Risk-Associated Temporal Clinical pathways in T2D Patients

A. Dagliati[1], L. Sacchi[1], P. Decata[2], P. Leporati[2], S.Curti[2], D.Segagni[2], L. Chiovato[2], R. Bellazzi[1]
[1]*Department of Electrical, Computer and Biomedical Engineering, University of Pavia - Italy*
[2]*Fondazione S.Maugeri Hospital, Pavia – Italy*

Abstract. *This work presents a framework enabling the integration and analysis of a temporal dataset of Type 2 Diabetes (T2D) patients. The main idea underlying our approach is to use a suite of temporal and process mining methods so to define coherent group within the analyzed population and derive similar healthcare pathways. The presented system gathers T2D patients' clinical and administrative data coming from Fondazione S.Maugeri (FSM) hospital and the Pavia local health care agency (ASL). It can be accessed by caregivers throughout a hospital's EMR and, using a temporal data mining module, allows for the stratification of patients on the basis of temporal profiles. This work is part of the MOSAIC EU project, funded by the 7th Framework Program (http://www.mosaicproject.eu/).*

Introduction. Data collected in hospital information systems, together with data recorded for administrative purposes, have been used in many clinical and epidemiologic studies over many years [1, 2, 3]. Thanks to the integration of these streams of data, patients' and clinicians' actions can be monitored during acute events, but also for routine processes. In this work, we aim at exploiting data integration to develop a framework that, by applying advanced data mining methods, is able to extract patients' subgroups with similar temporal profiles. This research has been performed within the EU project MOSAIC. The project gathers T2D patients' data coming from three European hospitals and a local health care agency. It aims at providing clinicians with a new approach to follow up of T2D chronic population with diabetes by offering clinicians and healthcare managers a novel way to look at patients' clinical histories, using a dashboard that is easily accessible from the Hospital Information System. The data analysis functionalities available in this framework are centered on the temporal nature of the clinical data. In particular, they are aimed at stratifying the risk of developing T2D-associated complications on the basis of significant behavioral patterns. The system relies on the data stored in a multidimensional data warehouse, which contains both raw data and abstract temporal patterns related to the main variables of interest. It is provided with a graphical interface (the Dashboard) that enables to select interesting groups of patients on the basis of their risk profiles and mine them to extract the most frequent temporal histories in the selected group, exploiting a module devoted to process mining. This module allows representing the most frequent behaviors registered in the event histories of the patients.

Methods. *Integrating clinical temporal data coming from different sources.* One of the main challenges in constructing a data repository able to represent the entire medical histories of chronic patients is heterogeneity of the data sources. To cope with this complexity, we implemented a system based on a state-of-the-art open source Data Warehouse (DW) system: the Informatics for Integrating Biology and the Bedside (i2b2) [4]. The i2b2 clinical data warehouse (CDW) is installed at Fondazione Salvatore Maugeri hospital (FSM) [5]. The CDW has been designed to collect multidimensional data coming from the hospital information system and the local health care agency; it aggregates them in a format suitable both for temporal abstraction (TA) and mining purposes. The system can be accessed throughout a dashboard, integrated with the EMR, which allows the stratification of patients on the basis of temporal profiles and the visualization of the mined frequent pathways. Thanks to this modular framework, is possible to guide caregivers and stakeholders through a set of drill-down actions on specific clusters of patients, focusing the analysis on the temporal aspects related to the evolution of the disease.
Pre-processing with temporal abstractions. On top of the CDW, the system contains a pre-processing layer (TA module) that is able to extract specific abstract concepts from raw quantitative data through the temporal abstraction framework [6]. Within our system, the TA module aims at representing a subset of the clinical and behavioral information in the form of qualitative events. In our implementation we exploited TAs to represent temporal patterns of glycemic control, weight changes and diet improvements. These TAs have been defined through a close collaboration with the medical experts, who provided the domain knowledge about the most important temporal patterns they usually look for in the data. Taking as input the raw time series of time-stamped data, already uploaded in the i2b2 DW, the dedicated pre-processing module gives as results the intervals where specific behaviors hold.
Risk stratification. Following the mentioned drill-down approach, one of the first objectives was to stratify patients according to their risk profile and disease stage so to analyze their clinical history according to these classes. The system allows selecting groups of patients on the basis of their demographic data (age and gender), BMI, cardiovascular risk (CVR) or disease stage. We calculated CVR by means of the "Progetto Cuore" algorithm [7], which is derived from the Framingham index [8] and readapted to the Italian population. The application of the

algorithm results in a set of continuous values representing the CVR for each time a clinical measure was taken. The "Progetto Cuore" indicates appropriate clinical threshold to stratify the score in three classes of risk: High, Moderate and Low. The complexity of diabetes pathophysiology, and the relatively high number of behavioral factors influencing the disease, often needs more exhaustive stratification over time. These reasons led us to classify patients also according to the four levels of complexity in terms of complications and related hospitalizations. Thus the patients' cohort can be characterized considering the temporal dimension of the disease.
Process mining. The contacts of patients with the health care system develop over time, creating temporal sequences of events representing the flow of care. If TAs are able to synthetically represent behaviors and clinical states as qualitative events, process mining methods can extract chained events so to explain how patient's continuous interactions with the health care system and disease complication could be related. We developed, and integrated into the system, a methodology for mining complex histories of clinical events. The method works by finding frequent histories connecting a set of events contained in event logs. Event logs can both include events represented in form of raw data or as TAs. The possibility to mine histories on the basis of pre-processed and concise events allows transforming unstructured process in sequences of qualitative episodes. In this way one of the main source of complexity, which affects process mining in healthcare environment [9, 10], can be reduced. The developed methodology is inspired by the techniques of frequent pattern mining [11, 12], thus exploiting concepts as the support to determine how frequent is each history that is built at each step.

Results and Conclusions. The framework has been applied on the data collected at FSM for the EU project MOSAIC. By the end of the project (December 2015), 1000 patients will be managed with the proposed framework at FSM. Currently, complete clinical histories of 444 patients have been introduced in the DW. Patients undergo at least two visits per year for at least 5 years. Preliminary analyses have been carried out on the currently available patients. For these patients, the most frequent temporal patterns have highlighted an intervention on diet, where 34% of patients had achieved improved eating habits. Furthermore, there was evidence of improved glycemic control in 39% of patients. The majority of the patients (399) start to be followed at FSM before developing diabetes-related complications. Out of these patients, 84 result complications-free at the last DW update (June 2014). The pathways of care related to drug treatments have been analyzed. From these analyses, it resulted that all the patients of our cohort had started drug treatment (either with Metformin or Insulin) before starting to be followed at FSM. This is expected, since diabetic patients are first treated by general practitioners and only later by a specific diabetology service. The proposed system offers the opportunity to explore a population of T2D patients of a specific center while evaluating its most frequent temporal patterns. These patterns assist in identifying groups of patients with similar disease progressions, giving caregivers the possibility of promptly managing the most critical situations and may help decision makers in the allocation of hospital resources in cases of the most demanding patterns.

References

[1] R.S. Evans, J.F. Lloyd, L.A. Pierce, Clinical use of an enterprise data warehouse, AMIA Symp AMIA Symp, 2012 (2012), pp. 189–198
[2] Klann JG, McCoy AB, Wright A, Wattanasin N, Sittig DF, Murphy SN. Health care transformation through collaboration on open-source informatics projects: integrating a medical applications platform, research data repository, and patient summarization. Interact J Med Res. 2013
[3] Segagni D, Tibollo V, Dagliati A, Malovini A, Zambelli A, Napolitano C, Priori SG, Bellazzi R. Clinical and research data integration: the i2b2-FSM experience. AMIA Summits Transl Sci Proc. 2013 Mar 18;2013:239-40. eCollection 2013.
[4] Murphy S.N., Weber G., Mendis M., Gainer V., Chueh H.C., Churchill S., Kohane I. (2010) Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc 17(2) 124-30.
[5] Segagni D, Tibollo V, Dagliati A, Malovini A, Zambelli A, Napolitano C, Priori SG, Bellazzi R. Clinical and research data integration: the i2b2-FSM experience. AMIA Summits Transl Sci Proc. 2013 Mar 18;2013:239-40.
[6] Y. Shahar, "A framework for knowledge-based temporal abstraction", Artificial Intelligence 90 (1997) 79-133
[7] http://www.cuore.iss.it/eng/assessment/risk_assessment.asp
[8] http://www.framinghamheartstudy.org/
[9] R.S. Mans, W.M.P. van der Aalst, R.J.B. Vanwersch, A.J. Moleman, Process Mining in Healthcare: Data Challenges when Answering Frequently Posed Questions. (2012) Proceedings of BPM 2012 Workshops.
[10] Lang M, Bürkle T, Laumann S, Prokosch H , Process mining for clinical workflows: challenges and current limitations. Stud Health Technol Inform. 2008
[11] Agrawal R., Srikant R. (1995) Mining Sequential Patterns. In: Yu PS, Chen ALP (eds) Proceedings of the 11th International Conference on Data Engineering. IEEE Computer Society, pp. 3-14
[12] Zaki M.J. (2001) SPADE: An Efficient Algorithm for Mining Frequent Sequences. Machine Learning 42(1-2) 31-60