Prague, Czech Republic

August 18, 2014

Proceedings of the 3rd International Workshop on

Artificial Intelligence and Assistive Medicine (AI-AM/NetMed 2014)

In conjunction with
ECAI 2014

21st European Conference on Artificial Intelligence
August 18-22, 2014
Clarion Congress hotel Prague
Czech Republic

Prague, Czech Republic

August 18, 2014

Proceedings of the 3rd International Workshop on

Artificial Intelligence and Assistive Medicine (AI-AM/NetMed 2014)

In conjunction with
ECAI 2014

Organization

Chairs

Constantine D. Spyropoulos	NCSR “Demokritos”, Greece
Aldo Franco Dragoni	Università Politecnica delle Marche, Italy
Stavros Perantonis	NCSR “Demokritos”, Greece

Program Committee

Panagiotis Bamidis	Aristotelian University of Thessaloniki, Greece
John Baras	University of Maryland, USA
Bruno Bouchard	Université du Qubec, Canada
David Buckeridge	McGill University, Canada
Eric Campo	LAAS-CNRS, France
Andrea Claudi	Università Politecnica delle Marche, Italy
Gabriella Cortellessa	Consiglio Nazionale delle Ricerche, Italy
Michel Dojat	Université Joseph Fourier(Grenoble I), France
Peter Flach	University of Bristol, United Kingdom
Werner Horn	Medical University of Vienna, Austria
Vangelis Karkaletsis	NCSR “Demokritos”, Greece
Ioannis Kompatsiaris	Centre for Research & Technology Hellas, Greece
Stasinos Konstantopoulos	NCSR “Demokritos”, Greece
Ilias Maglogianis	University of Piraeus, Greece
Fillia Makedon	University of Texas at Arlington, USA
Yuval Schahar	Ben Gurion University of the Negev, Israel
Paolo Sernani	Università Politecnica delle Marche, Italy
Luca Spalazzi	Università Politecnica delle Marche, Italy
Bert-Jan van Beijnum	University of Twente, The Netherlands

Organising Committee

Daniele Accattoli	Università Politecnica delle Marche, Italy
Andrea Claudi	Università Politecnica delle Marche, Italy
Paolo Sernani	Università Politecnica delle Marche, Italy

Table of Contents

Preface

The medical systems that the Governments operate to fulfil the healthcare requirements of the Society are always looking to offer higher quality at the lowest cost. It seems that the current systems are unsustainable as the rate of senior society members to the active employees is increasing and the curve of required medicine cost for the last 10 years of people's life is more than the corresponding medical cost used for the rest of their life. The question is then whether we can "flatter" or smoother the lifetime cost curve for the provided healthcare services.

This is a big challenge for AI as well as the always evolving digital technology. How can we be proactive rather than post-active in order to early detect, make diseases prognosis and give the right assistive medical support on time? Answer to this question can offer the exploitation of AI methods in combination with the sensor networks technology and the new technology devices like smart phones, tablets, digital TVs, web cameras and all the smart gadgets that appear in the market.

Thus, we organize the workshop on Artificial Intelligence and Assistive Medicine, hoping to contribute in answering this question by collecting works about AI-related techiniques in medicine. The papers accepted for this one-day workshop give special emphasis in:

- Ubiquitous real-time assistive healthcare
- Ambient assisted living
- Wearable and/or unobtrusive smart healthcare systems
- Multi-Agent architectures for patient monitoring and early diagnosis
- Fusion and interpretation of multimodal medical data and events
- Medical ontology modelling and evolution
- Semantically diagnosis modelling
- Reasoning with the uncertainty of medical data/knowledge
- Mining on medical data/knowledge
- Patient centric and evidence based decision support systems

Constantine D. Spyropoulos[1]
Aldo Franco Dragoni[2]
Stavros J. Perantonis[1]

Co-chairs of AI-AM/NetMed 2014

[1]**National Center for Scientific Research "Demokritos", Greece**
[2]**Università Politecnica delle Marche, Italy**

The Impact of Different Training Sets on Medical Documents Classification

Roberto Gatta[1], Mauro Vallati[2], Berardino De Bari[3], and Mahmut Ozsahin[3]

[1] Department of Radiation Oncology, Università Cattolica S. Cuore, Italy
roberto.gatta.bs@alice.it
[2] School of Computing and Engineering, University of Huddersfield, UK
m.vallati@hud.ac.uk
[3] Department of Radiation Oncology, Centre Hospitalier Universitaire Vaudois, Switzerland
name.surname@chuv.ch

Abstract. The clinical documents stored in a textual and unstructured manner represent a precious source of information that can be gathered by exploiting Information Retrieval techniques. Classification algorithms can be used for organizing this huge amount of data, but are usually tested on standardized corpora, which significantly differ from actual clinical documents that can be found in a modern hospital. The result is that observed performance are different from expected ones. Given such differences, it is unclear how should be the "right" training set, and how its characteristics affects the classification performance.
In this paper we present the results of an experimental analysis, conducted on actual clinical documents from a medical Department, which aims to evaluate the impact of differently sized and assembled training sets on well-known classification techniques.

1 Introduction

In modern hospitals a large amount of clinical documents are stored in a textual and unstructured manner; these documents are precious sources of knowledge that must be exploited rather than uselessly stocked. In order to exploit such knowledge, it is fundamental to classify the documents. Information Retrieval (IR) techniques provide an established way to distinguish the documents according to their general meaning (see, for instance [1]).

The traditional approach to IR envisages to exploit a large number of already classified documents — the ground truth — for training classification algorithms. Generally, we larger the training set, the better the expected performance. Usually, IR approaches are evaluated on standard corpora, that are significantly different from documents that can be found in real-world environments. In such environments, and especially in medical ones, several factors can affect the performance of IR classifiers, and limit the usefulness of extremely large training sets. Probably, the most critical one is the so-called documents obsolescence [4]. It refers to the fact that in a clinical context the turn-over of human resources, and the introduction of new techniques and methodologies, can quickly change the text style of medical reports; documents of the training set that include obsolete terms or structure can play the role of noise for the classification process.

Therefore, the usual approach based on exploiting large training sets could be not the best technique.

In this paper we perform an experimental analysis, on about 3,000 medical documents from a Radiotherapy Department, which aims to evaluate how classification performance are affected by (i) differently sized training sets, and (ii) the similarity of training documents with a given one.

2 Considered IR Algorithms

For the sake of this investigation, we considered three existing classifiers: Rocchio [6], ESA [4] and Naive Bayes [3]. Rocchio and Naive Bayse are well-known in literature, thus they represent the state-of-the-art. while ESA is a recent and somehow different classification algorithm.

Rocchio classifier uses a Vector Space Model (VSM) [7] to generate a multi-dimensional space where a document is represented as a vector, which components are functions of the frequencies of the terms. For each class of documents, a centroid is generated. New documents are classified as members of the class whose centroid is closer. Rocchio suffers of low accuracy while it has to classify documents that are close to the boundaries of a centroid. Our implementation adopted the tf-idf [8] technique to weight the terms in documents and used an Euclidean Distance metric to measure distances from centroids.

ESA is based on the idea of entropy, and exploits a two step training process. In the first step it selects a set of terms that better helps to predict the probability $p(t_i/c_j)$ that a document is classified as c_j given the fact that it contains the term t_i. In the second step, ESA calculates the entropy values associated to each term and discharges the terms which entropy is over a given threshold. For classifying a new document, the score $score(c_j)$ of each class is determined using Equation 1. The class with higher score is selected.

$$score(c_j) = \prod_{i=1}^{n} [1 - p(c_j/t_i)] \tag{1}$$

A Naive Bayes (NB) classifier uses a Bayesian approach to calculate the probability that a document is a member of every possible class. Even if it is based on the strong hypothesis of conditional independence between features, NB usually shows good performance; moreover it allows to estimate the uncertainty by evaluating the probability ratios between all the couples of possible classes.

3 Experimental Analysis

Clinical documents were collected from a Radiotherapy Department. It contributed with discharge "forms". Each form is composed by 21 different documents, which should be classified according to the aspect of the patient they describe; the usage of tobacco and alcohol, allergies, medications, treatment plan, etc. The total number of document is about 3,000, written in French, that were divided in 21 classes, as previously stated.

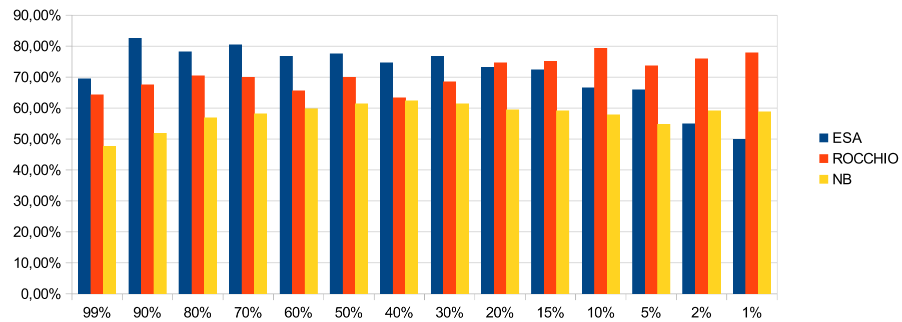

Fig. 1. Accuracy of ESA, Rocchio and Naive Bayes (NB), with regard to the considered percentage of ordered documents from the training set.

The documents are generally short (94 is the average number of words) and their structure can be significantly different, since no guidelines or "standard sentences" are proposed to physicians by the input system. We observed that different physicians wrote documents in very different ways, both from structural and syntactic point of view.

Out of the available 3,000 documents, 2,700 have been considered for training, while the rest for testing the different algorithms. Since the focus of the analysis is on the differences between large and small training sets, rather than on the testing accuracy itself, we decided to exploit a very large amount of available data for training. We considered different percentages of the learning set, ranging from 1% to 99%. In order to evaluate how quality of training instances affect classification performance, we decided to select training documents which are "close" to the given one. In other words, the given document plays the role of "centroid" for a kNN [5] extraction which aim is to build the training set. The rationale is that, in order to limit the detrimental effect of obsolescence on classification performance, looking at documents that have been already classified and are similar to the given one should provide useful information. Similar documents can either be written in the same period or from the same physician. Distance of documents have been quantified by evaluating the euclidean distance between the tf-idf of the documents on all the words. Higher percentages of considered training documents imply that less similar documents are exploited for training IR algorithms and, potentially, that more noise is introduced.

Figure 1 shows how the considered percentages of ordered (w.r.t. tf-idf) documents of the training sets affect the accuracy performance of the IR algorithms. Remarkably, the three algorithms show different behaviours. Rocchio shows the best accuracy while using only the 10% of the available training set; its performance are then monotonically decreasing when the number of training documents growth. Naive Bayes accuracy performance remain stable while considering training sets with a size between 1 and 60% of all the available training documents. After that the accuracy decreases quickly. Finally, ESA is the only considered approach in which accuracy proportionally increases with the size of the exploited training set. It is worth noticing that all the algorithms

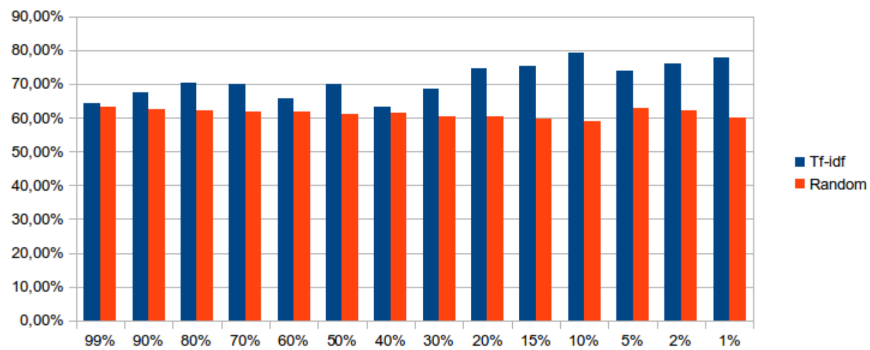

Fig. 2. Accuracy of Rocchio exploiting tf-idf based training documents selection and a random selection.

show somehow good performance, considering that the documents can be classified in 21 different classes, also when exploiting a very small number of training problems. This is probably due to the good quality of training sets, which derives from using the tf-idf technique for selecting them.

Interestingly, the average accuracy among the three algorithms monotonically increases between 1 and 20%, monotonically decreases from 70% upward, and remains almost the same in between. While a lower accuracy with very small training set was expected, it is surprising that also very large sets lead to reduced accuracy. This suggests that the approach "the more the better" should be revised and improved with better selection techniques of training sets.

For a better comprehension of the actual impact of a good selection of training instances, we compared the accuracy performance of Rocchio exploiting (i) the aforementioned tf-idf based technique and (ii) a random selection of documents from the training set. Figure 2 shows the results of such comparison. Remarkably, the performance gap is significant. A good selection of training instances lead to an evident performance improvement. Moreover, while exploiting a random selection of training documents, the size of the training set does not significantly affect the classification performance.

4 Conclusions

In modern hospitals a large amount of clinical documents, which represent a precious source of information, are stored in a textual and unstructured manner. In order to exploit such knowledge, it is fundamental to classify the documents using IR algorithms. Traditional IR approaches are tested and evaluated on standard corpora, that usually have characteristics which are very different from those of real-world documents. One of the aspects, observed in clinical documents but not in standard corpora, that has a remarkable impact on IR performance is the obsolescence, which refers to the fact that turn-over of human resources, and introduction of new techniques and methodologies, can quickly change the text style of reports. The presence of such sudden changes in real-word text corpora makes the standard learning approach — the larger the training set, the better — questionable; documents of the training set that include obsolete terms or structure, w.r.t. the current document to classify, can play the role of noise.

In this work we experimentally evaluated how differently sized and differently assembled training sets affect the classification performance of three IR approaches on clinical documents from a Radiotherapy Department. The take-home messages that can be synthesised are: (i) the size of the training set does not significantly affect the classification performance; (ii) a good selection of training instances can boost the accuracy, i.e. selecting training instances which are similar to the one to classify, according to some metrics. While the latter is intuitive, the former result is astonishing. It clearly indicates that focusing on collecting large amount of training documents is not always the best strategy for achieving good performance, at least on considered algorithms.

We see several avenues for future work. Concerning the documents, we are interested in performing a larger experimental evaluation on documents from different departments. It can be expected that, in departments that support physicians through guidelines or "standard sentences", selecting training instances will not have remarkable impact on IR performance. Moreover, we will extend the set of considered classification algorithms by including more well-known approaches (e.g., KNN [5]) and ensemble methods [2]. It will be useful to understand if such methods, which combine different classification algorithms, suffer noticeably documents obsolescence, and how different training sets affect their performance.

References

1. Bratsas, C., Koutkias, V., Kaimakamis, E., Bamidis, P.D., Pangalos, G.I., Maglaveras, N.: Knowbasics-m: An ontology-based system for semantic management of medical problems and computerised algorithmic solutions. Computer methods and programs in biomedicine 88(1), 39–51 (2007)
2. Dietterich, T.G.: Ensemble methods in machine learning. In: Multiple classifier systems, LBCS-1857. pp. 1–15. Springer (2000)
3. Frank, E., Bouckaert, R.R.: Naive bayes for text classification with unbalanced classes. In: In Proc 10th European Conference on Principles and Practice of Knowledge Discovery in Databases. pp. 503–510 (2006)
4. Gatta, R., Vallati, M., De Bari, B., Pasinetti, N., Cappelli, C., Pirola, I., Salvetti, M., Buglione, M., Muiesan, M.L., Magrini, S., et al.: Information retrieval in medicine: an extensive experimental study. In: the 7th International Conference on Health Informatics (HealthInf) (2014)
5. Guo, G., Wang, H., Bell, D.A., Bi, Y., Greer, K.: An knn model-based approach and its application in text categorization. In: In Proc 5th International Conference on Computational Linguistics and Intelligent Text Processing. pp. 559–570 (2004)
6. Rocchio, J.J.: Relevance feedback in information retrieval. In: The Smart retrieval system - experiments in automatic document processing, pp. 313–323. Englewood Cliffs, NJ: Prentice-Hall (1971)
7. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Commun. ACM 18(11), 613–620 (1975)
8. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. In: Information processing and management. pp. 513–523 (1988)

MAGPIE: An Agent Platform for the Development of Mobile Applications for Pervasive Healthcare

Albert Brugués[1,2], Stefano Bromuri[1], Josep Pegueroles[2], and Michael Schumacher[1]

[1] University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland
{name.surname}@hevs.ch
[2] Universitat Politècnica de Catalunya - BarcelonaTech (UPC), Barcelona, Spain
josep.pegueroles@upc.edu

Abstract. In this paper we present the Mobile computing with AGents and Publish subscribe for Intelligent pervasive hEalthcare (MAGPIE) platform. MAGPIE is an agent platform designed for the Android OS. The aim of the platform is to simplify the definition of Personal Health Systems (PHSs) to monitor chronic diseases. The agents running in the platform use a symbolic reasoning approach to formalize the events happening to the patient. We show the formalization of this reasoning for the particular case of monitoring Gestational Diabetes Mellitus (GDM).

1 INTRODUCTION

The new advances in medicine are contributing to an increase of life expectancy, which in turn increases the healthcare costs due to a major prevalence of age-related chronic diseases. Pervasive Healthcare [1] is a scientific discipline that tries to mitigate these issues by defining Personal Health Systems (PHSs). These systems shift the paradigm of healthcare services, by moving them from a centralized approach focused on doctors to a decentralized one focused on patients; that is a pro-active and preventive delivery model where people are active participants in their own well-being.

In the context of PHSs the use of mobile devices with sensors deployed on the body gives the vision of *healthcare to anyone, anytime and anywhere* [5]. In the recent years the market of smartphones and tablets has been well established. Nowadays the smartphones hardware components offer powerful computation capabilities that allow to perform the same tasks we do with a desktop computer. Another factor that contributed to the establishment of this new scenario for mobile computing is the apparition of operating systems specially designed for handheld devices, like Android [2]. Android as it is offered as an open source solution, can be used by different vendors in their products without adding additional costs. Moreover, application developers can create and publish applications for this operating system and target a wide range of devices. In the particular case of PHSs there is a key fact in the new generation of mobile devices that are the integrated sensors like accelerometers, GPS receiver, ambient light, etc. that can provide information that complements the one provided by the sensors deployed on the body.

In this work we present the implementation directions of MAGPIE, an agent platform for simplifying the development of mobile applications in Android with the aim of monitoring chronic diseases. The platform is based on the concept of agent environment as a first class abstraction [6], and it is designed with the aim of tackle some of the technological challenges arising from the development of PHSs like modeling the domain knowledge, their scalability and their personalization.

The agent environment concept, is becoming increasingly more important to simplify the definition and deployment of multiagent applications, by mediating the interaction between the agents and resources deployed in the system, by hiding to the agents the complexity of dealing with the state of resources external to the agent, and by providing standard interfaces and standard descriptions to resources so that the agents can utilize them for their own goals.

2 THE MAGPIE AGENT PLATFORM

The aim of the MAGPIE agent platform is to help on the development of mobile applications that can be used in a PHS for monitoring chronic diseases. In a PHS patients with one or more chronic diseases are monitored by means of sensors deployed on their body. In MAGPIE we link such sensors with the abstraction of agent environment in multiagent systems [6]. The agents deployed in the agent environment can perceive the events happening in the patient's environment, perform reasoning on these events and produce alerts of interest for the particular disease being monitored.

As shown in Figure 1 the MAGPIE agent platform consists on different components. The central element of the platform is the *environment* where we can deploy two main entities: *agents* and *context entities*. Agents are cognitive entities deployed on the agent environment and are composed by a declarative mind called agent mind, which is the component in charge of the agent's reasoning abilities. The mind of an agent is situated in the environment through another component called agent body. The agent body is the part of the agent that receives and produces events from/to the agent environment, so it acts as an interface between the agent mind and the agent environment.

Context entities are connectors linking the real environment with the agent environment. They encapsulate the communication with a source of information from the real world. The goal of a context entity is to throw to the agent environment events related with physical measurements from the real world, so that the agents can perceive them. There are different kinds of context entities for the three different sources of the information we have. First, measurements can come from Bluetooth sensors deployed on the body of the patient, which can measure physiological values like the heart rate. Second, measurements can

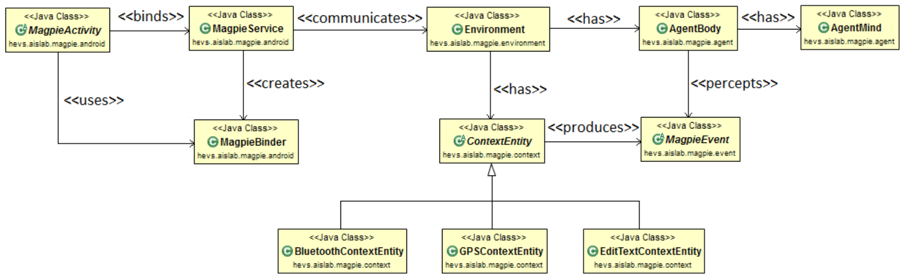

Fig. 1. MAGPIE class diagram

come from the sensors of the smartphone, which can provide for example the GPS position of the patient. Last, measurements can be provided by the patient itself through the user interface of a mobile application to report values that are difficult to measure with sensors, like the amount of carbohydrates of a meal.

The environment acts as a mediator for the interactions between the agents and the context entities. The events produced by the context entities are identified by the kind of measurement they represent, and the agent environment notifies the events to those agents interested on that particular measurement.

The agents and the environment have a lifecycle that takes into consideration the limited energy resources of smartphones. This limitation implies that in Android it is not possible to consider a full multithreading approach for agents as if multiple threads were to be run, then the battery life would decrease. The environment lifecycle takes care of mainly two things after the initialization of the Android application. Firstly, the environment dispatches events to the entities deployed in it. Secondly, the environment works also as a scheduler for the agents. For the sake of this contribution, the implementation of the environment entity has a sequential scheduler to execute the existing agents. The agent lifecycle is more complex. In a lifecycle an agent has to perceive the environment, to update its internal state and then, if no modification of the model is necessary, to perform actions in the environment, such as submitting alerts to the patient. When the agent perceives an event of model modification, then the current agent mind is discarded and modified with the new model, and the agent starts its cycle again from perceiving the environment. Contrary to the environment and the agents, the context entities are not active, so they are activated only when triggered by an event, performing a purely reactive behavior.

An important characteristic of the MAGPIE agent platform is its integration with the Android OS. We use two of the Android main components for that purpose: activities and services. An activity represents a graphical interface that the user can see on the screen, in MAGPIE activities are used as a communication channel between a mobile application and the patient. A service runs in the background to perform long-running operations that do not interact with

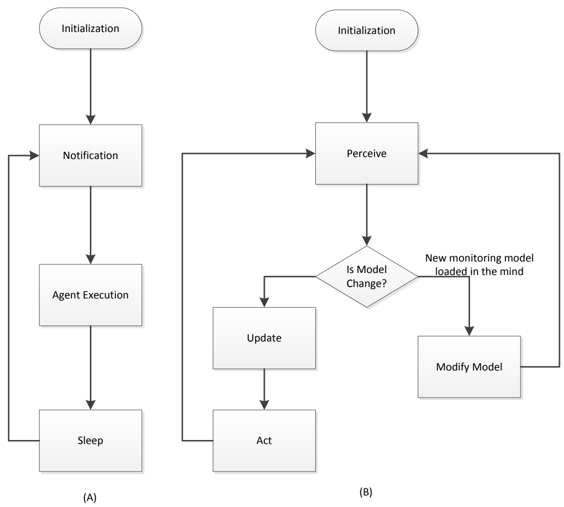

Fig. 2. Lifecycles of the (a) environment, and the (b) agents

the user, in MAGPIE the environment, the agents and the context entities run autonomously in a background service.

3 KNOWLEDGE REPRESENTATION

The knowledge used by the agents is based on the Event Calculus (EC) [4]. The EC is a formalism defined in Prolog for representing actions and their effects, so EC is suitable to model expert systems representing the evolution in time of an entity, by means of the production of events. In this case, our EC reasoner is embedded inside an agent, and models alerting rules applied by medical doctors through a web interface. More specifically, the events produced in the agent environment that are `LogicTuples` are automatically translated to a first order logic representation that can be interpreted by the reasoner residing the agent mind of MAGPIE agents.

In this paper we are motivated by the use case of monitoring Gestational Diabetes Mellitus (GDM); a condition affecting 3-4% of pregnant women due to increased resistance to insulin caused by the growth of the baby. Such a condition disappears just after delivery, but it is an indicator of the insurgence of diabetes type 2 (DT2) later in life: about 40% of the women affected by GDM also develop DT2 [3].

We give an example on how we define rules to handle the detection of repeated events such as hyperglycaemia events. The rule below is expressed in terms of the domain independent predicate `initiates_at/2`, which defines the conditions holding in the context of GDM

```
initiates_at(alert(postprandial_hyperglycaemia)=active, T) ←
    happens_at(glucose(V₁,P),T),
    last_week(Time7days,T),
    (P=after_breakfast;
    P=after_lunch;
    P=after_dinner),
```

```
V₁>=7,
count((happens_at(glucose(V₂,P),T₂),
    T₂>Time7days,
    T₂<T, V₂>=7),C),
C>3.
```

The predicate `count/2` specifies the amount of times that the condition taken in consideration holds. The rule states that an alert of hyperglycaemia after a meal is triggered when in the last week more than three times the value of glucose for postprandial periods was above 7 mmol/l.

4 CONCLUSIONS

In this paper we presented a prototype of MAGPIE, an agent platform to develop mobile Android applications in the context of chronic illnesses monitoring. MAGPIE allows the deployment of agents and context entities in an agent environment. Furthermore, MAGPIE allows to personalize for each patient the behaviour of their agents by means of alerting rules. These rules are deployed in the agent cognitive model in terms of an Event Calculus theory.

As future work we plan: to develop a web interface for the doctors so that they can define the monitoring rules for their patients remotely; study different strategies to minimize the energy consumption of the mobile application using MAGPIE; use a distributed event based system approach to notify the events generated by the platform to doctors and relatives of the patient.

ACKNOWLEDGEMENTS

This research has been funded by the Hasler Foundation

References

1. Bardram, J.E.: Pervasive Healthcare as a Scientific Discipline. Methods of Information in Medicine 47(3), 178–185 (2008)
2. Butler, M.: Android: Changing the Mobile Landscape. Pervasive Computing, IEEE 10(1), 4–7 (Jan 2011)
3. Kim, C., Newton, K.M., Knopp, R.H.: Gestational Diabetes and the Incidence of Type 2 Diabetes: A Systematic Review. Diabetes care 25(10), 1862–1868 (2002)
4. Kowalski, R., Sergot, M.: A Logic-based Calculus of Events. New Gen. Comput. 4(1), 67–95 (Jan 1986)
5. Varshney, U.: Pervasive Healthcare and Wireless Health Monitoring. Mobile Networks and Applications 12(2-3), 113–127 (2007)
6. Weyns, D., Omicini, A., Odell, J.: Environment as a First Class Abstraction in Multiagent Systems. Autonomous Agents and Multi-Agent Systems 14(1), 5–30 (Feb 2007)

The DemaWare Service-Oriented AAL Platform for People with Dementia*

Thanos G. Stavropoulos, Georgios Meditskos, Efstratios Kontopoulos, and Ioannis Kompatsiaris

Information Technologies Institute, Greece
{athstavr, gmeditsk, skontopo, ikom}@iti.gr

Abstract. This work presents DemaWare, an Ambient Intelligence platform that targets Ambient Assisted Living for people with Dementia. DemaWare seamlessly integrates diverse hardware (wearable and ambient sensors), as well as software components (semantic interpretation, reasoning), involved in such context. It also enables both online and offline processes, including sensor analysis and storage of context semantics in a Knowledge Base. Consequently, it orchestrates semantic interpretation which incorporated defeasible logics for uncertainty handling. Overall, the underlying functionality aids clinicians and carers to timely assess and diagnose patients in the context of lab trials, homes or nursing homes.

1 Introduction

This work introduces DemaWare, an integrated solution for enabling Ambient Assisted Living (AAL) applications for people with dementia. The infrastructure is Service-Oriented, providing remote, homogeneous access to the various components based on well-established web standards. It also enables both offline (pull-based) and online (push-based) data retrieval, endorsing information exchange to and from a semantic Knowledge Base (RDF triple store). The system supports a variety of pilot site scenarios such as lab trials, nursing homes and homes, providing the necessary means for data collection, patient support and clinician diagnosis.

DemaWare is applied in the Dem@Care project[1], which promises novel solutions for the holistic management of dementia, based on both medical knowledge and the latest advances in pervasive computing and sensor technologies. To this end, DemaWare aims to deliver a multi-parametric monitoring framework that will sustain context-aware, personalized and adaptive feedback mechanisms for the remote management of people with dementia. These include, among others, sensors for monitoring vital signs, location and lifestyle sensors, light and door sensors, as well as wearable and static cameras and microphones. Through fusion and aggregation of the different types of knowledge, DemaWare provides personalized feedback and care management services coupling clinical and domain knowledge with patients' contextual history and care plans.

* This work has been supported by the FP7 project Dem@Care: Dementia Ambient Care - Multi-Sensing Monitoring for Intelligent Remote Management and Decision Support (No. 288199)

[1] http://www.demcare.eu/

2 DemaWare Architecture

Overall, system requirements include an abstraction from data and functions, modularity for multiple components, orchestration of data transfer and reasoning, support for various roles and sites. The proposed architecture follows a layered approach (Figure 1) to address them. The hardware layer entails multi-modal sensors for collecting context information, each of which stores or handles data of specific format (video, audio, text or binary). Due to hardware constraints, some data have to be manually transferred offline by the clinicians. The analysis layer primarily addresses format heterogeneity, extracting higher-level information called *observations* to be stored in the Knowledge Base (KB). In detail:

- The SleepClock (SC)[2] logs residential patients' sleep state patterns (deep, shallow or no sleep) and summary (total deep/shallow sleep/awake time). Data is manually retrieved and parsed by the system's SC Library.
- The DTI2 Wristwatch (WW)[3] monitors physical activity (accelerometer), skin conductivity and temperature, ambient temperature and light. Binary WW data are collected offline and parsed by the system's WW Library.
- An ambient, Depth Camera is used for Complex Activity Recognition (CAR) [1] events related to patient location within zones of interest (e.g. "Kitchen", "Out of bed" etc.) or posture (e.g. "Standing", "Walking", "Sitting"). Multiple CAR nodes reside on linux-based mini-PCs with attached Depth Cameras in each residence.
- A second ambient, IP camera is used for Human Activity Recognition (HAR) [2] such as "Eat" or "Drink".
- Wearable Camera videos, processed by the Wearable Camera Processing Unit (WCPU) detect rooms and objects that the patient's encounter [3].
- Wearable wireless microphones capture audio for Offline Speech Analysis (OSA) [4], measuring various indicators for the progress of patient dementia.
- The KB Manager stores Observations to the KB in RDF triples (triple store).
- The Semantic Interpretation (SI) [5] performs analysis on KB-stored data and enriches it with new observations. SI also combines different sensor data (Fusion), detects various complex events in time (Complex Event Processing - CEP) and handles uncertainty.

The Service layer lifts platform heterogeneity based on the WSDL W3C standard[4] for remote access. It also allows using the XML/XSD-based Dema@Care Exchange Model to type-define observations, facilitating their mapping to KB constructs. Implementation-wise, the services wrap analysis components using Java JAX-WS[5]. Services (WSDL/SOAP) are pull-based and invoked as soon as Observations are available since most components perform offline processing (due to manual data transfer or time-demanding analysis). Meanwhile, CAR (non-WSDL service) processes streaming video and pushes observations.

[2] Gear4 Renew SleepClock: http://www.stage.gear4.com/
[3] Phillips Healthcare: http://www.healthcare.philips.com/
[4] http://www.w3.org/TR/wsdl
[5] JAX-WS: https://jax-ws.java.net/

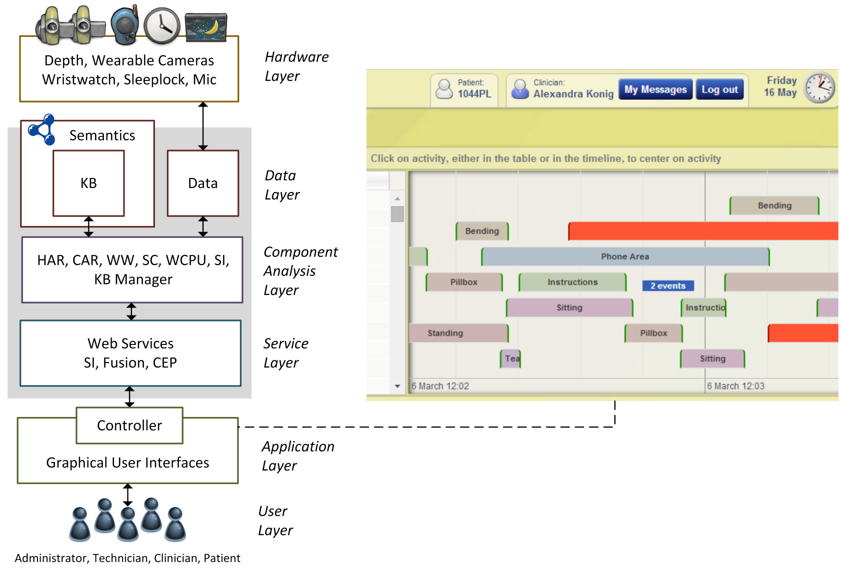

Fig. 1. The DemaWare architecture decomposition in platforms, layers and components.

The application layer consists of various Graphical User Interfaces (GUIs) as well as application logic (Controller). The Controller backend, resolves requirements related to information flow, as it orchestrates the retrieval of observations from components and stores them into the KB. It also performs certain hardware operations (e.g. start and end recordings) and gathers metadata for component invocation. GUIs are used to invoke analysis (Technician role) and view assessment results (Clinician, Carer roles).

3 Semantic Interpretation Layer

While individual sensing modalities monitor different perspectives, the Semantic Interpretation (SI) layer provides inferencing capabilities for the derivation of complex activities over the combination of those modalities. To do so, it encapsulates the ontology vocabularies for modeling the Dem@Care application context, such as activities, measurements, summaries, patients, locations and objects. The ontologies reuse the conceptual model provided by the SSN ontology [6] to model observations, measurements and sensors, as well as, relevant dementia-specific vocabularies. In detail they model:

– atomic activities and measurements detected by means of monitoring and analysis components (e.g. body temperature, luminance level, having meal, sleeping, etc.).
– problems and situations that the clinicians need to be informed about (e.g. missed meals, excessive napping, insufficient communication attempts, nocturia, etc.) .
– clinically relevant attributes and summaries (e.g. sleep efficiency and duration, number of daily telephone interactions, etc.).

SI's reasoning framework supports a hybrid combination of the OWL 2 reasoning paradigm and the execution of SPARQL rules in terms of a `CONSTRUCT` and a `WHERE` clause: the former defines the graph patterns, i.e. the set of triple patterns that should be added to the underlying RDF graph upon the successful pattern matching of the graphs in the `WHERE` clause. For example, the recognition of the `PrepareTea` activity is performed by fusing tea-related objects (detected from wearable camera) and the `PrepareDrink` intermediate activity (detected from static camera).

Since the proposed framework is required to handle data that is vastly heterogeneous, inherently uncertain and noisy, this work proposes *Defeasible Logics* [7] as an extremely suitable tool for handling this type of data. This can be illustrated by the following example that involves two complex activities $makeTea$ and $eatLunch$, which are respectively defined via the following sets of primitive observations as follows: $makeTea = \{kitchenZone, cup, kettle, teabag\}$, and $eatLunch = \{kitchenZone, cup, fork, dish\}$.

In multi-sensor environments with multi-modal and often incomplete information, where the absence of primitive observations is frequent, Defeasible Logics can offer a flexible and human-intuitive formalism for efficiently handling such situations. For instance, the following defeasible theory (written in Defeasible Logics) can handle some cases involving the above two activities:

$$
\begin{aligned}
&r_1 : kitchenZone \wedge cup \Rightarrow makeTea \\
&r_2 : kitchenZone \wedge cup \wedge fork \Rightarrow eatLunch \\
&r_3 : kitchenZone \wedge cup \wedge kettle \Rightarrow makeTea \\
&r_2 > r_1, r_3 > r_2 \; and \; C = \{makeTea, eatLunch\}
\end{aligned}
$$

Defeasible rule r_1 reads as "*if the user is in the kitchen and uses the cup then he is probably making tea*" and similar interpretations accompany defeasible rules r_2 and r_3). Moreover, rules r_2 and r_3 are superior to rules r_1 and r_2, respectively, meaning that they will prevail in potential conflicts - a conflict between two rules is initiated by complementary rule heads or heads with conflicting literals (i.e. pairs of mutually exclusive literals that cannot both be derived at the same time, see e.g. set C in the sample rule base above).

4 STATE OF THE ART

The work in [8], introduces openAAL, a general-purpose open source middleware for AAL, which provides context management, service matching, composition and workflow execution. However, while some components are similar to Dem@Care (e.g. KB Manager), openAAL does not yet handle hardware. FamiWare [9] implements a Publish/Subscribe approach, discovery, fusion etc., but targets limited hardware, e.g. Android smartphones and TinyOS sensors. Previous work in aWESoME-S [10] [11], AIM [12] and Hydra [13] have focused on energy and environmental sensors, excluding higher-level analysis. Work in [14] also provides context-sensing and user profiling. In contrast to those works, DemaWare unifies ambient and wearable devices, but also offers higher-level analysis e.g. speech, image recognition and interpretation.

5 CONCLUSIONS AND FUTURE WORK

The proposed system integrates both low and high level processes in the context of AAL for people with dementia i.e. sensor data retrieval, analysis and semantic interpretation under uncertainty. The framework is applicable to various pilot scenarios for patient monitoring and assessment. Current limitations include hardware constraints, e.g. manual data transfer, and the lack of even richer context information, for which we plan to investigate alternative sensors.

References

1. R. Romdhane, C.F. Crispim, F. Bremond, and M. Thonnat, 'Activity recognition and uncertain knowledge in video scenes', in *Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on*, pp. 377–382, (Aug 2013).
2. K. Avgerinakis, A. Briassouli, and I. Kompatsiaris, 'Recognition of activities of daily living for smart home environments', in *9th International Conference on*, pp. 173–180, (July 2013).
3. H. Boujut, J. Benois-Pineau, and R. Megret, 'Fusion of multiple visual cues for visual saliency extraction from wearable camera settings with strong motion', in *12th International Conference on Computer Vision - Volume Part III*, ECCV'12, pp. 436–445, (2012).
4. A. Satt, A. Sorin, O. Toledo-Ronen, O. Barkan, I. Kompatsiaris, A. Kokonozi, and M. Tsolaki, 'Evaluation of speech-based protocol for detection of early-stage dementia.', in *INTERSPEECH*, pp. 1692–1696. ISCA, (2013).
5. G. Meditskos, S. Dasiopoulou, V. Efstathiou, and I. Kompatsiaris, 'Sp-act: A hybrid framework for complex activity recognition combining owl and sparql rules', in *PerCom Workshops*, pp. 25–30, (2013).
6. Michael Compton, Payam Barnaghi, and Luis Bermudez et al., 'The SSN ontology of the W3C semantic sensor network incubator group', *Web Semantics: Science, Services and Agents on the World Wide Web*, **17**(0), 25 – 32, (2012).
7. D. Nute, 'Handbook of logic in ai and logic programming (vol. 3)', chapter Defeasible Logic, 353–395, Oxford University Press, Inc., New York, NY, USA, (1994).
8. P. Wolf, A. Schmidt, J.P. Otte, M. Klein, S. Rollwage, B. Konig-Ries, T. Dettborn, and A. Gabdulkhakova, 'openaal - the open source middleware for ambient-assisted living (aal)', in *AALIANCE conference, Malaga, Spain, March 11-12, 2010*, (2010).
9. N. Gámez and L. Fuentes, 'Famiware: a family of event-based middleware for ambient intelligence', *Personal and Ubiquitous Computing*, **15**(4), 329–339, (2011).
10. T. G. Stavropoulos, K. Gottis, D. Vrakas, and I. Vlahavas, 'awesome: A web service middleware for ambient intelligence', *Expert Syst. Appl.*, **40**(11), 4380–4392, (2013).
11. T. G. Stavropoulos, E. Kontopoulos, N. Bassiliades, J. Argyriou, A. Bikakis, D. Vrakas, and I. Vlahavas, 'Rule-based approaches for energy savings in an ambient intelligence environment', *Journal of Pervasive and Mobile Computing*, to appear.
12. A. Capone, M. Barros, H. Hrasnica, and S. Tompros, 'A new architecture for reduction of energy consumption of home appliances', in *TOWARDS eENVIRONMENT, European conference of the Czech Presidency of the Council of the EU*, pp. 1–8, (2009).
13. M. Eisenhauer, P. Rosengren, and P. Antolin, 'Hydra: A development platform for integrating wireless devices and sensors into ambient intelligence systems', in *The Internet of Things*, 367–373, Springer, (2010).
14. A. De Paola, S. Gaglio, G. Lo Re, and M. Ortolani, 'Sensor9k: A testbed for designing and experimenting with wsn-based ambient intelligence applications', *Pervasive and Mobile Computing*, **8**(3), 448–466, (2012).

Using 3D simulators for the Ambient Assisted Living

Paolo Sernani, Andrea Claudi, Paolo Calvaresi,
Daniele Accattoli, Roberto Tofani, and Aldo Franco Dragoni

Dipartimento di Ingegneria dell'Informazione (DII)
Università Politecnica delle Marche
Via Brecce Bianche 60131 Ancona Italy
{p.sernani, a.claudi, p.calvaresi, d.accattoli, a.f.dragoni}@univpm.it

Abstract. Ambient Assisted Living (AAL) and Ambient Intelligence (AmI) aim at building safe, smart and interconnected environments around people. Nevertheless, the testing phase of intelligent software systems managing smart homes, in real environments, requires an enormous effort in terms of time, work and money. With this paper, we discuss the possibility to use 3D virtual environments during the developing and preliminary tests of such systems, as it happens in robotics with 3D simulators. Even if this approach cannot totally replace the need of real user trials, it can speed up the implementation of prototypes, decoupling software from hardware. We also present a brief description of a proof-of-concept, showing some of the benefits described.

1 Introduction

The World Health Organization predicts that the number of people aged 65 or older will triple within 2050 (becoming the 16% of the World population) [1]. The Ambient Assisted Living (AAL) focuses on this demographic change and aims at extending the time people can live in their home environment. In this context Ambient Intelligence (AmI) builds a safe environment for the assisted person [2], by combining Home Automation, Internet of Things and Artificial Intelligence to implement smart homes: their main goal is to help people in their daily activities, composing around them an unobtrusive, interconnected, adaptable, dynamic, embedded and intelligent environment [3].

Nevertheless, the development and the following testing of smart homes for the AAL in a real environment (i.e. a daily-used home environment equipped with devices and sensors, where an assisted person can actually live) require several resources in terms of work, time and money. Assessing technologies for the AAL during their development will be almost impossible in real world because the trials should be conducted extensively:

– with real human inhabitants with different kinds of impairments;
– in different environmental situations;
– taking into account diverse economical conditions.

Such trials will be also extremely expensive, especially for the development of software prototypes that would require the hardware for a great number of tests; furthermore the result of this effort could be the need to modify or redesign some components (and thus to repeat tests).

A virtual 3D environment, simulating an interface with necessary sensors and actuators could be an effective tool to develop, simulate and test intelligent systems with the aim to extend the time people can (autonomously) live in their preferred environment.

This paper discusses the possibility to use 3D virtual environments and simulators for the AAL, analysing challenges and potential advantages. Using virtual environments to accurately reproduce real environments would allow a partial decoupling of the software development from the hardware development. Moreover it would improve the economical sustainability of the development of software to control smart homes, speeding up the implementation of prototypes and allowing the use of real environments only for release candidate versions. Finally it would provide an environment where researchers in the AAL could easily collaborate and combine their findings.

The paper is structured as follows: section 2 lists some current applications of virtual words in the AAL domain; section 3 discusses the challenges in the use of virtual environments to test intelligent software systems for the AAL; section 4 describes the implementation of a preliminary proof-of-concept, based on a robotics simulator; finally, section 5 concludes the paper, suggesting future research directions.

2 Related Works

Most of the current applications of virtual worlds in the AAL are addressed to motor and cognitive rehabilitation [4] and to improve social inclusion [5]. In particular serious gaming has been widely adopted in rehabilitation. Indeed, typical game features (as providing challenges and goals, stimulating curiosity, cooperation, competition [6]) enhance user engagement and intrinsic motivation: thus serious games can support users in developing their skills, in learning and in experiencing situations that are impossible (e.g. for economical reasons) in the real world [7, 8]. For these reasons, virtual reality and serious games are also used in therapies for pain management [9]. Beyond rehabilitation, a promising direction is the use of serious gaming for the acceptance of facility automation and smart homes [10]. Moreover the use of Interactive Scenario Visualization (ISV) can lead to the clarification of system functionalities, as well as to gain stakeholders' feedback [11], especially in the design phase.

3 3D Virtual Environments for the AAL

The use of 3D virtual environments to develop intelligent software systems to manage virtual homes is inspired by the robotics field, where a relevant number of 3D software simulators is available. The basic concept is to use a 3D simulator that exposes the interfaces to sensors and actuators (through suitable libraries) available in the market and that allows to create a virtual home environment: here, intelligent software systems could be tested to evaluate their behaviour, even when facing unexpected events. Ideally, the intelligent systems should be migrated transparently in a real environment with the sensors and actuators previously simulated.

To this point, one might argue that there is no need to use the 3D for AAL simulations. Nevertheless, the 3D feature gives not only the involvement typical of graphical effects, but should provide to system developers the possibility to interact directly with the simulations, for example moving objects on the fly. With a reliable physics engine (as those typical of computer games), developers can simulate unexpected events and thus understand if the system behaves correctly.

In a complete 3D simulator for the AAL, also human behaviours should be represented: the involved research topics are similar to the inclusion of human reasoning mechanism in computer games and serious games, by exploiting the results of the Artificial Intelligence (AI) field [12]:

- the selection of the suitable AI approach to model and imitate human behaviours especially taking into account the need of human like responsiveness and communication.
- The influence of the adoption of strong AI techniques to imitate human reasoning on the simulator design, since a well identified problem in computer game design is the introduction of the AI in the last phases.
- The definition of suitable goals and metrics to test intelligent software systems for the AAL.

A promising direction is the design and implementation of virtual characters as BDI agents (see, for instance, [13] for the development of BDI based Non-Player Characters (NPC) in computer games). Although this approach is ideal to implement virtual characters capable of carrying out long-term autonomous actions, it leads to several sub-problems: balancing between proactive and reactive behaviours, scheduling properly goals on the basis of the application domain, representing the environment in a symbolic manner and translating this representation in the interaction with the simulator engine.

To obtain the advantages described in the introduction section, a clear requirements analysis (as suggested in [14]) is needed: firstly system developers, as the prominent users of 3D simulators, and their needs have to be taken into account; moreover patients, their relatives, physicians and health operators with their needs about AAL systems and their personal experience may give a crucial contribution in the identification of the simulation scenarios.

4 A proof-of-concept example

To show some of the benefits, we implemented a proof-of-concept, using the multi-agent expert system described in [15]. To create the 3D environment, we used the outdoor multi-robot simulators *Gazebo* [16] and *Morse* [17]. Figure 1 shows the virtual domestic environment composed by four rooms.

Thanks to the *CameraSensor* and the *RaySensor* classes provided by Gazebo, we simulated two scenarios to test the behaviour of the multi-agent system controlling the home: in the first the expert system has to aerate the rooms by opening the windows, but keeping closed the one in the room where the assisted person actually stays; moreover it has to close and open windows coherently with the movements of the patient, using data provided by ray sensors. In the second scenario the system has to detect the fall of the patient, sending an alarm. To implement these scenarios, we added some agents to the architecture in [15]: two Ambient Agents (the Gateway and the Camera Agents, to control the ray and the camera sensors) and two Actuator Agents (the *WindowActuator* and

Fig. 1. Virtualized home in Morse: the architecture of the simulation allows to decouple the multi-agent software system controlling the environment from the sensor. Thus, maintaining the same protocols, the intelligent system can be tested in diverse simulators as well as in a real environment.

the *CameraActuator* Agents to send commands to the windows and the video cameras in the environment). We carried out tests also using the Morse simulator. We added also a wheelchair, temperature sensors, gas detectors and light sensors. We tested the control of the home through a mobile application using both touch and vocal commands[1].

Beyond the implemented scenarios, the communication between the software agents and the sensors in the 3D environment highlights the potential of the simulations: being based on the TCP/IP protocol, each agent creates a socket channel to send commands to the sensors. Thus, if virtual sensors provide the same interface as off-the-shelf sensors (as those implemented in Gazebo and in similar robotics simulators), the developed agents can be easily migrated in real environments. This is an advantage also when a real system is running: new agents can be implemented in virtual environments and added to the real one; moreover existing ones can be migrated in the virtual environment for maintenance.

5 Conclusions

We discussed on the opportunity to simulate virtual 3D environments to design and develop intelligent software system to control smart homes for the AAL. Even if such an approach cannot completely replace user trials before the everyday use, it can speed up the development of prototypes as well as simplify the maintenance and the update of existing systems.

Instead of using a robotics simulator, as in our simple proof-of-concept, a dedicated *AAL simulator* (with a physics engine) could be developed, in order to adhere to specific AAL needs: it could be an ideal platform to combine the efforts of the AAL research community for the development of assistive technologies.

[1] For a video of the Morse simulation and the mobile application see http://www.youtube.com/watch?v=zXEpShRNGuo

References

1. World Health Organization: Global health and ageing. (2011)
2. Sun, H., De Florio, V., Gui, N., Blondia, C.: Promises and challenges of ambient assisted living systems. In: Information Technology: New Generations, 2009. ITNG '09. Sixth International Conference on. (2009) 1201–1207
3. Sadri, F.: Ambient intelligence: A survey. ACM Computing Surveys **43**(4) (2011) 1–66
4. Rego, P., Moreira, P., Reis, L.: Serious games for rehabilitation: A survey and a classification towards a taxonomy. In: Information Systems and Technologies (CISTI), 2010 5th Iberian Conference on. (2010) 1–6
5. Cascado, D., Romero, S.J., Hors, S., Brasero, A., Fernandez-Luque, L., Sevillano, J.L.: Virtual worlds to enhance ambient-assisted living. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. (2010) 212–215
6. Malone, T.W., Lepper, M.R.: Making learning fun: A taxonomy of intrinsic motivations for learning. In: Aptitude, learning, and instruction 3. (1987) 223–253
7. Susi, T., Johannesson, M., Backlund, P.: Serious games: an overview. Technical Report HS-IKI-TR-07-001, School of Humanities and Informatics, University of Skövde (2007)
8. Connolly, T.M., Boyle, E.A., MacArthur, E., Hainey, T., Boyle, J.M.: A systematic literature review of empirical evidence on computer games and serious games. Computers & Education **59**(2) (2012) 661–686
9. Simmonds, M., Zikos, D.: Computer games to decrease pain and improve mood and movement. In: Proceedings of the the 7th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA 2014). (2014)
10. Leonhardt, S., Kassel, S., Randow, A., Teich, T.: Learning in the context of an ambient assisted living apartment: Including methods of serious gaming. In Motta, G., Wu, B., eds.: Software Engineering Education for a Global E-Service Economy. State of the Art, Trends and Developments. Progress in IS. Springer International Publishing (2014) 49–55
11. Van't Klooster, J.W., Van Beijnum, B.J., Eliens, A., Hermens, H.: Interactive scenario visualization for user-based service development. In: Proceedings of the 2012 International Conference on Collaboration Technologies and Systems, CTS 2012. (2012) 498–503
12. Dignum, F.: Agents for games and simulations. Autonomous Agents and Multi-Agent Systems **24**(2) (2012) 217–220
13. Palazzo, L., Dolcini, G., Claudi, A., Biancucci, G., Sernani, P., Ippoliti, L., Salladini, L., Dragoni, A.F.: Spyke3D: A new computer games oriented BDI agent framework. In: Computer Games: AI, Animation, Mobile, Interactive Multimedia, Educational & Serious Games (CGAMES), 2013 18th International Conference on, IEEE (2013) 49–53
14. Calvaresi, D., Sturm, A., Yu, E., Dragoni, A.F.: Exploring domain requirements and technology solutions: a goal modeling approach. In: Proceedings of the 7th International iStar Workshop (iStar14). (2014)
15. Sernani, P., Palazzo, L., Claudi, A., Dolcini, G., Biancucci, G., Trentalange, G., Dragoni, A.F.: Virtual carer: A BDI agent system for Ambient Assisted Living. In: 2nd International Workshop on Artificial Intelligence and NetMedicine (NetMed 2013). (2013) 14–23
16. Gazebo: Robot simulation made easy. Retrieved May 20, 2014, from http://gazebosim.org (2014)
17. Morse: the modular openrobots simulation engine. Retrieved May 20, 2014, from http://www.openrobots.org/wiki/morse/ (2014)

Today, how was your ability to move about?

Felip Miralles, Eloisa Vargiu, Eloi Casals,
José Alejandro Cordero, and Stefan Dauwalder

Barcelona Digital Technology Center,
{fmiralles, evargiu, ecasals, jacordero,
sdauwalder}@bdigital.org

Abstract. In this paper, we are interested in monitoring mobility activities in order to automatically assess quality of life of people. In particular, we are aimed at answering to the question "*Today, how was your ability to move about?*". To this end, we rely on a sensor-based telemonitoring and home support system. Although we are interested in assisting disabled people, we performed preliminary experiments with a healthy user, as a proof of concept. Results show that the approach is promising. Thus, we are now in the process to install the system in disabled people's homes under the umbrella of the BackHome project.

1 Introduction

Improving people's Quality of Life (QoL) is one of the expected outcomes of modern health applications and systems. Thus, several solutions, aimed at improving QoL of the corresponding users, have been investigated and proposed [2]. Among the huge kinds of proposed solutions, let us focus here on those that provide telemonitoring and home support [1], [3], [5]. TeleMonitoring and Home Support Systems (TMHSSs) help users (e.g., disabled or elderly people) to live normally at home keeping (or returning to) their life roles. On the other end, they support health care providers in the task of being aware of the status of their patients.

To assess users' QoL, in the literature several questionnaires have been proposed and adopted [6], [10], [4], [15], [8]. Users are asked to answer to a predefined set of questions about their mental and psychological status and feeling. Although they are largely adopted, as noted in [11], answering them could become tedious and annoying for users and could even be impossible in cases of severe impairment of the user.

In [13] we proposed a generic methodology aimed at automatic assessing QoL of users. Starting from that methodology, among all the items that may compose a QoL questionnaire, in this paper we focus on how to assess the ability to move about. In fact, we use information gathered from a sensor-based TMHSS to answer to the question "*Today, how was your ability to move about?*". Although several works study how to recognize activities [9] and behavior [7], to our best knowledge, this is the first attempt to use that information to automatically assess a (part of a) QoL questionnaire.

2 Materials and Methods

In [13] we proposed a generic methodology to assess and telemonitor QoL of individuals with a holistic bio-psycho-social approach, which intends to become the base for current and future telemedicine and teleassistance solutions. Since the overall proposal is very ambitious, in this paper, we focus on the task of assessing just one of the items of a given questionnaire. In other words, we show and discuss our implementation to assess movement ability.

2.1 The Methodology

To monitor QoL, we propose a sensor-based TMHSS able to monitor the evolution of the user's daily life activity, providing QoL automated assessment based on information gathering and data mining techniques [14]. Specifically, wearable sensors allow to monitor fatigue, stress, and further user's conditions. Environmental sensors are used to monitor –for instance– temperature and humidity, as well as the movements (motion sensors) and the physical position of the user (location sensors). Smart home devices enable physical autonomy of the user and help her/him carry out daily life activities. From the social perspective, an Internet-connected device allows the user to communicate with remote therapists, careers, relatives, and friends through email and social networks (i.e., Facebook and Twitter).

Starting from the standard EQ-5D-5L questionnaire, we propose and adopt a visual analogue scale QoL questionnaire. The proposed questionnaire is designed to assess the key QoL features of an individual, which correspond to the main features that we aim to monitor. In other words, we consider the user's QoL as the conjunction of the following items: *Mood*, *Health Status*, *Mobility*, *Self-care*, *Usual Activities*, and *Pain/Discomfort*. As already said, in this paper we focus only on user's movement ability, i.e., *Mobility*. In other words, we aim to automatically reply to the question "*Today, how was your ability to move about?*".

2.2 The Implemented Telemonitoring and Home Support System

The implemented TMHSS is able to monitor indoor and outdoor activities.

Indoor activities are monitored by relying on a set of home automation sensors. More precisely, we use motion sensors, to identify the room where the user is located (one sensor for each monitored room); a door sensor, to detect when the user enters or exits the premises; electrical power meters and switches, to control leisure activities (e.g., television and pc); pressure sensors, to track user transitions between rooms; and bed (seat) sensors, to measure the time spent in bed (wheelchair). From a technological point of view, the sensors are based on the z-wave wireless standard, which establishes a wireless mesh network of devices to send the measured data to a central unit located at user's home. That central unit collects all the data and sends them to the cloud where they are stored and analyzed. The system also comprises "virtual devices", which are software elements that fuse together information from two or more sensors in order to make some inference and provide new information. In so doing, the TMHSS is able to perform more actions and to be more adaptable to the context and the user's habits. In other words, virtual devices have been introduced to merge the information gathered by the installed real sensors.

Outdoor activities are monitored using the user's smartphone relying on Moves[1], an app for smartphones able to recognize physical activities (such as walking,

[1]http://www.moves-app.com/

running, and cycling) and movements by transportation. Moves is also able to store information about the location in which the user is, as well as the corresponding performed route(s). Moves provides an API through which is possible to access all the collected data.

2.3 How to Assess *Mobility*

Information gathered by the sensors is used as classification features to build a multi-class supervised classifier; one for each user. We considered the following features: (i) time spent on bed and (ii) maximum number of continuously hours on bed, extracted from the bed sensor; (iii) time spent on the wheelchair and (iv) maximum number of continuously hours on the wheelchair, extracted from the seat sensor; (v) time spent in each room and (vi) percentage of time in each room, extracted from the motion sensor; (vii) room in which passed the most of the time, inferred by the virtual device; (viii) total time spent at home, extracted from the door sensor; (ix) total time spent watching the TV and (x) total time spent using the PC, extracted from the corresponding power meters and switches; (xi) number of kilometers by transportation, (xii) number of kilometers by moving outdoors on the wheelchair and (xiii) number of visited places, given by Moves.
To train and test the classifier, the user is asked to answer to the question "Today, how was your ability to move about?", everyday. User's answer is an integer number in a scale from 1 to 5 that correspond to user's satisfaction in her/his movement ability. User's answers are then used to label the entries of the dataset for training and testing into three categories: "Low" (1-2), "Normal" (3) and "High"(4-5).

3 Preliminary Experiments and Results

The TMHSS presented in this paper is part of BackHome[2], an European R&D project that aims to provide a TMHSS using Brain Computer Interfaces (BCI) and other assistive technologies to improve autonomy and QoL of disabled people [12] [14].
The system is currently running in a healthy user's home in Barcelona. The corresponding user is a 40-year-old woman who lives alone. This installation is currently available and data continuously collected. According to the home plan, the following sensors have been installed: 1 door sensor; 3 motion sensors (1 living room, 1 bedroom, 1 kitchen); 3 switch and power meters (1 PC, 1 Nintendo WII, 1 kettle); and 1 bed sensor. Moreover, the user has installed in her iPhone the Moves app.
To test the feasibility of the approach, we considered a window of three months (February '14 – April '14) and made comparisons of results for three classifiers: decision tree, k-nn with k=1, and k-nn with k=3. During all the period, the user answered to the question "Today, how was your ability to move about?" daily at 7 PM. Answers have been then used to label the item of the dataset to train and test the classifiers built to verify the feasibility of the proposed QoL approach. Given a category, we consider as true positive (true negative), any entry evaluated as positive (negative) by the classifier that corresponds to an entry labeled by the

[2] http://www.backhome-fp7.eu/backhome/index.php

user as belonging (not belonging) to that class. Seemly, we consider as false positive (false negative), any entry evaluated as positive (negative) by the classifier that corresponds to an entry labeled by the user as not belonging (belonging) to that class. Results have been then calculated in terms of precision, recall, and F_1 measure.

Let us stress the fact that in this preliminary experimental phase, we are considering data coming from a healthy-user. Thus, while analyzing data, the following issues must be considered: tests have been performed with only one user; the user is healthy; and a window of less than 4 months of data has been considered. As a consequence, results can be used and analyzed only as a proof of concept of the feasibility of the approach.

The best results have been obtained using the decision tree. In fact, in that case, on average we calculated a precision of 0.64, a recall of 0.69 and a F_1 of 0.66. It is worth noting that, as expected (the user is healthy and not have difficulty in movements), the best results are given in recognizing "Normal" mobility. In fact, in this case we obtained a precision of 0.80, a recall of 0.89 and an F_1 measure of 0.84. The same behavior has been noted in the results with the k-nn classifiers, even if in that case average results are lower: precision 0.53, recall 0.56 and F_1 0.52. Also in that case, the best results are given in recognizing "Normal" mobility. In our opinion, results given by the decision tree are better than those given by the k-nn classifiers because of the very few number of data (a window of three months has been considered) and also because decision trees are more robust with respect to outliers.

4 Discussion

The methodology proposed in [13] is aimed at automatically assessing quality of life of disabled people. Relying on that methodology, in this paper, we considered the task of assessing the questionnaire item *Mobility*. A telemonitoring and home support system has been implemented to monitor both indoor and outdoor activities. Currently, the system is installed in a healthy user's home in Barcelona. Preliminary results show that the system is able to collect and analyze data useful to learn user's habits and it looks promising to assess the *Mobility* of a given user. Although, as mentioned above, preliminary results can be used only as a study of the feasibility of the approach, they are promising and encourage us to continue with the study. In fact, using also data coming from social activities (i.e., mailing, Facebook and Twitter), we started new experiments to assess also the questionnaire item "Mood".

As for the future work, the next step consists of experimenting the proposed approach under the umbrella of BackHome. Hence, we are currently setting up the proposed telemonitoring and home support system at BackHome real end-users' homes at the facilities of Cedar Foundation[3] in Belfast.

Acknowledgments

The research leading to these results has received funding from the European Communitys, Seventh Framework Programme FP7/2007-2013, BackHome project grant agreement n. 288566.

[3]http://www.cedar-foundation.org/

References

1. Carneiro, D., Costa, R., Novais, P., Machado, J., Neves, J.: Simulating and monitoring ambient assisted living. In: Proc. ESM (2008)
2. Celler, B.G., Lovell, N.H., Basilakis, J.: Using information technology to improve the management of chronic disease. Medical Journal of Australia 179(5), 242–246 (2003)
3. Corchado, J., Bajo, J., Tapia, D., Abraham, A.: Using heterogeneous wireless sensor networks in a telemonitoring system for healthcare. IEEE Transactions on Information Technology in Biomedicine 14(2), 234–240 (2010)
4. Hays, R., Morales, L., RAND Health, P., RAND, C.: The RAND-36 Measure of Health-related Quality of Life. RAND Health reprint, RAND (2001)
5. Mitchell, M., Meyers, C., Wang, A., Tyson, G.: Contextprovider: Context awareness for medical monitoring applications. In: Conf Proc IEEE Eng Med Biol Soc. (2011)
6. Murphy, B., Herrman, H., Hawthorne, G., Pinzone, T., Evert, H.: Australian WHOQoL instruments: User's manual and interpretation guide. Australian WHOQoL Field Study Centre, Melbourne, Australia (2000)
7. Ogawa, M., Suzuki, R., Otake, S., Izutsu, T., Iwaya, T., Togawa, T.: Long term remote behavioral monitoring of elderly by using sensors installed in ordinary houses. In: Microtechnologies in Medicine & Biology 2nd Annual International IEEE-EMB Special Topic Conference on. pp. 322–325. IEEE (2002)
8. O'Sullivan, S., Schmitz, T.: Physical Rehabilitation. G - Reference, Information and Interdisciplinary Subjects Series, F.A. Davis (2007)
9. Tapia, E.M., Intille, S.S., Larson, K.: Activity recognition in the home using simple and ubiquitous sensors. Springer (2004)
10. The Euroqol Group: Euroqol a facility for the measurement of health-related quality of life. Health Policy 16, 199–208 (1990)
11. Vargiu, E., Ceccaroni, L., Subirats, L., Martin, S., Miralles, F.: User profiling of people with disabilities - a proposal to pervasively assess quality of life. In: 5th International Conference on Agents and Artificial Intelligence (2013)
12. Vargiu, E., Miralles, F., Martin, S., Markey, D.: BackHome: Assisting and telemonitoring people with disabilities. In: RAatE 2012 - Recent Advances in Assistive Technology & Engineering (2012)
13. Vargiu, E., Fernández, J.M., Miralles, F.: Distributed Systems and Applications of Information Filtering and Retrieval, chap. Context-Aware based Quality of Life Telemonitoring. Studies in Computational Intelligence 515, Springer (2014)
14. Vargiu, E., Fernández, J.M., Torrellas, S., Dauwalder, S., Solà, M., Miralles, F.: A sensor-based telemonitoring and home support system to improve quality of life through bnci. In: 12th European AAATE Conference (2013)
15. Ware, J., Kosinski, M., Dewey, J.: How to Score Version 2 of the SF-36 Health Survey: Standars & Acute Forms. QualityMetric (2001)

SVM-based CBIR of Breast Masses on Mammograms

Lazaros Tsochatzidis, Konstantinos Zagoris, Michalis Savelonas, and Ioannis Pratikakis

Visual Computing Group,
Dept. of Electrical and Computer Engineering,
Democritus University of Thrace,
67100 Xanthi, Greece
{ltsochat,kzagoris,ipratika}@ee.duth.gr
msavelonas@gmail.com
http://vc.ee.duth.gr

Abstract. Mammography is currently the dominant imaging modality for the early detection of breast cancer. However, its robustness in distinguishing malignancy is relatively low, resulting in a large number of unnecessary biopsies. A computer-aided diagnosis (CAD) scheme, capable of visually justifying its results, is expected to aid the decision made by radiologists. Content-based image retrieval (CBIR) accounts for a promising paradigm in this direction. Facing this challenge, we introduce a CBIR scheme that utilizes the extracted features as input to a support vector machine (SVM) ensemble. The final features used for CBIR comprise the participation value of each SVM. The retrieval performance of the proposed scheme has been evaluated quantitatively on the basis of the standard measures. In the experiments, a set of 90 mammograms is used, derived from a widely adopted digital database for screening mammography. The experimental results show the improved performance of the proposed scheme.

Keywords: Content-Based Image Retrieval, Mammography, Support Vector Machines

1 Introduction

The use of content-based image retrieval (CBIR) schemes for computer-aided diagnosis (CAD) has been intensively investigated in the last decade [6]. Such an approach that facilitates searching for visually similar medical images, provides radiologists with visual aid and increases their confidence in incorporating CAD-cued results in their decision making [8].

There is only a limited amount of works devoted to CBIR-based CAD for breast masses in mammograms, although mammographic CAD is one of a mature and widely adopted CAD type [8]. An early attempt towards CBIR for breast masses in mammograms was the work of Alto et al. [1], who investigated the discriminant capability of compactness, fractional concavity, spiculation index and Haralick's textural features. In a recent work, Wang et al. [5] tested the relationship between CAD performance and the similarity level between the region of interest (ROI) of the query and the ROIs resulting as outputs of CBIR.

All the above works evaluated retrieval performance on the basis of discriminant capability between benign and malignant cases. It can be argued that a CBIR scheme is expected to retrieve cases on the basis of visual similarity, since by its own nature it cannot take into account accompanying clinical data. In a real clinical setting, the results of CBIR could be jointly assessed with all such data in the context of an integrated CAD scheme. Finally, it can be observed that all CBIR methods presented were based on simple similarity measures, which cannot optimally exploit the distribution of mammogram ROIs in the feature space.

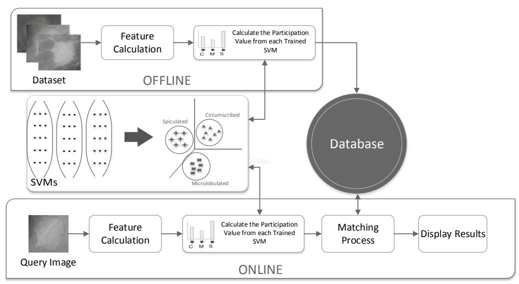

Fig. 1. The proposed system architecture.

In this work, we present a novel CBIR scheme, which utilizes a support vector machine (SVM) ensemble. The corresponding SVMs are capable of optimally exploiting the distribution of input samples in the feature space on the basis of breast imaging-reporting and data system (BI-RADS) classifications of breast masses [2], as performed by an expert radiologist. Then, based on each SVM's participation value, a new feature-vector is mapped to each feature set. The used dataset is formed by mammograms of the digital dataset for screening mammography (DDSM), which is widely adopted by the medical community.

The remainder of the paper is organized as follows: Section 2 describes the architecture of the proposed system. The experimental evaluation of the proposed CBIR scheme is presented in Section 3. Finally, conclusions and future perspectives of this work are discussed in Section 4.

2 Proposed CBIR scheme

Fig. 1 shows the proposed system architecture. At this point, it is worth to note that a mass detection and segmentation stage is applied prior to the proposed pipeline. Since those two stages are out of the scope of the proposed system, there will be no further discussion about the methodology used. For our experiments, the segmentation information is taken from the ground truth corpora.

Initially, the following features are extracted from the mass boundaries:

1. *Solidity:* $F_{solidity} = A/H$, where A and H denote the areas of the shape and its corresponding convex hull, respectively.
2. *Compactness:* $F_{compactness} = 1 - (4\pi A/P^2)$, where A and P denote the area and the perimeter of the shape, respectively.
3. *Discrete Fourier Transformation (DFT) coefficients of the Normalized Radial Length (NRL) function.* The Radial Length Function corresponds to the distance of each contour point (t) from the mass centroid (x_c, y_c): $r(t) = \sqrt{(x(t) - x_c)^2 + (y(t) - y_c)^2}$. This function is sampled to a fixed number of points (256 in this work) and is normalized before its DFT computation.

Thereafter, the above feature set is supplied to three different trained SVMs that correspond to three classes of breast masses, based on BI-RADS [2], namely spiculated, micro -lobulated and circumscribed.

The Support Vector Machines (SVMs) [3] are based on statistical learning theory and have been successfully applied to several classification problems because of their discriminant ability and the fact that do not require large training sets.

Instead of utilizing the sign of the SVM decision function, we propose to normalize it, based on [7], in order to calculate a participation value of each feature vector to each trained SVM, which corresponds to each BI-RADS class. The normalized decision function is calculated by the following equation:

$$R(x) = \begin{cases} \max\left\{\frac{1}{1+\frac{1}{3}e^{f(x)}}, \frac{1}{1+\frac{1}{3}e^{-f(x)}}\right\} & if\, f(x) > 0 \\ 1 - \max\left\{\frac{1}{1+\frac{1}{3}e^{f(x)}}, \frac{1}{1+\frac{1}{3}e^{-f(x)}}\right\} & if\, f(x) < 0 \end{cases} \tag{1}$$

where $f(x)$ denotes the SVM decision function. The output of the Eq. 1 represents the membership value of the data x to the corresponding class and ranges in the interval [0,1]. Finally, the outputs of the SVMs construct the new three-element feature vector, used in the remainder of the retrieval process. In the sequel, the euclidean distance between the query and each indexed samples is computed leading to a ranked list of similar objects.

3 Experimental evaluation

In this study, 90 regions of interest (ROI) were used, extracted from various mammograms of DDSM, which contain masses. Each case is accompanied with ground-truth delineations and additional information, such as the biopsy-proven pathology of the lesion, its shape and margin types, the overall breast density and the assessment, of an expert radiologist [4] based on the BI-RADS standards. The margin types that were taken into consideration in this work are circumscribed, micro-lobulated and spiculated, as they are highly correlated with the mass' pathology. For each margin type, masses of various shapes (oval, round, lobulated and irregular - Fig.2) were included. For each selected ROI, the contour of the depicted mass was acquired, by an expert radiologist indicating the exact position of its margin.

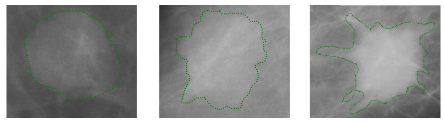

Fig. 2. Example ROIs for each margin type: A circumscribed (left), a micro-lobulated (center) and a spiculated (right) mass.

Two performance evaluation metrics are employed, which measure the system's ability to retrieve masses of similar margin type to the query. The first one is the Precision at Top 5 Retrieved items (P@5), which defines how successfully the algorithms produce relevant results to the first 5 position of the ranking list. The second metric used, is the Mean Average Precision (MAP) which is a typical measure for the performance of information retrieval systems and it is defined as the average of the precision value obtained after each relevant retrieved item.

The BI-RADS SVMs used a radial basis function (RBF) kernel, they were trained by the 2/3 and evaluated with the remainder portion of the dataset.

The proposed method was evaluated against a typical, unsupervised, state-of-the-art retrieval system employing the euclidean distance, calculated directly from the features instead of the participation values from each SVM. Comparative results are presented in Table 1 and show that the proposed method outperformed the typical unsupervised euclidean-based retrieval system.

Table 1. Experimental Results

Classes	**Unsupervised CBIR**		**SVM-based CBIR**	
	P@5	**MAP**	**P@5**	**MAP**
Circumscribed	0.9	0.915	0.9	0.92
Microlobulated	0.71	0.723	0.71	0.763
Spiculated	0.654	0.608	0.745	0.73
Average	0.75	0.743	0.78	0.8

4 Conclusions

This work introduced a CBIR scheme, which utilizes a support vector machine (SVM) ensemble. The retrieval performance of the proposed scheme has been evaluated on the basis of BI-RADS classifications of breast masses. The used dataset is formed by 90 cases of the DDSM, which is a dataset widely adopted by the medical community. The experimental results lead to the conclusion that the proposed CBIR scheme outperforms standard euclidean-based retrieval, while greatly reducing the feature vector dimension and, consequently, the computational cost.

Future perspectives of this work include: 1) the integration of the proposed CBIR scheme within the context of a mammographic CAD system, which will also consider accompanying clinical and textual data, 2) the development of a similar CBIR scheme to facilitate CAD of breast microcalcifications.

Acknowledgements

This work is funded by the Hellenic Republic, Ministry of Education and Religious Affairs, General Secretariat of Research and Technology (GSRT), and particularly the National Programme "SYNERGASIA 2011" (11SYN_10_1546) in the National Strategic Reference Framework (NSRF) 2007-2013.

References

1. Alto, H., Rangayyan, R.M., Desautels, J.E.L.: Content-based retrieval and analysis of mammographic masses. J. Electronic Imaging 14(2), 023016 (2005)
2. Berg, W.A., Campassi, C., Langenberg, P., Sexton, M.J.: Breast imaging reporting and data system: inter- and intraobserver variability in feature analysis and final assessment. Am. J. Roentgenol. 174(6), 1769–1777 (2000)
3. Cortes, C., Vapnik, V.: Support vector networks. Machine Learning 20, 273–197 (1995)
4. Heath, M., Bowyer, K., Kopans, D., Kegelmeyer Jr, P., Moore, R., Chang, K., Munishkumaran, S.: Current status of the digital database for screening mammography. In: Digital mammography, pp. 457–460. Springer (1998)
5. Wang, X., Park, S., Zheng, B.: Assessment of performance and reliability of computer-aided detection scheme using content-based image retrieval approach and limited reference database. Journal of Digital Imaging 24(2), 352–359 (2011)
6. Welter, P., Fischer, B., Günther, R.W., Deserno (Né Lehmann), T.M.: Generic integration of content-based image retrieval in computer-aided diagnosis. Comput. Methods Prog. Biomed. 108(2), 589–599 (2012)
7. Zagoris, K., Ergina, K., Papamarkos, N.: Image retrieval systems based on compact shape descriptor and relevance feedback information. Journal of Visual Communication and Image Representation 22(5), 378 – 390 (2011)
8. Zheng, B.: Computer-aided diagnosis in mammography using content-based image retrieval approaches: Current status and future perspectives. Algorithms 2(2), 828–849 (2009)

Employing time-series forecasting to historical medical data: an application towards early prognosis within elderly health monitoring environments

Antonis S. Billis[1] and Panagiotis D. Bamidis[1]
[1]Medical Physics Laboratory, Medical School, Faculty of Health Sciences, Aristotle University of Thessaloniki

Abstract. This work describes a first attempt to apply time-series forecasting analysis to health historical data in order to perform prediction of early pathological signs within telehealth applications, such as the Ambient Assisted Living environments for the elderly. A benchmark of state-of-the-art learning methods were applied to a set of artificial time-series data, simulating hypertensive patient profiles, based on blood pressure measurements. Results provided a fair proof of our initial hypothesis. Based on this first experimentation, our plans are to further investigate these findings in real –life or lab settings with seniors, thus proving the usefulness of time-series forecasting as a monitoring tool and an early prognosis mechanism in telehealth systems.

Keywords: telehealth, smart home, ambient assisted living, support vector machine, ARIMA models, time series forecasting, neural networks

1 Introduction

Senior citizens suffer nowadays from a wide variety of chronic conditions that reduce their independency level and deteriorate their health status [1]. This often happens due to the fact that they do not receive timely health assessment. In most of the cases this happens because any change is considered as part of the natural ageing process; seniors also tend to omit admitting they have a problem, or even of their fear of being institutionalized [2]. Early prediction of abnormal variation of health parameters may lead to early diagnosis of chronic diseases and subsequently to better medical decision-making and planning.

In this respect technology has much to offer the elderly. Recent technological advances resulted in the equipping of home environments with a plethora of sensors that aim to improve the seniors' quality of life and increase their independency by providing alerts in case of emergencies while increasing their socialization [3]. These approaches provided significant results in cases of fall detection, inappropriate use of electricity devices, estimation of participant’s functionality, etc. [4]. However, they have mainly focused on the detection of life-threatening acute events and they neglected the significance of slow-varying trends that may influence the health status of senior citizens [5] at a later time.

Time series analysis is a methodology that provides: i) pattern recognition of historical data (detection of frequent types of sequences) and .ii) forecasting of future values based on historical trends. There are several time-series forecasting methods known to the literature, such as exponential smoothing [6], Box-Jenkins seasonal ARIMA models [7] and neural networks [8].

In this paper, we aim to showcase the use of several state-of-the-art machine learning algorithms, such as Gaussian Processes, Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) and Box-Jenkins ARIMA models, with a set of artificially developed scenarios, relevant to patient models with high risk of hypertension. Our hypothesis is that one can take advantage of existing time-series forecasting methodologies to identify health trends over time and predict early signs of health deterioration, based on historical sets of health measurements and events.

The ARIMA model has advantages in its well-known statistical properties and effective (linear) modeling process. However, it may not work well in the presence of nonlinear relationships. In contrast, support vector machines and artificial neural networks time series models can capture the historical information by nonlinear functions and thus could prove to be efficient time series forecasting methods because of their flexible nonlinear mapping ability and tolerance to complexity in forecasting data.

2 Materials and Methods

In order to benchmark the forecasting methods a number of artificial scenarios were created. These concerned the simulation of day-to-day variation of diastolic blood pressure. Blood pressure is a prerequisite to monitoring seniors' physical health status progress or deterioration. This clinical parameter is very important since by monitoring it over a period of time it is possible to identify future risky health situations, such as hypertension. These situations may subsequently lead to a plethora of health problems, such as a sudden stroke episode or cardiovascular disease. Based on norms, an individual is hypertensive if he or she experiences repeatedly an elevated blood pressure exceeding 140 (systolic) and over 90 mmHg (diastolic).

The test data set consists of ten cases that reflect high risk profiles of seniors. The scenarios are produced randomly using a normal distribution with a mean diastolic blood pressure value nearly at the physiological margin of 80 mmHg and standard variation varying from 1 to 3.

The synthetic data were formed based on a priori knowledge derived from international norms, thus corresponding to hypertensive patient profiles. Specifically, the instances were modeled as a Gaussian process with a standard mean value and variability. The larger the variability the nosier the time-series is. The algorithms were tested under small, medium and high variability rates, so to provide cases of gradual forecast difficulty.

In order to perform time-series forecasting we have used well known machine learning methods and employed their existing implementations within publicly available AI suites such as WEKA [10] and Phicast [9]. More specifically, from the WEKA suite the following learning methods have been applied: GaussianProcesses (kernel: RBFKernel with gamma parameter set to default value 0.01), SMOreg (kernel: RBFKernel with gamma parameter set to default value 0.01) and finally MultilayerPerceptron (all parameters set to the default values). The parameters of the ARIMA model were selected as follows: p=1, d=1, q=1, resulting to a ARIMA(1,1,1) model.

3 Experimentation and results

The experiments compare predicted values with test data provided by the random artificial scenarios. Initial scenarios were split to two sets: one training set used for learning, which consisted of 50 instances and a second one which served as the test set for the evaluation of the forecasting methods. All instances represent measurements taken ideally on a daily basis.

Two experiments were conducted. Firstly, the ARIMA method alone was employed to forecast future value ranges and then the rest methods were applied to the same data sets to predict single future values. In the first case the metric used for the evaluation of the method is the accuracy of the prediction calculated as percentage of the correctly predicted ranges that the actual value was included and the total amount of the test set. In the second case, the following standard metrics are calculated: the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE).

An indicative result of the ARIMA prediction algorithm is depicted below (**Fig. 1**). Average prediction accuracy resulted to 91.2%.

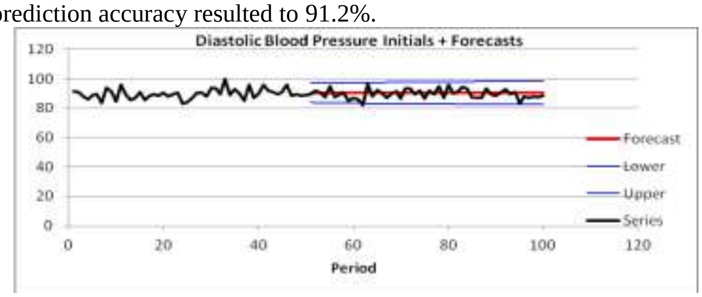

Fig. 1. Forecast Range vs Actual Values of Diastolic Data (95% prediction intervals)

Table 1 summarizes the evaluation results of the forecasting learning methods As shown GaussianProcesses learning algorithm provides the least error-prone analysis among the three. Neural Nets provided the worst forecasting results.

Table 1. Benchmark results of the tested forecasting algorithms

Learning Method	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)
GaussianProcess	1.798633	2.292558
SVM	2.005342	2.523508
Neural Network	6.061283	7.126667

4 Discussion

In this paper, time-series forecasting was employed to identify early pathological symptoms based on historical measurements. The clinical utility of forecasting (e.g.

vital signs) is of major importance in order to avoid hospitalization and alleviate the socioeconomical burden for caring, while it may improve significantly life quality of senior citizens.

Box-Jenkins ARIMA forecasting provides relatively accurate short-term forecasts regarding future ranges of the monitored variable (blood pressure). SVMs and GaussianProcessess can provide relatively fair near-future predictions. Preliminary results on artificial data partially support our initial hypothesis. However, there are several issues that set limits to the aforementioned results.

AI techniques were applied just in univariate time-series, whereas in most real-life health problems that elderly people suffer from, tend to be multifactorial. This means that forecasting a single parameter may not provide sufficient indications of early disease, since it needs to be investigated under its temporal relation with associative factors/parameters. Another limitation when it comes to real-life applications is that these methods require a larger amount of data than may be available within smart home environments; especially where a small set of unobtrusive sensors will be available. Also sensor failure may lead to missing data. In order to overcome, such problems several data imputation techniques may be applied [11].

Future plans of this research include among others: the setup of pilots, where actual sensors will be placed within a lab setting [13] and several seniors will be recruited to perform daily life activities for a long enough period so to gather enough historical data.

The forecasting output will be adaptive in terms of the size of the temporal window in order to facilitate the expert to estimate the possibility of both short-term events and long-term future trends. Furthermore, each actual/predicted instance will be mapped to normal or abnormal class and the classification accuracy will be calculated based on whether forecasting methods can predict the onset of a pathological condition at an early stage.

Time-series forecasting could be ideally employed within the context of an ambient assisted living environment, providing answers as whether we could detect transition patterns indicative of future health deterioration or timely estimate chronic alterations in the presence of outliers that may be either due to system (sensors) failure or to acute events. Combining seniors' health profile and time-series forecasting analysis, an intelligent Ambient Assisted Living environment would be able to propose the adoption of optimal lifestyle patterns according to the user's needs (acting proactively). This strategy could combine various non-pharmacological interventions (affective/cognitive/exer-games), alterations to the diet and/or some simple advice regarding a healthier lifestyle.

5 Conclusion

The reason that motivated us to use time-series forecasting methodologies is that telehealth monitoring should be able to detect apart from rapid deterioration states or life-threatening situations also slow-varying chronic conditions and provide early prognosis. More specifically, cases such as the geriatric depression or cognitive decline are characterized by a gradual impairment that may last for years and will be hardly noticed by the seniors or their carers, except for late stages where the outcomes of the disease remain irreversible [12]. Initial experiments with artificial data provided encouraging results. However, real-life experimentation will give us the chance to further fine-tune existing prediction algorithms and evaluate time series forecasting on the basis of its accuracy in the health monitoring field of use.

Acknowledgements. The work has been partially funded from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 288532. For more details, please see http://www.usefil.eu.

References

1. W. Lutz, B.C. O'Neill, and S. Scherbov, "Europe's population at a turning point", Science, vo. 28, no. 299, pp. 1991-1992, 2003.
2. Hayes TL, Pavel M, Kaye JA. An unobtrusive in-home monitoring system for detection of key motor changes preceding cognitive decline. *Proc. of the 26th Annual Intl. Conf. of the IEEE EMBS*, pp. 2480-2483, San Francisco, CA. (2004)
3. M. Popescu, G. Chronis, R. Ohol, M. Skubic, and M. Rantz: An eldercare electronic health record system for predictive health assessment. Paper presented at: Proceedings of the IEEE 13th International Conference on e-Health Networking, Applications and Services; Columbia, MO; June 13-15, pp. 193-196, 2011.
4. Rantz MJ, Scott SD, Miller SJ, Skubic M, Phillips L, Alexander G, et al. Evaluation of health alerts from an early illness warning system in independent living. Computers Informatics Nursing. 31(6):274–280, 2013.
5. T. Hayes, M. Pavel, and J. Kaye, An approach for deriving continuous health assessment indicators from in-home sensor data. In Technology and aging: Selected papers from the 2007 international conference on technology and aging, vol. 21, pp. 130–137, 2008
6. Winters, P. R. "Forecasting sales by exponentially weighted moving averages." Management Science, 6:324–342., 1960
7. Box GEP, Jenkins GM. Time Series Analysis: Forecasting and Control, 2nd edition. San Francisco, CA: Holden Day, 1976
8. Cortez, P., Rocha, M., and Neves, J. "Time Series Forecasting by Evolutionary Neural Networks." chapter III, Artificial Neural Networks in Real-Life Applications, Idea Group Publishing, USA, pages 47–70, 2005
9. Rob J. Hyndman, Anne B. Koehler, Ralph D. Snyder and Simone Grose. A state space framework for automatic forecasting using exponential smoothing methods http://www.elsevier.com/locate/ijforecast. (2002).
10. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten: The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1, (2009).
11. Shuai Zhang, Sally I. McClean, Member, IEEE, and Bryan W. Scotney Probabilistic Learning From Incomplete Data for Recognition of Activities of Daily Living in Smart Homes, IEEE Transactions On Information Technology In Biomedicine, Vol. 16, No. 3, 2012
12. Abellan van Kan, G., Rolland, Y., Nourhashémi, F., Coley, N., Andrieu, S., Vellas, B.,. Cardiovascular disease risk factors and progression of Alzheimer's disease. Dementia and geriatric cognitive disorders 27, 240–6 (2009)
13. Artikis, A., Bamidis, P.D., Billis, A., Bratsas, C., Frantzidis, C., Karkaletsis, V., Klados, M., Konstantinidis, E., Konstantopoulos, S., Kosmopoulos, D., Papadopoulos, H., Perantonis, S., Petridis, S., Spyropoulos, C.S.: Supporting tele-health and AI-based clinical decision making with sensor data fusion and semantic interpretation: The USEFIL case study. International Workshop on Artificial Intelligence and NetMedicine. p. 21 (2012).

Medical diagnostics based on combination of sensor and user-provided data

Maja Somrak[1], Anton Gradišek[1], Mitja Luštrek[1], Ana Mlinar[2], Miha Sok[3], and Matjaž Gams[1]

1 Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia
2 MESI d.o.o., Leskoškova cesta 9d, 1000 Ljubljana, Slovenia
3 Faculty of Medicine, University of Ljubljana, Vrazov trg 2, 1000 Ljubljana, Slovenia
{maja.somrak,mitja.lustrek}@ijs.si

Abstract. We present an approach that incorporates multiple machine-learning and data mining algorithms for prediction of the user's medical condition. The decisioning is based on vital signs data and user-provided input regarding the symptoms expressed. The presented method was trained and tested on virtual patients, generated using expert medical knowledge. We discuss future steps in the method development.

Keywords: medical diagnostics, questionnaire, sensor data, vital signs, user input

1 Introduction

Being able to diagnose common diseases at home can help the patient decide when to go to the doctor and at the same time reduce the burden on healthcare system. Here, we present a diagnostic method that incorporates information about user symptoms and vital signs readings to predict user's medical condition. The method was developed as a part of the diagnostic software for *HealthStation HOME*, a device competing at the Qualcomm Tricorder XPRIZE $10 million challenge [1]. We present initial results of experimental tests on virtual patients and outline possible improvements in future.

1.1 HealthStation HOME

The HealthStation HOME system [2] consists of a set of sensors that measure vital signs, such as heart rate, breathing rate, blood pressure, body temperature and blood oxygen saturation. The collected data is communicated to a mobile device for further diagnostics. The measurements can be interpreted as pathological symptoms (e.g. high blood pressure) that serve as an input for the diagnostic application. The user obtains an evaluation of his health condition by running the diagnostic application on the mobile device by selecting one of the starting options, *I feel pain* or *I feel unwell* (see Fig. 1). The application then guides the user through intelligently selected questions about the symptoms that are recognized as relevant. The result of the diagnostic method is the initial, home-based, medical condition assessment and a recommendation for further diagnostic testing (additional HealthStation HOME modules, such as blood or urine tests) to finally confirm or reject the diagnosis.

2 The Method

The HealthStation HOME diagnostic method is designed in form of a questionnaire with multi-modal inputs. The overview of the procedure is shown in Fig. 1. The initial input for the method consists of three types of data: (1) identified risk factors from the user profile data (e.g. smoking), (2) recognized pathological symptoms from vital signs measurements or other sensor data (e.g. high blood pressure) and (3) user selected pain symptoms in the application (e.g. chest pain). There are 60 predefined symptoms that the method can operate with, each of which can be treated as *unknown* or *known*, where *known* symptom is either *present* or *absent* in a patient. The *present* symptoms from the three inputs form the initial set of symptoms (4), which serve as the basis for automatically compiling a list of additional symptoms (5), from which the user is expected to select those he/she is experiencing. This list is generated to include both the symptoms that the user most probably experiences at the time and would probably want to report, and also the most relevant symptoms that would help the physician or the diagnostic method set a reliable diagnosis.

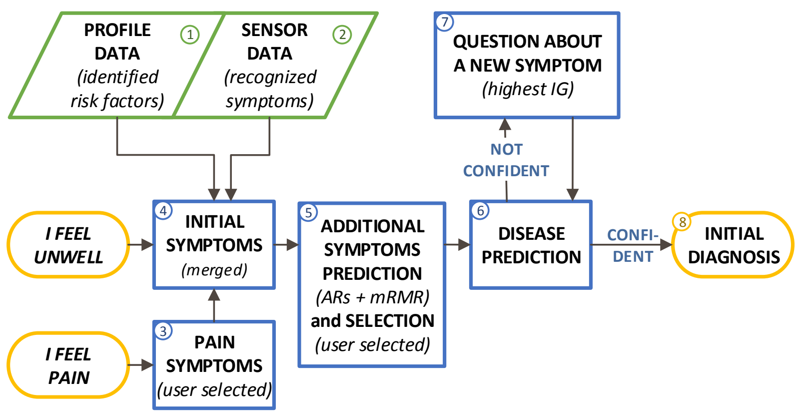

Fig. 1. Method overview. The numbers in circles indicate step numbers, as explained in the text.

If there is at least one symptom in the initial symptoms set, the method aims to find other symptoms that often emerge together with one of these initial symptoms. For this purpose, association rules (ARs) [3–5] of type *symptom A* $\rightarrow$ *symptom B* are searched, where symptom A is any of the initial symptoms and symptom B is any of the unknown symptoms. Rules with the highest confidence and minimal support condition satisfied are selected in order to produce a set of probable additional symptoms.

In addition to the ARs, the minimum-Redundancy-Maximum-Relevance [6] (mRMR) method is used to identify the most informative, mutually independent symptoms that have not been examined yet (unknown symptoms). This method works even if there are no symptoms in the initial symptom set. The mRMR resulting attributes subset is the subset of attributes (symptoms) that a) provide a lot of information about the class (medical condition) and b) are at the same time mutually uncorrelated. The criterion a) is measured with the mutual information between each attribute and the class, while b) is measured with the mutual information between the attributes. The mRMR rule used in our method is defined upon mutual information difference (MID) criterion [6] with the following equation

$$\max_{i \,\epsilon\, \Omega_S} [I(i,h) - \frac{1}{|S|} \sum_{j \,\epsilon\, S} I(i,j)]. \tag{1}$$

For selecting each additional symptom, an iteration of mRMR calculation over all unknown symptoms is repeated to find a symptom for which the value of the function is maximized (Eq. 1). $I(i,h)$ is mutual information between i, a symptom from the set of unknown symptoms Ω_S, and h, the classification variable − the medical condition. Likewise, j is a symptom from the set of known symptoms S, containing $|S|$ symptoms. Once a new symptom i is selected and added to the additional set of symptoms, it is treated as one of the known symptoms (the symptom is moved from Ω_S to S) in next iteration of mRMR calculation.

In the following step, the information about present and absent symptoms is first used for the disease prediction (6). The probabilities for predefined 15 different medical conditions (14 diseases and 'healthy') are evaluated using a set of J48 classifiers, one for each condition. There are two probability thresholds that represent medium and high chance for a certain medical condition, empirically selected to be 40% and 80%, respectively. If all condition probabilities fall below the medium threshold (improbable condition) or above the high threshold (very probable condition), the prediction is considered confident and the diagnostic procedure terminates, retrieving the diagnosis. However, if one or more medical condition probabilities lie between the medium and high threshold (neither very probable nor improbable condition), in the so called *gray zone*, the disease prediction is not considered confident. In this case, information about at least one additional symptom is needed to obtain a confident prediction. This is obtained by asking user a question about a new symptom (7). The additional symptom is chosen according to the highest information gain (IG), where the values are recalculated from a reweighted training set, such that the instances with the conditions from the *gray zone* are assigned higher weights. This approach, especially when incorporating reweighting, reduces the number of questions required for the probabilities to emerge out of the *gray zone*, as opposed to randomly selected questions. In case any condition probabilities still remain in the *gray zone* after a maximum number of question has been asked, the procedure terminates, selecting the medical condition with the highest probability for the diagnosis.

3 Experiments

We utilized expert medical knowledge to obtain the patient data sets. For this purpose, a table correlating 15 different medical conditions with over 60 symptoms was developed by physicians. The table was used for generating the training set containing 15000 virtual patients. Additionally, a test set of 1500 virtual patients was generated, where each medical condition was present in 100 patients. The tests demonstrated high sensitivity and specificity. For example, for otitis media, 94% of the patients with this medical condition were correctly identified while 84% of patients, diagnosed with otitis media actually had the disease. In the case of leukocytosis, the corresponding values were 59% and 60%, respectively [7]. On average for all medical conditions, the obtained values for sensitivity were 88.4%, for specificity 88.6% and for accuracy 88.3%. Currently, we are collecting the data of real patients for further testing, an example of a patient answering to the questions is shown in Fig. 2.

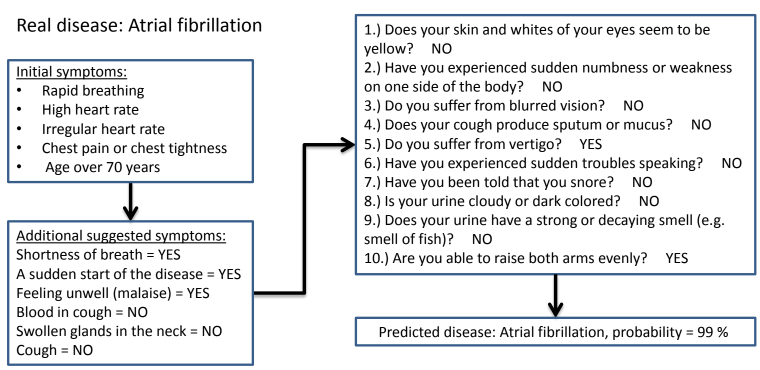

Fig. 2. Example of testing the diagnostic method on a real patient with atrial fibrillation.

4 Discussion and Conclusion

The initial results, based on both training and testing the method on virtual patients, show very high sensitivity, specificity, and classification accuracy (all over 80%). These values are probably too optimistic, according to the opinion of our medical associates. One of the main reasons for these results is that both the training and the testing set were generated from the same expert table. Because virtual patients are not biased when answering the questions, it is even more necessary to train and test the method on real patients, which we plan to do in the future. Moreover, the medical conditions were classified only into 15 different classes, which is far below the number of possible medical conditions in reality. The predefined medical conditions are also very distinctive in terms of symptom manifestation and it is therefore easier to distinguish between them (higher classification accuracy). The exceptions here are chronic obstructive pulmonary disorder, pneumonia, tuberculosis, and sleep apnoea; they are more frequently misclassified due to the similarity of their symptoms. In the future, we intend to incorporate hierarchical classification (e.g. additional class 'pulmonary disease'), when the data is insufficient for reliable differentiation between similar medical conditions. Current implementation of the method utilizes only the questionnaire for all of the symptom. In the future, we will use actual sensor input to determine the presence of a few specific symptoms. Additionally, we plan to include a larger number of medical conditions and implement intelligent methods for multilabel classification for discovering combinations of conditions.

Acknowledgments. We would like to thank MESI, DLabs, Gigodesign, Faculty of Medicine, and other partners, collaborating in the *MESI Simplifying diagnostics* team for the Qualcomm Tricorder XPRIZE competition.

References

1. Qualcomm Tricorder XPRIZE, `http://www.qualcommtricorderxprize.org/`
2. MESI, `http://www.mesimedical.com/home/`
3. McCormick T.H., Rudin C. and Madigan D.: A Hierarchical Model for Association Rule Mining of Sequential Events: an Approach to Automated Medical Symptom Prediction. Annals of Applied Statistics. (2012)
4. Soni S. and Vyas O.P.: Fuzzy Weighted Associative Classifier: A Predictive Technique For Health Care DataMining. International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol. 2, No.1, (2012)
5. Soni S. and Vyas O.P.: Using Associative Classifiers for Predictive Analysis in Health Care Data Mining. International Journal of Computer Applications, Vol. 4, No. 5, (2010)
6. Peng H., Ding C.: Minimum Redundancy Feature Selection from Microarray Gene Expression Data. Journal of Bioinformatics and Computational Biology, Vol. 3, No. 2, (2005)
7. Somrak M., Luštrek M., Šušterič J., Krivc T., Mlinar A., Travnik T., Stepan L., Mavsar M., Gams M.: Tricorder: Consumer Medical Device for Discovering Common Medical Conditions. Informatica 38, Ljubljana (2014)

Realtime depression estimation using mid-term audio features

Theodoros Giannakopoulos, Christos Smailis, Stavros Perantonis, and Constantine Spyropoulos

National Center for Scientific Research DEMOKRITOS, Athens, Greece
{tyianak,xsmailis,sper,costass}@iit.demokritos.gr
http://www.iit.demokritos.gr

Abstract. This paper presents a method towards estimating a clinical depression-specific score, namely the Beck Depression Inventory (BDI) score, based on analysis of mid-term audio features. A combination of support vector machines and semi-supervised learning has been applied to map the mid-term features to the BDI score. The method has been evaluated on the AVEC 2013 depression dataset. The overall system has been implemented in Python achieving a 20$x$realtime computational complexity on an average computer.

Keywords: Depression; Audio Analysis; Regression; Support Vector Machines

1 Introduction

One of the most common mood disorders is clinical depression. Therefore, automatic estimation or detection of its presence has gained research interest during the last years, through means of audio-visual signal analysis. This work focuses on the audio domain in order to estimate the Beck Depression Inventory (BDI) score of an individual, given a long audio recording of his/her voice.

Some of the previous studies in the field of depression analysis focus on utilizing acoustic features for the task of either predicting the Beck Depression Inventory (BDI) score of an individual, or classifying a person as depressive or not depressive. For example in [1] the performance of different audio features is explored for the task of classifying audio files contained within the AVEC 2013 dataset [2], as either depressed or nondepressed. In [3] several novel Canonical Correlation Analysis (CCA) based feature selection methods are presented, in order to reduce the massive dimensionality introduced for the AVEC 2013. Results revealed that using only 17% of original features, a relative improvement of 30% decrease of RMSE over the baseline on challenge test set was obtained. In [4] audio features that reflect changes in coordination of vocal tract motion have been utilized.

Other studies have adopted different approaches by attempting to fuse audio as well as video modalities in order to boost their performance. For example in [5] the authors made use of Motion History Histogram (MHH) in the video modality, to capture the movement of each pixel (texture variation) within the face area. In the audio modality the authors used a set of spectral Low-Level Descriptors (LLD). MHH was once again used for extracting change information of the vocal expression. For each modality, the Partial Least Square (PLS) regression algorithm is applied and predicted values of visual and vocal clues were further combined at decision level. Another multimodal approach is presented in [6]. This study presents a multimodal approach using a GMM-UBM system with three different kernels for the audio subsystem and Space Time Interest Points in a Bag-of-Words approach for the vision subsystem. These are then fused at the feature level to form the combined AV system. Key results include the strong performance of acoustic audio features and the bag-of-words visual features in predicting an individuals level of depression.

2 Feature Extraction

In most audio analysis and processing methods, it is rather common that the input audio signal is divided into short-term frames before feature extraction. In particular, the audio signal is broken into (non-)overlapping frames and a set of features is extracted for each frame. The result of this procedure is a sequence of feature vectors per audio signal. Another common technique used as a second step in audio feature extraction is the processing of the feature sequence on a mid-term basis. The audio signal is first divided into mid-term segments and then, for each segment, the short-term processing stage is carried out. At a next step, the feature sequence, which has been extracted from a mid-term segment, is used for computing feature *statistics*. In practice, the duration of mid-term windows typically lies in the range $1 - 10$ secs, depending on the application domain. In this work, extensive experimentation has led to selecting a 5 second mid-term window and a 50 msec short-term frame. In both cases, 50% overlap has been adopted.

2.1 Short-term audio features

In this section we describe the adopted short-term features. These features have been used in several general audio analysis methods and speech processing applications [7], [8], [9] and they cover a wide range of audio signal properties achieving discrimination abilities in several classification and regression tasks.

Energy The energy feature is computed as a sum of squared signal values (in the time domain), normalized by the window length. Short-term energy usually exhibits high variation over successive speech frames, since speech signals contain weak phonemes and short periods of silence between words.

Zero Crossing Rate The Zero Crossing Rate (ZCR) of an audio signal is the rate of sign-changes of the signal divided by the duration of that signal. ZCR has been interpreted as measure of the noisiness of a signal, therefore it usually exhibits higher values in the case of noisy signals.

Entropy of Energy The entropy of energy is a measure of abrupt changes in the energy level of an audio signal. It is computed by firstly dividing each frame in sub-frames of fixed duration. Then, for each sub-frame j its energy is computed and divided by the total energy. Finally, the entropy of that sequence of (normalized) sub-energies e_j is computed as the final feature value. This feature has lower values if there exist abrupt changes in the energy envelop of the respective signal.

Spectral Centroid and Spread The spectral centroid and the spectral spread are two basic spectral domain features that quantify the position and shape. The spectral centroid is the center of gravity of the spectrum, while spectral spread is the second central moment of the spectrum.

Spectral Entropy Spectral entropy is computed in a similarly to the entropy of energy, however it is applied on the frequency domain.

Spectral Flux Spectral flux is a measure of spectral change between two successive frames and it is computed as the squared difference between the normalized magnitudes of the spectra of the two successive frames.

Spectral Rolloff Spectral rolloff is the frequency below which a certain percentage of the magnitude distribution of the spectrum is concentrated. It can be treated as a spectral shape descriptor of an audio signal and it has been used for discriminating between voiced and unvoiced sounds.

MFCCs The Mel-Frequency Cepstrum Coefficients (MFCCs) have been very popular in the field of speech analysis. In practise, MFCCs are the discrete cosine transform coefficients of the mel-scaled log-power spectrum. MFCCs have been widely used in speech recognition, speaker clustering and many other audio analysis applications.

Chroma vector This is a 12-dimensional representation of the spectral energy of an audio signal. This is a widely used descriptor, mostly in music-related applications, however it has also been used in speech analysis. The chroma vector is computed by grouping the spectral coefficients of a frame into 12 bins representing the 12 equal-tempered pitch classes of western-type music.

2.2 Final feature vector extraction

The process described in 2.1 leads to a sequence of short-term feature vectors of 21 dimensions (this is the total number of short-term features described above). As a next step, statistics are calculated in a mid-term basis as described in the beginning of this Section. In particular, the following statistics are computed:

(a) Average value μ, (b) Standard deviation σ^2 and (c) σ^2/μ ratio. This leads to several mid-term feature vectors of 63 elements. The number of these mid-term vectors depends on the overall duration of the audio signal. Each of these vectors are fed as input in the next regression step that produces the final depression estimate decision.

3 Depression Estimation

Each mid-term feature vector described in the previous section is used in the context of a Support Vector Machine regression technique to estimate the Beck Depression Index. Note that since one decision is calculated per feature vector, the final decision (per audio recording) is extracted by averaging the mid-term BDIs. This rationalle helps in generating a sufficient number of samples for the SVM regression model training phase and in addition it manages to handle mid-term vocal characteristics. Finally, a semi-supervised dimensionality reduction technique has been adopted in order to give weight to feature dimensions that are discriminative in terms of depression estimation. In particular, recordings that share similar BDIs have been grouped together using a simple k-Means clustering approach. In the sequel, a Linear Discriminant Analysis step has been performed on the initial feature space to extract a depression-discriminant subspace. During this process, the BDI clusters have been used as indices in the LDA process.

4 Experiments

4.1 Dataset

In order to evaluate the presented method, we have used the AVEC 2013 depression dataset [2]. This includes 340 video recordings of 292 subjects performing a human-computer interaction tasks while being recorded by a audio-visual sensors. The average age is 31.5 years and a range of 18 to 63 years. The length of each recording varies from 20 to 50 minutes, with an average duration of 25 minutes per recording. The total duration of all recordings is 240 hours. The behaviour within the clips consisted of different tasks such as: sustained vowel phonation; speaking loud while solving a task, counting from 1 to 10, read speech, singing, telling a story from the subject's own past, and telling an imagined story. The 16-bit audio was recorded at a sampling rate of 41KHz.

4.2 Results

A cross-validation procedure on the AVEC dataset has been conducted in order to evaluate the presented scheme via the following performance measures: (a) Root mean square error (RMSE) (b) 4-class classification accuracy: in order to calculate this measure, the final decisions are discretized to four "depression levels" according to the following categorization: 0-13: minimal depression, 14-19: mild depression, 20-28: moderate depression and 29-63: severe depression.

Then the overall accuracy of the classification task is computed. (c) 2-class classification accuracy: similarly to the 4-class case, the problem is binarized using $BDI = 14$ as a threshold. The cross-validation procedure estimated an RMSE of 9.5, 4-class accuracy rate equal to 52% and binary accuracy rate of 70%.

5 Conclusions and future work

We have presented a method for detecting a subject's clinical depression score using audio information. A wide range of audio features has been extracted in a mid-term basis, while a combination of SVMs and a semi-supervised dimensionality reduction method has been used in the recognition process. Results demonstrate a 20% drop in the error regarding the baseline performance of the adopted dataset. We conduct ongoing research on using temporal analysis techniques to enchance the signal representation process by selecting representative areas of the recording with high discrimination ability (in terms of depression analysis). In addition, we plan to implement visual shape modelling methods in the context of a fusion system. Finally, it is rather important to conduct collaboration with a specialist in mental health in order to extract useful correlations between low-level audio features and mid-level depression-related characteristics.

References

1. Lopez-Otero, P., Docio-Fernandez, L., Garcia-Mateo, C.: (A study of acoustic features for the classification of depressed speech)
2. Valstar, M., Schuller, B., Smith, K., Eyben, F., Jiang, B., Bilakhia, S., Schnieder, S., Cowie, R., Pantic, M.: Avec 2013: the continuous audio/visual emotion and depression recognition challenge. In: Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge, ACM (2013) 3–10
3. Kaya, H., Eyben, F., Salah, A.A., Schuller, B.: (Cca based feature selection with application to continuous depression recognition from acoustic speech features)
4. Williamson, J.R., Quatieri, T.F., Helfer, B.S., Horwitz, R., Yu, B., Mehta, D.D.: Vocal biomarkers of depression based on motor incoordination. In: Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge, ACM (2013) 41–48
5. Meng, H., Huang, D., Wang, H., Yang, H., AI-Shuraifi, M., Wang, Y.: Depression recognition based on dynamic facial and vocal expression features using partial least square regression. In: Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge, ACM (2013) 21–30
6. Cummins, N., Joshi, J., Dhall, A., Sethu, V., Goecke, R., Epps, J.: Diagnosis of depression by behavioural signals: a multimodal approach. In: Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge, ACM (2013) 11–20
7. Giannakopoulos, T., Pikrakis, A.: Introduction to Audio Analysis: A MATLAB Approach. Academic Press (2014)
8. Theodoridis, S., Koutroumbas, K.: Pattern Recognition, Fourth Edition. Academic Press, Inc. (2008)
9. Hyoung-Gook, K., Nicolas, M., Sikora, T.: MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval. John Wiley & Sons (2005)

Author Index