pHealth 2016
N. Maglaveras and E. Gizeli (Eds.)
IOS Press, 2016

doi:10.3233/978-1-61499-653-8-123

Designing smart analytical data services for a personal health framework

Lefteris KOUMAKIS[a,1], Haridimos KONDYLAKIS[a], Maria CHATZIMINA[a], Galatia IATRAKI[a], Panagiotis ARGYROPAIDAS[a], Eleni KAZANTZAKI[a], Manolis TSIKNAKIS[a], Stephan KIEFER[b], Kostas MARIAS[a]
[a] *Computational BioMedicine Laboratory, FORTH-ICS, Heraklion, Crete, Greece*
[b] *Fraunhofer Institute for Biomedical Engineering, St. Ingbert, Germany*

Abstract. Information in the healthcare domain and in particular personal health record information is heterogeneous by nature. Clinical, lifestyle, environmental data and personal preferences are stored and managed within such platforms. As a result, significant information from such diverse data is difficult to be delivered, especially to non-IT users like patients, physicians or managers. Another issue related to the management and analysis is the volume, which increases more and more making the need for efficient data visualization and analysis methods mandatory. The objective of this work is to present the architectural design for seamless integration and intelligent analysis of distributed and heterogeneous clinical information in the PHR context, as a result of a requirements elicitation process in iManageCancer project. This systemic approach aims to assist health-care professionals to orient themselves in the disperse information space and enhance their decision-making capabilities, to encourage patients to have an active role by managing their health information and interacting with health-care professionals.

Keywords. Data Analysis, data mining, personal health record, visual analytics, knowledge discovery

1 Introduction

Decision-making is a key concept in the healthcare domain while at the same time the ability to share personalised medical information, including Personal Healthcare Record (PHR), is becoming increasingly vital. Moving across different healthcare services (EHRs, PHRs, mHealth applications), patients' data must be available in a unified way, discoverable, accessible, and understandable. This will allow, on the one hand, healthcare professionals to quickly access well organized medical information especially in case of emergency or critical situations, thus enhancing the quality of care and increasing patient safety. On the other hand, patients move from a passive to active role in their treatment and can receive an integrated care, with enhanced quality, from different points of care.

The iManageCancer project supports chronic cancer treatment via a cancer disease self-management platform designed according to the specific needs of patient groups and focusing on the wellbeing of the cancer patient with special emphasis on psycho-emotional evaluation [1] and patient empowerment [2], [3], [4]. The platform is centred

[1] Corresponding Author: Lefteris Koumakis. CBML ICS-FORTH, N. Plastira 100 Vassilika Vouton, GR-700 13 Heraklion, Crete, Greece Email: koumakis@ics.forth.gr

in a Personal Health Record [5] that exploits recent advances on Health Avatars for the individual cancer patient surrounded by mHealth applications designed to encourage the patient, enhance clinician-patient communication, maximise compliance to therapy, inform about drug interactions, and contribute to the management of pain and other side-effects of cancer treatment. Among others, iManageCancer aims to design and assess public-private-partnership based service and business models around an ecosystem oriented to a Health Data Cooperative [6] to make sustainable analytical services available on the internet.

This paper focuses on the research activities within the iManageCancer project for the design of the smart analytical framework. First, it presents the setting and the methods for the requirement analysis performed in Section 2. According to the requirement analysis, it proceeds further to define the architectural design and the workflow for the data analysis and mining scenario presented in Section 3. Finally, in section 4 we conclude this paper and discuss our future plans.

2 Methods

The iManageCancer platform collects multidisciplinary data covering areas from the medical, the environmental and the lifestyle domains. The objective of the iManageCancer smart analytical services is to extract information from the diverse data of iManageCancer and transform it into an understandable structure for better knowledge and further use. To do so, an effective analytical framework is needed which will be based on real user requirements. At the first step we gathered information related to users requirements in order to understand the needs from such tools not only for the statisticians but also for the physicians and the patients. For that reason in iManageCancer we conducted an in depth user requirements analysis along with a survey for the patients and a workshop where physicians, developers and patients discussed and concluded to the user needs and the smart analytical services scenario.

2.1 Requirements elicitation

The iManageCancer platform supports different types of data including lifestyle data, clinical data and vital signs which will be continuously evaluated against the personal health record and history, and feedbacks towards individuals will be automatically generated at the point of need. For a successful data analysis framework, apart from the data mining algorithms, access to integrated data and knowledge of the underlying data and data structures is needed. The following sub-sections describe the results of the requirements elicitation for the data, the data access and the data analysis in iManageCancer.

2.1.1 Data

Smart analytical tools can take advantage of all the available data in row and structured format. Specifically, the available data and their structure are as follows: (i) Medical data: Well structured, text (using ontologies) & images with rich metadata, (ii) Environmental and sensor data from medical sensors (e.g. blood glucose, temperature) with intervals for normal-abnormal and activity sensors with numeric values and (iii) Personal profile data from the PHR which support, among others, free text (not always in English). Such data can be used taking advantage of Natural Language Processing frameworks for annotation in the biomedical domain [7].

2.1.2 Access to Data

The heterogeneity and scale of clinical, environmental and lifestyle data raises the demand for seamless data access along with the availability of powerful and reliable data analysis operations, tools and services [8]. Obviously the amount of information available, the heterogeneity of the information and the wide range of biomedical ontologies dictate the identification of a solution able to handle all this information. iManageCancer implements interlinking of several ontologies into a global schema to integrate all internal and external data, called Smart Access Layer (see **Figure 1**). The heterogeneous eHealth streams of data to come, might affect the performance and scalability. For that reason, iManageCancer goes beyond the state of the art by exploring the ontology-based data integration of new storage approaches (such as NoSQL databases) and addressing the challenges occurring in such a setting.

2.1.3 Data Analysis

The data analysis and data mining tools aim to extract information from the diverse data of iManageCancer and transform it into an understandable structure for enhancing knowledge extraction. Smart data analytics will provide mechanisms able to identify patterns or trends in data, screen pre-frailty states and provide different views of data for new management plans. Data mining consists of various methods and algorithms [9] which have been applied to many research areas and the healthcare domain is not an exception. Methodologies which take advantage of data mining in the clinical domain, such as the one proposed by G. Potamias et al [10], will be evaluated and possibly adopted and deployed following the service-oriented model. Handling of the diverse and large amount of data will be supported by feature elimination algorithms used in the demanding domain of bioinformatics [11]. In order to advance data mining within the iManageCancer context, special efforts will be given in the utilization of main data mining standards such as the Predictive Model Mark-up Language [12] and open source environments and libraries like Weka [13].

2.2 Online survey

In order to take into account the opinion of the patients, an online survey was created and promoted to the cancer community via ecancer.org[2] along with a host of other platforms to distribute this in English, Italian, German and Greek. People from all across Europe with 226 surveys submitted their opinion. The report of the surveys can be found in the public deliverable of iManageCancer D2.2 ‘Scenarios and use cases including the ethical and legal aspects’ from the web site of the project[3] and the questionnaire is provided as Appendix in the deliverable. Among others, the following conclusions related to the analytical services are driven from this survey:

- More than 80% of responders want to have tools for analysing their health data.
- Around 80% are willing to provide their health data for research with only 50% giving consent online.

---

[2] http://ecancer.org
[3] http://imanagecancer.eu/resource-centre

- Security for sharing is most important for all respondents. Interestingly the security is seen more important in sharing health data than in online banking

Based on the results of the survey, an intensive review of existing scenarios and use cases and during a two days' workshop, the consortium concluded to five categories of scenarios including a scenario for data analysis. The scenario and the architecture of the smart analytical services are described in the following section.

3 Results

An outline of the reference architecture for smart analytical services is shown in **Figure 1** where, the basic operational modules of the system are also shown. Smart analytical services will try to go much further than traditional statistics by examining the raw data and then attempting to hypothesise relationships within the data. Such systems are able to produce quite complex characterisations of data relationships and attempt to discover humanly understandable concepts. As shown in **Figure 1**, data analysis and data mining in iManageCancer is an iterative approach, which combines data from the semantic layer of iManageCancer, pre-processes the data, performs the analysis and provides the results for visualization based on the data distillation model [14]. The loop closes with the interaction of the end user who can refine the results and continue with a drill-down analysis to extract knowledge from cohorts with specific criteria.

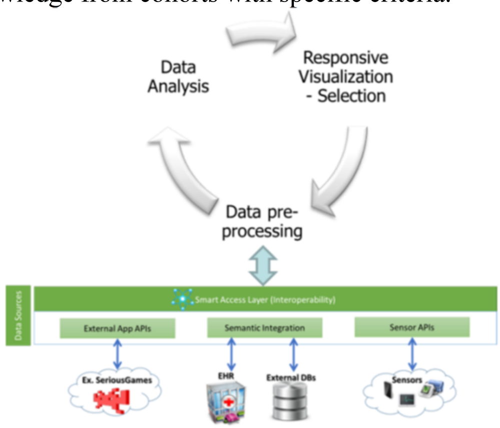

Figure 1. Architecture of the analytical services over personal health infrastructure

End users of the smart analytical services are physicians, data miners, statisticians and data managers. The workflow for the data analysis and mining scenario consists of the following steps:

- **Step 1 (Overview of the data)**: The end user logs-in to the system. An overview of the iManageCancer data for the whole population using various visualization techniques, such as charts, plots, bubble charts, parallel coordinates, is available. The user can view statistics for the active daily living and the frailty state of the iManageCancer population.
- **Step 2 (Create and analyze cohorts):** The user can select, using the interactive visualization techniques, specific features and set inclusion/exclusion criteria in order to create a new cohort, e.g. select only male patients and age > 60. Alternatively the user can select specific features, e.g. age and disease, and request

from the system to propose new cohorts. The proposed cohorts will be extracted using clustering techniques along with feature selection algorithms for the identification of relationships in the data. Then the user selects/approves one or more new cohorts and requests to analyse and view the results for these cohorts. The system applies data analysis for the new cohort(s) using machine learning methodologies such as classification, regression, clustering and association rules. Analysis results are presented to the end user using the same techniques as of step 1.

- **Step 3 (Monitor specific patient):** The user can select one or more patients and (i) View patient's statistics, or (ii) Plot patient's data in conjunction with average values of specific cohort(s), e.g. daily active living of population, in order to highlight and identify deviations.
- **Step 4 (Alerts):** Finally, the user can select and create a new alert for specific cohorts (from step 1 or 2) or specific patients (step 3). The user will be able to select features and add an automatic alert/indication. Such an alert could be to compare the side effects of 2 different treatments. The system will monitor the specific feature(s) for the selected cohorts and if significant differences appear the researcher will be informed via email.

4 Discussion and Conclusions

The iManageCancer platform has been designed on clinical evidence, in close collaboration of clinical experts, IT specialists and patients. The data analysis and data mining tools aim to extract information from the diverse data of the healthcare domain and transform it into an understandable structure for better knowledge discovery. Smart data analytics will provide mechanisms able to identify patterns or trends in data, screen pre-frailty states and set different views of data for new management plans. Analysis tools will be implemented for extracting new knowledge, by the effective integration of data mining and expert knowledge. Visual analytics will make use of information from iManageCancer data sources, and bring together valuable information in visual form to support exploration. Such a system is expected to successfully overcome the limitation of traditional intelligent data analysis that works only with a small number of well-defined and well trained cases.

Pilots will assess the added value on health and quality of life of the decision support and analysis tools and the platform as a whole. iManageCancer platform will be evaluated in two healthcare domains for children and adults. The patients that will be enrolled in the children pilot are 60 nephroblastoma patients and for the adult's pilot 100 prostate cancer patients. Both pilots will have duration of at least 6 months and will start the second half of 2016.

Our approach to predictive modelling and data mining will leverage the available community knowledge. We currently evaluate and extend existing models able to handle the large datasets which the iManageCancer project will support [15]. These updated models together with the new models that we develop will be integrated in the iManageCancer platform and used to provide support to clinicians. The environment will enable the continuous evaluation of the prediction and decisions taken as new knowledge and data becomes available in the platform. The data driven tools to be implemented will analyse the information in the iManageCancer database and draw conclusions related to the usage of the self-management platform, reported adverse events and several health issues. Smart analytical services will provide to physicians a global view of the end users data and monitor the evolvement over time. Furthermore, the platform will provide specific services for data-driven analysis services on anonymised clinical information for public health research.

5 Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643529 (iManageCancer project).

6 References

[1] H. Kondylakis, E. Kazantzaki, L. Koumakis, I. Genitsaridi, K. Marias, A. Gorini, K. Mazzocco, G. Pravettoni, D. Burke, G. McVie, M. Tsiknakis, Development of interactive empowerment services in support of personalised medicine., *ecancermedicalscience*, 8, 2014.

[2] H. Kondylakis, L. Koumakis, E. Genitsaridi, M. Tsiknakis, K. Marias, G. Pravettoni, A. Gorini, K. Mazzocco, IEmS: A collaborative environment for patient empowerment, in *IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE)*, 2012, pp. 535-540.

[3] H. Kondylakis, L. Koumakis, E. Kazantzaki, M. Chatzimina, M. Psaraki, K. Marias, M. Tsiknakis, Patient Empowerment through Personal Medical Recommendations., *MedInfo*, 2015, p. 1117.

[4] H. Kondylakis, L. Koumakis, M. Tsiknakis, K. Marias et. al. Smart recommendation services in support of patient empowerment and personalized medicine. In Multimedia Services in Intelligent Environments (pp. 39-61). 2013, Springer International Publishing.

[5] I. Genitsaridi, H. Kondylakis, L. Koumakis, K. Marias, M. Tsiknakis, Towards Intelligent Personal Health Record Systems: Review, Criteria and Extensions., *Procedia Computer Science*, vol. 21, pp. 327-334, 2013.

[6] E. Hafen, D. Kossmann, A. Brand, Health data cooperatives–citizen empowerment, *Methods Inf Med*, vol. 8, no. 53, 2014.

[7] P. Sfakianaki, L. Koumakis, S. Sfakianakis, G. Iatraki, G. Zacharioudakis, N. Graf, K. Marias, M. Tsiknakis, Semantic biomedical resource discovery: a Natural Language Processing framework, *BMC medical informatics and decision making*, vol. 15, no. 77, 2015.

[8] K.J. Cios, G.W. Moore, Uniqueness of medical data mining, *Artificial intelligence in medicine4.*, vol. 26, no. 1, pp. 1-24, 2002.

[9] G.M. Weiss, B.D Davison, *Data Mining. Handbook of Technology Management*, H. Bidgoli, Ed. 2010

[10] G. Potamias, L. Koumakis, V. Moustakis, Mining XML clinical data: the healthobs system, *Ingénierie des systèmes d'information*, vol. 10, no. 1, pp. 59-79, 2005.

[11] G. Potamias, L. Koumakis, V. Moustakis, Gene selection via discretized gene-expression profiles and greedy feature-elimination, *Methods and Applications of Artificial Intelligence*., 2004, pp. 256-266.

[12] A. Guazzelli, M. Zeller, W.C. Lin, G. Williams, PMML: An open standard for sharing models, *The R Journal*, vol. 1, no. 1, pp. 60-65, 2009.

[13] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten., The WEKA data mining software: an update., *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10-18, 2009.

[14] F. Frankel, R. Reid, Big data: Distilling meaning from data., *Nature*, 455(7209), 30-30..

[15] H. Kondylakis, L. Koumakis, M. Tsiknakis, K. Marias, S. Kiefer: Big Data in Support of the Digital Cancer Patient. ERCIM News 2016(104) (2016).