TOWARDS KNOWLEDGE-BASED INTEGRATION OF PERSONAL HEALTH RECORD DATA FROM SENSORS AND PATIENT OBSERVATIONS

John Puentes
Institut Telecom, Telecom Bretagne, Dpt. Image et Traitement de l'Information, Inserm U650 LaTIM, Brest, France

Jaakko Lähteenmäki
Knowledge Intensive Services, VTT Technical Research Center of Finland, Otaniemi, Finland

Keywords: Body monitoring, Personal health record, Physiological sensors, Heterogeneous data integration, Knowledge model, Data understanding.

Abstract: Personal Health Records (PHR) containing physiological data collected by multiple sensors are being increasingly used for wellness monitoring or disease management. These abundant complementary raw data could be nevertheless disregarded given the challenges to understand and process it. We propose a knowledge-based integration model of PHR data from sensors and personal observations, intended to facilitate decision support in scenarios of cardiovascular disease monitoring. The model relates knowledge at three data integration layers: elements identification, relations assessment, and refinement. Details on specific elements of each layer are provided, along with a discussion of use and implementation guidelines.

1 INTRODUCTION

An increasing amount of physiological data produced by multiple wearable body monitoring devices, is gradually becoming available to individuals (Jovanov et al., 2009). Depending on user requirements - wellness monitoring or disease management - these data streams can be either used separately, or be stored with personal observations in the Personal Health Record (PHR).

Whereas in wellness monitoring a particular signal as weight or heart rate is periodically measured and analyzed according to a goal, in disease management several sensor inputs are studied in order to continuously account for abnormal parameters variation. The second scenario implies significant additional work for the physician, compelled to handle and interpret complementary voluminous data, as well as for the patient asked to acquire data on a regular basis. Such supplementary common effort bears a major promise: both physicians and patients expect a return in terms of improved follow-up and decision support (DS).

Notwithstanding its importance, these data acquired by body monitoring devices and personal observations could be quickly neglected, given its significant volume and the numerous challenges to make sense out of it automatically (Garg et al., 2010), unless it could be seamlessly integrated to the PHR for further use after acquisition. This paper defines and analyses a knowledge-based integration model of PHR data from sensors and personal observations, adapted to use cases of cardiovascular disease. It focuses on the role of data for information and knowledge discovery, by means of data processing to provide pertinent DS (Figure 1).

We intend to explore the question of how to combine relevant complementary data sources in the PHR, enabling data utilization for DS, independently of the concerned devices and data features. The resulting integration model relies on a knowledge infrastructure capable of handling meaningful connections between sensors data, observations, and information processing algorithms. It focuses particularly on knowledge about sensors output, annotations meaning, and related data structures.

Even though such integration is required in various healthcare related contexts applications (Kulkarni and Öztürk, 2007, Stuntebeck et al., 2008, Martínez-López et al., 2008), it has not been

Puentes J. and Lähteenmäki J..
TOWARDS KNOWLEDGE-BASED INTEGRATION OF PERSONAL HEALTH RECORD DATA FROM SENSORS AND PATIENT OBSERVATIONS .
DOI: 10.5220/0003162502800285
In *Proceedings of the International Conference on Health Informatics* (HEALTHINF-2011), pages 280-285
ISBN: 978-989-8425-34-8
Copyright © 2011 SCITEPRESS (Science and Technology Publications, Lda.)

addressed within a knowledge framework, to facilitate conformity with multiple acquisition devices, DS oriented data mining, and flexible user queries. This work defines a knowledge-based model to enable integrated PHR data understanding and processing. In section 2, PHR functionality is described, before presenting in section 3 the main components of the knowledge-based model. Section 4 provides details on specific elements of each layer. Section 5 discusses the use and implementation guidelines of the proposed model. Section 6 presents the conclusions.

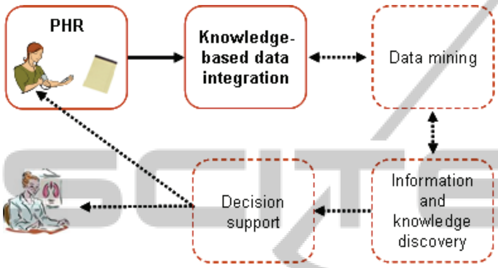

Figure 1: PHR knowledge-based data integration with respect to subsequent data processing modules, as a means to provide patient and physician DS.

2 PERSONAL HEALTH RECORD DATA

As a complement to health care professionals oriented Electronic Health Record (EHR), PHR endorses individuals' active role in their own healthcare providing means to acquire, store, and exchange health data like personal observations related to specific events and physiological measurements collected by sensors. Multiple definitions as well as a variable range of complexity characterize the approaches to build a functional PHR (Tang et al., 2006). Furthermore, PHR has been rarely incorporated to medical care flows, and its adoption raises a wide variety of questions and challenges (Halamka et al., 2008).

For our purpose, a PHR is mainly composed of cardiovascular disease physiological sensor measurements like activity, weight, temperature, blood pressure, heart rate, and blood glucose, which can be acquired by the patient throughout the day or night, without the intervention of medical personnel, as part of cardiovascular disease follow-up. The PHR also contains personal observations related to the occurrence of events like dizziness, weakness, dyspnea, arrhythmias, or other anomalies, conveying additional elements that properly documented have the potential to reinforce DS. Supplementary data like medications, laboratory tests, medical history, and allergies that can be either accessed from the patient EHR or copied from it to the PHR are not taken into account for this analysis.

3 KNOWLEDGE-BASED DATA INTEGRATION MODEL

Seamlessly DS requires data processing and knowledge discovery to be absolutely independent of available data resolution, imperfection, heterogeneity, and formats. Besides, DS components must be aware of the probable inadequacy of data before activating the inference engine. Applying the algorithms directly could produce severe errors.

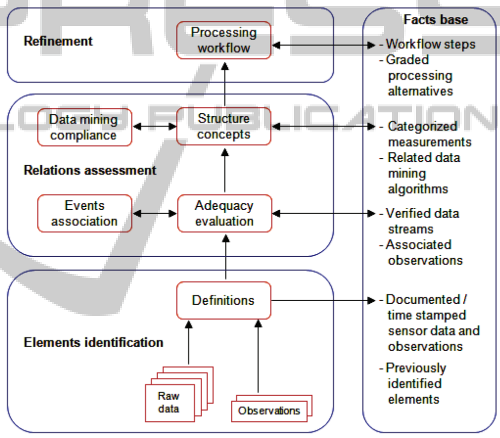

Figure 2: Main components of the knowledge-based model to integrate data from physiological sensors and patient observations.

The conceptualized knowledge-based integration of data from physiological sensors and observations encompasses therefore (Figure 2): data definitions, data structure concepts, association of reported events with the sensors measurements, evaluation of the obtained relations adequacy, verification of the integrated data and relations compliance with data mining requirements, and the respective workflow generation. The model can be abstracted as the interaction of three layers, i.e. elements identification, relations assessment, and refinement.

3.1 Elements Identification

Raw data obtained from sensors e.g. vital signs, blood glucose, or activity, are not likely to have the same format, and even if the streams are standardized the DS system still needs to properly

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

identify every pertinent element. This level contains knowledge in the form of definitions making possible to understand the: file format (XML, CSV, text, binary, etc.), file source (sensor type, constructor, and model), data structure (values ordering, variables meaning, time stamps, and units), and function to read the file. Equivalent separate definitions permit to comprehend patient observations. This kind of files is not generated by the sensors, but created by the control application running on the patient's mobile phone or computer.

3.2 Relations Assessment

Once all the uploaded data streams elements have been identified, events are associated and multiple relations assessed. Data streams coherence is examined looking for incomplete, invalid, or contradictory measurements and observations that could hinder subsequent processing, impeding as a consequence to obtain proper DS. This step is guided by knowledge about valid values, anomalies, outliers, and accuracy criteria. Observations about unusual events are added to the facts base according to the corresponding time stamps.

Inference rules are then applied to define a DS-oriented data structure. Compliance with related predefined data mining algorithms is verified, determining the valid data streams of available previous and current measurements, depending on DS requests. Knowledge at this stage mainly categorizes measured values according to computed characteristics. Additionally, common time intervals of combined measurements are detected.

3.3 Refinement

Further detail about integrated data usability for knowledge discovery, is defined by assembling the appropriate processing workflow to answer a DS query. Taking into account the categorized measurements variability and the analysis goal extent, the significance of intended processing alternatives is graded against specific criteria. This mechanism is necessary to produce a unified workflow of processing steps, depending on the quality of available data. The grading criteria determine hence up to what point and how, the DS request can be answered with validated data. To that end, knowledge in the form of rules about processing algorithms features linked to DS goals is embedded in this layer.

4 MODEL ELEMENTS

Schematic descriptions of the three knowledge layers components are presented in the next sections, to illustrate how the different model elements are interrelated.

4.1 Definitions

Basic reference knowledge about the observed variables' characteristics is stored in the first layer of the model, permitting to have a global evaluation of available data quality. The following definitions were considered to initially assess a database:

- Time reference to be used when analyzing the integrated data – date, hour, minutes, and concerned measurement sessions. It is taken as the starting point in time to evaluate a group of measurement sessions.
- Time measurement definitions and related units – instant and interval. Values can vary from a specific moment, to intervals of minutes, hours, days, weeks, or months.
- Elements of a session - measurement, user and session identification, date, start time, duration, sensor, or observation report.
- Activity definition and units – amount of steps per time unit. It depends on the pedometer characteristics and can be expressed in additional terms like distance, speed, etc.
- Pedometer model – brand, model, activity definition, measured values ordering, accuracy, and file type.
- Blood pressure definition and units – systolic and diastolic blood pressure measured at the upper arm in mmHg.
- Blood pressure meter model – brand, model, blood pressure definition, measured values ordering, accuracy, and file type.
- Heart rate definition and units – pulse in beats per second.
- Heart rate monitor model – brand, model, heart rate definition, measured values ordering, accuracy, and file type.
- Skin temperature measurement definition and related units – heat in degree Celsius. It depends on the thermometer characteristics and can be expressed in other units.
- Thermometer model – brand, model, temperature definition, measured values ordering, accuracy, and file type.
- Weight measurement definition and units – sub-

ject's body mass in kilograms and grams.

- Scale model – brand, model, weight definition, measured values ordering, accuracy, and file type.
- Glucose measurement definition and related units – sugar concentration in mmol/L or mg/dL.
- Glucometer model – brand, model, glucose measurement definition, values ordering, accuracy, and file type.
- Observations source description – source application, time stamp, and file type.
- Observations definition and description – dizziness (confusion, loss of stability and perception), weakness (discomfort, fatigue), dyspnea (distress, breathlessness), quantified to levels such as "none", "occasionally" and "frequently".

These definitions can be extended if more detailed descriptions are used, based for instance on a standard (IEEE Health Informatics, 2009).

4.2 Adequacy Evaluation and Events Association

The second layer of the model contains knowledge about relations and validations to be assessed, with the objective of evaluating adequacy and data mining compliance of the stored data streams. After database elements identification, the following relations are examined:

- Time stamps during a measurement session can be sequential or fragmented in continuous sub-intervals. When the second case is detected a list of sub-intervals is generated.
- Valid steps counts - 0 to 180 steps per minute.
- Anomalies in steps counts - negative values, time slots without measurement, or long sequences of constant values (except 0 steps).
- Outliers in steps counts - more than 180 steps per minute for a normal subject.
- Valid blood pressure values - normal systolic 120/140 mmHg and normal diastolic 80/90 mmHg.
- Anomalies of blood pressure values – lower (systolic < 50 mmHg / diastolic < 35 mmHg), or higher (systolic > 230 mmHg / diastolic > 140 mmHg) than physiological limits.
- Outliers of blood pressure values – for systolic between 150-230 mmHg, or 90-50 mmHg; for diastolic between 100-140 mmHg, or between 70-35 mmHg.
- Valid heart rate values - at rest 40-60 bpm (beats per minute), in moderate activity 60-80 bpm, walking 76-108 bpm, during exercise 109-160. These magnitudes vary depending on age, weight, height, and clinical condition.
- Anomalies of heart rate values – lower than 40 bmp or higher than 160 bpm.
- Outliers of heart rate values - punctual sets of rather low or high bpm compared to the rest, within a session.
- Valid skin temperature values – from 10 to 40 C°.
- Anomalies of skin temperature measurements - continuous repeated changes from low to high values (or inversely).
- Outliers of temperature measurements - continuous low or high temperatures periods.
- Valid weight measurements – table of values according to age, height, and clinical condition.
- Anomalies of weight measurements - negative values, very low or high values with respect to valid measurements, or drastic changes.
- Outliers of weight measurements – lower or higher points than expected in a sequence.
- Valid glucose measurements – 3.83 to 8.88 mmol/L, or 69 to 160 mg/dL.
- Anomalies of glucose measurements - very low or high punctual values.
- Outliers in glucose measurements - continuous low or high glucose periods.
- Observation report content – observations definitions presented in section 4.1, rank on a scale given by the patient to the sensation, circumstance that provoked it, frequency, source application, and comments.
- Valid observation report. Applications to create observations reports must comply with the required observation report content.
- Anomalies in observations. Although, applications to create observations reports are designed to prevent input errors, files can be corrupted during transfer. Therefore, it is necessary to verify all the parameters validity.
- Outliers in observations. Depending on the pathology and patient profile, some observations could be considered as outliers because of the severity, duration, frequency, and/or circumstances. Those particular events need to be identified and displayed separately.

Findings about verified data are stored in the facts base to be used at the refinement stage.

4.3 Structure Concepts and Data Mining Compliance

As part of the data exploration procedure, it is also necessary to obtain additional information, about available data sets compliance with the sought DS request. The suggested list of knowledge components is not exhaustive, because DS features and processes are not examined in this work. However, a basic scheme is proposed, in order to illustrate how the model is expected to prepare a structure of concepts. Some of these knowledge components are:

- Description of data streams depending on the presence of outliers and abnormal measurements.
- Statistical description of data streams. Multi-parameter statistical description of the valid measurements (continuous sequences or scattered values of punctual values).
- Common reference profiles. In several cases, measurement patterns could appear. A set of rules describes the reference profiles.
- Characterization of data streams according to differences with respect to pre-defined common reference profiles.
- Data distribution and grouping at different degrees of temporal resolution. Identify data and observations that correspond to adjustable time intervals, from minutes to months.
- Temporal reasoning. Temporal relations between measurements and observations, focused on simultaneity and sequencing.
- Generate a global annotated description of the examined data streams and observations.

At this point, it is possible to carry out an evaluation to examine the compliance of available data with pre-defined data mining algorithms, associated to the respective DS request. Such evaluation verifies if the global annotated description does not contain penalizing facts, which will permit to conclude before describing the processing workflow that collected values are not reliable. For instance, significant amount of anomalous values, outliers, poor statistical descriptors, and/or incoherent temporal relations, will prevent any further processing.

4.4 Processing Workflow

When evaluated data complies with the algorithms of a given DS request, the knowledge-based model defines to which data sets those algorithms can be applied. Again, the proposed list of knowledge components is not exhaustive given its intrinsic relation to the wanted DS:

- Identification of candidate algorithms using a set of rules.
- Evaluation of the differences between characteristics of expected input for each candidate algorithms and available data.
- Grading of the detected input differences according to a significance scale.
- Construction of a processing workflow list, containing the ordered algorithms and associated data sets.

Depending on the asked DS, some processing workflows are likely to be more elaborated than others. Thereafter the knowledge-based data integration model output is applied by a procedural-oriented system. Addition of other sensors like for instance, oxygen saturation or respiratory rate, will require expanding the elements of each knowledge layer accordingly, as well as enlarged procedures to integrate sensors data and observations.

SCIENCE AND TECHNOLOGY PUBLICATIONS

5 DISCUSSION

The proposed model is a first attempt to integrate data from multiple sensors and patient observations at the acquisition and processing stages, by means of a knowledge-based framework. It assumes that these data are part of the PHR, easily accessed by patients and clinical personnel alike, and is capable of handling DS requests made by any of them. Answers to those specific requests should be obtained according to an optimal processing workflow produced by the model, making use of pre-defined algorithms and the relative significance of longitudinal data.

It is important to note that data values are not likely to be continuous, predictable or synchronized, as in more conventional approaches like data streaming management or the so-called wide-area vigilance network (Han and Kamber, 2006). In our particular case, the integration of asynchronous and incomplete patient observations with partially unreliable data from sensors, require to go beyond separate data sources storage and database queries. Furthermore, relevant information extraction depends fundamentally on previously validated knowledge, instead of blind processing routines. Still, that knowledge base should evolve dynamically, depending on the patient changing condition, the respective variable volume of sensors data, its quality, and the corresponding variable user needs.

Different strategies could be applied to implement the proposed knowledge-based model. Some criteria for evaluating alternatives include among others: easy representation of prior knowledge statements, simple structure and maintenance of facts bases, straightforward definition of rules and reasoning mechanisms, portability, interoperability of involved software components, feasible scalable development, rich data visualization resources, and web services deployment. The decision may involve using different specialized programming languages in order to optimize specific modules, and in that case the model implementation would be an embedded engine of a larger distributed system.

Envisioned applications and services that could make use of the proposed data integration model are mainly related to wellness monitoring and disease management. Given the richness of possible integrated data analysis in these contexts, DS queries must be controlled, precise, and constrained to verified processing methods. For this reason, available validated data sometimes may not be relevant to answer a DS request, and the system should be capable, thanks to the knowledge-based model, to opportunely inform about such situation. However, knowledge definitions are flexible enough to be independently adapted at the three levels of the model, and completed with other rules.

6 CONCLUSIONS

Regardless of the multiple ways to collect and structure physiological sensors data and patient observations, integration of these data for combined analysis and interpretation requires embedded knowledge at different levels. We formulated a preliminary proposal towards the integration of complex relationships between various physiological sensors data and patient observations, in order to produce goal oriented processing workflows. The resulting knowledge-based data integration model is intended to define how to process PHR data, applying comprehensive dedicated knowledge clusters, to validate goal oriented inferences. Further work concerns the development and evaluation of a complete prototype including clinical data from the EHR, as well as the assessment of DS results for specific use cases. Being an initial attempt to formulate a general methodology, multiple open questions remain beyond the conception of a working system, like acceptance, performance, and usability.

ACKNOWLEDGEMENTS

This work was supported in part by Telecom Bretagne and in part by VTT and the Finnish Funding Agency for Technology and Innovation (Tekes) in the framework of the ITEA2/Care4Me project.

REFERENCES

Garg, M. K., Kim, D-J., Turaga, D. S., Prabhakaran, B., 2010. Multimodal analysis of body sensor network data streams for real-time healthcare. In *Proc. ACM SIGMM International Conference on Multimedia Information Retrieval*, Philadelphia, USA, pp. 469-478.

Halamka, J. D., Mandl, K. D., Tang, P. C., 2008. Early experiences with personal health records. *Journal of the American Medical Informatics Association*, vol. 15, no. 1, pp. 1-7.

Han, J., Kamber, M., 2006. Data Mining: Concepts and Techniques. Morgan Kaufmann Series in Data Management Systems, 2nd edition, pp. 467-488.

IEEE Health Informatics, 2009. Personal Health Device Communication, Part 11073-10441: Device Specialization - Cardiovascular Fitness and Activity Monitor. *IEEE Engineering in Medicine and Biology Society*, 11073™ Standard Committee, 85 pp.

Jovanov, E., Poon, C., Yang, G-Z., Zhang, Y. T., 2009. Body sensor networks: from theory to emerging applications. Guest Editorial. *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 6, pp. 859-863.

Kulkarni, P., Öztürk, Y., 2007. Requirements and design spaces of mobile medical care. *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 11, no. 3, pp. 12-30.

Martínez-López, R., Millán-Ruiz, D., Martín-Domínguez, A., Toro-Escudero, M. A., 2008. An architecture for next-generation of telecare systems using ontologies, rules engines and data mining. In *Proc. CIMCA, IAWTIC, and ISE International Conferences*, Vienna, Austria, pp. 31-36.

Stuntebeck, E. P., Davis II, J. S., Abowd, G. D., Blount, M., 2008. HealthSense: classification of health-related sensor data through user-assisted machine learning. In *Proc. 9th workshop on Mobile Computing Systems and Applications*, Napa Valley, USA, pp.1-5.

Tang, P. C., Ash, J. S., Bates, D. W., Overhage, J. M., Sands, D. Z., 2006. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. *Journal of the American Medical Informatics Association*, vol. 13, no. 2, pp. 121-126.