e. at any point in time the system has available many possible fragments of stories and has to decide which one is the right one for every specific dynamic user context. We are not prescriptive on which type of tangible interactive experiences will be offered to visitors: so far we have experimented with a magnifying glass, a book-like device, a haptic ovoid shell, a smart plinth and an interactive display case. The framework allows to experiment with personalised content in all these settings and many more. The personalisation features come from a wide range of equally possible options; e.g. where the user physically is and where they have been; what is nearby; the type of device they interact with; what type of gesture/movement they make; their age, knowledge and interest; the content available and its media type; and many more [5]. Which set of features is more suitable for a given tangible interaction setting is what needs to be experimentally determined via fast prototyping. Further challenges are the computational constraints posed by small computational units and the increasing number of sensors for tangible interaction [4].

In our proposal small chunks of self-contained content are controlled by a set of triggers, high-level expressions of any kind such as interpretation of sensor data or the visitor's selection of a specific theme. Specifically, seven elements model content, conditions and how to activate the first on the basis of the second:

- *Point of Interest* (POI): A POI marks a place in a cultural heritage site or an object in a museum. A POI is used to aggregate multiple content nodes or as a station in a multi-stage narrative.
- *Content Node*: A content node delivers its media if its activating condition is matched. The condition is defined by a set of pairs trigger-value. A content node may have predecessors, nodes that must be used before this one is activated, e.g. providing in-depth information.
- *Trigger*: A trigger is a personalisation feature with a discrete set of possible values. Triggers can map physical settings, people's choices, stages in the story, locations, or anything else that can be a condition.
- *Context*: A context is a set of active triggers. It contains the current value of each trigger for which sensor data is available.
- *State*: The state holds the current context as well as the set of all visited nodes, so it captures the visit so far and the condition for the next selection.
- *Active Set*: the set of content nodes that can be selected, i.e. they don't have any predecessors or the predecessors have been used already.
- *Inactive Set*: the set of content nodes for which the predecessor nodes are still to be used.

The openness of this framework is clear as both the content (nodes) and the conditions (triggers) can be anything we want to try. The creation of an interactive narrative prototype follows these stages: locations or objects that are part of an interactive story are associated to a POI; then a set of content nodes is defined for each POI to capture different layers or types of information; for every node the corresponding activation condition is defined, that is to say the set of triggers with their activation values must be listed; if any content had to be delivered before, a predecessor list is created.

The selection of which content node should be delivered at every point in time starts with the sorting of content nodes into two sets: Active or Inactive. At the start the active set contains all content nodes without predecessors while the inactive set contains all the nodes with a predecessor. This sorting reduces the computational time needed to test the context against nodes that cannot be activated anyway, a substantial advantage when the nodes are counted in hundreds. As the visit progresses, the context is continuously updated adding or removing triggers and their value; for example if a POI is reached, the location trigger is included in the context with the value corresponding to this specific POI. Every time the context changes it is checked against all the nodes in the Active Set and a node is selected for delivery (if there is more than one candidate the selection is random). At this point the State is updated to include the newly activated node and the nodes in the Inactive set that have the selected node as predecessor are moved into the Active set (Fig. 1).

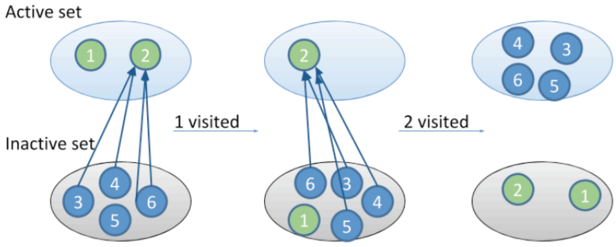


Figure 1. Shift of nodes between Active/Inactive sets following the delivery of content.

The activation mechanisms are completely independent from the content nodes and the triggers: defined both using XML, they can be extended or modified at any point in time without having to change the implementation at all thus supporting the prototyping of new interactive stories. A simulation mode allows to test at the desk what a specific visitor would receive when visiting in person; the only input required is the sequence of POIs the hypothetical visitor will follow.

The following section illustrates the framework using implemented prototypes.

3 Examples of Personalised Tangible Interactions

The *Companion Novel* (Fig. 1) is a book-like device complemented by a set of Bluetooth speakers located at POIs. Every page is a theme and the visitor can choose one (among 4 available) by placing a magnetic bookmark on a page. Designed for an outdoor setting, the Novel tracks the visitor walking the ground: when a POI is detected the Novel instructs the loudspeaker to play an attraction sound; if the visitor get closer to the POI a content specific to the theme currently selected is played.

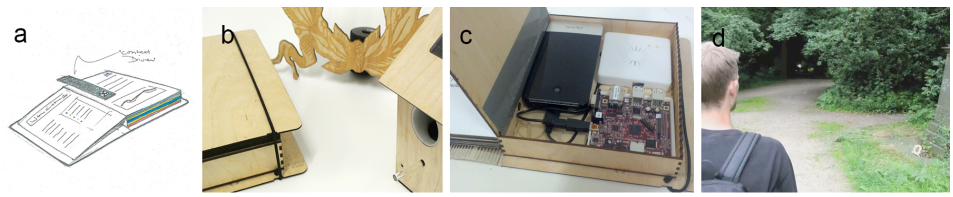


Figure 2. The Companion Novel delivers audio files to a visitor of an outdoor heritage.

The Novel has 7 POI and 4 pages so 4 parallel themes. The content nodes for this prototype are audio only, but of 2 different types: the attraction sound and the story. Each POI holds 4 attraction sounds (one for each theme), and 4 stories (played when the visitor is close by) (Fig. 2). The triggers defined for the Novel are: themes (values 1-4) mapping the pages of the book; location (the 1-7 POI); proximity ('far' or 'close'). When composed in a context the triggers capture the visitor's choice (the theme); the visitor's position (the location); their movement (proximity). At the beginning the visitor choses the theme, for example Nature, and 'theme: 1' is put in the context; when the visitor enters the area of POI 5 the context is updated adding 'location: 5' and 'proximity: in'. The trigger set of each node is checked against the current context: the node with triggers [theme:1 POI:5 proximity:in] is played. The visitor approaches POI-5; the sensor detects 'proximity: close' and updates the context accordingly starting a new node selection process for [theme:1 POI:5 proximity:close].

The *Way Detector* leads visitors to a very specific POI via haptic feedback: the closer to the target POI the stronger the vibration. The bottom round shape of the Way Detector fits the hand; on its top flat face are 4 small buttons, each one associated to a different POI. By selecting each button in turn the visitor is taken through the heritage stopping at four different stations where the visitors scans the Detector (which holds an NFC tag) and access the content nodes at the POI location: the content nodes are images and text displayed on tablets installed at each POI. By pressing button 3 the visitor has selected 'POI:3' and following the vibration they reach the location; the scan of the Detector updated the context in the tablet to [stage:3]; however the content nodes are played only if POI 1 and 2 have been already visited; this is achieved by the set of predecessors that create a chine of content nodes. If the order has not been followed, only the content node that hints to follow the order is displayed.

The two very different examples show how our framework allows to prototype different personalised interactive narratives that are instantiated in very different tangible contexts as both the content nodes and the triggers are defined by the applictaion.

4 Acknowledgements

This research is part of meSch, Material Encounters with Digital Cultural Heritage funded by the European Commisson's Seventh Framework Programme 'ICT for access to cultural resources' (ICT Call 9: FP7-ICT-2011-9) Grant Agreement 600851.

5 References

1. Ardissono, L., Kuflik, T., Petrelli, D. *Personalization in Cultural Heritage: The road travelled and the one Ahead*. UMUAI, 22 (1-2) Springer, 73-99, 2012.
2. Brown, T. *Change by Design*. Harper, 2009.
3. Floyd, C. *A Systematic Look at Prototyping*. In Budde, R. et al. (eds) "Approaches to Prototyping", Springer Verlag, 1-18, 1984.
4. Kubitza, T., Schmidt, A. *A First Set: Physical Components for the Creation of Interactive Exhibits*. Proc. TEI14 – Work in Progress, 16-19 February 2014.
5. Not, E., Petrelli, D. *Balancing Adaptivity and Customisation: In Search of Sustainable Personalisation in Cultural Heritage.* UMAP 2014 – Short Paper, 7-11 July 2014.
6. Petrelli, D., Ciolfi, L., van Dijk, D., Hornecker, E., Not, E., Schmidt, A. *Integrating Material and Digital: A New Way for Cultural Heritage*, ACM Interactions, July+August, 2013.

Effects of Search Interface and Decision Style on Learning Material Search Behavior and Reaction

Jincheul Jang, Mun Y. Yi, and Wan Chul Yoon

Department of Knowledge Service Engineering,
Korea Advanced Institute of Science and Technology
Daejeon, Republic of Korea
{jcjang, munyi, wcyoon}@kaist.ac.kr

Abstract. Fueled by the popularity of online learning and the widespread adoption of the Internet, educational materials are abundant online. However, it is unclear how the materials should be intelligently presented to different types of users. In this study, we examine how alternative search interfaces (list type vs. scatterplot type) influence users with different decision styles and consequently change their search behaviors and reactions to the interfaces. The two alternative interfaces were applied with different strengths while providing the same access to the underlying learning materials. A user study was conducted in a lab experiment setting, in which search interface was manipulated as a within-subject variable and presentation order as a between-subject variable. The findings indicate that the interface type has significant effects on the user behaviors and reactions with different decision styles.

Keywords: User interface; Decision style; Information search; Learning material

1 Introduction

The advancement of information technology has changed how people educate themselves. Online learning is widespread and popular as learning materials are abundant and mostly freely available on the Internet. Consequently, searching or sharing learning materials is an important aspect of learning that is supported by many commercial websites. Recognizing the importance of this phenomenon, IEEE Computer Society announced that "Supporting New Learning Styles" is a crucial trend in 2014 [1].

In processing and representing information, an emerging topic is the volume of data that can be visualized to reinforce human cognition. From this perspective, an important objective of information visualization is to project high-dimensional data onto low dimensional interfaces [4]. Users can understand complex information intuitively with well-designed visualizations, which can also reduce the amount of cognitive effort required of users to find relevant information. Thus, from the perspective of information visualization, it would be beneficial to provide a set of information to users in a better form than in a textual list.

In this study, we examined the effects of two alternative search interfaces on user search behaviors and reactions. The interfaces, designed for supporting online learning activities, consist of (1) a list type that provides a list of learning materials, mostly in the form of textual descriptions and (2) a scatterplot type that plots learning materials represented by icons in a two-dimensional graph. In addition to the main effects of interface differences on user behaviors and reactions, we examined the possibility that those effects vary with respect to individual differences such as decision style so that we can develop a deeper understanding of the interface effects.

2 Test System

In this system, lists of educational materials are visualized in two modes: list and scatterplot types (see Figure 1). The two types of view can be used to visualize the same set of materials. The list type page comprises a list view and a tag search window, which can be commonly seen in general search engines and online commerce websites, while The scatterplot type has a two-dimensional view, with difficulty on the horizontal axis and popularity on the vertical axis. These two alternative interfaces provide access to the same underlying learning materials with different strengths. The text-driven list type provides search results in the form of textual descriptions, whereas the graph-driven scatterplot type in the form of nodes located inside a scatterplot.

Fig. 1. List (left) and scatterplot (right) views shown to users in our study

3 Evaluation Method

Total fifty-three volunteers served as the experimental participants. All subjects went through two experimental blocks, in which task execution was followed by evaluation (one block for the list type and the other block for the scatterplot type). To prevent ordering effects, one group performed the list type experiment before the scatterplot type experiment, whereas the other group performed the scatterplot type experiment before the list type experiment. The subjects were assigned randomly to each of the two experimental conditions. 29 subjects firstly performed list type and then used scatterplot type interface, while 24 subjects performed with opposite order of interface.

The overall experimental procedure is shown in Figure 2. In the pretest session, we obtained demographic information, decision styles, and web search self-efficacy. After each experiment, the participants evaluated each task and information format (list and scatterplot types) using a questionnaire. Participants answered the questions about the cognitive decision effort, perceived ease of use, perceived usefulness, satisfaction of using the interface, and continuous intention to use to measure the task performance. All of the items used a 7-point Likert-type scale.

Data bookmarked by the participants were recorded in the system's database. To analyze the bookmarked data, we recruited 3 graders and let them judge the coincidence of the bookmarked materials with the given search topics. Every bookmarked record was graded by a 3-point scale where 0 = incorrect, 1 = partially correct, and 2 = correct. The majority of grades between three graders determined the final precision grade. If the three graders all had different opinions about the grade, a discussion among the three graders was additionally required to reconcile the differences. The average number of bookmarked materials and the precision score for each user were counted to measure the task performance.

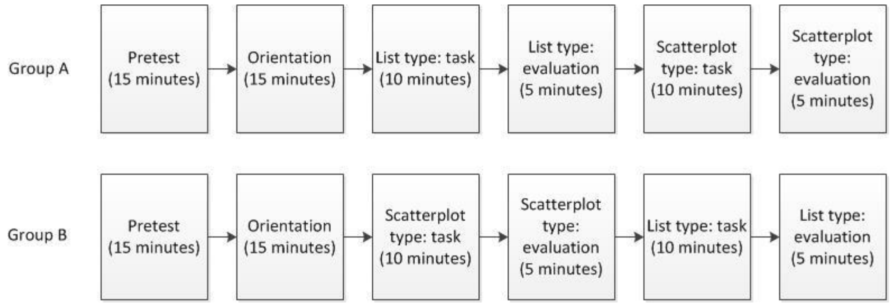


Fig. 2. Experimental procedure

4 Results and Conclusion

We performed quantitative comparisons of the user reactions and performances by analyzing the questionnaires. A paired samples t-test was performed to examine the effects of all information formats in the two interfaces Perceived ease of use ($t = 2.88$, $p < .01$), satisfaction ($t = 2.14$, $p < .05$), and the number of bookmarks ($t = 2.67$, $p < .05$) differed significantly between the two interfaces. In addition, the cognitive decision effort ($t = -1.84$, $p = .07$) and continuous intention to use ($t=1.80$, $p = .08$) had effects on borderline-significance between the two interfaces. This means that users perceived the list type much easier to use than the scatterplot type because the cognitive decision effort, perceived ease of use, and satisfaction scores were all higher for the list type than the scatterplot type.

To analyze the relationships between users' decision styles and reactions to the two interfaces, we performed a repeated measures ANOVA test to detect any moderating effects of decision making styles, measured along the sensing-intuition and thinking-feeling dimensions. Based on the decision making style questions, we found that there were 34 sensing-style subjects and 19 intuition-style subjects along the sensing-intuition dimension, while there were 25 thinking-style subjects and 28 feeling-style subjects along the thinking-feeling dimension.

For the sensing-intuition decision style, the perceived ease of use (F = 6.67, p < .05), satisfaction (F = 7.67, p < .01), continuous intention (F = 11.24, p < .01), and bookmarking precision score (F = 5.53, p < .05) differed significantly between the two interfaces. In addition, the cognitive decision effort (F = 3.55, p = .07) and perceived usefulness (F = 2.92, p < .09) had effects on the marginal significance between two interfaces. In case of the sensing style, the list type produced a lower cognitive decision effort score and higher perceived ease of use, usefulness, satisfaction, continuous intention, and bookmarking precision scores than the scatterplot type did. The opposite was true for the intuition style. This means that sensing-style users more easily and effectively used the list type interface, but intuition-style users found it easier and more effective to use the scatterplot type. This result confirms that sensing-style users performed better with the list type rather than scatterplot type interface, as opposed to intuition-style users who preferred the scatterplot type.

For the thinking-feeling decision style, the cognitive decision effort (F = 4.70, p < .05) differed significantly between the two interfaces. Moreover, the perceived ease of use (F = 3.62, p = .06) and bookmarking precision score (F = 3.03, p = .09) had marginally significant effects on the two interfaces. Specifically, thinking group users perceived that the scatterplot type demanded a lower level of cognitive decision effort than the list type did, in contrast to the feeling group users. At the same time, the thinking group users perceived that the scatterplot type was easier to use than the list type, in contrast to the feeling group users. Thus, people who prefer a thinking style theoretically tend to make decisions in a more reasonable, logical, and considered manner. In contrast, people with a feeling style tend to reach decisions by preferring a low tolerance for ambiguity and focusing on affections and intuitions [2]. Our results showed that feeling-style users perceived less cognitive decision effort and higher ease of use with the list type interface than with the scatterplot type.

In summary, we performed a laboratory experiment to determine their effectiveness across users with different decision making styles. Prior research [3] found that users' cognitive style did impact their search behavior; the current study extends those prior studies by examining the effects of cognitive styles on user reactions and behaviors in two alternative user interfaces. The study results can facilitate the improved design of online search interfaces by web designers. For example, our study show that a scatterplot type interface is a preferable choice for intuition-style users while a list type interface is a preferable choice for the remaining, highlighting the need to develop an adaptive system that can accommodate different users' decision styles and their preferences.

References

1. IEEE computer society, Top Technology Trends for 2014. http://www.computer.org/portal/web/membership/Top-10-Tech-Trends-in-2014.
2. Leonard, N.H., Scholl, R.W., Kowalski, K.B.: Information processing style and decision making. Journal of Organizational Behavior 420, 407–420 (1999).
3. Palmquist, R., Kim, K.-S.: Cognitive style and online database search experience as predictors of Web search performance. Journal of Organizational Behavior 51, 558–566 (2000).
4. Steele, J., Iliinsky, N.: Beautiful Visualization. O'Reilly Media, Inc. (2010).

A Peer to Peer Architecture for a Distributed User Model

Fotis Paraskevopoulos and Gregoris Mentzas

National Technical University of Athens
{fotisp,gmentzas}@mail.ntua.gr

Abstract. The current existence and continuous introduction of new social networks, provide a constant growth of the available information which can be used to create richer user profiles. Due to the disparity of this information centralized approaches to generate user models result in complex systems which can adapt with difficulty to the constant reshaping of information. For this reason there is a trend to create decentralized approaches for distributed user models. There are still two apparent constraints in these, one is the flexibility of the schema used to distribute the user model, and the other is the need for a centralized entity which facilitates communication and interoperability. We propose a peer to peer architecture to achieve a distributed user model. Each peer in the system acts as a standalone user model, it extracts and translates information into pre-defined templates, which are used to ensure interoperability when communicating with other peers, through an internal communication broker.

1 Introduction

Currently there is a plethora of social networks available, and new ones are being introduced constantly. The information found in these, range from information the users share to that generated from the interaction of the user with the system. Information from a single social network (referred as to *information source* from this point on) may be abundant enough to generate rich user profiles, which can be used for personalization services, recommendation services, etc. There is however an advantage in merging information from multiple information sources, to create richer user profiles [1].

In order to cope with this multitude of information sources, user modeling approaches evolved from centralized monolithic ones such as GUMO [8], GeniUS [6], UserML [7] to decentralized ones. A basic constraint of decentralized approaches is the interoperability of the information exchanged between parts of the system [3,4]. This is currently handled either by introducing a central component which handles the information exchange [1][10][9], or by homogenizing the schema to which the information is mapped [2].

In this paper we propose a peer to peer architecture to provide a fully decentralized system to generate and maintain distributed user models. Each peer in the system is made up of a User Model Translator(UMT) which encompasses the necessary logic to extract and translate information from a social network into *key:value* pairs which are grouped into *Templates*. *Templates* are uniquely identifiable by a Universal Unique Identifier(UUID) which is a hash generated from the combination of the *key* parts of the *key:value* pairs. Peers communicate through a *Broker* which can exchange information between any peers in the system provided they support *templates* with matching UUIDs. Our vision is that for any number of information sources available, there will be a matching number of peers.

The structure of this paper is as follows. In section 2 we provide a brief survey of the related work. In section 3 we describe our approach in detail. In Section **??** we go over a "Proof of Concept" which we developed in order to examine the feasibility of such system. Section 4 concludes this paper by going over our planned future work.

2 Related Work

In this section we will go over several of the previous approaches to create decentralized distributed user models.

In [1] the authors introduce a framework which can achieve interoperability of a distributed model via a *centralized* server called *Mypes*, where a single vector based user model is built per user by aligning the information from several information sources using hand crafted rules called *alignments*. PersonisAD [2] is defined as a distributed, active, *scrutable model*, which is capable of gathering information from different sensors about different users and combine their preferences to provide a richer experience. MobiTribe [10] is a system which focuses on mobile devices and its advantage is the fact that the the model itself is distributed in nature. The exchange of information between the devices and applications using the user modeling information is mediated by a centralized content management system. In [5] the authors present a vision of a distributed user model by creating single function stand alone agents which are responsible for storing a single attribute of a holistic vector based user model. Finally in [9] the authors present a decentralized architecture for sharing and re-using life logs from different systems. These are gathered by agents, which then forward the information to a centralized broker.

3 The P2P Distributed User Model

We propose a *P2P* architecture to generate and maintain a distributed user model, which uses pre-defined information exchange templates to communicate the user model. The information provision and exchange is controlled by standalone peers which contain the necessary logic to process, and store the information from a single information source.

Each peer is self contained and can act as stand-alone user model which only handles information from a single source. The objective however is to leverage information from multiple sources, without the need to handle the workload of extracting, storing and processing this information individually. A peer not only extracts information, but it also translates it into pre-determined templates, in order to communicate it with other peers in the system. A peer can act either as a producer, a consumer or both, depending on its desired effect on the entire system.

A peer in the system is composed of three parts, the User Model Translator(UMT), the *Templates* it supports, and its internal *Broker*.

The UMT needs to be developed individually for each peer, and contains the necessary logic to extract information from a particular information source, and consequently translate it into *key:value* pairs for each type of information like in the following example {*name: John*},{*lastname:Smith*} [1].

By translating the information into a basic exchangeable format, it can be grouped to correspond to any of the *Templates* which the agent supports. A *Template* is the specifications of a group of *key:value* pairs. The previous *key:value* examples could be part of the "Basic Profile" template. Each template is identified by a Universally Unique Identifier (UUID) which is generated using the *key* part of the *key:value* pairs. The process to generate the UUID starts by concatenating the string values of templates' *keys* into a single string and subsequently we generate a hash of that value. The resulting hash is the UUID of the template and it determines its *syntactical signature* and uniqueness, not its semantic one. There is no requirement for any peer in the system to support the exact same types or number of templates. Any peer can introduce new templates into the distributed model, which are defined by their UMT. These can be discovered and supported by other peers, through the internal *Broker*.

We need to determine the Level of Interoperability(LI) between any two peers. The *LI* between any **n** number of peers is determined by the number of matching *template UUIDs* they support. The *LI* is determines whether two peers can communicate and exchange information at all.

Each peer contains its own *broker*. Through the *broker* a peer can determine the *LI* between other peers, the list of available templates in each peer, and the rules and pre-conditions in place to access the information from a particular peer. Not all *templates* of a peer are publicly accessible, only those that have been specified in the broker configuration.

Currently there is no *LI* threshold set in place to determine whether any two peers can communicate, but rather the *brokers* themselves analyze the lists of *UUID*s which are exchanged during the initial handshaking process. It is possible for *brokers* to request the syntax of a *template* pertaining to a UUID which it does not currently support.

4 Future Work

In this paper we presented a P2P architecture to achieve a decentralized distributed user model. We believe this approach allows for the continuous expansion of the user model, in the form of a genealogy of peers. The information exchanged between peers can be further reused to create connections and inferences which are not immediately present from a single information source, without the constraints and overhead of extracting and analyzing the information from multiple information sources by a single pre-defined non-extensible service.

There are two aspects of the peer which need to be extended, the first is to provide a method by determining the emphsemantical signature of the templates, and the second is to extend the brokers contain the necessary functionality to enforce cross-boundaries policies, and expiration policies to ensure the privacy of the users information[11].

We have deployed a live version of the P2PUM which has four active peers in p2pum.imuresearch.eu.

References

1. Abel, F., Herder, E., Houben, G.J., Henze, N., Krause, D.: Cross-system user modeling and personalization on the social web. User Modeling and User-Adapted Interaction 23(2-3), 169–209 (Apr 2013)
2. Assad, M., Carmichael, D.J., Kay, J., Kummerfeld, B.: PersonisAD: distributed, active, scrutable model framework for context-aware services. In: LaMarca, A., Langheinrich, M., Truong, K.N. (eds.) Pervasive Computing, pp. 55–72. No. 4480 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (Jan 2007)
3. Carmagnola, F.: From user models to interoperable user models. AI Communications 22(2), 121–123 (2009)
4. Carmagnola, F., Cena, F., Gena, C.: User model interoperability: a survey. User Modeling and User-Adapted Interaction 21(3), 285–331 (Aug 2011)
5. Dim, E., Kuflik, T.: User models sharing and reusability: a component-based approach. In: UMAP Workshops (2012)
6. Gao, Q., Abel, F., Houben, G.J.: GeniUS: generic user modeling library for the social semantic web. In: Pan, J.Z., Chen, H., Kim, H.G., Li, J., Wu, Z., Horrocks, I., Mizoguchi, R., Wu, Z. (eds.) The Semantic Web, pp. 160–175. No. 7185 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (Jan 2012)
7. Heckmann, D., Krueger, A.: A user modeling markup language (UserML) for ubiquitous computing. In: Brusilovsky, P., Corbett, A., Rosis, F.d. (eds.) User Modeling 2003, pp. 393–397. No. 2702 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (Jan 2003)
8. Heckmann, D., Schwartz, T., Brandherm, B., Kröner, A.: Decentralized user modeling with UserML and GUMO. In: PROCEEDINGS OF THE WORKSHOP ON DECENTRALIZED, AGENT BASED AND SOCIAL APPROACHES TO USER MODELLING (DASUM 2005. pp. 61–65 (2005)
9. Iyilade, J., Vassileva, J.: A decentralized architecture for sharing and reusing lifelogs. In: UMAP Workshops (2013)
10. Thilakarathna, K., Petander, H., Mestre, J., Seneviratne, A.: {MobiTribe}: cost efficient distributed user generated content sharing on smartphones. IEEE Transactions on Mobile Computing PP(99), 1–1 (2013)
11. Wang, Y., Kobsa, A.: Privacy in cross-system personalization. In: Intelligent Information Privacy Management Symposium, Stanford, CA (2010)

Evaluation of Personalized Concept-Based Search and Ranked Lists over Linked Open Data

Melike Sah and Vincent Wade
Centre for Global Intelligent Content, Trinity College Dublin, Dublin, Ireland
{Melike.Sah, Vincent.Wade}@scss.tcd.ie

Abstract. Linked Open Data (LOD) provides a rich structured data. As the size of LOD grows, accessing the right information becomes more challenging. Especially, the commonly used ranked lists presentation of current LOD search engines is not effective for search tasks in unfamiliar domains. Recently, combination of clustering and personalized search gained more attention for this purpose. In this paper, we evaluate the impact of personalized concept-based search in terms of task assistance and user satisfaction, while comparing with the ranked lists. A user study was conducted with 32 subjects. Results showed that the personalized concept-based search enabled users to be more effective and efficient at performing both information gathering and fact finding tasks.

Keywords: Click through, personalized search, linked open data, ranked list.

1 Introduction

As the size of LOD grows, providing efficient search mechanisms becomes important. However, current LOD search engines (e.g. Sindice [1], Watson [2]) use ranked lists presentation, which is not efficient for information gathering search tasks in unfamiliar domains [3]. This is an important drawback since it is estimated that ~80% of Web queries are informational [11]. Faceted search overcomes this issue by assisting users with topics for interactive search and browsing [4]. However, facet creation depends on specific data and schema properties of underlying metadata and it can be difficult to generate useful facets to large and heterogeneous LOD [5]. In traditional Information Retrieval (IR), results clustering and personalized search are two popular methods for enhancing search efficacy. In clustering search, results are organized into categories for assisting users in results exploration and in query disambiguation. Results categorization is used widely such as Google categories or Yahoo Directories. Alternatively, personalized search aims to enhance retrieval by adapting results to context/interests of the user. On the other hand, [6] and [7] combine both results clustering and personalization in order to improve retrieval accuracy. [6] uses hierarchical clustering and user's manual selections on these clusters to filter out relevant results. Whereas, [7] assumes that search results are clustered (such as using ODP) and they propose a personalized ranking algorithm for results re-ranking. Similarly, [9] combines results categorization and personalized IR techniques to provide a personalized concept-based search mechanism over LOD. In this approach, results are categorized based on a conceptual ontology, UMBEL (umbel.org) [8]. Then, based on the user's click through data on concepts and results, the search results are adapted; for instance they apply concept re-ranking, results re-ranking, query expansion or concept suggestions [9]. The common aspect of [6, 7, 9] is that all approaches receive initial search results from a search engine and apply clustering (concepts) as the basis of the personalization. Although these approaches [6, 7, 9] individually evaluated the impact of their concept-based personalization, they only focused on retrieval precision and ignore the user aspects of such personalization. In this paper, we focus on this problem and evaluate personalized concept-based search against the ranked lists in terms of task assistance and user satisfaction. Then, we provide some preliminary conclusions.

2 Linked Open Data Search Approaches

Current LOD search engines use ranked lists to present search results [1] [2]. In this approach, results order depends on relevancy to the user query. Ranked lists work well if the user has a specific information need. However, by the nature of the LOD, resources are structured and often resources contain scattered data that is linked to other resources. This makes it even harder, for unfamiliar users, to search and explore the LOD. Faceted search [4] and personalized search [9] approaches are proposed to overcome this problem. In this work, we use [9] to compare with the ranked lists. In particular, results are grouped based on their concept categorizations [8]. Then, initial results are presented with no adaptation, where result categories are ranked according to relevance to the query. The following personalization is applied when the user interacts with the search system: (i) When a user selects a concept for exploration; all concepts are re-organized according to their semantic and syntactic similarity. In addition, within the selected concept, more relevant results are included using results re-ranking and query expansion as well as relevant concepts are suggested for results exploration. (ii) When the user clicks on a result, within the interacted concept, immediately personalization is applied. Such as, relevant results and concepts are added by query expansion and results re-ranking according to last *N* clicks of the users [9].

3 Evaluations and Analysis

3.1 Experimental Setup

Dataset: Our evaluation is based on information seeking tasks in a tourism domain, particularly "tourism in Killarney Ireland". We selected a tourism domain since it suits well for data gathering and informational queries as well as users can issue specific queries as they learn the topic. We use the benchmark dataset as explained in [9].

Ranked List vs Personalized Concept-based Search: As a comparison, a purpose-built non-adaptive ranked list search system was created. For a fair comparison, the ranked list (i) uses the same underlying indexing and retrieval models, (ii) operates across the same dataset, and (iii) results are ranked according to relevancy to the query as with the personalized concept-based search. The only difference is that the personalized system categorizes and adapts the results to the click through data [9].

Tasks: The search tasks were inspired from Google popular search queries in our domain. In order to test different types of queries, 4 search tasks were prepared with varying level of specificity, such as from fact finding tasks to information gathering tasks [10]. The questions were deliberately very specific, this enabled users to decide if they were satisfied to complete the task. Task 1 and 3 contained a mixture of open-ended and fact finding questions. Task 2 and 4 contained open-ended questions.

Experiment: We used 32 participants from School of Computer Science and Statistics of Trinity College Dublin (1 master student, 20 PhD students, 9 post-docs and 2 academics). For a fair comparison, users were divided into two groups, such as Group A and Group B. Group A users performed either Task 1 or Task 3 using the ranked list first and then performed Task 2 or Task 4 using the personalized search. Similarly, Group B users performed Task 1 or Task 3 using the personalized system first and then performed Task 2 or Task 4 using the ranked list. Thus, all tasks were equally tested and both systems equally (and anonymously) presented as the first system.

3.2 Results and Analysis

Task Assistance: It is desirable that a system requires users to invest the least amount of effort in order to find relevant information as quickly as possible. The results from the task completion times revealed that the personalized concept-based search outperformed the ranked list with an average of 6.50 (m:ss) versus 10.48. Moreover, t-tests confirm that the results are indeed significant for each task (for Task 1 $p=0.037$, for Task 2 $p=0.047$, for Task 3 $p=0.03$ and for Task 4 $p=0.003$) as shown in Figure 1. It is also shown that Task 2 and especially Task 4 took considerably longer to complete using the ranked list compared to the personalized search. The reason for this could be that both Task 2 and Task 4 contained open-ended questions, compared to Task 1 and Task 3. Similarly, users formulated fewer queries across all tasks using the personalized system (Figure 1). Again t-tests confirm that the results are significant (for Task 1 $p=0.045$, for Task 2 $p=0.031$, for Task 3 $p=0.017$ and for Task 4 $p<0.001$). On average, users issued 6.46 queries using the ranked list to complete the tasks compared to average of 3.03 queries for the personalized system. Another aspect of task assistance is the number of viewed pages. It is desirable that the search system provides the best resources in the top results, thus users require few page views. The results showed that users consistently required more page views across all tasks using the ranked list (average of 10.56 versus 5.34 page views as shown in Figure 1). T-tests also confirms the significance (Task 1 $p=0.048$, Task 2 $p=0.048$, Task 3 $p=0.038$ and Task 4 $p=0.013$). This gain was mainly obtained by the personalization; results re-ranked as well as more relevant results were automatically pushed on top of the search list using category/results re-ranking and query expansion techniques using our system.

User Satisfaction: The findings were backed up by post-questionnaires as shown in Figure 2. These results are indeed statistically significant with $p<0.001$ for Q1-Q12. In addition, users strongly recognized and valued categorization (Q8 - with an average of 4.43 versus 1.78) and personalization aspects (Q12 - with an average of 3.96 versus 1.68). Moreover, personalized concept-based search achieved an average SUS usability score of 88.90 compared to 75.15 of the ranked list. This is an interesting finding, especially considering the familiarity of users with the traditional ranked lists. These results are even more encouraging with personalization features like re-ranked lenses, re-ranked results and category suggestions, users thought that the personalized concept-based search was easy to use and better than the ranked list in terms of usability.

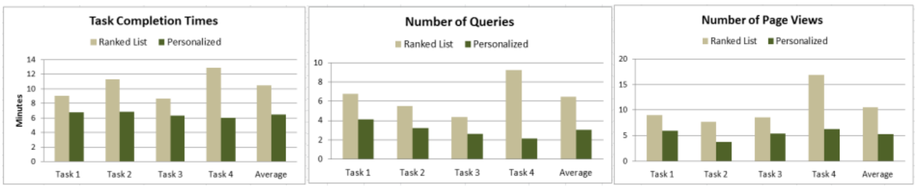
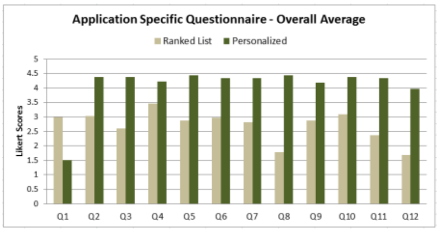


Figure 1. Task completion times, number of queries and page views across tasks



Q1: I had to search a lot before I found interesting content.
Q2: I spent less time querying and more time browsing.
Q3: I was less exposed to irrelevant content.
Q4: I did well on tasks.
Q5: The results structure and content was helpful to solve the tasks.
Q6: I am satisfied with the system performance, guidance and assistance.
Q7: I found the presentation of the search results helpful.
Q8: I found the categorization and grouping of search results helpful.
Q9: The result structure and content matched my expectations accurately.
Q10: The content composition generated by the system was easy to navigate.
Q11: I felt guided to relevant results.
Q12: The system guided me towards more personally relevant content.

Figure 2. Users' perceptions of search, result presentation and personalization

Analysis: Personalized concept-based search enabled users to be efficient and effective at performing both information gathering and fact finding tasks as shown by task completion times, number of issued queries and viewed pages. Questionnaires backed up these findings. Results are indeed statistically significant for all metrics.

Acknowledgements. This research is supported by the Science Foundation Ireland (Grant 12/CE/I2267) as part of CNGL (www.cngl.ie) at Trinity College Dublin.

4 References

1. Delbru, R., S., Campinas, G., Tummarello: Searching Web Data: an Entity Retrieval and High-Performance Indexing Model. Journal of Web Semantics, 10, 33-58, 2012.
2. D'Aquin, M., E., Motta, M., Sabou, S. Angeletou, L., Gridinoc, V., Lopez and D., Guidi: Toward a New Generation of Semantic Web Applications. IEEE Intelligent Systems, 2008.
3. Tummarello, G., R. Cyganiak, M. Catasta, S. Danielczyk, R. Delbru and S. Decker: Sig.ma: live views on the Web of Data, Journal of Web Semantics, 8(4), 355-364, 2010.
4. Heim, P., T. Ertl and J. Ziegler: Facet Graphs: Complex Semantic Querying Made Easy, Extended Semantic Web Conference, LNCS, 6088, 288-302, 2010.
5. Hogan, A., J. Umbrich, A. Harth, R. Cyganiak, A. Polleres, S. Decker: "An empirical survey of Linked Data conformance. Journal of Web Semantics, 14, 14–44, 2012.
6. Ferragina, P., and A., Gulli: 2005. A Personalized Search Engine Based on Web-Snippet Hierarchical Clustering. In: International conference on World Wide Web, 801-810, 2005.
7. Lee, H.C., and Borodin, A: Cluster Based Personalized Search. Algorithms and Models for the Web-Graph, 5427, 167-183, 2009.
8. Sah, M., and V. Wade: A Novel Concept-based Search for the Web of Data using UMBEL and a Fuzzy Retrieval Model. In: Extended Semantic Web Conference, 7295, 103-118, 2012.
9. Sah, M., and V. Wade: Personalized Concept-based Search and Exploration on the Web of Data using Results Categorization. In: Extended Semantic Web Conference, 7882, 532-547, 2013.
10. User study tasks can be accessed from https://www.scss.tcd.ie/melike.sah/tasks.html
11. Jansen, B.J., D.L., Booth, A. Spink: Determining the User Intent of Web Search Engine Queries. International Conference on World Wide Web, 1149-1150, 2007.

Exploiting Wikipedia Categorization for Predicting Age and Gender of Blog Authors

K Santosh, Aditya Joshi, Manish Gupta, Vasudeva Varma

International Institute of Information Technology
Hyderabad, India

Abstract. For privacy reasons, personally identifiable information like age and gender of people is not available publicly. However accurate prediction of such information has important applications in the fields of advertising, forensics and business intelligence. Existing methods for this problem have focused on classifier learning using content based features like word n-grams and style based features like Part of Speech (POS) n-grams. Two major drawbacks of previous approaches are: (1) they do not consider the semantic relation between words, and (2) they do not handle polysemy. We propose a novel method to address these drawbacks by representing the document using Wikipedia concepts and category information. Experimental results show that classifiers learned using such features along with previously used features help us achieve significantly better accuracy compared to the state-of-the-art methods. Indeed, feature selection shows that our novel features are more effective than previously used content based features.

1 Introduction

In recent years, exponential increase in textual information has sparked interest in automatically predicting personally identifiable information (PII) such as age and gender of users. Automatic prediction of age and gender has various applications in the fields of forensics, business intelligence and security.

Research in identifying author's age and gender started with extensions of the earlier works on categorization and classification of text. Koppel et al. [2] exploited combinations of lexical and syntactic features to infer the gender. Koppel et al. [3] explored differences in writing style and content between male and female bloggers as well as among authors of different ages. Meina et al. [4] used an ensemble based classification method to determine age and gender. They used various content and style based features. The overview paper of the PAN Author Profiling task [5] discusses various approaches used by their participants. It states that participants used content based (bag of words, word n-grams, slang words, etc.), style based (POS, readability measures, punctuations etc.) features, and that the ensemble of all features performed better. However, the two major issues with the content based features used in above works are: (1) they do not consider the semantic relation between words, and (2) they do not handle polysemy. Our method addresses these issues by representing a document in the feature space of Wikipedia concepts and categories.

2 The Proposed Approach

People of different gender and age have different interests. Hence, there is a lot of contextual difference between blogs written by different people. In our approach we explore these contextual differences to predict age and gender of an author of a text. Our approach consists of two phases: Semantic representation of documents, and age and gender prediction.

2.1 Semantic Representation of Documents

We extracted Wikipedia concepts related to the entity mentions in the text for each document in the training corpus. For every Wikipedia concept, we found its categories in Wikipedia. In order to get an exhaustive list of categories, we recursively collected the categories up to five levels. We refer to the list of categories at level i as $Cat\ L_i$. Our final document representation thus consists of a collection of Wikipedia concepts and categories. We refer to these features as Wikipedia semantic features.
Preprocessing Data: The text from blogs is preprocessed to remove HTML tags and unwanted boilerplate content like advertisements to get the clean data.
Entity Linking: We used TAGME API [1] to find Wikipedia concepts in the text. TAGME uses anchor text found in Wikipedia as spots (sequence of terms which are ambiguous) and the pages linked to them in Wikipedia as their possible senses. TAGME tackles the ambiguity and polysemy problems in the potentially many available anchor-page mappings by finding the collective agreement among them via scoring functions which are both fast to compute and effective.
Finding Parent Categories for Wikipedia Concepts: For all the Wikipedia concepts extracted in the previous step, their parent categories up to five levels are extracted. We created a Wikipedia category network using Wikipedia's category corpus and the networkx library[1] and traversed up to five levels on this network to obtain all parent categories. Semantically related words get mapped to similar set of Wikipedia categories at various levels; thus semantic relations between the words get captured in our approach.

2.2 Age and Gender Prediction

To predict the author's profile, i.e., age group and gender of the author of a document, we used two machine learning classification models namely, K-Nearest Neighbors (KNN) and Support Vector Machines (SVM).
KNN: Given a test document q, we represented the documents in terms of Wikipedia semantic features as mentioned in Subsection 2.1. We used Okapi-BM25F [6] distance metric to compute k nearest neighbors to the test document. While computing Okapi-BM25F, we considered Wikipedia concepts and category at different levels as the fields.
SVM: We also learned SVM classifiers for age and gender prediction using Wikipedia semantic features, and content and style features.

[1] http://networkx.github.io/

3 Experiments

Dataset: We used PANTraining and PANTesting datasets provided by the 2013 PAN Author Profiling Task[2]. The class labels are Male and Female for gender, and three groups of age (10s: 13-17 yrs, 20s: 23-27 yrs and 30s: 33-47yrs). Dataset details are shown in Table 1. We divided PANTraining into two parts: 70% for training, 30% for validation.

KNN Classifier: For KNN classifier, we learned the boost factor for each field c using the validation set as $boost_c = \frac{AccWith_c}{AccWithout_c}$ where, $AccWith_c$ is the accuracy obtained by using field c alone, and $AccWithout_c$ is the accuracy obtained by using all the other fields except c. Figures 1 and 2 show that each of the features are important for the prediction task.

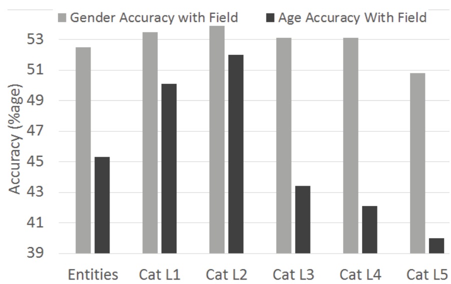


Fig. 1. Accuracy with Particular Field Considered

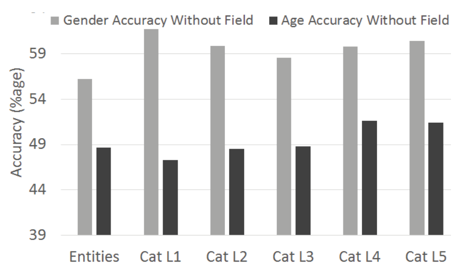


Fig. 2. Accuracy with Particular Field Ignored

On validation data, we obtained best accuracy at k=5 for gender prediction and k=7 for age prediction. Hence, we use these values of k while testing.

SVM: For SVM, along with Wikipedia semantic features, the following features are also used. (a) **Content based features**: These features analyse the content of the blogs. Koppel et al. [2] used unigrams as content features. In this work, we use unigrams, bigrams and trigrams as content based features. (b) **Style Features**: These are features which capture people's writing styles. In this work, we use POS n-grams (upto trigrams) as style features. Various combination of above mentioned features are used for building classifiers. The size of the feature vector for these feature sets are listed in Table 2.

Age	#Train Instances	#Test Instances
10s	17200	1776
20s	85800	9174
30s	133600	14408

Table 1. Dataset Details (Equal Distribution for Males and Females)

	Gender	Age
word n-grams	50000	61781
POS n-grams	16000	18000
Wikipedia Semantic	300000	306910

Table 2. Number of Features for Gender and Age Classifiers

We learned the parameters of SVM using 10-fold cross validation on PANTraining data. Our experiments did not find other kernels to perform any better than the linear kernel. Table 3 compares accuracies of approaches using different combination of word n-grams, POS n-grams, our Wikipedia Semantic features and state-of-the-art method

[2] http://www.webis.de/research/corpora/corpus-pan-labs-09-today/pan-13/pan13-data/

on the PANTesting data. We also compared with Meina et al. [4]'s method that obtained the best accuracy in the PAN Author Profiling Task at CLEF 2013.

Features	Classifier	Gender	Age
Wikipedia semantic	KNN	56.42	61.38
Wikipedia semantic	SVM	56.61	61.85
Word n-grams	SVM	53.21	56.79
POS n-grams	SVM	54.56	57.37
Wikipedia semantic + word n-grams	SVM	57.27	62.67
Wikipedia semantic + POS n-grams	SVM	58.39	63.29
Wikipedia semantic + word n-grams + POS n-grams	SVM	**62.12**	**66.51**
Meina et al. [4]		59.21	64.91

Table 3. Accuracy Comparison of Various Approaches on PANTesting Data

Accuracy comparisons in Table 3 show that our Wikipedia semantic features are better than the word n-gram based features, and combination of all features yields the best accuracy.

4 Conclusions

We studied the problem of age and gender prediction. We leveraged the document representation using Wikipedia concepts and category information as features for KNN and SVM classification. Experimental results show that the proposed approach beats the best approach for a similar task at CLEF 2013. By enhancing the entity linking part of the proposed system, overall accuracy of the age and gender prediction can be further improved. In the future, we would like to limit our reliance on entity linking and also explore other learning algorithms and robust features that can help in predicting the age and gender of the author of a document.

References

1. P. Ferragina and U. Scaiella. TAGME: On-the-fly Annotation of Short Text Fragments (by Wikipedia Entities). In *Proceedings of the* 19^{th} *ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1625–1628. ACM, 2010.
2. M. Koppel, S. Argamon, and A. R. Shimoni. Automatically Categorizing Written Texts by Author Gender. *Literary and Linguistic Computing*, 17:401–412, 2003.
3. M. Koppel, J. Schler, S. Argamon, and J. Pennebaker. Effects of Age and Gender on Blogging. In *Proceedings of the AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*, 2006.
4. M. Meina, K. Brodzinska, B. Celmer, M. Czokow, M. Patera, J. Pezacki, and M. Wilk. Ensemble-based Classification for Author Profiling Using Various Features. *The Notebook for 2013 PAN at the Conference and Labs of the Evaluation Forum (CLEF)*, 2013.
5. F. Rangel, P. Rosso, M. Koppel, E. Stamatatos, and G. Inches. Overview of the Author Profiling Task at PAN 2013. *The Notebook Papers of the Conference and Labs of the Evaluation Forum (CLEF) 2013 LABs and Workshops*, pages 23–26, 2013.
6. H. Zaragoza, N. Craswell, M. J. Taylor, S. Saria, and S. E. Robertson. Microsoft Cambridge at TREC 13: Web and Hard Tracks. In *Proceedings of the 2004 Text REtrieval Conference (TREC)*, 2004.

Predicting Player Type in Social Network Games

Dereck Toker, Ben Steichen, Max Birk*

Department of Computer Science,
University of British Columbia
2366 Main Mall, Vancouver, BC, Canada

*Department of Computer Science,
University of Saskatchewan
110 Science Place, Saskatoon, SK, Canada

{dtoker, steichen}@cs.ubc.ca, max.birk@usask.ca

Abstract. This paper presents preliminary work towards personalizing and recommending social network games based on a user's player type. In particular, we present research aimed at supporting such personalization through the prediction of player type from automatically collected user data. We first provide a brief overview of player types, and then outline several data sources that we gathered from a popular social network game to study the feasibility of player type predictions. Finally, we perform a preliminary analysis using one of these sources, namely music interests.

Keywords: Games User Research, Personalization, Machine Learning

1 Introduction

As of 2013, over 250 million users regularly play games each month on Facebook [1]. With the explosive growth and widespread popularity of social network gaming, players are faced with the increasingly difficult challenge of discovering games that are enjoyable and remain engaging. To address this concern, the long-term goal of our research is to improve users' social network gaming experience through personalization. Previous research pertaining to user differences in online gaming environments has typically focused on characterizing player behaviors and preferences using the Big-5 personality traits [2, 3, 4]. However, a potentially more relevant trait that ought to be considered in games is *player type*, which characterizes users based on the games they prefer, as well as their individual playing styles [5]. By taking into account player type, personalizing a user's game experience *within* a game could be implemented in order to improve player engagement by offering, for example, additional in-game content like extra quests or virtual goods. Furthermore, matching player types *across* games would also be advantageous to both players and advertisers/traffic-exchangers, since players could be offered games that more strongly match their individual player type. Unfortunately, administering questionnaires to ascertain player type is costly, and can have a potentially negative impact on players by intruding their game experience, and may even make them feel uncomfortable. In this paper we provide a brief overview of player types, and outline two data sources we collected to study the feasibility of automatically inferring *player type* from: i) general user interests available through the Facebook social graph; and ii) user behavior data.

As a first step to investigate the feasibility of utilizing these data sources, we present a preliminary prediction analysis using users' Facebook music interests.

2 Player Type

Player type can be measured using the BrainHex player typology [6], which provides a model based on prior findings from player research and neurobiology. This model has been applied successfully to adapt persuasive games to player types [7], and to develop design guidelines for personalized game systems [8]. It distinguishes between the following 7 different types of players: '*Achievers'* are often satisfied by completing tasks or collecting things (e.g., badges). They are goal-oriented and motivated by the reward of achieving long-term goals [5]. '*Conquerors'* enjoy struggling and the defeat of difficult opponents before achieving victory. They show forceful behavior, and channel their anger to face and overcome difficult challenges. '*Daredevils'* are excited by the thrill of taking risks and enjoy playing on the edge. The enjoyment of game activities such as navigating dizzying platforms, rushing around at high speeds while still in control characterizes the Daredevil. *'Masterminds'* enjoy solving puzzles, devising strategies to overcome difficult puzzles, and making efficient decisions. *Seekers* enjoy exploring things and discovering their surroundings. They are curious, have sustained interest, and love sense-stimulating activities. *'Socialisers'* enjoy interacting with others. For instance, they like talking, helping, and hanging around with people they trust. *'Survivors'* love the experience associated with terrifying scenes and the thrill of escaping from scary situations.

3 Data Sources & Collection

Our pool of users was obtained via a Facebook game, Pot Farm, where players perform actions such as planting and harvesting crops, collecting gold, completing quests, and unlocking achievements (see [9] for an overview of 'Ville style games). In order to gather ground-truth values for users' player types, we collected 3487 BrainHex player type questionnaires over a period of one week. Participation was optional and players were offered a reward of premium in-game content for completion. Given that reliability is always a concern with online surveys, data was cleaned via several methods and resulted in 2009 valid users.

Facebook Profile Interests: For each of the users who completed the BrainHex player type questionnaire, we collected preference data available to us according to the permissions enabled when a user installs Facebook games. This data consisted of three categories: Music interests, Movie/Television interests, and Game interests. In total, we obtained profile data from 1899 users (i.e., approx. 55% of surveyed users).

This data source is particularly interesting because it can be obtained independent of any specific game, and thus game personalization and recommendation can be performed without requiring any prior game experience. In fact, it has been reported that one third of users opting to play games on Facebook are entirely new to gaming, meaning that social network gaming is their first exposure to gaming [10]. This suggests that many of these new users may have either a minimal or limited understanding regarding their own player type, which highlights an even greater need to offer personalization by detecting player type based solely on other existing interests.

Game Telemetry: In addition to gathering general Facebook profile information, we also collected users' in-game behaviour such as logins, achievements, level, etc. While this data source is richer in terms of games-related behavior, a potential limitation is that it is restricted to events from one game. Therefore, while results will be very useful for in-game personalization for this game, they might not apply to other games. Nevertheless, our intuition is that there likely exist certain types of actions that will be indicative of general player type(s), therefore it could be transferrable to the personalization and/or recommendation of other games.

4 Prediction Experiments Based on Music Interests

In the previous section, we outlined multiple data sources and our intention is to evaluate each of them as candidates to predict player type. As a preliminary investigation, we selected Facebook profile music interests as a first step to understanding the feasibility of such predictions. Research has already shown that music preference relates to personality [11], and personality in turn has been linked to online games [2, 3, 4]. Therefore, since the player type constructs have roots in psychological typology, our supposition is that music interests could be a promising data source to predict player type. Additionally, we also had intuitions regarding the connection between playing style with music style. For example, *Daredevils* or *Survivors* may prefer musical moods that are more aggressive (e.g., heavy metal or industrial/trance).

In order to capture the general mood of a user’s music interests, we first obtained a set of mood tags for each music interest from a popular music website Allmusic.com. Mood tags are used to describe musical qualities based on the emotional or aesthetic properties of a given band/artist (e.g., cheerful, rowdy, sensual). Next, we counted the frequency of each mood tag, and used these values as features for our classifier. Since there are 7 player type constructs, and each user can be classified to one or multiple player types, we ran a separate classifier on each type. Our classification experiments use the WEKA data mining toolkit for model learning and evaluation. Using 10-fold cross validation, we evaluated several algorithms using correlation-based feature selection method. R values for the top two classifiers are reported in Table 1.

Table 1. R values for player type prediction results.

	Seeker	Survivor	Mastermind	Conqueror	Socialiser	Daredevil	Achiever
Linear Regression	0.12	0.07	-0.05	0.11	-0.13	0.01	0.01
Decision Table	0.16	0.07	-0.02	0.04	-0.11	0.04	0.11

These initial results are not very strong and indicate the need for either an improved theoretical approach, or that music interests and player type are not directly related. One interpretation is that the relationship between everyday life personality and music interests does not necessarily match their personality or player type in online games.

5 Conclusions and Future Work

The main goal of our work is to personalize the social network gaming experience by considering a user's player type. In this paper, we provided an overview of different player types, and presented several potential data sources for automatic predictions. Finally, we performed a preliminary prediction analysis using one of these sources, i.e., music interests. While these initial results were rather weak, our future work will consist of the following: in addition to performing classification experiments using movies/television and games interests, we will investigate several features presented by Golbeck et al. [12], who successfully predict personality from a user's Facebook profile using methods that incorporate linguistic and structural features. In our future work, we also aim to use a more theoretical approach, for example, by categorizing the genres of music interests into the four music-preference dimensions identified in [13], which has already been shown to relate to personality. Lastly, we plan on analyzing the game telemetry data, which consists of several months worth of recorded in-game user behaviour.

Acknowledgements. We gratefully acknowledge the support for this work from the Mathematics of Information Technology and Complex Systems (MITACS) Accelerate Program, as well as East Side Games Studio. A very special thanks to Alex Miller for providing programming support.

References

1. Inside Social Games 2013. Facebook reveals new growth stats; 20 percent of web users play games every day. *http://www.insidesocialgames.com/*
2. Yee, N., et al. Introverted Elves & Conscientious Gnomes: The Expression of Personality in World of Warcraft. CHI'11, pp. 753-762, (2011).
3. Shen, J., et al.: Inferring personality of online gamers by fusing multiple-view predictions. UMAP'12, pp. 261-273, (2012).
4. Peever, N., Johnson, D., & Gardner, J. Personality & video game genre pref. IE'12 (2012).
5. Nacke, L.E., Bateman, C., Regan, M.L. BrainHex: A neurobiological gamer typology survey. Entertainment Computing. (2013).
6. BrainHex. *http://blog.brainhex.com/*
7. Orji, R., Mandryk, R. L., Vassileva, J., & Gerling, K. M. Tailoring persuasive health games to gamer type. CHI'2013, pp. 2467-2476, (2013).
8. Ferro, L. S., Walz, S. P., & Greuter, S. Towards personalised, gamified systems: an investigation into game design, personality and player typologies. IE'13. p. 7, (2013).
9. Tyni, H., Sotamaa, O., & Toivonen, S. Howdy pardner!: on free-to-play, sociability and rhythm design in *FrontierVille*. MindTrek'11, pp. 22-29 , (2011).
10. NPD 2010. 20 percent of the U.S. population, or 56.8 million U.S. consumers, reports having played a game on a social network. *https://www.npd.com/*
11. Collingwood, J. (2008). Preferred Music Style Is Tied to Personality. *Psych Central*. *http://psychcentral.com/lib/preferred-music-style-is-tied-to-personality/0001438*
12. Golbeck, J., Robles, C., & Turner, K. Predicting personality with social media. CHI '11, pp. 253-262, (2011).
13. Rentfrow, P.J., Gosling, S.D. The do re mi's of everyday life: the structure and personality correlates of music preferences. J. of Personality and Social Psych. pp. 1236-1256, (2003).

Modeling Mobile User Activity Planning Targets

Chad A. Williams[1] and Sean T. Doherty[2]

[1] Department of Computer Science, Central Connecticut State University, New Britain, CT, USA
[2] Department of Geography & Environmental Studies, Wilfrid Laurier University, Waterloo, ON, Canada

Abstract. Modeling the information relevant to mobile users has primarily focused on the current activity location and occasionally the immediate next activity. We suggest a more forward-looking approach may provide additional insight as to what is relevant to a user beyond their immediate surroundings. We empirically demonstrate modeling what activities are currently being planned and the time frame for how far in the future the activity will be executed can be made with high precision and recall.

1 Introduction

Modeling what is relevant to a user has been an area of continual interest in recommender systems. Within mobile systems, the majority of models do not extend beyond the immediate environment except perhaps trying to predict the next location. Some works have had current context include data such as time availability, real-time weather or traffic updates and real-time events such as accidents [1–3]. These types of approaches overlook that while sometimes people make spur of the moment decisions on what to do next in the immediate area, for a large number of activities the decisions have been made hours or even days earlier through preplanning [4]. We believe benefit can be gained by extending these models to take into account this planning behavior.

We propose to predict what is currently being planned and for how far in the future, or the *planning target*. Recognizing that a plan is being made for a specific activity at a specific time window may provide information that results in a significant improvement of determining relevant immediate recommendations and thus improving the overall mobile user's experience beyond just considering their immediate surroundings.

2 Planning Targets/Horizons

An important aspect of planning targets is the length of time between when an activity is planned and its actual execution, referred to as the *planning horizon* [5]. Planning horizon can range from far in advance, such as a visit to the doctor planned several weeks ahead, to spur of the moment, such as an urgent stop to get gas. A typical person's day has a mix of routine and actively planned activities with a more finite planning horizon. For most people, plans are continuously made and finalized throughout the day for varying planning horizons [4]. For example, a person might be making reservations for dinner the next morning, followed by an impulsive decision to grab a snack, followed by making plans for meeting a friend for lunch in an hour.

3 Methodology

For this work, we assess how well a person's planning targets can be predicted at a point in time during a person's day against empirical data derived from the Computerized Household Activity Scheduling Elicitor (CHASE) survey conducted in Toronto in 2002-03 [6]. The CHASE survey captured a detailed accounting of the activity scheduling process of adult members of 271 households over a one-week period. This included capturing the time frame for when each activity was planned, in addition to observed activity attributes such as start time, end time, location, involved persons, and category of the activity. The activities were broken down into 11 categories: active recreation; drop-off/pick-up; entertainment; household obligations; meals; basic needs; other; services; shopping; social; and work/school. The planning time frame included routine, X number of days ago, more than 2 hours before, 1-2 hours before, less than one hour before, and just prior.

To model the planning targets for a participant the activities and their planning horizons were used to construct a time line for each day. The day was discretized into 6 time segments. For each of these time segments a series of possible planning target entries were created that consisted of the combination of a specific activity type and one of three planning horizons (1-2 hours prior, more than 2 hours prior, or 1-2 days before) for a total of 33 possible planning targets for each time segment. The data set was then encoded such that for each user, if one of the 33 possible targets was indicated it was marked as being active, otherwise it was encoded as a negative example. The current activity type and user demographics of age, employment status, and gender were then selected based on information gain to train C4.5 classifiers for whether for a given time a planning target for that activity was active. For measuring prediction performance, we use the information retrieval metrics of precision and recall to evaluate the identified planning target vs. the actual planning target.

4 Results

Below we compare the precision and recall across the activity types captured in CHASE for the various planning horizons. All results reported are based on a 10-fold cross validation methodology. As Figure 1 shows by training on each of the different planning horizons the planning activity can be predicted well across all of the activity types, ranging from .80 to just over .89 precision. Interesting to note are the activities types where a longer planning horizon was more precise,

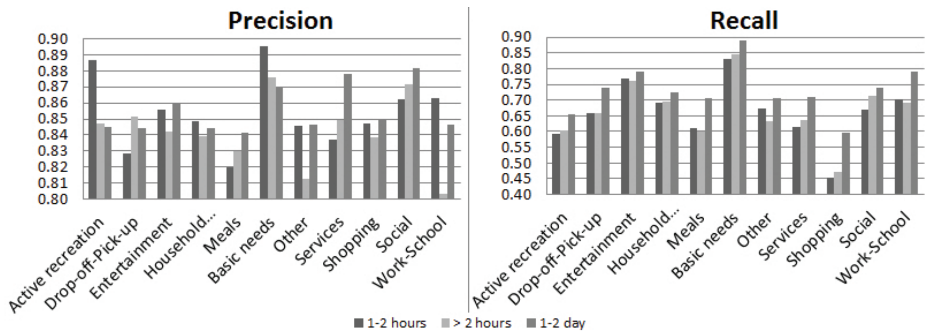


Fig. 1. Precision and recall of activity types across varying planning horizons

including 'services', 'meals', and 'social'. This is likely explained by the need for more coordination of these activities as they often involve other people, making the timing for when that coordination would occur more predictable. For example some service activities require appointments with others such as doctors, hair stylist, etc. which are more predictable when looking 1-2 days in the future.

Also interesting to note are activities such as 'basic needs' and 'active recreation' where precision was highest for the shortest planning horizon, but lower as the planning horizon became longer. A notable similarity between these activities is they often don't involve other people and require relatively little planning. The other trait observed in the C4.5 trees that stood out from other activity types was that these classifiers were more highly dependent on the immediately surrounding activities. This may reflect the planning of these activities being consistently occurring (if not triggered) after an activity type(s) occurs.

A somewhat unexpected pattern was observed for 'Work-School' that had a pronounced dip in precision in the mid-range ¿2 hour period, but higher precision in the short and longer horizons. This may reflect the haphazard way work plans made 2 hours in the future on the same day are made. However, additional investigation is needed to determine the cause for this difference with more confidence. Figure 1 displays the recall comparison of these same planning horizons.

Figure 1 displays the recall comparison of these same planning horizons. Across nearly all activity types, being able to recall a greater portion of planning activity was possible with a longer planning horizon. Intuitively this makes sense as being able to identify a higher proportion of the time when an activity type needs to be planned a day or two in the future seems easier than the same task within a day when considering the additional level of granularity being considered when windows as small as 1-2 hours are being considered.

5 Conclusions

Being able to anticipate when activities are being planned and the planning horizon for those activities may provide valuable insight in what data is the most relevant to the user. As the results demonstrate, with some basic information about the user and their current activity, the activities that are currently being planned can be modeled to identify these planning interests highly successfully. Of further interest, the predictability of this planning behavior varies across activity type and planning horizon. These findings and the potential reasons behind these differences advance the understanding of what may be relevant to a mobile user throughout the course of their activities and travels beyond just their immediate area. A broader implication of this work is that the field's current approach to identifying the information relevant to a mobile user is likely missing a significant opportunity to broaden the information returned while maintaining a high degree of relevancy. One potential way to leverage this data would be to combine the predicted activity and desired time frame with a weighting scheme based on the utility of adding different options to the user's plans through an approach such as that suggested by Horvitz et al. [7]. Because with mobile users it is rare that a single activity is considered in isolation due to travel times etc., in future work we plan to explore methods of recommending a bag/group of activities that match multiple predicted planning targets based on the combined utility of each activity need and convenience of combining trips.

References

1. Tatomir, B., Rothkrantz, L.J.M., Suson, A.C.: Travel time prediction for dynamic routing using ant based control. In: Winter Simulation Conference (WSC'09). WSC '09, Winter Simulation Conference (2009) 1069–1078
2. Hinze, A., Junmanee, S.: Travel recommendations in a mobile tourist information system. In Kaschek, R., Mayr, H.C., Liddle, S., eds.: Proceedings Fourth International Conference on Information Systems Technology and its Applications (ISTA'05). (2005) 86–100
3. Vanajakshi, L., Subramanian, S., Sivanandan, R.: Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses. Intelligent Transport Systems, IET **3**(1) (March 2009) 1 –9
4. Doherty, S.T.: How far in advance are activities planned? Measurement challenges and analysis. Journal of the Transportation Research Board **1926** (2005) 41–49
5. Mohammadian, A., Doherty, S.T.: Modeling activity scheduling time horizon: Duration of time between planning and execution of pre-planned activities. Transportation Research Part A: Policy and Practice **40**(6) (July 2006) 475–490
6. Doherty, S., Nemeth, E., Roorda, M., Miller, E.: Design and assessment of the toronto area computerized household activity scheduling survey. Journal of the Transportation Research Board **1894** (2004) 140–149
7. Horvitz, E., Koch, P., Subramani, M.: Mobile opportunistic planning: methods and models. User Modeling 2007 (2007) 228–237

GroupCollaborate2: Interactive Community Mapping*

L. Ardissono(1), M. Lucenteforte(1), A. Savoca(1), and A. Voghera(2)

(1) Dipartimento di Informatica, Università di Torino, Italy
liliana.ardissono@unito.it, maurizio.lucenteforte@unito.it
(2) DIST, Politecnico di Torino, Italy
angioletta.voghera@polito.it

Abstract. This paper presents GroupCollaborate2, a prototype Participatory GIS for the management of 3D community maps which support the shared design of public policies by offering a virtual representation of the territory and by enabling the crowdsourcing of heterogeneous types of contributions, including documents, 3D models and comments, within focus groups. The paper sketches user requirements and personalization opportunities for this type of application.

Keywords: Participatory GIS, 3D User Interfaces, Community Maps.

1 Introduction

Public Administrations use web-based crowdsourcing platforms to extend participation in public policy making beyond representatives of major stakeholders. For this purpose, Participatory GIS support the publication of information and the collection of people's feedback in geographical maps, as this is an immediate visualization format for geo-referenced data. The concept underlying such systems is the Community Map, intended as a way to represent people's view of a certain area and value attached to places or elements of their living space by gathering and presenting site-specific data; see [1].

Most participatory GIS handle bidimensional maps and collect textual feedback from users. However, the provision of virtual representations of the territory and the integration of other types of contributions, such as 3D information, are important to (i) improve the understanding of the intended effects of planned policies and (ii) enable people to make synthetic and expressive proposals. Moreover, the integration of advanced search features can enhance the exploration of the information space.

Our view is therefore that community maps should be enriched with interactive features to provide a flexible support to information sharing, access and communication. With this idea in mind we developed GroupCollaborate2, a Participatory GIS which enables the on-line sharing and editing of geo-localized documents and 3D models. The system supports the establishment of discussion groups similar to those used in participatory processes and it enables group members to communicate, share information and search for information within a community map which provides a virtual representation of the territory, of the collected proposals and of the emerging opinions about them. A

* This work was partially supported by the University of Torino, grant "Supporto intelligente e flessibile all'esecuzione di attività collaborative complesse".

key aspect of our system is the provision of information filtering features supporting the visualization of maps which reflect specific interests.

We designed GroupCollaborate2 in a user-centered development model, involving domain experts and generic internet users as these are expected to use this type of Participatory GIS. A very preliminary test with a restricted number of users has provided initial, promising results which we will further validate. Moreover, it has provided a few personalization requirements to be investigated.

2 GroupCollaborate2

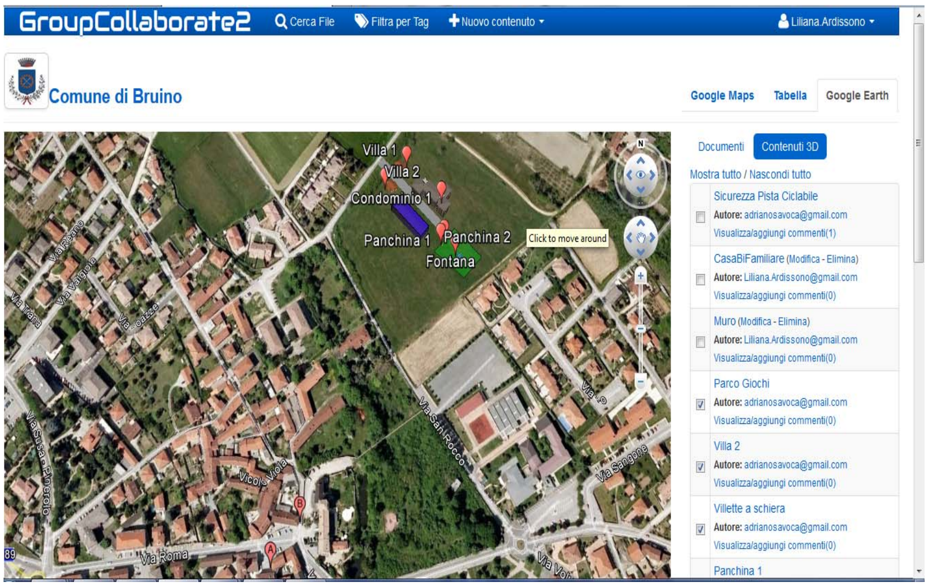


Fig. 1. Community map displaying geo-referenced documents (markers labeled with letters in the bottom part of the map) and 3D models (located in the upper part of the map).

The system enables to create open and closed discussion groups supporting the collaboration to the development of shared plans. Group members can communicate with each other and send messages to the whole group via e-mail. Within a group, users interact with a community map which represents the entry point to shared information items and which can be visualized as a bi/three-dimensional map. The system enables users to share and collaboratively edit geo-referenced documents of various types (e.g., text, drawings, spreadsheets); moreover, it supports the sharing of 3D models and the sketching of drawings (lines, polygons) in the map.

All the geo-referenced items shared within a discussion group can be visualized in the community map and users can publish comments about them. Moreover, the system supports the dynamic generation of community maps reflecting individual information needs by offering a tag-based classification and filtering of items, as well as the possibility to search documents by name or by included words.

Figure 1 shows the user interface of the system (in Italian) and in particular the three-dimensional community map for a sample discussion group named “Bruino”.

– The top of the page shows the links supporting the content-based search for documents (“Cerca File”), the tag-based filtering of items (“Filtra per Tag”) for restricting the set of items visualized in the map. Moreover, it shows the (“Nuovo contenuto”) link for: (i) creating or uploading a document; (ii) uploading an existing 3D model from a repository (e.g., a KMZ model), or (iii) drafting a new 3D item in the map. Items have to be enriched with metadata (title, author, description) for presentation purposes and can be tagged according to user-generated tags.
  New 3D items can be drafted by means of an editor which enables the user to draw broken lines and polygons (or to resize them) by double clicking their vertexes in the map. The editor also allows the selection of the color, height and orientation of items, and the thickness of lines.
– The community map shows the items satisfying the search criterion (all items, or the selected ones). It displays documents as markers; e.g., A and B at the bottom of the map in Figure 1. Moreover, it displays 3D models as shapes; e.g., the figure shows, among the other, two houses (“Villa 1” and “Villa 2”, uploaded as 3D models), a blue polygon drafted on the map to represent a building, and a green area delimiting a playground with benches and fountain.
  Each marker/3D model can be clicked to view its metadata. Moreover, markers can be clicked to view the content of the associated documents which can be edited or not, depending on their format and permissions. Furthermore, maps can be zoomed.
– The right portion of the page displays items as a checkable list which allows the user to further refine the elements to be shown. For each item, a row reports (i) its metadata, (ii) a link to revise or remove the item (if the user has permission), and (iii) a link which displays the number of comments associated to the item and that enables the user to view/add new ones.

Domain experts who tested the system strongly appreciated the integration of search, filtering, access, modification and commenting features in a community map, as this enables them to analyze and discuss ongoing proposals using a unified environment which provides immediate visualization of geo-data. Moreover, they suggested to introduce new functions. For instance, they proposed to enable users to handle personal views on content based on concept selection (e.g., only scholastic buildings, sport and leisure facilities, etc.) and on the role of users in participatory processes (e.g., generic citizen, Public Administration, etc.), as well as to introduce subgroup management features aimed at supporting focused discussions among selected representatives of the population. These aspects open research paths on data representation (to classify content by concepts), user modeling (to understand the user’s interests and model user groups) and manual/automatic maps adaptation to derive personal views focused on specific interests. Moreover, there are interesting research avenues in the analysis of people sentiment towards specific public policies and about participation culture.

From the viewpoint of usability, the user interface has a neat layout to address basic W3C accessibility guidelines. Various features could be added to support different types of interaction with users. E.g., the design of simple 3D items might be supported by introducing libraries of shapes to be dragged and dropped. Moreover, sophisticated tools might be proposed to draft complex polygons with irregular shapes; e.g., see [2].

GroupCollaborate2 is a Java web-based application and uses open APIs for the integration of various functions; e.g., Google Map and Google Earth APIs for the representation of the community map; the OAuth protocol for authenticating users, and the Google Drive APIs for data storage. The user interface of the system is developed in HTML5 using JavaScript for interacting with the maps and AJAX to speed up the visualization of the user interface. Moreover, the Google Earth plug-in is used to simulate 3D environments in the user's browser. As the plug-in for mobile phones is not available, GroupCollaborate2 is not accessible from mobile devices. However, device-dependence should be overcome in the next future thanks to the integration of HTML5 with the WebGL standard for graphic user interfaces, currently under definition.

3 Related Work

A few Participatory GIS projects support 3D information management. E.g., LIVE+GOV (http://liveandgov.eu/) combines AR and VR techniques with social networks in order to enable internet users to upload and receive geo-localized information about a city, as well as to participate in polls and discussions. Min Stad (http://minstad.goteborg.se/minstad/index.do) integrates GIS with social networks enabling users to upload 3D contents and to publish comments. In comparison, GroupCollaborate2 lacks the support to deliberation provided by polls and a connection to existing social networks. However, it improves crowdsourcing support by enabling users to share and collaboratively edit heterogeneous types of contents in thematic discussion groups with consequent information hiding. Moreover, it supports tag-based and content-based search for information thus enabling the generation of customized community maps.

4 Conclusions

GroupCollaborate2 is an attempt to integrate community mapping, communication, information sharing and filtering in a Participatory GIS supporting focus group discussion. While the current prototype is devoted to basic user collaboration, the next steps in its development will focus on extending it with personalization features supporting the provision of adaptive community maps. We thank Giuseppe Scaramuzzino for his work on the first version of the system.

References

1. Ellul, C., Francis, L., Haklay, M.: Engaging with local communities: A review of three years of community mapping. In: Urban and Regional Data Management, UDMS Annual 2011 - Proc. of the Urban Data Management Society Symp. 2011, Delft, Netherland (2012) 165–177
2. Isikdag, U., Zlatanova, S.: Interactive modelling of buildings in google earth: A 3d tool for urban planning. In: Developments in 3D Geo-Information Sciences Lecture Notes in Geoinformation and Cartography, Berlin Heidelberg New York (2000) 52–70

YummyKarachi: Using Real-Time Tweets for Restaurant Recommendations in an Unsafe Location

M. Atif Qureshi[123], Arjumand Younus[123], Muhammad Yousuf[34], Abdul Moiz[34], Muhammad Saeed[34], Nasir Touheed[3], Colm O'Riordan[1], and Gabriella Pasi[2]

[1] Computational Intelligence Research Group, Information Technology, National University of Ireland, Galway, Ireland
[2] Department of Informatics, Systems and Communication, University of Milan Bicocca, Milan, Italy
[3] Web Science Research Group, Faculty of Computer Science, Institute of Business Administration, Karachi, Pakistan
[4] Department of Computer Science, University of Karachi (Umaer Basha Institute of Information Technology)
{muhammad.qureshi@nuigalway.ie}

Abstract. In this paper the YummyKarachi system is presented; it is a combined restaurant and safe routes recommender system for an unsafe city such as Karachi. YummyKarachi utilizes a hybrid model combining service recommendations with route recommendations in a map-based user interface. It utilizes tweets arriving in real-time to search for danger zones within the city, and it generates restaurant recommendations while taking into account security conditions of the city.

1 Introduction

Recently Karachi was declared as the most dangerous megacity of the world by the US Foreign Policy Magazine[5], and rightly so as the city has often been struck by extreme incidents of shootings and blasts over recent years. The nature of dangers in this metropolis of Pakistan is peculiar in that danger in one part of the city does not affect other parts of the city due to the huge size and population of the city. In this scenario, citizens are often confronted with the problem of finding out the security conditions of various zones within the city, and this is particularly the case when planning how to reach restaurants[6]. Currently the citizens of Karachi rely on local news services and sms technology to receive updates about a particular location, and to be informed on the overall security situation of the city. However, these solutions do not provide updates in real-time and hence, they cannot be relied upon.

[5] http://www.foreignpolicy.com/articles/2013/09/03/cooking_in_karachi_meth_pakistan?page=0,0
[6] Note that this does not imply insensitivity on the part of citizens of Karachi but a necessary outcome of the huge size of the metropolis.

As a solution we propose YummyKarachi, which is a restaurant recommendation system that provides recommendations of interesting places to eat/drink while taking into account the security situation of the entire city. The proposed system processes tweets arriving in real-time to discover the dangerous zones of the city. Furthermore, the system implements a novel approach that utilizes various rating levels for the restaurant recommendation problem.

2 Related Work

Tourism serves as a primary application area when it comes to mobile environments, and an increasing number of mobile services have been proposed to aid the traveller before, during and after the travel [7]. In this context, the information overload problem tends to make it more and more difficult for travellers to find the right information that is needed to complete a particular task (e.g., choosing a movie, or planning a trip); hence, mobile recommendation systems offer information filtering and decision-making support in such situations [5]. The tasks supported by mobile recommendation systems involve tourist recommendations including service recommendations (i.e., restaurants, transportation services etc.) [1, 4], route recommendations [6], and information recommendation [3]. Within YummyKarachi, we propose a hybrid model that combines service recommendations with route recommendations in a map-based user interface. Recently, the focus of route recommendation services has moved towards the suggestion of safe routes, and as an example Kim et al. propose to use the sentiment expressed in tweets for determining the crime hotspots thereby recommending routes that exclude the crime areas [2]. Similar to Kim et al., we propose the use of tweets for detection of danger zones in a city, but instead of utilizing sentiments we use explicit *"alert terms"* extracted from tweets.

Fig. 1: A Snapshot of Map-Based UI Depicting Danger Zones

3 Methodology and System Overview

YummyKarachi includes a back-end tweet processing module that scans tweets for detecting what we call *alert terms*. These *alert terms* are extracted by mining tweets related to terror incidents. To the aim of defining a base of "alert terms" we have first used the Twitter search API to search for tweets related to "Karachi"; note that we search for tweets with hashtag #khialerts since citizen journalists and social media activists use it for reporting about events in Karachi. This step has then been followed by a manual annotation of tweets relating to a danger event such as a shooting incident or bomb blast with the annotation marking whether or not the tweet is relevant to the event[7]. We have then used the tweets marked as relevant to extract *alert terms* by asking the annotators to provide the strongest term indicative of a danger, and the terms chosen by all three annotators have been used as *alert terms*. A similar methodology is followed for terms indicating a traffic blockage. Unlike the approach in [2], YummyKarachi does not rely on geo-tagged tweets[8]; instead its tweet-processing module makes use of the Twitter API to fetch tweets with hashtag #khialerts along with tweets issued by security analysts and journalists of Karachi, while finally scanning fetched tweets for names of Karachi zones extracted from Wikipedia. Corresponding to each zone a danger level is computed by counting the number of tweets in which any of the *alert terms* occur. The danger zones are then shown in a map-based user interface with varying levels of intensity of the red color (see Figure 1). At the same time, YummyKarachi also shows the zones with heavy traffic, and this feature is included to further facilitate the citizens of Karachi. Finally, YummyKarachi recommends a navigation path to the user's recommended restaurants by taking into account the danger zones i.e., it recommends the routes with no danger zones or with the minimum amount of danger zones.

The recommendation module of YummyKarachi introduces a novel three-level rating module to produce restaurant recommendations. We adopt a standard collaborative filtering methodology and we apply the matrix factorization module of singular value decomposition to produce restaurant recommendations. However, the user-item ratings are combined on three levels based on the following user preferences: i) cuisine, ii) food item, and iii) restaurants. We use a linear combination for the user similarity scores generated from each of the above three ratings. The categories and food items available in each restaurant are obtained by crawling the menus of these restaurants that are available on food portal web sites. This offers the opportunity to take into account user preferences on a multi-granular level. As an example, consider a group of users who express preference for the cuisine type "Italian" and the food items of "pizza" and "pasta"; a user who expresses ratings similar to this group is likely to be recommended a restaurant that serves authentic Italian cuisine thereby leading to a higher user satisfaction.

[7] A total of three annotators were used.

[8] This is done due to the scarce amount of geo-tagged tweets available.

4 Preliminary Evaluations

The demonstration video for YummyKarachi is available at http://bit.ly/1seBXLJ. We performed a set of preliminary evaluations for YummyKarachi, where we asked 10 users to utilize the application for restaurant recommendations. Our evaluation questionnaire required the users to express their level of satisfaction with the danger zones shown by YummyKarachi along with their level of satisfaction with the generated recommendations and likert scales from 1 to 5 were used for the survey. We asked the users to rate the satisfaction level of danger zones with respect to the truth value of the information which we asked them to verify through news sites. The mean likert scale values for level of satisfaction with the danger zones is 4.2 while for generated recommendations it is 3.9. This reflects that generally users consider YummyKarachi as a useful source of information for planning their restaurant trips.

5 Future Directions

As a future work we aim to incorporate within YummyKarachi the ability to predict likely zones that may become dangerous in the near-future. We aim to do this through incorporation of a predictive text mining module within YummyKarachi. Furthermore, we plan to extend the YummyKarachi framework to recommendations of other location types in order to enable citizens of Karachi to plan all their travels in a safe manner.

References

1. M. D. Dunlop, B. Elsey, and M. M. Masters. Dynamic visualisation of ski data: A context aware mobile piste map. In *Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '07, pages 375–378, New York, NY, USA, 2007. ACM.
2. J. Kim, M. Cha, and T. Sandholm. Socroutes: Safe routes based on tweet sentiments. In *Proceedings of the 23rd International Conference on World Wide Web Companion*, WWW '14 Companion.
3. H. J. Lee and S. J. Park. Moners: A news recommender for the mobile web. *Expert Systems with Applications*, 32(1):143–150, 2007.
4. M.-H. Park, J.-H. Hong, and S.-B. Cho. Location-based recommendation system using bayesian users preference model in mobile devices. In J. Indulska, J. Ma, L. Yang, T. Ungerer, and J. Cao, editors, *Ubiquitous Intelligence and Computing*, volume 4611 of *Lecture Notes in Computer Science*, pages 1130–1139. Springer Berlin Heidelberg, 2007.
5. F. Ricci. Mobile recommender systems. *Information Technology & Tourism*, 12(3):205–231, 2010.
6. G. Tumas and F. Ricci. Personalized mobile city transport advisory system. In W. Hpken, U. Gretzel, and R. Law, editors, *Information and Communication Technologies in Tourism 2009*, pages 173–183. Springer Vienna, 2009.
7. H. Werthner and F. Ricci. E-commerce and tourism. *Commun. ACM*, 47(12):101–105, Dec. 2004.

A value-sensitive mobile social application for families*

Alex Kayal[1], Willem-Paul Brinkman[1], Hanna Zoon[2], Mark A. Neerincx[1], and M. Birna van Riemsdijk[1]

[1] Delft University of Technology, The Netherlands
{a.kayal,w.p.brinkman,m.a.neerincx,m.b.vanriemsdijk}@tudelft.nl
[2] Fontys Hogeschool Eindhoven, The Netherlands

Abstract. We investigate how to create a mobile social application to support families with elementary school children, assisting them in exploring their social and geographical environment. While existing social applications provide this functionality to some extent, these kind of applications can negatively impact important user values such as freedom and privacy while intending to promote others such as family security. We propose as a solution that users express rules of behavior to the application using *norms* as concretized values. Norms can thus be used to produce tailored behaviour capable of fulfilling certain values while posing minimal risk to others. We demonstrate a prototype of a mobile, socio-geographical support application we built based on this concept.

1 Introduction

Usage of social media platforms such as Facebook, Foursquare and Twitter as well as mobile applications for supporting family life such as Life360[3] have become an integral part of our life. In our research we investigate how to create a mobile social application to support families with elementary school children, assisting them in exploring their social and geographical environment (for example, by helping them staying safe, making new friends or getting to know their neighbourhood).

Existing social applications such as mentioned above already provide functionality that supports these tasks to some extent. However, research in value-sensitive design and ethics in technology has shown that while this kind of applications may promote a number of user values such as family security and comfort, they can negatively impact equally important values such as freedom, responsibility and privacy [3]. For example, when parents use GPS tracking to know where their children are, this may be beneficial for family security, but it negatively impacts children's privacy and freedom. On the other hand, ad hoc sharing of locations such as done through Foursquare does not provide the potential safety benefits that GPS tracking would provide, while it does promote freedom.

We propose to address this issue by making social applications more *adaptive*. Recent works have proposed to make social applications more adaptive through collecting

* This publication was supported by the national Dutch program COMMIT.
[3] An application that allows families to locate eachother 24/7 on a map, see https://www.life360.com.

and analysing user data (for example, see [5,2]). We take a complementary approach, proposing that users can express rules of behavior to the social application using *norms*. Norms have been proposed (see [1] for an overview) to regulate the behavior of autonomous agents for achieving a better overall system performance, inspired by the way social norms regulate people's behavior in society. Research in philosophy and normative systems shows that values can be promoted and demoted by norms [8,4]: since norms are considered action guiding (by obligating or forbidding actions), they can be used to produce a tailored behaviour capable of fulfilling certain values while posing minimal risk to others.

In this paper we describe our prototype of a mobile, socio-geographical support application that we built based on this concept.

2 System description

This prototype runs on the Android platform and it permits its users (elementary school children and their parents) to share *check-ins* in certain locations with other users of the system, similar to Foursquare. This feature was selected based on the analysis of previous user data [6], where values such as *family security*, *social recognition*, and *independence* were found to be most relevant for this target group. Knowing where family and friends are is connected with these values.

2.1 Basic preferences, creating locations, and checking-in

In a similar way as on popular social platforms, with our prototype users can place other users of the system in the user group *family* or *friends* or in neither, in which case the application places them in the group *others*. Users can select with which groups they share their check-ins, and from which groups they receive shared check-ins. Users can add or remove users from either group, and change sharing and receiving preferences at any time.

Users can create locations in two possible ways: 1) through selecting a specific point (corresponding to a GPS position) on an integrated Google map, and then assigning to it a name of their choice, and 2) through detecting the current position automatically if a GPS signal was available, and then assigning a name. In both cases, a location is added to a list of available user locations, defined by a name, a GPS position, and a square area of a side length of 50 meters centred around that GPS point. Locations can be removed by the user at any time.

When a user would like to check-in, the list of locations that fall within a radius of 300 meters (according to the currently detected GPS position) are displayed, with the option of adding a location using the second method described above, in case the current location was not yet on the list. The user then can select a location, and confirm their check-in, which will be shared with the users that belong to the groups which our user is sharing with, according to the basic preferences in the previous subsection. Users who, accordingly, receive this shared check-in, will get a pop-up with the sharer's name and location information (viewable also on an integrated Google map), if they have selected to view check-ins from the group to which the sharer belongs in their own basic preferences. An “event log” is available, which shows a user’s own latest check-in information, as well as the five most recent check-ins seen from others.

2.2 Social commitments

As discussed, our prototype allows for additional, norm-based behavior customization through norms. While basic preferences are set by the user of an application, norms can come from others in the user’s social context. For example, a parent may want to make an agreement with a child about when the child’s location is shared with the parent and when check-ins can be received by the child. Models for such agreements have been studied in research on normative systems. In particular, we draw from the social commitments framework in [7] to create the following commitment model that we use for expressing agreements between different users about the behavior of the application:

A commitment has a source (creator of the commitment), a target (who is asked to comply with the commitment), a triggering condition that activates the commitment, a normative effect (an obligation or a prohibition of sharing or viewing a check-in, from someone or a group of people), and the deadline by which (in the case of an obligation), this obligation should be fulfilled.

In our prototype, this translates to the following feature: a user can create an agreement (i.e., a commitment) with another user consisting of a specific normative effect (to share or view a check-in from one or a number of users) if a certain condition (based on time or geographical location) is active, and the target of the commitment accepts it. For example, a parent x (source) can create the following commitment with his/her child y (target): 1) I want my child to share his/her check-ins with me (normative effect), if s/he enters school (triggering condition). Another example would be a parent x (source) creating the following commitment with his/her child y (target): 2) I want my child to “not receive” check-ins from the group “friends” (normative effect) after 9 pm (triggering condition).

When the source user creates a commitment, it is sent to the target user, who can either directly accept it, or “decide later”. In case the latter option was chosen, the target user can later decide whether to accept or reject the proposed commitment. Users can, at any time, review the list of commitments they created or received, delete commitments they created or received, and accept received commitments that are still pending. A user action such as accepting or deleting a commitment notifies the other user involved with that action.

In this version, conflicts between basic preferences and an accepted, active commitment are solved in favor of the commitment. For example, if parent x is in child y’s family list, and child y opted in their basic preferences to “not share check-ins with family”, accepting commitment 1) above means the child’s check-in will be shared with parent x if they enter school. Similarly, conflicts between two accepted, active commitments would be solved in favor of the commitment most recently accepted. We refer to the literature for research on reasoning with norms, e.g., [9].

In the current version of the prototype, the deadline “as soon as possible” is used for all obligation-type commitments. Also, in this version commitments do not expire automatically, but they can be removed manually by users. In future work we will investigate extensions that add expressivity to the commitment model with respect to deadlines and expiration of commitments. Moreover, in this prototype commitments are created explicitly. One may also consider adding a component for learning norms, which could for example learn by observing user behavior that no check-ins should be received during dinner.

A 3-minute tutorial video (with subtitles) can be seen at `http://bit.do/ePartner`.

3 Discussion

The use of social commitments places our prototype application somewhere inbetween Foursquare (which is similar to the basic check-in functionality of our prototype) and Life360 (where location information is shared continuously). Its flexible commitment model allows parents and children to make agreements on sharing location information in a targeted way, tailored to that particular family in a particular situation. We hypothesize that in this way, the application can promote family security without violating (or with minimal impact on) a child's freedom and privacy. In future research we will perform a user study to test this hypothesis.

References

1. G. Andrighetto, G. Governatori, P. Noriega, and L. van der Torre, editors. *Normative Multi-Agent Systems*, volume 4 of *Dagstuhl Follow-Ups*. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2013.
2. S. Berkovsky, J. Freyne, and G. S. 0003. Personalized network updates: Increasing social interactions and contributions in social networks. In J. Masthoff, B. Mobasher, M. C. Desmarais, and R. Nkambou, editors, *UMAP*, volume 7379 of *Lecture Notes in Computer Science*, pages 1–13. Springer, 2012.
3. A. Czeskis, I. Dermendjieva, H. Yapit, A. Borning, B. Friedman, B. Gill, and T. Kohno. Parenting from the pocket: value tensions and technical directions for secure and private parent-teen mobile safety. In *Proceedings of the Sixth Symposium on Usable Privacy and Security (SOUPS'10)*, pages 15:1–15:15. ACM, 2010.
4. S. Hansson. Norms and values. *Crítica*, 23(67):3–13, 1991.
5. M. Kaminskas and F. Ricci. Location-adapted music recommendation using tags. *User Modeling, Adaption and Personalization*, pages 183–194, 2011.
6. A. Kayal, W.-P. Brinkman, R. Gouman, M. A. Neerincx, and M. B. van Riemsdijk. A value-centric model to ground norms and requirements for epartners of children. In *Coordination, Organization, Institutions, and Norms in Agent Systems IX*. Springer-Verlag, 2014. To appear.
7. M. Singh. An ontology for commitments in multiagent systems. *Artificial Intelligence and Law*, pages 97–113, 1999.
8. T. van der Weide. *Arguing to motivate decisions*. PhD thesis, Utrecht University, 2011.
9. M. B. van Riemsdijk, L. Dennis, M. Fisher, and K. V. Hindriks. Agent reasoning for norm compliance: a semantic approach. In *Proceedings of the twelfth international joint conference on autonomous agents and multiagent systems (AAMAS'13)*, pages 499–506. IFAAMAS, 2013.

Turning Learners into effective better Learners: The Use of the askMe! System for Learning Analytics

Christian Saul[1] and Heinz-Dietrich Wuttke[2]

[1] Data Representation and Interfaces Group,
Fraunhofer Institute for Digital Media Technology (IDMT), Ilmenau, Germany
christian.saul@idmt.fraunhofer.de
[2] Integrated Communication Systems Group,
Ilmenau University of Technology, Ilmenau, Germany
dieter.wuttke@tu-ilmenau.de

Abstract. Learning analytics is defined as the measurement, collection, analysis and reporting of data about learners in order to optimize their learning. It was recognized very early that the field of learning analytics offers promising possibilities for education and assessment. In this paper, the interactive and personalized e-assessment system *askMe!* is presented that makes use of learning analytics in order to turn learners into effective and better learners. The paper also shows how the system addresses future challenges in the field of learning analytics. Finally, a brief summary of the use and evaluation of the system in a real-life setting completes the paper.

Keywords: e-assessment; adaptivity; personalization; interactivity; learning analytics; askMe!

1 Introduction

During the last decade, Learning Analytics (LA) has emerged as a significant area of research in the field of technology-enhanced learning. It has been considered as one of the fastest growing areas of research related to education and technology [1]. In contrast to educational data mining, LA exclusively focuses on the learning process and tries to collect, manage, interpret and purposefully use (large) data sets in education. In this way, it offers promising possibilities for education and assessment. The focus of this paper is the presentation of an e-assessment system that makes use of LA in order to support and improve students' individual learning process.

2 Learning Analytics

According to the 1st International Conference on Learning Analytics and Knowledge (LAK'11), LA is defined as the "measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs". Simply put, the broad goal of LA is assisting students' learning process by giving feedback. The feedback is based on the results obtained by analyzing the data gathered by monitoring and measuring the individual learning process. It was recognized very early that the field of LA offers promising possibilities for education and assessment. LA provides a variety of information that can be used to adapt the students' learning and/or assessment to their individual strengths and weaknesses. Moreover, it can also supports students' self-regulation of learning by providing feedback in forms of visual data analyzes or visualizations. This allows students taking control of and evaluating their own learning process and behavior.

There are a lot of challenges that drive the research in this field. This includes the appropriate collection, protection and use of the large data sets, but also issues related to data protection, ownership and privacy are of particular relevance [1]. In 2012, Ferguson [2] identified four significant challenges that research in this field must address: Due to the fact that LA emerged from the fields of analytics and data mining, the first challenge is to build strong connections with the learning science. The second challenge is to support and incorporate a wider range of data sets. The third challenge is to focus on the perspectives of the learner and finally, the fourth challenge is to take decisions regarding the ownership and stewardship of the data into account.

3 The askMe! system

askMe! is a web-based e-assessment system that covers the whole life-cycle of e-assessments starting from creating questions, presenting them to the students up to preparing the results and presenting them to teachers, tutors, etc. The questions and tests can consider individual aspects so that e-assessments and their feedback can perfectly be tailored to students or groups of students [3]. Moreover, the author of the adaptive tests is not limited to traditional question types such as multiple-choice, but can use Interactive Content Objects (ICOs) to create sophisticated (interactive) e-assessments [4]. The latter aspect takes into account the assumption that learning is the result of interaction and more specifically, the result of engagement with the subject matter [5]. In order to deal with the different ICOs located elsewhere in the Web, a communication mechanism based on the Experience API (xAPI) specification has been developed. In this way, the *askMe!* system allows integrating and mining a wide range of data sets from multiple sources and thus, already addresses the second above-mentioned main future challenge of LA.

When a student has completed a test in the *askMe!* system, he/she will not be confronted with an abstract score, but will get detailed feedback on his/her strengths and weaknesses, which allows him/her to efficiently address specific deficits afterward. This information is presented in his/her *knowledge dashboard* (cf. Figure 1). This component not only provides students with a tabular and graphical overview of his/her testing results, but also with a detailed overview

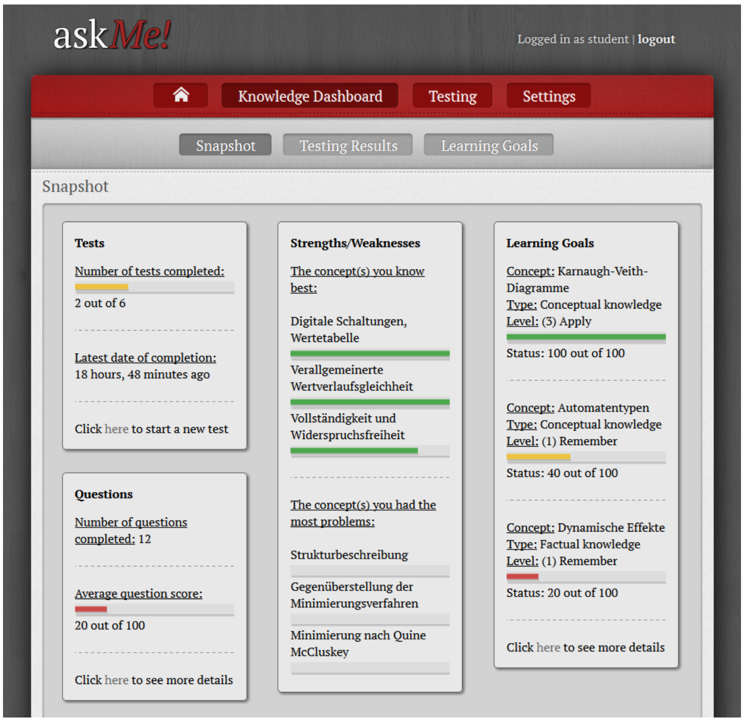


Fig. 1. askMe! knowledge dashboard (snapshot)

about his/her knowledge level according to the topics addressed by the respective question. In this way, the *askMe!* system assists students' learning process by giving feedback in form of expressive visualizations. This also increases students' self-awareness significantly. In addition, the system considers their needs when reporting and visualizing analytics data. This is done by letting students decide whether they want to have a compact view of the information at a glance (dashboard snapshot) or whether they want to have all information in detail. The system captures this behavior and automatically presents the view mostly selected by the respective student next time. With respect to the dashboard snapshot, it is also planned to let students freely decide which widgets they want to see and how they want to arrange them. Consequently, the *askMe!* system provides the degree of individualization and personalization needed to address the third main future challenge of LA.

The presentation of statistics in *askMe!* is not limited to students, but is also provided to authors, tutors, etc. The information for this user group is presented in *user and test statistics*. This component presents an overview of students' testing results to (adaptive or non-adaptive) tests as well as their individual learning progress. In addition, the system uses statistical techniques in order to predict, which students are struggling with the content. Tutors and teachers can use this indicator (i.e., green, yellow and red) to intervene either online or face-to-face. Basis of the predictions are students' grades, but also their learning progress over time according to the learning goals set.

In order to get feedback whether the *askMe!* system is applicable for its intended use, it was used and tested in a real-life setting at the Ilmenau University of Technology [6]. The study was performed in the course digital systems design for about 80 students. The system was made available for students preparing for their final exam. All in all, 101 tests were completed and 714 question were answered over a period of 12 weeks. As a result, it can be stated that the system was rated very well by the test persons. In addition, the study has shown that the average grade of students that used the *askMe!* system for test preparation was much better than students who did not use system. Furthermore, the failure rate of students that did not use the system was four times higher than students that used the system for test preparation.

4 Conclusion

This paper has presented the *askMe!* system, a web-based e-assessment system that has been developed at the institute of the main author in cooperation with the institute of the co-author. The system aims at evaluating, but also supporting and improving students' individual learning process by providing real-time feedback to students and tutors/teachers. It has also been shown that the system addresses two main key challenges in the field of LA and thus, provides a decisive contribution to the research in this field. Finally, the use and evaluation of the system in a real-life setting has proven the educational benefit of LA in general and of the *askMe!* system in particular.

References

1. P. Broadfoot, S. Timmis, S. Payton, A. Oldfield, and R. Sutherland, "Rethinking Assessment - Learning analytics and technology enhanced assessment (TEA)," University of Bristol, Tech. Rep., 2012.
2. R. Ferguson, "Learning analytics: drivers, developments and challenges," *International Journal of Technology Enhanced Learning*, vol. 4, no. 5/6, pp. 304–317, 2012.
3. C. Saul and H.-D. Wuttke, "E-Assessment meets Personalization," in *Proceedings of the IEEE Global Engineering Education Conference (EDUCON 2013)*, 2013, pp. 200–206.
4. ——, "Towards a high-level Integration of interactive Tools with E-Assessments," in *Proceedings of the 12th IEEE International Conference on Advanced Learning Technologies (ICALT 2012)*, 2012, pp. 597–598.
5. K. Mattingly, M. Rice, and Z. Berge, "Learning analytics as a tool for closing the assessment loop in higher education," *Knowledge Management & E-Learning: An International Journal*, vol. 4, no. 3, pp. 236–247, 2012.
6. C. Saul and H.-D. Wuttke, "Assessment 3.0 meets Engineering Sciences," in *Proceedings of the 16th International Conference on Interactive Collaborative Learning (ICL 2013)*, 2013.

Making It Game-Like: Topolor 2 and Gamified Social E-Learning

Lei Shi and Alexandra I. Cristea

Department of Computer Science, University of Warwick
CV4 7AL, Coventry, United Kingdom
{lei.shi, acristea}@dcs.warwick.ac.uk

Abstract. This paper briefly introduces Topolor 2, a social personalised adaptive e-learning environment with novel gamification features, aiming at reducing undesirable 'noise' effects of social interaction and at further improving the learning experience. The goal of this paper is to showcase the main gamified social interaction features.

1 Introduction

Topolor is a social personalised adaptive e-learning environment, designed to address a particular aspect of adaptive systems and adaptive hypermedia [2], which is that of *social interaction for adaptive e-learning systems*. It is under iterative implementation and evaluation. The first version of Topolor [10] was launched in November 2012, and it has been used as an online learning environment for undergraduate and postgraduate students in Western & Eastern Europe, and Middle Eastern universities. It was designed based on the hypothesis that *extensive social features, personalised recommendations and Facebook-like appearance would make a system more familiar to the learners, and subsequently increase the learning experience*. Evaluations were conducted via real-life learning sessions, targeting various perspectives (e.g., learning behaviour patterns [12]), based on usage data, questionnaire and oral feedback, aiming at investigating the granularity of social interactions and how adaptations can support these, towards the ultimate goal of enhancing learning experience and efficiency. The evaluation results illustrate high satisfaction from the students, as well as a high level of student engagement [9] which indicates that our approach is promising. Nevertheless, some side effects of the extensive social interaction features were also detected, such as 'noise', i.e., students' off-topic conversations through 'chitchat' socialisation [11]. Undeniable is the important role that the informal 'chitchat' plays in motivating and scaffolding peer learning [7] in a social e-learning context - positive social dialogue, e.g., greetings, may help students to relieve anxiety or promote participating in discussions, but reducing side effects whilst maintaining a reasonable scale of informal 'chitchat' - and thus improving learning experience and efficiency in a social e-learning context - is still a crucial challenge to address.

Gamification is "the use of game design elements in non-game contexts" [4] to engage users and promote desirable behaviours. It increasingly attracts researchers' attention in the education intelligence area, and its benefits have been reported in the recent literature [5, 13]. Considering that gamification and social e-learning have various mechanics in common, such as collaboration, discovery, achievement, loyalty and virality, their appropriate combination may enhance e-learning environments. Therefore, this research introduces a specific blend of light gamification, aimed at reducing side effects and further improving learning experience and efficiency. This paper focuses on the design of these gamified social interaction features.

2 Main Gamified Social Interaction Features

We adopt a *light* gamification approach that applies self-determination theory (SDT) [8] and flow theory [3] to promote intrinsic motivation in existing social e-learning environments, rather than a full-fledged approach that may "over-gamify" the existing mechanics, or even replace the social learning communities that have already formed.

2.1 Peer-reviewed Posting

Topolor 2 introduces a new blend of powerful tools for querying, sharing and filtering the learning resources, as shown in Fig. 1 (a) and (b). It has finer categories especially for sharing, i.e., text, image, quote, link, audio and video (e.g., in Topolor 1, students can only 'share a learning status' as a text). In fact, these categories are widely used in Web 2.0 tools, e.g., Tumblr, and some online teaching/learning platforms recommend teachers to use these external Web 2.0 tools for delivering learning materials, but it is seldom that they are seamlessly integrated in an e-learning system. Additionally, students can express like/dislike for any of these categories of posts, including for comments on a post and the answers to a question. This was introduced for quality control, i.e., to prevent students from abusing social interactions, e.g., by writing an irrelevant comment on a course video. This also encourages them to improve their reputation - a part of a user model, i.e., a learner with higher reputation has benefits, e.g., greater weight in determining peer posts' quality. Additionally, posts can be filtered and sorted based on their perceived quality (as the difference between 'like' and 'dislike' votes from students). More importantly, this method can potentially improve the *quality* of user modelling by filtering out low quality data, as well as reduce the burden of the user modelling process, and thus improve its *efficiency*.

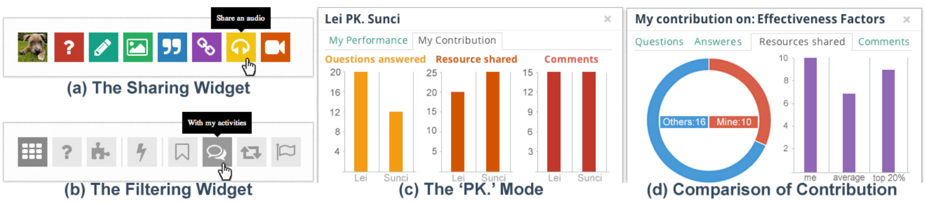


Fig. 1. User Interfaces in Topolor 2

2.2 Visualised Social Status

Topolor 2 additionally provides student profile pages as another information and interaction ‘hub’, which leads to various features of recommendation, adaptation, personalisation and social interaction. For example, by clicking on a student’s avatar in a post list, a pop-up view appears, containing statistics of her learning status, a shortcut to send her a message or to go to her profile page to see her learning status and activities in detail. In a profile page, several gamified social interaction features are provided. For instance, by clicking on the button ‘PK.’ (‘Player Killer’, a naming convention taken from games), a pop-up view shows the comparison of performance (e.g., quiz score trends) and contribution (e.g., the number of questions answered, as shown in Fig. 1 (c)) to the learning community between its current viewer and the profile page’s owner. Apart from the student profile pages, the graphic and interactive view of contribution and performance allows students to operate *multi-context comparisons* (i.e., in the context of a specific course or a specific topic) and *multi-group comparisons* (i.e., compare to another learner, top 20% learners, or all other learners), as shown in Fig. 1 (d). This can capture learner motivation by triggering competitive instincts [6].

2.3 Adaptive Leaderboard

Leaderboards are embedded into different contexts. They adapt to the students and the learning content by adjusting the way of ordering and displaying student information. For instance, in a course page, the students can be shown based on how many topics in this course they have learnt, while in a topic page, they can be shown based on how many questions related to the topic they have answered correctly. Students can adjust the order, and Topolor remembers their preference for the next time. Each item on the leaderboard can be separately viewed as a student ‘info-card’, containing her learning status information, buttons for sending her a message or seeing her profile page. Additionally, the information on the item is device-adaptive, e.g., for a certain size of the browser, smaller icons replace big ones and text information. In Topolor 2, leaderboards create a sense of community and provide opportunities for students to directly interact with others and compare their learning progress to others, because students see their status publicly and can be instantly recognised.

3 Conclusion and Future Work

In this paper, we have presented the main gamified social interaction features in Topolor 2, a social personalised adaptive e-learning environment. We have adopted a specific blend of *light gamification* approach that applies motivational theories and symbiotically builds gamification mechanics upon social interaction features, in order to promote intrinsic motivation in existing social e-learning environments, without replacing the social e-learning community that has already formed.

The preliminary evaluations using online survey were performed and showed both high usability and appreciation of the new gamified social interaction features introduced (the *SUS* [1] *score* of Topolor 2 was 73.9% with σ=13.7, median=75). The oral feedbacks received also showed that the students wanted to have more lessons in Topolor. Decisive in this, we believe, was the fact that gamification mechanics made the social interaction enjoyable that is essential to consider in designing such systems.

Additionally, we have collected usage data from Topolor's logging mechanism when the students were using Topolor in their online lesson sessions, and we have already started analysing these usage data to evaluate each of the new gamified social interaction features in detail, in order to investigate the effect of each of them on learning experience and learning efficiency. Noteworthy is the fact that, though most comparisons hide personal data and deal with averages, the popular "PK." mode, where a 'player' can compare with one other 'player' may raise ethical issues which further need evaluated. Therefore, our future work also seeks to solve this issue, e.g., by introducing a privacy management mechanism to allow learners to expose data to different groups in different ways.

References

1. Brooke, J.: SUS-A quick and dirty usability scale. Usability Eval. Ind. 189, 194 (1996).
2. Brusilovsky, P.: Adaptive hypermedia. User Model. User-Adapt. Interact. 11, 1-2, 87–110 (2001).
3. Csiksczentmihalyi, M. et al.: Flow: The psychology of optimal experience. Aust. Occup. Ther. J. 51, 1, 3–12 (2004).
4. Deterding, S. et al.: From game design elements to gamefulness: defining gamification. Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. pp. 9–15 ACM (2011).
5. Domínguez, A. et al.: Gamifying learning experiences: Practical implications and outcomes. Comput. Educ. 63, 380–392 (2013).
6. Johnson, R.T. et al.: Comparison of computer-assisted cooperative, competitive, and individualistic learning. Am. Educ. Res. J. 23, 3, 382–392 (1986).
7. Lin, J.-W. et al.: The impacts of network centrality and self-regulation on an e-learning environment with the support of social network awareness. Br. J. Educ. Technol. (2014).
8. Ryan, R.M., Deci, E.L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am. Psychol. 55, 1, 68 (2000).
9. Shi, L. et al.: A social personalized adaptive e-learning environment: a case study in Topolor. IADIS Int. J. WWWInternet. 11, 2, (2013).
10. Shi, L. et al.: Social Personalized Adaptive E-Learning Environment: Topolor - Implementation and Evaluation. In: Lane, H.C. et al. (eds.) Artificial Intelligence in Education. pp. 708–711 Springer Berlin Heidelberg (2013).
11. Shi, L. et al.: To build light gamification upon social interactions: requirement analysis for the next version of Topolor. Proceedings of the 6th York Doctoral Symposium on Computer Science and Electronics. (2013).
12. Shi, L. et al.: Towards Understanding Learning Behavior Patterns in Social Adaptive Personalized E-Learning Systems. Proceedings of The 19th Americas Conference on Information Systems. pp. 1–10 Association for Information Systems. (2013).
13. Simões, J. et al.: A social gamification framework for a K-6 learning platform. Comput. Hum. Behav. 29, 2, 345–353 (2013).

Adaptive Visualization of Plans

Nava Tintarev, Roman Kutlak, Judith Masthoff,
Kees Van Deemter, Nir Oren, and Wamberto Vasconcelos

Department of Computing Science*
University of Aberdeen
{n.tintarev}@abdn.ac.uk

Abstract. It is often difficult for humans to understand what course of action is proposed in a plan or workflow. This is particularly the case for long plans, or plans with multiple actors. Our contributions are a) the ability to present plans as both text and graphics and b) a method of filtering and highlighting, in both modalities, which focuses the information presentation to the portion of the plan which is relevant to a particular user – i.e., a view based on their roles and capabilities.

Keywords: Adaptation, Visualization, Natural Language

1 Introduction

The output from A.I. planners (e.g. PDDL) and business workflows (e.g. YAWL, BPMN, etc.) can often be difficult for humans to understand. In particular for large plans with multiple actors, it may be difficult for a human to understand what they need to do, or focus on. The aim of the SAsSy project[1] is to reduce the opacity of such plans. To this end, this paper describes ways of adapting the presentation of plans. We show how plans can be presented as either graphics or text (modality adaptation) and how these can be highlighted and filtered to focus on the most relevant information (view adaptation). This paper follows Shneiderman's Information Seeking mantra: *"Overview first, zoom and filter, then details-on-demand."* [1], focusing specifically on "zoom and filter".

We illustrate the functionality of our system using an example from the delivery logistics domain, originally taken from the International Planning Competition[2]. Our sample plan describes how four objects (a truck, a piano, a table and a drum) are delivered to different locations (cities and airports). The plan also contains a number of resources (trucks and airplanes). In addition to being able to present in two modalities (text or graphics), the system can use highlighting and filtering as a means to focus on the most relevant information to

* This research has been carried out within the project "Scrutable Autonomous Systems" (SAsSy), funded by the Engineering and Physical Sciences Research Council (EPSRC, UK), grant ref. EP/J012084/1.

[1] `http://www.scrutable-systems.org/`, retrieved April 2014

[2] `http://ipc.icaps-conference.org/`, retrieved April 2014

a user (given their role and capabilities). The aim is to reduce the amount of information the user needs to deal with, and thereby reduce their cognitive load.

Previous work on visualizing plans has looked at filtering graphs by content [5], and applying fish-eye views to grow or shrink parts of a graph [6]. There is also research on verbalizing plans generated by A.I. planning systems [4]. However, these approaches do not consider that individual differences in user needs, abilities, and preferences can have a large impact on user performance and satisfaction when using visualizations [2].

Our approach differs in that it proposes filtering based on a user model, and introduces a method for including all the required steps, i.e. including dependencies. Our method also works for *both* graphics and natural language. This makes it possible to study the effect of tailoring modality, which may be useful when a user has strong visual working memory, but poor verbal working memory.

2 System Description

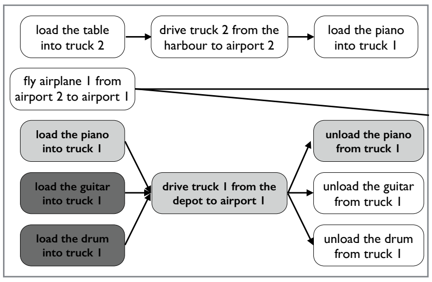
Our system is developed in Python and is available under the BSD license[3]. [7] describes the reasoning in the system, based on argumentation theory, which supports explanation mechanisms. The system can present plans in two modalities (as either graphics or natural language), and can highlight/filter plans based on user views (based on role or capability).

2.1 Modality

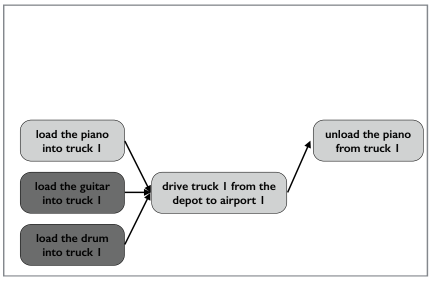
Natural Language Generation. (NLG) is the study of algorithms which produce texts in English or other human languages, from non-linguistic representations of information. Instead of presenting a plan as a sequence of tasks as produced by an A.I. planner (e.g., `(load-truck obj12 tru1 pos1), (load-truck obj11 tru1 pos1), (load-truck obj13 tru1 pos1)`) our system presents the plan as text: e.g., *Load the piano, the guitar and the drum into truck 1.*
We use NLG techniques to supply a summary (e.g. *"The workflow has 21 tasks. The workflow has 0 choices. "*). We also use aggregation (combining simple sentences together for better presentation) and referring expression generation (e.g., using pronouns when referring to past entities) to improve the presentation of the full plan.

Graphical presentation. The plan can also be represented as a graph, such as the one in Figure 1. Here, each action is a node, and edges are transitions to other possible actions. For simplification, the example plan in this paper assumes that there are no decisions points – there is only one recommended course of action. In addition, we make the simplifying assumption that parallel actions need to be completed before the next step in the plan can be executed. The algorithm below can be generalized to plans with choice points without too much difficulty.

[3] `https://bitbucket.org/rkutlak/sassy`



(a) ***Highlighting only (cropped).***



(b) ***Filtering and highlighting.***

Fig. 1: Graphical presentation of a plan in the logistics domain. All parallel steps are required for successful completion of the plan. Highlighting for the piano object is applied. Steps which directly involve the piano have a light gray background, while co-dependent steps have a dark gray background.

2.2 View

Algorithm 1

$o \leftarrow$ *selected object*
$T \leftarrow$ *all tasks*
$S \leftarrow \{t \in T | manipulates(t, o)\}$

for $s_1, s_2 \in S$ **do**
  $paths \leftarrow all\_path(s_1, s_2)$
  add tasks in paths to S
end for

for $s \in S$ **do**
  $T' \leftarrow required\_tasks(s)$
  $S \leftarrow S \cup T'$
end for
return S

This section describes an algorithm which aims to focus the information presented to a user using filtering and highlighting. The algorithm can be applied to both text and graphics to focus on those parts of the plan that are most relevant to a given user. Highlighting emphasizes steps that are relevant to the user (Figure 1a), while filtering hides the steps that would not be highlighted, i.e. portions of the plan not relevant to the user (Figure 1b).

Relevance can be determined by the role of someone enacting a plan (e.g. air-traffic controller), or their capabilities (e.g. the person who can operate a fork-lift). For example, a fork-lift operator may only want to know about actions relating to their fork-lift, while an air-traffic controller may only want to know what happens to *all* airplanes (not just one). Note that while the example below is filtered by a particular object (e.g. piano), filtering by object type (e.g. vehicle), or multiple objects is also supported.

Algorithm 1 works as follows: Given an object, for example the piano, the algorithm first selects all tasks that operate on the given object. These are the tasks `load the piano into truck 1` and `unload the piano from truck 1` colored in light gray. The algorithm then finds all paths between each pair of the selected tasks. All tasks on these paths are then added into the list of selected tasks. This corresponds to the task `drive truck 1 from the depot to airport 1` also colored light gray. Lastly, the algorithm inspects all the selected tasks and checks if any of them require completion of other tasks (indicated by multiple arrows arriving at a node in the workflow). In the example, `drive truck 1 from the depot to airport 1` requires the completion of loading all three objects so the two tasks `load the guitar into truck 1` and `load the drum into truck 1` are also selected (colored dark gray).

This example requires all steps to be completed. In a plan with choices, the algorithm includes *all* the paths between a set of actions. An alternative, more aggressive approach, would require a stricter definition of "required" nodes (e.g. include exactly one of the paths).

3 Next Steps

We have introduced a system which can present a plan as both text and graphics. It can also filter and highlight parts of a plan according to areas of relevance based on a user's role or capabilities. The system therefore allows us to conduct experiments testing the efficacy of tailoring to modality and view. The next steps are to conduct a series of experiments comparing the effect of filtering versus highlighting on cognitive load, while asking questions about participants' awareness of steps currently out of view (i.e., situational awareness [8]). We also plan to conduct studies using simple user models. These models will support compound filters such as filtering by several objects, e.g. truck 1 and airplane 1; filtering by other actors, e.g. what the other truck driver is responsible for; negation, e.g. do not show ship 1. We also plan to test different methods of filtering in plans with competing branches. Our research agenda includes a continued collaboration with industrial partners in the hydrocarbon exploration and unmanned vehicle domains.

References

1. Shneiderman, B.: The eyes have it: A task by data type taxonomy for information visualizations. In: Symposium on Visual Languages. (1996) 336–343
2. Conati, C., Steichen, B., Tory, M., Buono, P., eds.: Workshop on User-Adaptive Visualization (WUAV) in association with UMAP. (2013)
3. Monroe, M., Lan, R., Plaisant, C., Shneiderman, B.: Temporal event sequence simplification. IEEE Trans. Visualization and Computer Graphics **19** (2013) 2227–36
4. Mellish, C., Evans, R.: Natural language generation from plans. Computational Linguistics **15**(4) (December 1989) 233–249
5. Henry, T.R.: Interactive graph layout: The exploration of large graphs. PhD thesis, The University of Arizona (1992)
6. Sarkar, M., Brown, M.H.: Graphical fisheye views of graphs. In: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM (1992) 83–91
7. Tintarev, N., Kutlak, R.: Sassy - making decisions transparent with argumentation and natural language generation. In: Workshop on interacting with Smart Objects in association with Intelligent User Interfaces. (2014)
8. Endsley, M., Jones, D.: Designing for Situation Awareness. CRC Press (2012)

Interaction Model to Predict Subjective–Specificity of Search Results

Kumaripaba Athukorala ‡, Antti Oulasvirta †, Dorota Glowacka ‡, Jilles Vreeken †, Giulio Jacucci ‡

‡ Helsinki Institute for Information Technology HIIT
‡ Department of Computer Science, University of Helsinki
† Max Planck Institute for Informatics and Saarland University

Abstract. Exploratory search is becoming more common as the web is used more increasingly as a medium for learning and discovery. Compared to traditional known-item search, exploratory search is more challenging and difficult to support because it initiates with poorly defined search goals, while the user knowledge and information needs constantly change throughout the search process. Modeling the user behavior in exploratory search is a hard problem to solve. In spite of a large amount of research on personalization, little attention has been devoted to personalization in the context of exploratory search taking into account the evolving information needs of the user. We propose a formal model—motivated by Information Foraging Theory—for predicting specificity of search results with respect to the evolving knowledge and information needs of the user in exploratory search.

Keywords: User Modeling, Exploratory Search, Scientific Information-Seeking, Search Result Specificity, Information Foraging Theory

1 Introduction

Nowadays the web is used increasingly more as a source for learning and exploratory discovery [1]. Most of the existing information retrieval (IR) systems provide adequate support for well defined information needs [2]. However, there is still room for improvement for current IR systems to support users in situations where the search goal is ill-defined and changes as the search progresses, users lack the knowledge to formulate queries that express their information needs clearly, and users struggle in complex information spaces [2]. Researchers from diverse communities such as IR, machine learning or human computer interaction (HCI) have been working on designing search engines, user interfaces and user models to better support this kind of searches commonly referred to as "exploratory search."

Over the last decade many techniques have been proposed to provide better support for exploratory search, such as results clustering [3], relevance feedback [4], faceted search [5], as well as novel visualizations to support the exploration of unfamiliar information spaces [6]. However, evidence from user studies suggests that results clustering, faceted search, and relevance feedback based methods are rarely used due to the high cognitive overload of selecting relevant results and providing feedback for a large number of items [5, 4]. In response, a number of new techniques were designed to visualize search results and capture user feedback. Some of them include rich user interfaces combined with learning algorithms to support users to comprehend the search results [6], and visualization and summaries of results [7]. All these solutions are giving users more control, however, they fail to take the moment-by-moment information-needs of the user into consideration [8].

Exploratory search involves many different phases. For example, users begin exploring an unfamiliar information space by formulating imprecise queries because they lack knowledge to express their information needs [1]. Then, through several successive iterations of exploring the retrieved information and reformulating queries, the scope of the information need might narrow down [2]. This iterative and evolving nature of exploratory search makes it difficult for IR systems to identify the constantly changing information needs of the user and different phases of exploration. This is where user modeling can greatly improve existing approaches to exploratory search.

There already exist some research on modeling query formulation and interaction strategies to predict the user knowledge and information needs to personalize search results. For example, [9] presents a model which predicts the user knowledge from eye movement patterns. Even though such a model is useful in identifying domain novices and experts, it cannot predict how the information need of a user changes in a search session. Research on the correlation between the length of search queries and specificity of the user information needs [10] suggest that the length of a search query is positively correlated with the specificity of the user information-need. Such a model is useful to predict whether the information needs are too specific or broad. However, users may express very specific information needs with narrower queries having specific keywords at all query lengths and such a correlation based model cannot predict this scenario.

Systems that suggest or expand queries, provide interactive keyword visualizations, cluster results to better support exploratory search need to "know" whether the results generated from such suggestions are too broad/narrow for the information needs of the user. Hence, in exploratory search it is important to predict whether the search engine result pages (SERPs) are too narrow/broad for the evolving information needs of the user. There exist user models for navigational/transactional searches, that predict user satisfaction and relevance of SERPs from behavioral signals such as search result clicks, query refinements, gaze distribution, and dwell times [11]. Even though they provide implicit relevance feedback to the search engine, they do not predict whether the future search results should be narrower/broader. To the best of our knowledge, there has been no work on predicting the specificity (whether the results are broad or narrow) of SERPs to user knowledge and information needs.

One way to address this problem is by understanding user behaviors with queries that retrieve SERPs with varying subjective specificity in exploratory searching, which, in turn, will allow us to build a user model to predict whether given SERPs are too broad or too narrow for the current information needs of the user.

In this short paper, we will briefly discuss the following issues:

– Relationship between result click rate and specificity of SERPs to user information needs;
– A formal model to predict the specificity of SERPs to user information needs at three levels: broad, intermediate, specific;
– Empirical validation of the model.

2 Interaction Model Overview

We designed a model of user interaction by combining insights from research into exploratory search and Information Foraging Theory (IFT) [12]. According to IFT, information gain can be modeled as a linear function of time when the results are not ordered by relevance to the query. Further, IFT states that this information gain function will qualitatively shift towards a diminishing returns curve if results are ordered by relevance to the query. The gradient of this information gain function can be further improved by introducing new interface elements, such as result clustering. Hence, IFT shows how information gain is affected by the user interface or system changes.

Our research is motivated by this model. We use the term *subjective–specificity* to refer to the specificity of SERPs to the user knowledge and information need. By information need we mean the type of information that the user is actually interested in. If we keep the user interface constant, the information gain function should change according to the subjective–specificity of search results. We define subjective–specificity at three levels: broad, intermediate, and narrow. If a user issues search queries that retrieve SERPs covering many diverse topics, then we refer to them as having *broad* subjective–specificity. For example, consider an undergraduate who has just begun to follow a course in data mining issuing "data mining" as her first query to explore this domain. However, for a broad subject like data mining users would benefit from visualizations that provide an overview of the information space [13]. If a user issues search queries retrieving SERPs that are referring to a sub-topic in this domain, such as "pattern mining" under the subject data mining, then we refer to it as having *intermediate* level of subjective–specificity. In exploratory search users passively obtain cues about new keywords and repetitively reformulate queries based on these cues [2]. However, the SERPs that the user retrieves with these new keywords might be too specific for the users information need. For example, consider the same user issuing the query "subgroup discovery" based on the keywords the search engine suggested or s/he has noticed in the previous SERPs. The SERPs for this query might cover a very narrow topic, containing technical details that are less comprehensible for a novice in that area. We refer to such SERPs as having *narrow* subjective–specificity. Generally, for such a narrow search query, a novice user would benefit from more introductory material about the topic such as Wikipedia articles, book chapters, and literature reviews as well as more guided support through the specific information space [14]. The key idea is that the same search result can have very different information content for a user depending on how well it matches their current information needs. If a search engine can predict the subjective–specificity of SERPs (i.e. as broad or narrow) according to such changes in user's information need then it would be very useful to provide more related and personalized results to the user.

Our model captures how information gain in exploratory search is affected by this subjective–specificity. We define the information gain function as given in Equation 1:

$$g(n) = \lambda \ln(n) - \alpha \tag{1}$$

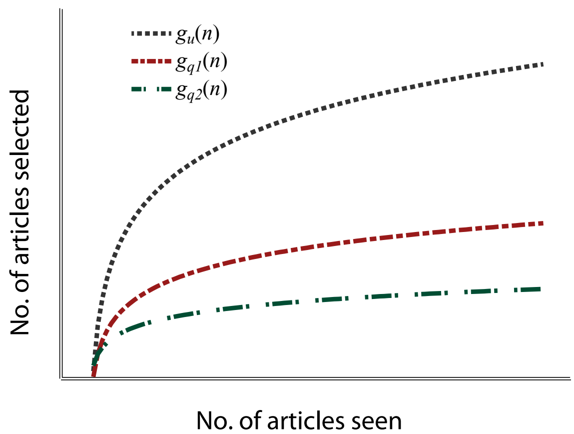


Fig. 1. Hypothetical example of information gain as a function of the number of articles Seen–Selected. $g_u(n)$ is the user-specific effective information gain function. $g_{q1}(n)$ and $g_{q2}(n)$ show how the gradient of the Seen–Selected graph reduces when the subjective–specificity of SERPs is higher than the user's information need.

Here, we define information gain as the number of results *selected* by the user. We refer to the action of clicking a search result in SERPs as selecting. We express the information gain (g) curve of a user as a function of number of result items from SERPs seen by the user (n). Following IFT, we expect this gain function to take the shape of a diminishing returns curve as shown in Figure 1. We refer to this graph as the *Seen–Selected* curve. In this gain function λ determines the slope of the information gain curve. We expect the gradient, λ, to decrease if the SERPs are narrower than the actual information need of the user. If SERPs are broader then the gradient of the Seen–Selected curve, (g), will be high. Note that α is a case-specific term which affects the maximum gain—it is determined by several factors, such as subjective–specificity of search results and case-specific factors like the search task, and the maximum number of search results the user is expecting to gain.

This model can predict the subjective–specificity of SERPs to the current information need of the user. It allows to compare the gradient of the Seen–Selected graph based on the user's selection behaviour on the current SERP with that of the user's baseline Seen–Selected graph. Such a baseline graph can be constructed by observing the everyday interactions of a user with a search tool. Then, if this user formulates a particular query to explore a topic, the gradient of the new Seen–Selected graph can be compared against the gradient of her baseline graph, and so the system can predict whether the SERPs derived from this query is too narrow or broad for her information-need—and adjust the behaviour of the system accordingly.

3 Empirical Evaluation

In order to empirically validate this model, we conducted a user study where 24 computer science student (masters and PhD level) searched for scientific information in research topics that are not very familiar to them. The task for the participants was to collect scientific articles for a scientific essay writing task in a given topic. We used six experts in six different computer science disciplines to define six unique tasks. The experts defined three search queries in each topic which retrieved SERPs from Google Scholar at three levels of specificity: broad, intermediate, and narrow. Prior to the study, we provided a questionnaire to the participants and made sure that the subjective–specificity of the queries were in comply with the participants' knowledge. We asked the participants to scan these SERPs and click articles that they find useful for the given task. We randomized the order in which they get the tasks. We logged the click interactions and plot the Seen–Selected graph as in Figure 1.

In order to confirm that, in accordance with our model, the gradients of the Seen–Selected curves decrease with the increase of the subjective–specificity of SERPs, and that they follow a natural logarithmic distribution, we analysed the overall distribution of the user information gain over information seen for the three types of SERPs. As our model predicts, the gradient of the Seen–Selected curve decreases as the subjective–specificity SERPs increase (see Table 1).

We used Wilcoxon signed-ranked test to statistically compare the gradients of the predicted models of each type of SERPs. The gradients of the broad SERPs (Mdn 3.56) were significantly greater than the gradients of intermediate (3.08) and narrow SERPS (2.04). The gradients of the predicted models of the intermediate SERPs were significantly greater than that of narrow SERPs.

This empirical evaluation shows that our model captures the effects of subjective–specificity of SERPs to the information-need of the user.

An important future challenge is to investigate in a real exploratory search scenario the performance of the formal model that we developed to predict the subjective–specificity of search results. In the future, we will incorporate our

Table 1. Logarithmic regression models and model fit (R^2) for number of articles Seen–Selected. Breakdown per Broad, Intermediate and Narrow SERPs.

Query	**Model**	**Fit** (R^2)
Broad	$3.83\ln(n) - 3.59$	0.97
Intermediate	$2.40\ln(n) - 2.06$	0.97
Narrow	$2.05\ln(n) - 1.96$	0.97

model in a running IR system and further validate its usefulness in enhancing performance of exploratory search tasks.

References

1. Marchionini, G.: Exploratory search: from finding to understanding. J. Comm. ACM, 49(4), 41–46 (2006)
2. White, R. W., Kules, B., Drucker, S. M.: Supporting exploratory search. J. Comm. ACM, 49(4), 36–39 (2006)
3. Ferragina, P., Gulli, A.:A personalized search engine based on Web-snippet hierarchical clustering. J. Software: Practice and Experience, 38(2), 189–225 (2008)
4. Kelly, D., Fu, X.: Elicitation of term relevance feedback: an investigation of term source and context. In SIGIR, 453–460 (2006).
5. Yee, K. P., Swearingen, K., Li, K., Hearst, M.: Faceted metadata for image search and browsing. In: CHI, (2003)
6. Chau, D. H., Kittur, A., Hong, J. I., Faloutsos, C.: Apolo: making sense of large network data by combining rich user interaction and machine learning. In CHI, pp 167–176 (2011)
7. Kki, M.:Findex: search result categories help users when document ranking fails. In CHI, pp 131–140 (2005)
8. Teevan, J., Alvarado, C., Ackerman, M. S., Karger. D. R.: The perfect search engine is not enough: a study of orienteering behavior in directed search. In CHI, pp 415–422 (2004)
9. Cole, M. J., Gwizdka, J., Liu, C., Belkin, N. J., Zhang, X.: Inferring user knowledge level from eye movement patterns. J. Inf. Proc. Manag (2012)
10. Phan, N., Bailey, P., Wilkinson, R.: Understanding the relationship of information need specificity to search query length. In SIGIR, ACM, pp 709–71 (2007)
11. Joachims, T.: Optimizing search engines using click-through data. In SIGKDD, pp 132–142 (2002)
12. Pirolli, P., Card, S.: Information foraging in information access environments. In: CHI, pp 51–58, (1995)
13. Głowacka, D., Ruotsalo, T., Konyushkova, K., Athukorala, K., Kaski, S., Jacucci, G.: Directing exploratory search: Reinforcement learning from user interactions with keywords. In: IUI (2013)
14. Athukorala, K., Hoggan, E., Lehtio, A., Ruotsalo, T., Jacucci, G.: Information-seeking behaviors of computer scientists: Challenges for electronic literature search tools. In: AGM of Assoc. Inf. Sci. Tech. (2013)

STS: A Context-Aware Mobile Recommender System for Places of Interest

Matthias Braunhofer, Mehdi Elahi, and Francesco Ricci

Free University of Bozen-Bolzano, Bozen-Bolzano, Italy
{mbraunhofer,mehdi.elahi,fricci}@unibz.it,
http://www.unibz.it

Abstract. In this demo paper we present a novel context-aware mobile recommender system for places of interest (POIs). Unlike existing systems, which learn users' preferences solely from their past ratings, it considers also their personality - using the Five Factor Model. Personality is acquired by asking users to complete a brief and entertaining questionnaire as part of the registration process, and is then exploited in: (1) an active learning module that actively acquires ratings-in-context for POIs that users are likely to have experienced, hence reducing the stress and annoyance to rate (or skip rating) items that the users don't know; and (2) in the recommendation model that builds up on matrix factorization and therefore can be trained even if the users haven't rated any items yet.

1 Introduction

Tourist's decision making is the outcome of a complex decision process that is affected by "internal" (to the tourist) factors, such as personal motivators or past experience, and "external" factors, e.g., advices, information about the products, or the climate of the destination [12]. Context-aware recommender systems can represent and deal with these influencing factors by extending the traditional two-dimensional user/item model that computes recommendations based only on the ratings given by a community of users to a catalogue of items. This is achieved by augmenting the collected ratings with data about the context of an item consumption and rating [1]. For example, the types of place of interest (POI) that users like can differ significantly depending on whether they are visited on a cold or sunny day. If the system stores, together with the rating, the situation in which a POI was experienced, it can then use this information to provide more appropriate recommendations in the various future target contextual situations of the user.

The first challenge for generating context-aware recommendations is how to identify the contextual factors (e.g., weather) that are truly influencing the ratings and hence that are worth considering [3]. Secondly, acquiring a representative set of in-context ratings (i.e., ratings under various contextual conditions) is clearly more difficult than acquiring context-free ratings. Finally, extending traditional recommender systems to really exploit the additional information brought by in-context ratings, i.e., building more accurate recommendations, is the third challenge for context-aware recommender systems.

In this demo paper, we describe an operational context-aware recommender system, called STS (South Tyrol Suggests). STS is an Android-based mobile application that recommends POIs in South Tyrol (Italy) by exploiting various contextual factors (e.g., weather, time of day, day of week, location, mood) and an extended matrix factorization rating prediction model. STS can generate recommendations adapted to the current contextual situation, for example, by recommending indoor POIs (e.g., museums, churches, castles) on bad weather conditions and outdoor POIs (e.g., lakes, mountain hikes, scenic walks) on good weather conditions. The user's preference model is learned using two different sources of knowledge: (1) personality, in terms of the Five Factor Model, that the system acquires with a simple personality questionnaire, and (2) in context ratings that the system actively collects from the user. This allows - and this is the novel aspect of STS - to personalize recommendations and rating requests even for the new users, by leveraging their personality, which is known to be strongly correlated with their tastes and interests [11].

2 Interaction with the System

This section describes a typical interaction with STS and shows some of its functions. Let's assume a tourist or a citizen who is looking for a POI to visit near to Bozen - Bolzano, Italy. The first run of STS opens the registration screen where the user can enter a username, password, birthdate and gender. After registering into the system, the user is asked to fill out the Five-Item Personality Inventory (FIPI) questionnaire [9], in order to allow the system to assess her Big Five personality traits (i.e., openness, conscientiousness, extroversion, agreeableness, neuroticism) (see Figure 1, left). As an alternative to the FIPI questionnaire, other popular personality questionnaires, such as the 120 or 240 item International Personality Item Pool Representation of the NEO PI-R (IPIP-NEO; see [8]), could have been used. These allow a more accurate and reliable personality assessment. However, these questionnaires are time-consuming, taking at least 10-20 minutes to complete, and hence they are ill-suited for mobile interaction models which are usually short in time or on the move.

The entered birthdate, gender and calculated personality are then used by an active learning component [7], which identifies and requests the user to rate a series of POIs whose ratings are estimated to provide the largest improvement of the quality of the subsequent recommendations (see Figure 1, right). We note that this active learning component is able to provide personalized rating requests, without completely relying on explicit feedback (e.g., ratings) or implicit feedback (e.g., item views) which is usually not available for newly registered users.

After that the system is ready for usage, and the user can browse her personalized recommendations through the main application screen (see Figure 2, left). This screen displays a list of POIs that are considered as highly relevant, consid-

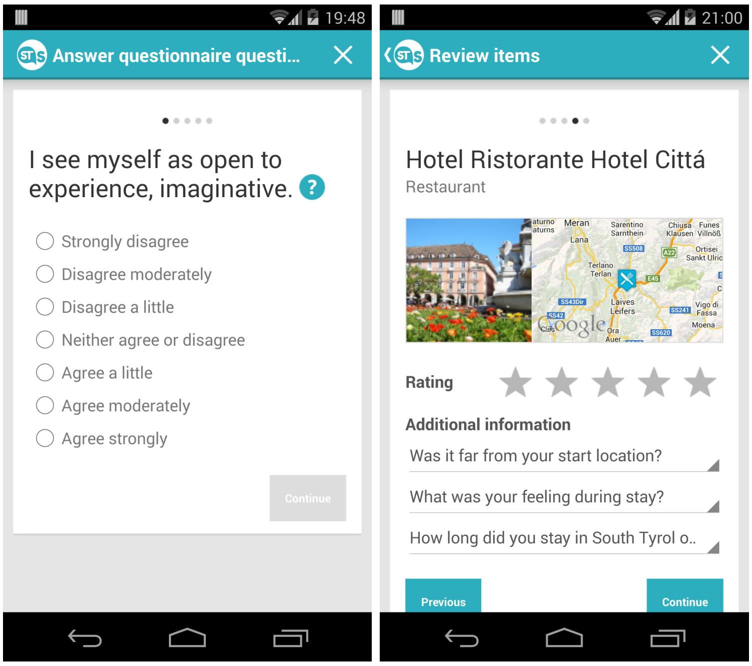


Fig. 1. Screenshots of STS: (left) Five-Item Personality Questionnaire, and, (right) Active Learning

ering the current user's and items' contexts. We note that some context data is automatically acquired by the system (e.g., user's distance to the POIs, weather conditions at the POIs, whereas others can be specified by the user through an appropriate system screen (e.g., user's mood and companion) [5, 4]. In the event the user is interested in one of the POIs, she can click on it and access the POI details window (see Figure 2, right). This window presents various information about the POI, such as a photo, its name, a description, its category as well as an explanation of the recommendation based on the most influential contextual condition. Other supported features include, among others, the ability to write a review for the POI, to view the POI on the map and to bookmark the POI, which then makes it easy to access the POI description.

3 Software Architecture and Implementation

STS implements a rich client always-online architecture, i.e., the client has been kept as thin as possible and it works only in a limited way offline. The client application has been developed using the open-source Android platform, and implements the presentation layer (GUI and a presentation logic). The server application is based on Apache Tomcat server and PostgreSQL database. It

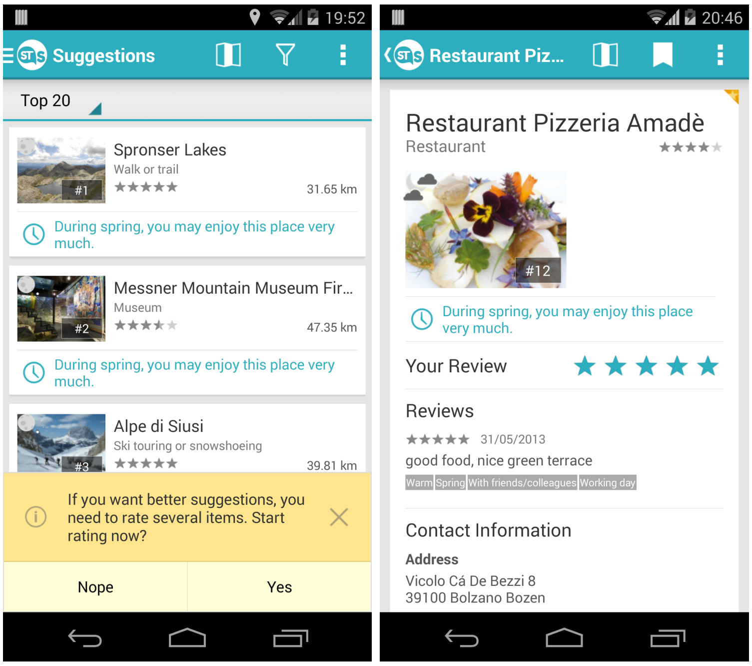


Fig. 2. Screenshots of STS: (left) Recommendation (right) POI description page

implements the data and business logic (recommendation) and makes use of web services and data storages provided by the Regional Association of South Tyrol's Tourism Organizations (LTS[1]), the Municipality of Bolzano[2] and Mondometeo[3] in order to obtain the graphical/textual descriptions as well as weather forecast information for a total of 27,000 POIs. All the server's functionality is exposed via a RESTful web service that accepts and sends JSON objects and that provides several types of resources (e.g., suggestions, POIs, reviews/ratings, user profiles).

4 Recommendation Logic and Evaluation

In order to take into account the current contextual conditions when generating POI recommendations, we have extended the context-aware matrix factorization approach proposed by Baltrunas et al. [3]. This model incorporates baseline parameters for each contextual condition and item (or item category) pair, besides the standard parameters (i.e., global average, item bias, user bias and user-item interaction), in order to capture the deviation of the rating for an item produced

[1] LTS: LTS: http://www.lts.it
[2] Municipality of Bolzano: http://www.comune.bolzano.it
[3] Mondometeo: http://www.mondometeo.org

by the contextual conditions. Since the original context-aware matrix factorization model fails to provide personalized recommendations for users with no or few ratings (i.e., new user problem), we also enhance the representation of a user u by incorporating the set of known user attributes $A(u)$ (i.e., age group, gender and the discretized scores for the Big Five personality traits), analogously as in [10]. A distinct factor vector y_a corresponds to each attribute to describe a user through the set of user-associated attributes $\sum_{a \in A(u)}$. This allows to model the user preferences, even in cases where implicit and explicit feedback are absent.

The resulting model computes a rating prediction for user u and item i in the contextual situation described by the contextual conditions $c_1, ..., c_k$ using the following rule:

$$\hat{r}_{uic_1,...,c_k} = \bar{i} + b_u + \sum_{j=1}^{k} b_{ic_j} + q_i^\top \cdot (p_u + \sum_{a \in A(u)} y_a), \tag{1}$$

where q_i, p_u and y_a are f dimensional real-valued factor vectors representing the item i, the user u and the user attribute a, respectively. $\bar{i}$ is the average rating for item i, b_u is the baseline parameter for user u and b_{ic_j} is the baseline for contextual condition c_j and item i. Model parameters are learned offline, once every five minutes, by minimizing the associated regularized squared error function through stochastic gradient descent.

This recommendation model as well as the implemented active learning strategy for eliciting ratings have been evaluated in two live user studies [5, 7], with the following findings: (1) the recommendation model successfully exploits the weather conditions at POIs and leads to a higher user's perceived recommendation quality and choice satisfaction; and (2) the active learning strategy increases the number of acquired user ratings and the recommendation accuracy in comparison with a state-of-the-art active learning strategy.

5 Conclusions and Future Work

In this demo paper we have illustrated a novel mobile context-aware recommender system, named South Tyrol Suggests (STS), that learns users' preferences from their past ratings as well as their personality. Users's personality is acquired through a brief five-item questionnaire that is subsequently used for actively eliciting ratings for POIs that were estimated to be experienced by users. Finally, this information is exploited for generating high quality recommendations for POIs under the target contextual situation. We have described the implementation of STS, in terms of design, recommendation logic, user interface, and features.

For future work, we plan on making several improvements to STS. Firstly, we intend to provide push recommendations to the user when the current situation seems appropriate, without relying on explicit user's request. Additionally, we would like to exploit in the recommendation process human emotion and the knowledge of the current user activity that can be derived from wearable devices such as smart-watches and smart-bands. Finally, we plan to develop ways to determine the user's personality, without explicitly asking the user to fill a questionnaire by inferring the personality traits of users from their Facebook profiles [2] or their mobile phone usage [6].

References

1. G. Adomavicius, B. Mobasher, F. Ricci, and A. Tuzhilin. Context-aware recommender systems. *AI Magazine*, 32(3):67–80, 2011.
2. Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell. Personality and patterns of facebook usage. In *Proceedings of the 3rd Annual ACM Web Science Conference*, pages 24–32. ACM, 2012.
3. L. Baltrunas, B. Ludwig, S. Peer, and F. Ricci. Context relevance assessment and exploitation in mobile recommender systems. *Personal and Ubiquitous Computing*, 16(5):507–526, 2012.
4. M. Braunhofer, M. Elahi, M. Ge, and F. Ricci. Context dependent preference acquisition with personality-based active learning in mobile recommender systems. In *Human-Computer Interaction (HCII)*. Springer, 2014 (to be appeared).
5. M. Braunhofer, M. Elahi, F. Ricci, and T. Schievenin. Context-aware points of interest suggestion with dynamic weather data management. In *21st Conference on Information and Communication Technologies in Tourism (ENTER)*. Springer, 2014.
6. R. de Oliveira, A. Karatzoglou, P. Concejero Cerezo, A. Armenta Lopez de Vicuña, and N. Oliver. Towards a psychographic user model from mobile phone usage. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 2191–2196. ACM, 2011.
7. M. Elahi, M. Braunhofer, F. Ricci, and M. Tkalcic. Personality-based active learning for collaborative filtering recommender systems. In *AI*IA 2013: Advances in Artificial Intelligence: XIIIth International Conference of the Italian Association for Artificial Intelligence, Turin, Italy*. Springer, 2013.
8. L. R. Goldberg, J. A. Johnson, H. W. Eber, R. Hogan, M. C. Ashton, C. R. Cloninger, and H. G. Gough. The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1):84–96, 2006.
9. S. D. Gosling, P. J. Rentfrow, and W. B. Swann Jr. A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528, 2003.
10. Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.
11. P. J. Rentfrow and S. D. Gosling. The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of personality and social psychology*, 84(6):1236, 2003.
12. J. Swarbrooke and S. Horner. *Consumer behaviour in tourism*. Routledge, 2007.

Understanding Usages by Modeling Diversity over Time

Amaury L'Huillier, Sylvain Castagnos, and Anne Boyer

KIWI Team, LORIA - Université de Lorraine
Campus Scientifique, B.P. 239
54506 Vandœuvre - France
{amaury.lhuillier,sylvain.castagnos,anne.boyer}@loria.fr

Abstract. Let's imagine a system that can recommend the kind of music (among other application domains) you like to listen when you are at work, without having to know your location, IP address or even to ask your current mood. In this paper, we bring this dream closer by proposing a model that can automatically understand the user's current context. This model, called DANCE, analyzes the attributes of the items in your recent history and monitors the relative diversity brought by your consultations over time. We validated our approach with a music corpus of 100 users and a global history of 204,758 plays.

Keywords: User Modeling, Analysis of usages, Diversity, Recommenders.

1 Introduction

Recommender systems aim at improving human-computer interactions between the general audience and online services such as e-commerce websites, VOD, music streaming or search engines. These systems rely on various machine learning techniques to understand users' behaviors, to build preference models, and to predict future interests or needs so as to personalize information access. After two decades of researches and uses in industry, recommender systems have been proven to be useful both for companies (for knowing their customers, increasing revenues) and users (gain of time, usefulness of recommendations). However, most of recommenders only focus on the precision of the recommendations, now reaching excellent performances. They do not take enough into account human factors involved in the decision making process, such as context, confidence, explanations and need for diversity. These factors are yet crucial to maximize users' satisfaction, as highlighted by Jones in [7]: while a difference of 10% of the RMSE between two algorithms cannot be perceived by users, recommendations made at the right time in the proper way can double users' acceptance rate.

Our work-in-progress follows on from the experiment we conducted in [3], which aimed at identifying the steps of a user's decision process, when facing a recommender system on a e-commerce website. In that study, the diversity of recommendations has been shown to be correlated with the active user's confidence level and his intention to buy. Our user study strongly suggested to favor approaches where similarity between recommendations progressively increases at the product brokering stage, and where diversity is re-introduced in the end of the session (when decision is close to being made). However, as far as we know, no temporal model has ever been proposed in the literature to adapt recommendations to the active user's evolving need for diversity over time.

As a first step towards this objective, we propose a new way to analyze usages by modeling the diversity brought by each consulted item relatively to a short user's history. We took inspiration from natural language processing techniques [1] to survey the diversity over time within a sequence of consultations, and introduce generic multi-attribute diversity metrics. In that way, we can automatically understand the current context of the active user in real time. We called our model DANCE, which is the acronym for "Diversity And Natural Context Elicitation". DANCE offers the advantage of being highly scalable and generic, preserves privacy and does not require an ontology to put words on the context. We validated our approach on a music corpus and have been able to detect skipped songs and ends of sessions within users' listening sequences.

This paper is organized as follows: Section 2 is an overview of the state of the art of diversity in recommender systems. Section 3 is dedicated to the presentation of our model and Section 4 presents and discusses its performances.

2 Related Work

2.1 Usages of diversity

Recommender systems usually rely on similarity measures between users' preferences, contents of items or past usages to predict future interests as accurately as possible, even if accuracy is insufficient to maximize user's satisfaction [9]. Several recent research papers encourage the use of diversity, in addition to classic evaluation metrics, to measure the quality of algorithms so as to discredit those focusing on certain parts of the item spectrum [6].

The diversity has been defined by Smyth and McClave [11] as the opposite dimension to similarity. More recent works refined this definition as the measure that quantifies the dissimilarity within a set of items [4]. McGinty and Smyth [8] have been the first to show that diversity improves the efficiency of recommendations. They see diversity as a contribution in response to a bad recommendation, to provide new exploration strategies. Other usages include intrinsic diversity which avoids redundancy between the items to be recommended [5], and extrinsic diversity which aims at alleviating the uncertainty due to data ambiguity or sparsity in user preference models by recommending a large set of items [10]. Nevertheless, all these works try to manage diversity at a given moment, but do not integer the time dimension.

In [3], we aimed at identifying the different steps that mark out the decision-making process to get deeper into details, and pay attention to why and when diversity should be provided. It comes out that the user's need for diversity evolves over time, and is particularly high when the user is close to reach a decision or to change his mind. Starting from this conclusion, the core idea of this paper consists in modeling the evolution of diversity over time within sequences of consultations so as to automatically detect and understand the user context. Time-aware and context-aware recommender systems are receiving increasing attention, since it has been proven to be an effective approach in the Netflix Prize competition [2]. However, current context-aware systems often need an *a priori* knowledge of various possible contexts (work vs. home, holidays or not, season, moment of the day, etc.), with various representations such as ontologies, to filter information. Our long-term goal is to detect changes of context and common features/patterns inside these contexts thanks to diversity, without any prior context knowledge nor user involvement.

2.2 Metrics of diversity

In the literature, most of diversity metrics rely on the similarity between items: the more the items are similar, the lower is the diversity between them [11]. Similarity between two items is defined as the weighted sum of the similarities on their attributes (see (1)). A is the set of attributes, with $card(A) = n$.

$$sim_A(i_1,i_2) = \frac{\sum_{j=1..n} w_j * sim_{attribute=j}(i_1,i_2)}{\sum_{j=1..n} w_j} \tag{1}$$

Starting from this similarity metric, Ziegler *et al.* defined an Intra-List Similarity metric (ILS) which computes the average similarity – and by opposition the average diversity – within a class a class C, made up of m items. Smyth and McClave [11] has introduced the same kind of metric, called $Diversity$, which computes the average dissimilarity within the class C. Additionally, they define the notion of relative diversity ($RelDiversity$) as the added value in terms of diversity of an item i on a class of items C for a user u, while considering all items' attributes simultaneously (see (2)).

$$RD^u(i,C) = \begin{cases} 0 \text{ if } C = \{\}, \\ \frac{\sum_{j=1..m}(1 - sim_A(i,c_j))}{m} \text{ otherwise.} \end{cases} \tag{2}$$

3 Our Model

As explained above, our model intends to analyze usages so as to detect the changes of user context. We called this model DANCE, acronym for "Diversity And Natural Context Elicitation", since it took inspiration from natural language processing techniques and focus on diversity. We consider the active user's history of consultations as a contiguous sequence of items. As in a k-order Markov model, we assume that the consultation of an item i only depends on the k previously accessed items. These k items are called *history*. We then measure the relative diversity brought by i relatively to the k previous consulted items. Our model can compute a global score of relative diversity by taking all the attributes of items into account together, or can monitor the evolution of relative diversity for each attribute separately.

Within the framework of this model, we define the notion of *context* as a period of time during which consulted items share common features. During this interval, items have similar values for one or several attributes, which means that the relative diversity evolve slowly over time for these attributes (and potentially for the global relative diversity). Otherwise, when the relative diversity of the current item increases more than a given threshold, the user context is about to change. Consequently, in order to detect these changes of context, our model looks for local maximums on the curve of relative diversity (i.e. where there is an horizontal tangent after a significant increase of the relative diversity in comparison with previous time steps).

In the following, we note $I = \{i_1, i_2, ..., i_m\}$ the global set of items available for the active user. Each item has a fixed number of attributes ($card(A) = n$). The set of items actually consulted by the active user u is $C_u = \{c_1^u, ..., c_k^u\}$, but we do not need to remember it since we only observe the recent *history* of size k. The latter can be decomposed under the form of a tuple, written $< c_{t-k}^u, ..., c_{t-2}^u, c_{t-1}^u >$, where t is the current time step. At last, the value of the attribute a for the item c_t^u is written $c_t^u.a$.

The DANCE model consists in computing the relative diversity brought by c_t^u in comparison with the *history*, using the formula (2) above at each time step t, that is to say: $RD^u(c_t^u, < c_{t-k}^u, ..., c_{t-2}^u, c_{t-1}^u >)$. The function of relative diversity RD^u, and at the same time the function of similarity (see (1)) requires to define sub-functions of similarity per attribute.

If the attribute a can take values in (R), this sub-function is computed using Equation (3), where min_a and max_a respectively correspond to minimum and maximum values of a on the set of items I.

$$sim_{attribute=a}(c_t^u, c_{t-1}^u) = 1 - e^{-10*\left(\frac{c_t^u.a - c_{t-1}^u.a}{max_a - min_a}\right)^2} \quad (3)$$

If the attribute a contains a list of terms, the sub-function of similarity per attribute is given by Equation (4).

$$sim_{attribute=a}(c_t^u, c_{t-1}^u) = 1 - \frac{card(c_t^u.a \cap c_{t-1}^u.a)}{min(card(c_t^u.a), card(c_{t-1}^u.a))} \quad (4)$$

For a given user and a fixed size of *history*, our algorithm has a complexity in constant time. It can perform computations in real time, while preserving his privacy since we only need his short *history* and attribute values of items.

4 Performance Analysis

We based our evaluation on a music corpus, due to its numerous advantages. First, music items can be characterized by a large number of attributes and information about the songs are easy to retrieve. Besides, even though very similar to analyses of sequences in other application domains (e-commerce websites, VOD), it avoids some bias. Songs are short enough to be fully listened in a sequence contrary to movies, and we can record list of songs that have been skipped before the end. The duration of a consultation is independent from the reading speed of the subject, by opposition to textual websites. At last, users of music streaming services like to discover new songs.

We used the API of Last.fm[1] to collect data of 100 users for a period between June 2005 and October 2013. For each user, we own a set of music plays which contains the names of the songs, the names of the artists and timestamps corresponding to the times he listened each song. We extracted values of attributes about these songs using the Echonest[2] API. Information about the various item attributes and possible associated values are summarized in Table (1).

	music							artist	
attribute	duration	tempo	mode	loudness	energie	hotttness	danceability	hotttness	familiarity
max	4194	239	1	41.76	0.99	0.91	0.98	0.99	0.91
min	12	0	0	0.08	0.00002	0.000782	0.039	0.109	0.051
average	218.09	128.25	0.57	7.68	0.76	0.33	0.43	0.62	0.63
deviation	84.25	30.31	0.49	4.19	0.21	0.12	0.17	0.12	0.13

Table 1. Characteristics of the corpus.

To test the ability of our model to detect the changes of context in users' sequences, we decided to mark the ends of sessions for each user, and to record the list of songs that have been skipped. In order to do so, we fixed the maximum duration without any music play before considering a session as ended to 15 minutes. We obtained 18,640 sessions composed of 204,758 plays (40,923 distinct songs and 5,571 distinct artists), with an average of 10.98 songs per session (39.92 minutes). After that, we put the whole list of plays for each user in a single sequence (test data), and tried to retrieve these changes of context (see Table (2). We tested our model with four-gram (i.e. with *history* size $k = 3$).

	Test data	Changes detected	% of detections
Ends of sessions	18,640	13,551	74.30 %
Number of skipped songs	28,849	2,752	9.53 %
Total number of local maximums in RD^u		25,193	

Table 2. Detecting the changes of context with the DANCE model.

Taking in consideration all the attributes with the same weight, we were able to retrieve 74.30% of the ends of sessions contained in our dataset. Moreover, only 9.53% of the skipped songs were detected. These results are not very surprising, since we are more likely to detect differences between sessions (making them

[1] `www.lastfm.fr/`

[2] `http://developer.echonest.com/`

more identifiable), rather than to detect anomalies with skipped songs: a user can skip a song even if this one is not diverse (e.g. because he is fed up).

5 Conclusion and Perspectives

In this paper, we showed that modeling diversity over time is an elegant and efficient way to detect users' changes of context when analyzing usages in real time. The biggest strength of our model is to be able to automatically detect similarities within contexts, and dissimilarities between contexts. In Table (2), we can notice that there are more local maximums obtained with RD^u (25,193 peaks of diversity) than the number of sessions. This means that, inside a session, we can find several contexts and take advantage of them. These preliminary results constitute a first step to adapt recommender systems to users' context, by explaining recommendation with common session features, and by varying the level of diversity to users' needs and expectations. In the same time, we plan to optimize the weights of attributes and size of history.

References

1. Brun, A., Bonnin, G., Boyer, A.: History Dependent Recommender Systems based on Partial Matching. In: UMAP, pp. 343–348 (2009)
2. Campos, P.G., Díez, F., Cantador, I.: Time-Aware Recommender Systems: A Comprehensive Survey and Analysis of Existing Evaluation Protocols. Journal on User Modeling and User-Adapted Interaction, Volume 24, pp. 67–119 (2014)
3. Castagnos, S., Jones, N., Pu, P.: Eye-Tracking Product Recommenders' Usage. In: Proc. of the 4th ACM Conference on Recommender Systems, pp. 29–36 (2010)
4. Castagnos, S., Brun, A., Boyer, A.: When Diversity Is Needed... But Not Expected! In: IMMM, pp. 44–50 (2013)
5. Clarke, C.L.A. et al.: Novelty and Diversity in Information Retrieval Evaluation. In: Proc. of the 31st International ACM Conference on Research and Development in Information Retrieval, pp. 659–666 (2008)
6. Jannach, D., Lerche, L., Gedikli, F., Bonnin, G.: What Recommenders Recommend – An Analysis of Accuracy, Popularity, and Sales Diversity Effects. In: Proc. of User Modeling, Adaptation, and Personalization, pp. 25–37 (2013)
7. Jones, N.: User Perceived Qualities and Acceptance of Recommender Systems: The Role of Diversity. PhD Thesis, EPFL, Lausanne, Switzerland (2010)
8. McGinty, L., Smyth, B.: On the Role of Diversity in Conversational Recommender Systems. In: Int. Conf. on Case-Based Reasoning, pp. 276–290 (2003)
9. McNee, S.M., Riedl, J., Konstan, J.A.: Being Accurate Is Not Enough: How Accuracy Metrics Have Hurt Recommender Systems. In: Proc. of Computer-Human Interaction, pp. 1097–1101 (2006)
10. Radlinski, F. et al.: Redundancy, Diversity and Interdependent Document Relevance. Journal SIGIR Forum, Volume 43(2), pp. 46–52 (2009)
11. Smyth, B., McClave, P.: Similarity vs. Diversity. In: Proc. of the 4th International Conference on Case-Based Reasoning, pp. 347–361 (2001)
12. Ziegler, C.N., McNee, S.M., Konstan, J.A., Lausen, G.: Improving Recommendation Lists Through Topic Diversification. In: Proc. of the 14th international conference on World Wide Web, pp. 22–32 (2005)

An Attractiveness Evaluation of Picture Books Based on Children's Perspectives

Min-Yuan Ma [1], Chun-Chun Wei [2], Yang-Cheng Lin [3]

[1] Department of Industrial Design, National Cheng Kung University, Tainan, Taiwan
E-mail: mamy@mail.ncku.edu.tw
[2] Department of Industrial Design, National Cheng Kung University, Tainan, Taiwan
E-mail: p38991123@mail.ncku.edu.tw
[3] Department of Arts and Design, National Dong Hwa University, Haulien, Taiwan
E-mail: lyc0914@mail.ndhu.edu.tw

Abstract. In the contemporary publishing market, changes in publishing formats have provided diverse reading experiences and have attracted readers' interest and attention. Accordingly, the objective of this study is to examine various design forms of picture books and identify the emotional preferences and attractiveness factors for children. Miryoku Engineering is an intelligent system to explore how consumers select picture books and their corresponding reading experiences, determine the attractiveness of picture books, and understand the qualitative attractiveness factors by using a quantitative method. The results of this study show how various picture book forms appeal to children, and it can provide a reference for designing and selecting picture books.

Keywords: Picture Books, Design Forms, Miryoku Engineering, Attractiveness Evaluation, Evaluation Grid Method.

1 Introduction

In the growing market of picture book publications, readers possess enhanced sensory demands when reading. Because picture books are designed for children, they cater to children's nature and possess characteristics that are relevant to language development [2], cognitive engagement [4], and artistic thinking, and fun [5]. Subsequently, the traditional reading behaviour of turning pages has been abandoned for picture books. Accordingly, the formatting of picture books has been redesigned for various media to provide different reading models and increase users' interest and attention. For example, *The Very Hungry Caterpillar*, a bestselling picture book written by the renowned author Eric Carle has been published in various forms, including a toy book, pop-up book, and e-book. In this era of emotional consumption, understanding and catering to diverse and changing consumer preferences and demands before adopting appropriate book forms is crucial for book publishers and picture book designers.

However, few studies of picture book design forms have been conducted, especially research that analyses children's emotional preferences. Thus, the primary objective of this study is to examine how various picture book forms influence children's emotional preferences and the relevant book attractiveness factors. In order to explore how consumers select picture books and the attractiveness factors, we apply Evaluation Grid Method (EGM) of Miryoku Engineering which is an intelligent system that converts personal feelings and images into architecture of product designs. These results can be used as a reference for designing and selecting picture books.

2 Evaluation Grid Method of Miryoku Engineering

Miryoku Engineering is a research method that captures and tabulates personal perceptual concepts in a particular order [1,8]. Sanui modifies Kelly's repertory grid technique to create a new approach is known as Evaluation Grid Method (EGM) [6]. By employing this approach, researchers can establish a three-layer hierarchical structure that comprises abstract reasons, original evaluation items, and proposed concrete conditions. Subsequently, a personal evaluation hierarchy can be created. EGM is a critical research method of Miryoku Engineering that has been successfully employed in numerous fields such as product design and service design [7,9]. EGM operation steps are as follows:

1. Preparing interview materials:
   Determine interview participants and materials depending on the research topics and use products or pictures as stimulant samples for interviews.
2. Identifying original evaluation:
   Conduct interviews based on stimulant samples and compare pairs of stimulant samples. Based on the perspectives of "satisfied/dissatisfied" or "most preferred/least preferred," interview questions are provided for participants to determine their satisfaction or most preferred product characteristics. These characteristics are the original evaluation items.
3. Laddering:
   Use supplementary questions to identify the abstract reasons (ladder up) and concrete conditions or features (ladder down) proposed in the original evaluation items identified in Step 2. Subsequently, the three hierarchical levels of each evaluation item can be identified: the abstract reasons (ladder up: abstract value judgment), original evaluations (medium: comprehension of feelings), and proposed concrete conditions (ladder down: objective and concrete understanding).
4. Organizing a personal evaluation hierarchical map:
   Repeatedly conduct Steps 2 and 3 on all evaluation items provided by participants, and compile all the evaluation items into a three-layer hierarchy to form a personal evaluation hierarchical map.
5. Organizing an overall evaluation hierarchical map:
   Compile the personal evaluation hierarchical maps of all the participants and calculate the number of overlapped evaluations to plot an overall evaluation hierarchical map.

The results of conducting the aforementioned processes should clearly indicate the attractive product elements commonly perceived by numerous participants. Subsequently, these results can be applied to product development and design concepts. In addition, the results of such qualitative evaluation combined with questionnaire survey methods can be quantitatively analysed. Based on the results of such quantitative analysis, designers can understand consumer values and needs and offer affective and user-friendly designs that closely fulfil consumer demands. We argue that the unique attractiveness of picture books is a critical factor that appeals to child readers. EGM provides a theory-based method for analysing product attractiveness factors and eliminates the potential influence of researcher subjectivity on the quality and results of qualitative research. Thus, EGM can be used to effectively systemize data and accurately interpret the affective needs and thoughts of children.

3 The Procedure of Experiment

The experimental design is divided into two stages. In the first stage, the results of personal interviews are used to connect the ladder-up and ladder-down concepts to the original evaluations obtained using EGM. Subsequently, an evaluation hierarchical map of attractiveness is established. In the second stage, the attractive elements in the evaluation hierarchical map are converted into question items for the questionnaire survey. We use Quantification Theory Type I (QTTI) to analyse the attractive element items, category influences, and weight relationships, thereby compiling the attractive characteristics of various picture book forms and identifying children's preferences regarding picture book attractive elements. Moreover, independent samples t-test is conducted to determine whether significant between gender group differences existed.

3.1 Experimental Samples and Participants

According to the sensory perception methods used when reading the books, as determined by a focus group of three picture book designers. picture books are then classified into four types: (a) Conventional books; (b) Pop-up books (c) Talking books; and (d) e-books, thereby enlivening and enriching the content of picture books (*The Very Hungry Caterpillar*) has been published in various forms [3].

The main readers of picture books are elementary students. Thus, children aged between 8 and 10 years old are selected as the primary research participants in this study. In the first stage of the EGM interviews, the children who exhibit a considerable interest in picture books and have previous experience with all four types of picture book design formats are classified as the high-involvement group. In the second stage, a questionnaire survey regarding the attractiveness of picture books is conducted. However, to achieve the statistical requirements, 30 participants are assigned to each group (with an equal number of male and female respondents).

3.2 Evaluation Hierarchical Map Analysis

Based on the interview results, we plot the overall evaluation hierarchical map, calculate the number of overlapped evaluations, and establish an original attractiveness evaluation. A total of 12 items are identified (shown Fig.1); specifically, six items for *Conventional Books* (resistance to damage, simple design, easy to carry, enhanced concentration, convenient storage, and physical books), three items for *Pop-up Books* (enhanced concentration, 3D pictures, and varied reading experiences), and five items for *Talking books and e-books* (multifunctional book, worth more than the purchase price, easy to carry, readers are easily involved in the protagonist's situation, and the story can be easily understood). Subsequently, three experts who possess more than 5 years of graphic design experience participate in a focus group discussion, using the Kawakita Jirou (KJ) method to integrate similar factors from the 12 items. Subsequently, six major attractiveness factors of picture books are identified.

There are three factors for conventional books, including “Simple and Comprehensible Layout Design”, “Thin and Lightweight Dimensions”, and “The Sense of Ownership”. One is “Various Page Displays” for *Pop-up Books*, “Vivid Story Plots” for *Talking books*, and “Multifunctional Reading Methods” for *e-books*. Figure 2 shows the corresponding ladder-up vocabularies and ladder-down vocabularies when the attractiveness factor of *Pop-up books* is the factor of “the various page displays” in *Pop-up Books*.

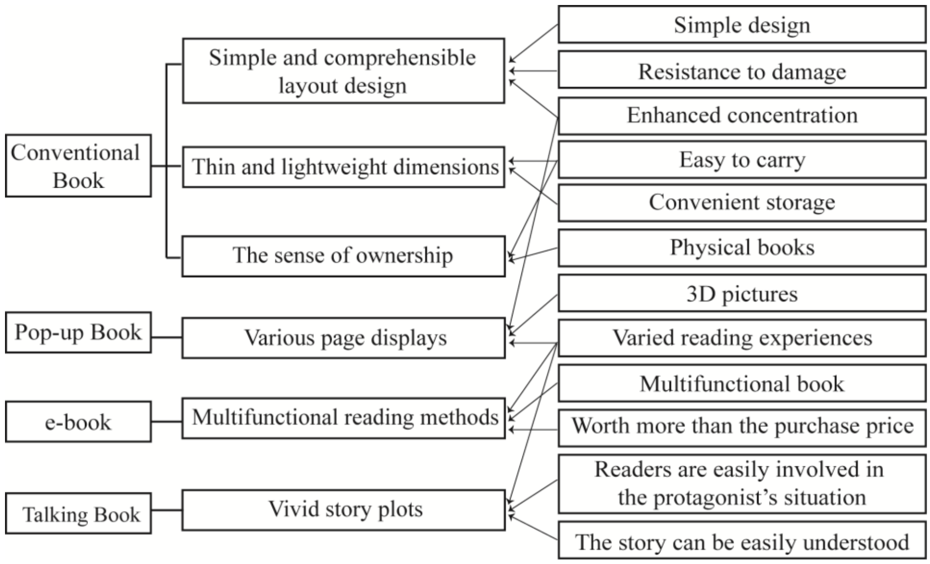


Fig. 1. KJ method to integrate similar factors of the 12 items

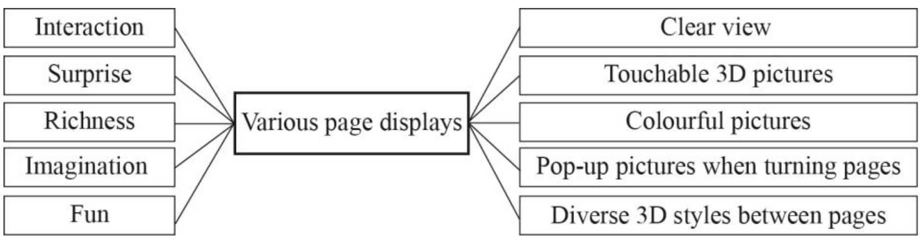


Fig. 2. Evaluation Hierarchical Map

4 Experimental Analysis with QTTI and t- test

Based on QTTI, we examine the factors that influence children’s affective preferences for picture books in the forms of *Conventional Books, Pop-up Books, Talking books,* *and e-books*. Table 1 shows the *Pop-up Books* of results of QTTI. According to the QTTI results, under “The various page displays” of *Pop-up Books'*, the children rate ‘Pop-up pictures during page-turning’ as the most influential factor, followed by ‘Colourful pictures’. To understand whether differences are existed between the male and female perceptions of picture book attractiveness, we conduct an independent samples t-test to examine the two independent groups of children. And their perceptions regarding “Multifunctional Reading Methods” item of *e-books* is significantly differed ($p = .024 < .050$). Because the t value was negative ($t = -2.830$), the negative evaluations by the male participants are lower than those of the female participants, indicating that the female participants preferred “Multifunctional Reading Methods” (shown Table 2).

Table 1. The Pop-up Books of QTTI results

Attractiveness Factors		no	Form Type (ladder-down vocabularies)	Form Type Grade
Pop-up books	Various page displays	1	Clear view	0.22
		2	Touchable 3D pictures	-0.11
		3	Colourful pictures	0.48
		4	Diverse 3D styles between pages	-0.85
		5	*Pop-up pictures when turning pages	0.76

Table 2. Independent Samples Test of “Multifunctional Reading Methods”

Levene's Test for Equality of Variances			t-test for Equality of Means			
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference
Equal variances assumed	.31	.58	-2.38	28	.024	-1.35

5 Discussion and conclusion

In this paper, we have developed the evaluation hierarchical map of the attractiveness factors for picture books, using EGM based on Miryoku Engineering. The results of this study are shown as follows:

1. The primary attractiveness factors for *Conventional Books* are “Simple and Comprehensible Layout Design”, “Thin and Lightweight Book Dimensions”, and “The Sense of Book Ownership”; for *Pop-up Books*, the main factor is “Various Page Displays”. “Vivid Story Plots” is for *Talking books,* and a “Multifunctional Reading Methods” factor is for *e-books*. And Table 3 shows the preference form type of children in various books.

2. Regarding the differences between the male and female attractiveness evaluations, the results show that that the female participants preferred “Multifunctional

Reading Methods “of *e-books*. According to the interview of EGM, we can know Female attach importance to many different content of books can be choose from e-book, rather than animation and sound effects of stimulation.

In summary, the results can be used as an important reference for picture book designers to design new picture books, and for children to select appropriate ones. In addition, the method proposed in this study can be used to facilitate the design process that match to children’s affective preferences and needs while stimulating their purchase intentions.

Table 3. the preference form type of children in various books

	Attractiveness factors	Form type
a	Simple and comprehensible layout design	Absence of accessories Suitability for people of all ages
	Thin and lightweight dimensions	Traditional book format
	The sense of ownership	The sense of ownership obtained while holding
b	Various page displays	Pop-up pictures when turning pages
c	Vivid story plots	Using talking books to tell bedtime stories by mother
d	Multifunctional reading methods	Various reading methods

Reference

1.Asano, H. (2001). Miryoku Engineering Practice the procedure of popular product. Tokyo: Kaibundo shuppan.

2.Bus, A. G., van IJzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: a meta-analysis on intergenerational transmission of literacy. Review of Educational Research, 65(1), 1–21.

3.Carle, E. (1984). The very hungry caterpillar. New York: Philomel Books.

4.Elia, I., van Den Heuvel-Panhuizen, M., & Georgiou, A. (2010). The role of pictures in picture books on children’s cognitive engagement with mathematics. European Early Childhood Education Research Journal, 18(3), 125–147.

5.Hsiao, C. Y. (2010). Enhancing children’s artistic and creative thinking and drawing performance through appreciating picture books. International Journal of Art and Design Education, 29(2), 143–152.

6.Kelly, G. (1955). The Psychology of Personal Constructs: Vol. 2; Clinical Diagnosis and Psychotherpy (pp.278). New York: WW Norton.

7.Ma, M. Y., Chen, Y. C., & Li, S. R. (2011). How to Build Design Strategy for Attractiveness of New Products. Advances in Information Sciences & Service Sciences, 3(11), 1-10.

8.Sanui, J. (1996). Visualization of users' requirements: Introduction of the Evaluation Grid Method. Proceedings of the 3rd Design & Decision Support Systems in Architecture & Urban Planning Conference (Vol. 1, pp. 365-374).

9.Wei, C. C., Ma, M. Y., & Lin, Y. C. (2011). Applying Kansei engineering to decision making in fragrance form design. Intelligent Decision Technologies (pp. 85-94). Berlin Heidelberg: Springer.

A Virtual Reality Environment for Prospective Memory Training

Antonija Mitrovic[1,*], Moffat Mathews[1], Stellan Ohlsson[2], Jay Holland[1], Audrey McKinlay[3], Scott Ogden[1], Anthony Bracegirdle[1], Sam Dopping-Hepenstal[1]

[1]Intelligent Computer Tutoring Group, University of Canterbury, New Zealand
tanja.mitrovic@canterbury.ac.nz
[2]Department of Psychology, University of Illinois at Chicago
[3]Department of Psychology, University of Melbourne

Abstract. Prospective Memory (PM), or remembering to perform tasks in the future, is of crucial importance for everyday life. Stroke survivors often have impaired prospective memory, which can interfere with their independent living. In 2011, we started working on computer-based training for improving prospective memory in stroke patients. The primary goal of our project is to develop an effective PM treatment that could be used without the input of clinicians. Our approach combines the use of visual imagery with practice in a Virtual Reality (VR) environment. In this paper, we present the VR environment and the user modelling approach implemented.

Keywords: prospective memory training, virtual reality environment, constraint-based modeling,

1 Introduction

People with brain injury (including stroke) have severely impaired prospective memory in comparison to healthy people [1, 2]. Prospective memory, or remembering to perform actions in the future, is of crucial importance for everyday life [3]. Prospective memory failure can interfere with independent living, as it can result in forgetting to take medication, switch off the stove or missing doctor's appointments. It is a complex cognitive ability, which requires coordination of multiple cognitive abilities: spatial navigation, retrospective memory, attention and executive functioning [4].

There are two critical aspects of PM: it is closely related to retrospective memory (remembering what was learnt and experienced previously), as it is necessary to know what the task is in order to actually perform the task. The other aspect is the retrieval of the intention at the time appropriate for the action. There is a distinction between event- and time-based prospective tasks. In the case of a time-based task, a certain action needs to be performed at a certain time (e.g. having a doctor's appointment at 4pm). In event-based tasks, an action needs to be performed when a certain event happens (like asking a friend a question when we see them next time).

Prospective memory is very difficult to assess using neuropsychological tests as conventional tests consist of simple, abstracted activities that are very different from real-world tasks. In the last decade, many research projects have used Virtual Reality

(VR) in neuroscience research and therapy [5], ranging from the use of VR for assessing cognitive abilities, over neuro- and motor rehabilitation to psychotherapy, such as treatment of phobias. VR environments are computer-generated environments that simulate real-life situations and allow users to interact with them. They provide rich, multisensory simulations with a high degree of control and rich interaction modalities. They can also have a high level of ecological validity. VR has been used for assessment of prospective memory in patients with traumatic brain injury (TBI) [4] and stroke patients [2]. VR is suited for prospective memory as it supports complex, dynamic environments that require coordination of many cognitive abilities.

Although there has been some research done on how to assess PM, there is very little available on rehabilitation strategies for PM. Yip and Man [6] involved 37 participants in 12 sessions of prospective memory training using non-immersive VR. The participants were asked to perform a set of event- and time-based PM tasks in parallel with an ongoing task. The PM training was based on remedial and process approaches. The remedial approach provides repetitive exercise within the VR environment. The process approach, on the other hand, aims to support multiple facets of PM, and supports encoding of intention, retention and performance interval and recognition of cues. Participants were given a list of four shopping items they needed to memorize, and their recall was tested before entering the VR environment, where they needed to perform the tasks. The VR training showed significant improvement in participants' immediate recall of PM tasks, performance on both time- and event-based tasks as well as ongoing tasks, and also a significant improvement in self-efficacy.

In our previous work, we have developed many successful Intelligent Tutoring Systems (ITSs) using Constraint-Based Modeling (CBM) [7, 8]. In this paper, we present the VR environment we developed for PM training, and describe how we utilize CBM for tracking the user's PM skills in this environment. The participant will first be administered a set of psychological tests, followed by a set of sessions in which he/she will be trained on using visual imagery to remember PM tasks. After the training, the participant will practice in the VR environment, presented in Section 2. We have recently started an evaluation study, the goal of which is to determine the effectiveness of the developed PM treatment.

2 VR environment

We have used the Unity[1] game engine to develop a VR environment, which represents a house with common household objects, and a garden. Figure 1 shows two scenes from the environment. The user is given a problem, which consists of several PM tasks he/she needs to visualize first, and then perform in the VR environment. The user can perform various actions on objects in the VR environment, such as turning the TV set on or off. To perform an action, the user first selects the object, and then specifies the desired action from a menu. The user can view a clock whenever they choose, which is necessary for time-based tasks. The tasks vary in complexity: the ones in early sessions consist of a cue and a single action, such as *Turn on the*

[1] https://unity3d.com/unity

radio at 3pm. In later sessions, the user will be given more complex tasks, such as *When the oven timer beeps, take the roast out of the oven and put it on the dining table.* Some tasks, such as taking the roast out of the oven, involve other objects, which are added to the inventory. Other tasks require inventory items to be collected beforehand. Consider the task *Take the red shirt from the bedroom and put it into the washing machine*. The first step involves collecting the inventory item *red shirt*, while the second step involves operating the washing machine. The user can view the inventory at any time. The problems range in complexity: the initial ones contain only three simple tasks, and they become more complex as the user practices in the environment.

Fig. 1. Two scenes from the VR environment

The system maintains the list of active tasks. Tasks should only be attempted from a point known as 'cue discovery'. Time-based tasks become active several minutes before the stated time. For example, if the task is *Turn on the radio at 3pm*, the user can start to move towards the radio a few minutes earlier in preparation. Event-based cues only begin when the stated event occurs. Consider the task: *When the courier truck arrives, take the parcel and leave it in the study*. For this task, the user has no way of knowing when the courier will arrive, and so he/she cannot perform the action before the cue is discovered.

For every task, there is a finite amount of time for which the task can be completed before it becomes obsolete or impossible. However, this alone is not the only factor in determining which tasks are more important. Some tasks, such as turning off the stove, have worse outcomes for failing to complete than other tasks do, such as turning on the radio. Each task therefore has a priority level, which is an integer from 0 (lowest) to 5 (highest). Tasks with a level 5 priority are tasks with a very real chance of injury or household damage if they are not completed on time. A typical priority 5 task is *When the timer beeps, turn off the stove top*. By contrast, a priority 0 task may be: *When you are finished all other tasks, watch television.* From cue discovery, the user has a fixed time to complete the task before it becomes obsolete.

3 Using CBM for PM Training

We have defined a set of constraints that enable us to evaluate the participant’s actions and provide feedback. As originally proposed by Ohlsson (1992), each constraint has two components: a relevance condition and a satisfaction condition. The relevance condition specifies features of situations for which the constraint is relevant, while the satisfaction condition details what must be true for the constraint to be satisfied. A constraint can be described as: *If <relevance condition> is true, then <satisfaction condition> had better also be true, otherwise something has gone wrong*. If a constraint is violated, the user needs some means of knowing that he/she has made a mistake, and they need to know what needs to be done differently next time. This is the role of feedback: it informs the user on what tasks need to be performed, and what objects need to be interacted with.

We have developed 15 constraints that deal with navigation, prioritization of tasks, selection of objects to perform actions on, remembering/selecting actions to be performed and general skills of interacting with VR (such as selecting objects, selecting items from the menu or crouching). In order to be able to specify relevance and satisfaction conditions, we have defined a set of functions and predicates. For example, the *OnRouteTo* predicate takes the current position of the user (i.e. the room the user is currently in), the target position needed in order to perform the current task, and returns *True* if the current position is on a path to the target position.

A constraint contains three feedback messages. When a constraint is violated for the first time, the user will be given a general message, in order to remind them that they have missed something. For example, if the user is going in the opposite direction from the target destination, he/she will be given feedback "*You're going the wrong way*!" If the user continues down the wrong path, the feedback for the second violation of the same constraint becomes more specific: "*Perhaps you should be going to the [goalRoom]*" (*[goalRoom]* is a function which returns the position for the current task). This culminates on their third violation of the constraint with "*You should be going to the [goalRoom] and use the [goalObjects]*". This is the bottom-out feedback which instructs the user what to do.

Three constraints check whether the user is working on the correct task. Tasks with only one minute left should be done before tasks with more than one minute left, even if that task with more than one minute left is of higher priority. In this way the user can still complete all the tasks. It is also important to bear in mind that higher priority tasks will reach the point of only having one minute left a lot sooner than a lower priority task. If there are multiple tasks with less than one minute left, the user should choose the highest priority one. The next threshold is at five minutes. Users must do tasks with less than five minutes left before they attempt tasks with more than five minutes left. As discussed in the previous section, tasks are first stratified according to time left into less than one minute, less than five minutes, more than five minutes. From there they are ranked according to priority. If any tasks have equal time strata and priority, they can be done in any order, otherwise the user must pick the top one.

Figure 2 illustrates a situation when the user is interacting with the wood burner, but there is another task that is about to expire (*Once it starts raining, bring in washing*). The constraint relevant to that situation is:

If the user is interacting with Object X and there is a task with less than 1 min left,
Then Object X should be related to that task.

The feedback from the violated constraint shown in Figure 2 informs the user that there is a more pressing task. If the user cannot recall the other task, the next feedback message will be more specific, and will provide a hint to the user about the object he/she needs to interact with. In the case of the third violation of the same constraint, the user will be told which task needs to be performed.

Fig. 2. Feedback from a violated constraint

In addition to receiving feedback, the user can also press the H key for more help. If the user has received feedback in the last 30 seconds, the same feedback is displayed again as a reminder. Otherwise the default message is displayed. If there are no tasks left to do, the default feedback informs them of this. Otherwise it gives them increasingly specific hints as to what they should be doing.

In our previous work with ITSs, constraints are evaluated when the student submits the solution, therefore explicitly requiring feedback from the system. The timing of constraint evaluation in the VR environment differs, as the system needs to be able to evaluate constraints when appropriate. The constraints that deal with task prioritization are evaluated at intervals of 0.5s. Other constraints are evaluated in the appropriate contexts: for example, navigation constraints are evaluated every time the user changes room, while constraints that deal with objects are evaluated when the user selects an object or an action.

We have conducted a case study with a stroke survivor, who used the VR environment for 30 minutes. The case study identified a few usability issues and further improvements to the timing and duration of feedback. We then had a domain expert interact with the system. The domain expert was able to compare the feedback generated by constraints with the feedback they expected from the system. All constraints were satisfied or violated as expected. At some points, the feedback actually led to the domain expert making more errors. In such situations, the user was alerted that they should be doing one of several tasks, and told all the tasks currently available. When the user completed the lowest priority of these tasks, they violated the constraint that they should be doing the most high priority tasks. This led to the recommendation that feedback messages should only suggest the single most important task at the current time. The findings were then used to improve the constraint set and the system.

4 Conclusions

In our previous research, we have shown that constraint-based modeling is an effective student modeling approach applicable in a wide range of instructional domains. In this paper, we describe how we use CBM to track the user's prospective memory. We present a VR environment in which stroke survivors can improve their memorization skills. The contribution of this research is in extending CBM from modeling cognitive skills to modeling PM skills. We have developed a constraint set that allows us to track the user's behavior in the VR environment. The constraints identify whether the user is prioritizing tasks correctly, whether there are any problems with navigation, identifying cues (time or event ones), interacting with objects and specifying actions. The pilot study performed with one stroke survivor was promising. We also had a domain expert evaluate the feedback from the VR environment, which resulted in further improvements made. We are currently conducting an evaluation study with stroke patients.

Acknowledgments

This research was supported by a grant UOC1004 from the Marsden Fund of the Royal Society of New Zealand.

References

1. Mathias, J. L., Mansfield, K. M. Prospective and declarative memory problems following moderate and severe traumatic brain injury. Brain Injury. 19(4), 271-282 (2005).
2. Brooks, B. M., Rose, F. D., Potter, J., Jayawardena, S., Morling, A. Assessing stroke patients' prospective memory using virtual reality. Brain Injury. 18(4), 391-401 (2004).
3. Titov, N., Knight, R. G. A procedure for testing prospective remembering in persons with neurological impairments. Brain Injury. 14(10), 877-886 (2000).
4. Knight, R. G., Titov, N. Use of virtual reality tasks to assess prospective memory: Applicability and evidence. Brain impairment. 10(01), 3-13 (2009).
5. Bohil, C. J., Alicea, B., Biocca, F. A. Virtual reality in neuroscience research and therapy. Nature reviews neuroscience. 12(12), 752-762 (2011).
6. Yip, B. C., Man, D. W. Virtual reality-based prospective memory training program for people with acquired brain injury. NeuroRehabilitation, 32(1), 103-115 (2013).
7. Ohlsson, S. Constraint-based student modeling. Artificial Intelligence in Education, 3(4), 429-447 (1992).
8. Mitrovic, A. Fifteen years of constraint-based tutors: what we have achieved and where we are going. *User Modeling and User-Adapted Interaction*, *22*(1-2), 39-72 (2012).

Life-Logging for Healthcare Proactive Advisory Systems

Thai Son Nguyen[1], Francesco Ricci[1], Floriano Zini[1], and Marcello Granconato[2]

[1] Free University of Bozen-Bolzano, Faculty of Computer Science
Piazza Domenicani 3, 39100 Bolzano, Italy
thaison.ngn@gmail.com, fricci@unibz.it, floriano.zini@unibz.it
[2] Stallergenes Italia
marcello.granconato@gmail.com

Abstract. Proactive Advisory Systems (PASs) are an extension of Recommender Systems (RSs) that push suggestions and advice even if not explicitly requested, while adapting such advice to the specific contextual situation of the user. They support the user in decision-making processes more complex than those considered by traditional RSs. In this paper, we illustrate the first results of the *LiloPAS* research project, targeted to the design, development, and evaluation of PASs for healthcare. We present *Smart Allergy Taming*, a prototype mobile PAS for allergic patients, and give the results of its evaluation.

Keywords: proactive advisory systems, life-logging, user modeling, healthcare.

1 Introduction

Proactive Advisory Systems (PASs) [1, 14] are decision and task support systems that take the initiative to timely suggest actions and help users to make decisions. PASs require the extension of the traditional computational and interaction model adopted in Recommender Systems (RSs). Other than providing recommendation when explicitly requested, PASs should also be able to proactively push or recall suggestions and advice, while adapting such advice to the specific contextual situation of the user. In addition, PASs should support user's decision-making processes that are generally more complex and risky than those supported by traditional RSs, which are mostly focused on suggesting simple items such as movies to watch or products to purchase.

We are interested in implementing PASs for healthcare as mobile applications. In particular, our case study application supports patients suffering from allergic rhinitis in the proper management of their immunotherapy and in the adoption of behaviors that reduce the allergy impact on their quality of life. Allergic rhinitis is one of the most common chronic diseases in the world and affects up to 30% of people [9]. Allergen immunotherapy offers to patients more than symptomatic treatments, as its benefits persist several years after discontinuation of the treatment. However, immunotherapy requires a very careful management, as the treatment often needs to be administrated for a period of 3–5 years to be effective, and the patients need to strictly adhere to a precise schedule of medication intakes. Patients suffering from allergic rhinitis could take advantage of a mobile PAS in several ways. For example: receiving notifications that improve the adherence to the treatment schedule; or being suggested actions that reduce the allergy impact, such as closing the car windows when they are driving in an area with high concentration of pollens to which they are allergic.

The development of PASs brings new research challenges. In order to be effective in the proactive support of the user, as allergic patients require, PASs should be able to continuously monitor the user's activities and environment and build, automatically or with minimal user intervention, a meaningful user model. The model could capture, for example, that the allergic patient is keen on physical exercise in the open. The learned model must then be exploited to provide relevant advice fired by the current users situational context. For example, the sporty allergic patient could be advised to exercise in the open when the concentration of the pollen to which she is allergic is low.

The goal of the *LiloPAS* project (*Life-logging for Proactive Advisory Systems*) is to address the above mentioned research challenges and investigate techniques and algorithms for healthcare PASs, to be embedded in a mobile prototype that provides advice to patients affected by allergic rhinitis. As a first step towards a fully operational PAS, we designed and implemented *Smart Allergy Taming*, a proof-of-concept we used to quickly collect feedback useful to drive the further stages of the project.

In the rest of this paper, we briefly give an account of the background research on which our project is based and describe the design and the first evaluation of *Smart Allergy Taming*.

2 State of the Art

Personalized ehealth [10], and in particular the field of mobile-phone-based personalized health interventions, is receiving increasing attention. In their rich review [7], the authors survey systems developed in health science and human-computer interaction (HCI) and highlight the importance of understanding the patient context in order to improve the intervention strategies.

Previous research on life-logging focused on the acquisition, elaboration, and access of users' life data. Some researches concentrated on the development of toolkits that help to automatically acquire users' data, either capturing desktop activities [5], or logging from mobile phone sensors and applications [3, 11], or doing context-triggered sampling of experiences by directly asking users [3]. Other researches [2, 8] investigated how to elaborate the acquired data so that meaningful higher-level context can be discovered. How the users can access their logs is addressed in [2, 13].

Nowadays there is an increasing number of commercial wearable trackers[3], equipped with sophisticated sensors that record data such as steps, heart rate,

[3] See `http://www.pcmag.com/article2/0,2817,2404445,00.asp` for some examples.

burned calories, or how many times one wakes up during the night. Wearable trackers can be exploited, in addition to smartphones, as life-logging devices.

Finally, it is worth mentioning that life-logging research has its roots in HCI. Users need to interact with their (mobile) life-logging applications, which must be appealing and usable. For example, in [4] the authors evaluate various user data input interfaces for a mobile health application that collects data relevant to the treatment of insomnia.

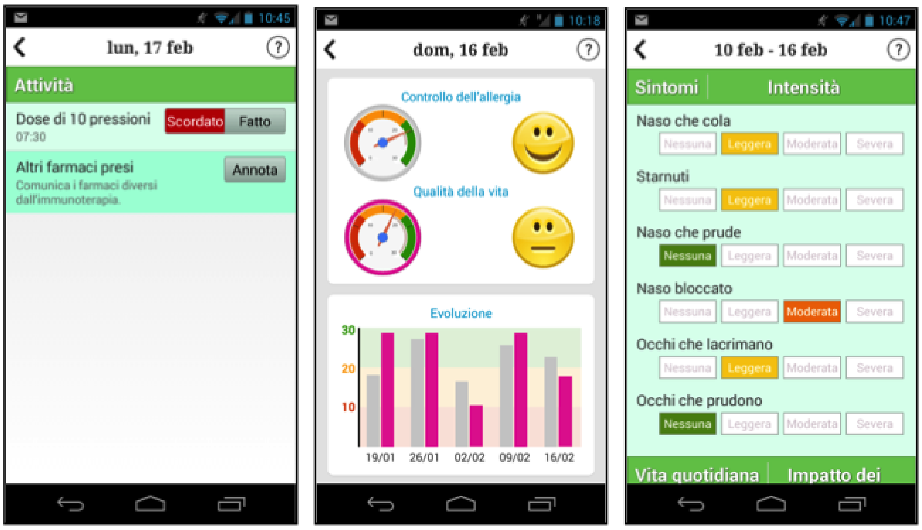
3 Smart Allergy Taming

Smart Allergy Taming is a mobile PAS for patients suffering from allergic rhinitis. The functionalities of the prototype were designed with the cooperation of experts of the partner company Stallergenes Italia[4]. They were selected to cover the main aspects (support for a correct intake of the treatment and monitoring of the impact of the treatment on the patients' health) that the expert hypothesized as important for a mobile system supporting a successful sublingual immunotherapy. The function of *Smart Allergy Taming* is:

1. Digital agenda of the activities the allergic patients have to do to correctly adhere to their immunotherapy;
2. Proactive reminders of the tasks in the agenda, e.g., treatment intake, fill out of health surveys (Figure 1-a);
3. Simple proactive tips about how to manage and keep the vials containing the medicines for the immunotherapy;
4. Timer that helps the patients when they take the dose of medicine, which has to stay two minutes under the tongue;
5. Indicators for the current and recent level of control of the allergy, and for the impact of the allergy symptoms on the patients' quality of life (Figure 1-b);
6. Log of the symptoms experienced by the patients and their impact on the patients' quality of life, by means of digital survey administrated every week (Figure 1-c); the collected information is processed to calculate the indicators mentioned above.

In addition to the above mentioned functionalities, we implemented in *Smart Allergy Taming* a simple mechanism, showing predicted values for the user replies to the various items in the questionnaire. We conjectured that such a mechanism could help the patients better remember the intensity of symptoms and their impact on the daily life in the past week. In this way, we wanted to heal the well-known discrepancy between experienced well-being (more precise) and remembered well-being (influenced by memory and personality biases) [6], that can limit the reliability of the health report filled by the patients. The system weekly collects the user's replies to the various questions, hence building time

[4] Stallergenes Italia (`www.stallergenes.it`) is a leading company for the production of treatments for allergy.



(a) (b) (c)

Fig. 1. Some function of *Smart Allergy Taming* (Italian version): (a) Reminder; (b) Indicators; (c) Survey.

series, and then predicts the next week user input, offering it as default value in the questionnaire. The predictions are calculated using exponential smoothing[5], so that the recent answers have a greater weight than the older ones.

3.1 Evaluation

We are evaluating *Smart Allergy Taming* with the goal to acquire feedback, in the early stage of the project, on the effectiveness of the implemented function, on the usability of our design, and on the usefulness of the remembering support mechanism we have developed.

We conducted a pilot test involving two domain experts from Stallergenes (different from those involved in the design of the prototype), one HCI expert, and four computer science students of our University. We decided not to involve patients in this phase to speed up the evaluation, as recruiting patients would have delayed the pilot. The users were given a paper sheet describing the functions of the system and a typical usage scenario, they were asked to execute. The task consisted in reporting the immunotherapy and other medicines intake, simulating the immunotherapy intake, filling the weekly health survey, and visualizing the allergy control and quality of life indicators.

At the end of the task the users were asked the *Single Ease Question (SEQ)*, a 7-points rating scale to asses how difficult a user finds a task, with 1 indicating

[5] `http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc43.htm`.

"very difficult" and 7 indicating "very easy" [12]. In addition, they were asked to indicate up to 5 usability problems they encountered and up to 5 liked features of the application.

The average SEQ score of our sample was 5.43 with a standard deviation of 0.53. The 98% confidence interval for the score, calculated using the t-distribution, ranges from 4.80 to 6.06. We compared our result with the average SEQ score across over 200 tasks performed by 5000 users (included in a reference benchmark), which hovers between about 4.8 and 5.1[6]. Even though our sample contains a small number of subjects, we can conclude that *Smart Allergy Taming* easiness of use is on a par with the benchmark.

By examining the users' answers, we discovered that they generally liked the system function. Some users appreciated the tips and indicated that additional contextual advice about how to manage the disease could be useful. The users also appreciated the look and feel of the application, especially the graphical elements. Most of them considered the navigation through the various views as easy and intuitive. Some users also highlighted usability problems such as minor navigation issues between application views, colors and contrast not suitable for people with sight problems, and audio notifications not always audible. In summary, the pilot test demonstrated the goodness of the system functionalities and the design of the GUI.

We are currently improving *Smart Allergy Taming* in order to eliminate the usability problems highlighted by the pilot test. The improved application will be evaluated in a between-subjects experiment involving real allergic patients. The objective of the experiment is to assess the usefulness of our memory support mechanism to fill in the survey, comparing a version of *Smart Allergy Taming* embedding the support mechanism with a version without support. The metric we are going to use in the evaluation is the accuracy of the reported intensity of symptoms and their reported impact on the patient's quality of life. Our hypothesis is that the memory support mechanism produces a higher accuracy. In order to assess our hypothesis, we plan to use a digital diary that the patients will fill every day of the experiment. In the diary they will indicate the intensity (on a scale from 1 to 4) of the symptoms that have experienced in the current day and which was the impact of these symptoms (on a scale from 1 to 4) on various aspects of their daily life. The answers recorded in the diary will be averaged in order to produce the gold standard to be compared with the answers given in the weekly survey using the two versions of the mobile application.

4 Conclusions

Smart Allergy Taming is a PAS prototype that was developed to investigate which functionalities are suitable for supporting allergic patients in the the execution of their immunotherapy, to early evaluate the design of a GUI for these functionalities, and to evaluate a memory support mechanism for manual logging of the health status of the patients. The results of the pilot test we have

[6] `http://www.measuringusability.com/blog/seq10.php`.

conducted indicate that the selected functions the GUI are appropriate. We are currently assessing the memory support mechanism. The results of the evaluation of *Smart Allergy Taming* will drive the further stages of the *LiloPAS* project.

References

1. H Al Tair, MJ Zemerly, M Al-Qutayri, and M Leida. Architecture for context-aware pro-active recommender system. *International Journal Multimedia and Image Processing (IJMIP)*, 2(1/2):15–133, 2012.
2. Petros Belimpasakis, Kimmo Roimela, and Yu You. Experience explorer: a life-logging platform based on mobile context collection. In *Next Generation Mobile Applications, Services and Technologies, 2009. NGMAST'09. Third International Conference on*, pages 77–82. IEEE, 2009.
3. Jon Froehlich, Mike Y Chen, Sunny Consolvo, Beverly Harrison, and James A Landay. Myexperience: a system for in situ tracing and capturing of user feedback on mobile phones. In *Proceedings of the 5th international conference on Mobile systems, applications and services*, pages 57–70. ACM, 2007.
4. Daniel Gartenberg, Ross Thornton, Mortazavi Masood, Dustin Pfannenstiel, Daniel Taylor, and Raja Parasuraman. Collecting health-related data on the smart phone: mental models, cost of collection, and perceived benefit of feedback. *Personal and ubiquitous computing*, 17(3):561–570, 2013.
5. Jim Gemmell, Gordon Bell, and Roger Lueder. Mylifebits: a personal database for everything. *Communications of the ACM*, 49(1):88–95, 2006.
6. Daniel Kahneman and Jason Riis. Living, and thinking about it: Two perspectives on life. *The science of well-being*, pages 285–304, 2005.
7. Predrag Klasnja and Wanda Pratt. Healthcare in the pocket: Mapping the space of mobile-phone health interventions. *Journal of biomedical informatics*, 45(1):184–198, 2012.
8. Henar Martín, Ana M Bernardos, Josué Iglesias, and José R Casar. Activity logging using lightweight classification techniques in mobile devices. *Personal and ubiquitous computing*, 17(4):675–695, 2013.
9. Ruby Pawankar, GW Canonica, ST Holgate, and RF Lockey. *WAO white book on allergy*. Milwaukee, WI: World Allergy Organization, 2011.
10. C. C. Y. Poon, M. D. Wang, P. Bonato, and D. A. Fenstermacher. Editorial: Special issue on health informatics and personalized medicine. *Biomedical Engineering, IEEE Transactions on*, 60(1):143–146, 2013.
11. Reza Rawassizadeh, Martin Tomitsch, Katarzyna Wac, and A Min Tjoa. Ubiqlog: a generic mobile phone-based life-log framework. *Personal and ubiquitous computing*, 17(4):621–637, 2013.
12. Jeff Sauro and James R Lewis. *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann, 2012.
13. Teruhiko Teraoka. Organization and exploration of heterogeneous personal data collected in daily life. *Human-centric Computing and Information Sciences*, 2(1):1–15, 2012.
14. Wolfgang Woerndl, Johannes Huebner, Roland Bader, and Daniel Gallego-Vico. A model for proactivity in mobile, context-aware recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys '11, pages 273–276, New York, NY, USA, 2011. ACM.