Application of Process Mining and Semantic Structuring Towards a Lean Healthcare Network

Dario Antonelli, Giulia Bruno

Application of Process Mining and Semantic Structuring towards a Lean Healthcare Network

Dario Antonelli, Giulia Bruno

Politecnico di Torino, Department of Management and Production Engineering,
Corso Duca degli Abruzzi 24, 10129 Torino, Italy
{dario.antonelli,giulia.bruno}@polito.it

Abstract. Modern healthcare systems are evolving towards a complex network of interconnected services. The increasing costs and the conversely increased expectations for high service levels leveraged the birth of healthcare monitoring activities and the proposition of numerous performance evaluation indicators. Generally, the adopted performance measures allow to draw a picture of quality, equity, appropriateness and efficiency of the medical care at different levels: caregiver, hospital, local health authority, region. The role of network organization and its impact on the performances is largely underestimated. It is difficult to build a Value Stream Mapping of the healthcare network because of the number and complexity of care and diseases followed. The study tries to overcome this issue. Starting from a database of the accesses to the services in a local health agency, the activity flow diagram is produced by using a process mining software, Disco. A knowledge structured by means of an ontology allows to describe the logic behind the health service provision. The resulting process flow chart is the base for the identification and amendment of redundant and non value added flows among services.

Keywords: Process mining, healthcare network, ontology.

1 Introduction

Worldwide there is an increasing number of initiatives aimed at introducing a standardized and centralized information management in healthcare organizations (hospitals, medical centers, drugstores) through digitalization of medical data. It is proven that digital tools like the electronic health records provide benefits to both patients and physicians by improving health care efficiency [1]. The availability of the medical history of the patient's accesses to medical centers will allow both the physicians to express meaningful analyses at the patient level (e.g., searching for similar patients based on their medical history or predicting future events in care pathway) and the system manager to operate at the organizational level (e.g., discovering which are the most accessed resources or which are the anomalous managements of patients) [2]. Therefore the analyses are intended to improve the quality of services offered to citizens while to reducing costs and wastes. Since the data volume is very high, and it is expected to grow dramatically in the years ahead, for healthcare organizations it is vitally important to acquire the available tools,

infrastructure, and techniques to leverage big data effectively. The enormity and complexity of collected medical data present great challenges [3].

The format of medical data and the data base structure is a serious issue preventing their use for operation managements applications. Data describe the healing pathway of the patient and do not give direct evidence of the process flow in terms of process times, queues, unproductive times, etc. To have a better insight of the process flow, data should be elaborated by process mining.

Process mining aims at extracting process knowledge from event logs which may originate from different kinds of systems, e.g., enterprise information systems or hospital information systems [4]. Typically, these event logs contain information about the start/completion of process steps together with related context data (e.g. actors and resources). Previous works addressed the problem of analyzing single entities (e.g., a hospital department) by applying techniques of data/process mining and simulation [4,5,6,7]. In these cases, existing processes are compared with the medical knowledge to determine whether the pathway of a patient within the structure is correct. However, the analysis of a single entity is limitative, because it does not consider the previous history of the patient and thus it is not able to evaluate the quality of the healthcare system as a whole.

Before applying the process mining techniques, it is important to merge data of different nature to collect the patient movements inside the network. In order to merge heterogeneous data, we need a controlled vocabulary in term of set of ontologies to give data a meaning despite the different original data structure. After the merge of data, it is possible to apply a process mining tool, to automatically reconstruct the movements of patients inside the network. This analysis is useful both to analyse the changes in patient flows depending on age or gender of patients and to discover the bottleneck and waste of the system, toward a lean restructuring of the healthcare process.

The overall objective of this paper is to contribute in giving medical managers an accurate and deep understanding of the healthcare network functioning. There are several contributions: the first contribution is the definition of an ontology of the healthcare network: general concepts and relationships. Then, starting from the data organized in the model, the process mining analysis is used to extract information for the network evaluation. To make the methodology more concrete, the real data collected by an Italian Healthcare Territorial Agency (HTA) is exploited as a case study. The preliminary results we obtained proved the applicability and the usefulness of the proposed approach.

2 State of the Art

There is an evolving trend in recent years that has modified the healthcare system from a few nodes hospital based organization to a branched network of service centers spread on the territory [8]. Furthermore the approach to disease treatment is now based on integration among the different agents of the care system. Therefore there is a convergent trend towards a network of centers that delivery integrated care [9].

The integration of different aspects of healthcare system has been a subject of study by many authors in the field of operations management [10-12] and several authors highlighted the benefits of integration both in terms of quality of care as in terms of lean organization [13-15]. Drawbacks are equally reported but are mainly due to lack of cooperation and commitment of the healthcare personnel [16-17].

The analysis of healthcare network by using the operations management models is not effective as in the industrial environment where the process activities and the flow of material are utterly defined. The production is substituted by care pathway, that is far less deterministic. Products are substituted by patients that are free to move along the process flow at their will (or even to abandon care and consequently interrupt the flow of activities). To extract performance variables to be a guidance in the process management, it is necessary to have recourse to other methods, like data mining techniques [18,19]. If time is not a monitored output, the focus is on pattern extraction in order to detect the most frequent medical treatments undergone by patients [20-22]. These techniques do not give a comprehensive view of the processes in act in the healthcare systems.

On the other hand, process mining can be applied to healthcare data to identify the processes and derive meaningful insights from the complex temporal relationships existing between activities and resources involved in processes [23]. For example, process mining was applied to a hospital emergency service in a public hospital in Portugal to identify regular behavior, process variants, and exceptional medical cases in [24], and to a hospital in Belgium to model the activities related to breast cancer treatment in [25]. It was also applied to perform a comparative analysis across four hospitals in Australia [23]. We follow these previous works, but we applied process mining to extract the movements of patients among the different centers of a Healthcare Territorial Agency (HTA), in order to perform a comparative analysis among different patient segments.

3 Healthcare Network Ontology

The ontology used in our work has two aims: firstly to model the entities and relationships needed to collect data coming from a healthcare network, and secondly to provide the controlled vocabulary in order to merge data coming from heterogeneous sources. For the first aim, we define a model as a UML class diagram, after the carefully analysis of the data available in Italian HTA. For the second aim, we reused the controlled vocabularies previously developed relevant for our purpose, i.e., ICD10 [26] for disease classification, MDC [27] for the major diagnostic categories, DRG [28] for diagnosis related group and ATC [29] for drug classification.

The UML class diagram representing the healthcare data collected by the HTA is reported in Figure 2. A service is any kind of healthcare service provided to a citizen, form examinations to drugs to hospitalizations, while a user is any person who access the healthcare system. A prescriber is a physician who can do prescription of services to the users, while a provider is a structure that provides one or more services. For

each provider, the list of services it provides is known. A prescription represents the information of the specific services that a prescriber prescribes to a user, and a provision stores the information of services provided by providers to the users. This is a very general model that is valid for all the healthcare territorial agency [30].

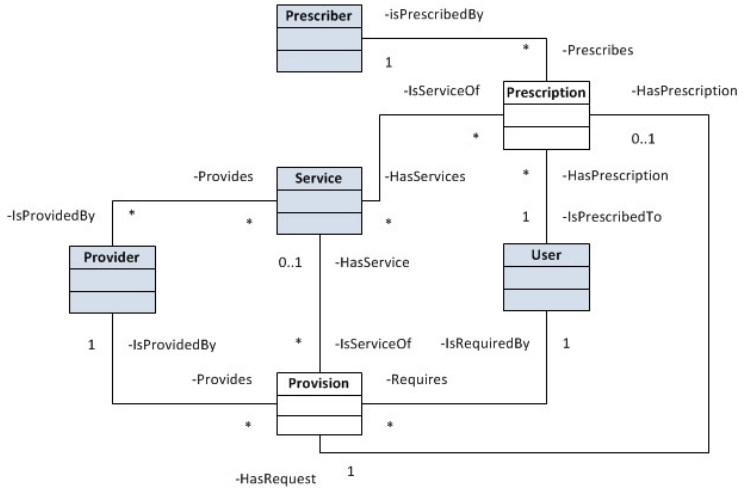

Fig. 1. Main entities of the healthcare network ontology.

This UML class diagram can be represented in the form of an ontology, where each entity is a class, with the corresponding relationships linking the classes. Then, for each entity, it is possible to specify the hierarchical tree of concepts at increasing detail levels which represent the knowledge related to the services, providers, users and prescribers of the HTA [31]. This information constitutes the HTA ontology that will be used to select the data of interest for the following analysis.

In a HTA, three types of providers can be identified: (i) the medical centers without hospitalization capacities, (ii) the hospitals and (iii) the drugstores. The existence of three providers determine the existence of three type of services, i.e., (i) the examinations, (ii) the hospitalizations and (iii) the drugs. For each kind of service, a further hierarchy can be defined. The examinations can be further specified based on the medical branch they belong, the hospitalizations based on the diagnosis-related group (DRG) and the major diagnostic category (MDC), the drugs based on the anatomical therapeutic chemical (ATC) classification. The prescribers can be divided in two categories, i.e., the general physician and the specialist physician. The users can be divided based on their pathology, represented by their exemptions code. These hierarchical structures are shown in Fig.2.

The information stored in the ontology is used to extract from the database the subset of data relevant for the analysis.

Fig. 2. Hierarchical levels of the healthcare network ontology, built with the Protégé ontology editor (http://protege.stanford.edu).

4 Process Mining

The goal of process mining is to extract process models from event logs. It includes a family of a-posteriori analysis techniques exploiting the information recorded in the event logs. Typically, these approaches assume that it is possible to sequentially record events such that each event refers to an activity (i.e., a well-defined step in the process) and is related to a particular case (i.e., a process instance). Furthermore, some mining techniques use additional information such as the performer or originator of the event (i.e., the person/resource executing or initiating the activity), the timestamp of the event, or data elements recorded with the event (e.g., the size of an order).

Process mining was already applied in healthcare systems, trying to answer the following questions: (i) what are the most followed paths and what exceptional paths are followed? (ii) are there differences in care paths followed by different patient groups? Standard paths are the activities that are typically executed by patients and the order of them. Exceptional paths are anomalous activities due to the way of working of medical specialists or related to specific patient characteristics or not. The comparison of the behavior of different patient groups is another interesting issue. This comparison may not only be interesting for patient groups within a hospital but also for similar patient groups in different hospitals.

In this paper, we applied process mining techniques to reconstruct the actual movements of patients among the providers of a Healthcare Territorial Agency. To this aim, we exploit the software Disco (https://fluxicon.com/disco/), a commercial process mining tool, freely available under an academic license. It exploits the Fuzzy Miner algorithm for process mining [32], which uses significance/correlation metrics to interactively simplify the process model at desired level of abstraction.

The core functionality of this tool is the automated discovery of process maps by interpreting the sequences of activities in the imported log file. According to the process mining paradigm, at least the following three elements have to be identified in the file log: case id, activity, and timestamp. In our analysis, since we are interested in analyzing the movements of patients among the providers, the case id corresponds to the patient id. The activity is an event of the process, thus in our case it is the provider visited by the patient. The timestamp is the date in which the patient visits the provider.

5 Data Analysis

The database considered as a case study contains data collected by an Italian HTA in the 2007-2012 years. Indeed, the HTAs collect data about the supplied services for cost accounting. Supplied service data represent an essential resource in planning and monitoring the Regional Healthcare System. Data for the analysis of diagnostic pathway are selected from the data warehouse: Hospital Discharge Records, Ambulatory Care Records, Emergency Department Records, Ambulatory Care Records. They are composed of personal data and clinical data section, so that it was

possible to merge anonymized personal and clinical data, to collect all databases in a single MySQL database by means of a PHP routine that automatically import data.

In order to focus our study on a specific pathology, we extracted from the database the log file of all the patients suffering from asthma for year 2007. The aim is to analyze the mobility of patients across the different medical centers placed on the territory. The data refer to a total of 451 asthma patients who accessed medical centers, divided in 207 males and 244 females. Regarding the age of patients, 155 are younger than 36 years, 333 are between 36 and 65 years old, and 63 are older than 65 years.

Before attempting any process mining it was necessary to preprocess the data in order to simplify them, by skipping unnecessary low level activities and by merging the significant low level activities in singular high level ones. We used the knowledge deriving from the ontology of section 3 to preprocess the data. An example is the log describing the accesses to a laboratory for executing analyses. Whether the analyses are executed on the same patient, the same day, for the sake of operations management, they can be safely converted in a singular access to one activity.

After preprocessing, we performed two alternative segmentations of the dataset, one based on gender and the other one based on age, in order to highlight the effect of both factors on the process flow diagram. Each segment was imported in Disco to perform the process mining. For readability reasons, only the 30% of activities involved and the 10% of the path between activities are shown in the results. These percentage grant to cover at least the 90% of the data, since there are many centers that are accessed only one or two times, and thus are not relevant for the analysis.

From the analysis of the first segmentation, it can be noticed that the some similarities exist between the two obtained graphs (Fig.3 and Fig.4). The first one is that in both cases there is a “hub center” which is the most accessed by both genders, since the majority of patients perform examinations only in this center. Another similarity is that the processes usually do not involve more than two different centers (the hub center and one of the other centers). Despite these similarities, the graphs also show some differences. First of all, the number of visited centers is different: male patients visit 12 different centers, while female patients visit 21 different centers, thus showing a higher mobility of female patients.

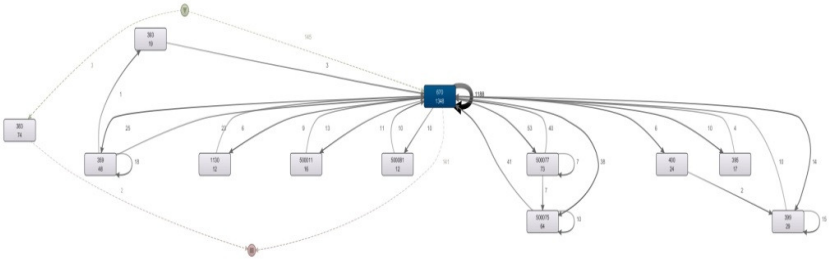

Fig. 3. Process flow extracted by Disco on the male segment, all the ages considered.

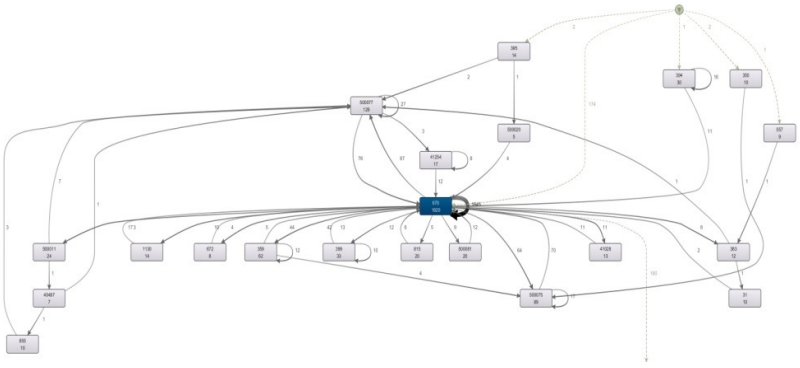

Fig. 4. Process flow extracted by Disco on the female segment, all the ages considered.

Also the frequency of accesses is different. Fig.5 reports the number of accesses to each center for the male segment and the female segment (the hub center is not reported since its accesses are significantly higher than the others: 1348 for male and 1820 for female). Furthermore, the majority of male processes follow the path "hub center - other center - hub center", while the female process follow various paths, starting for a center, then passing the hub center, followed by another center.

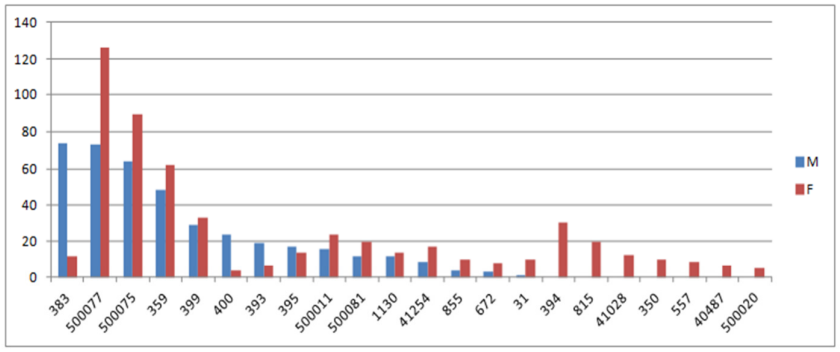

Fig. 5. Different numbers of accesses to medical centers by male patients (blue bar at the left) and female patients (red bar at the right).

From the analysis of the second segmentation (Fig.6-Fig.8), it can be noticed that the process extracted from the adult segment is the one which includes the higher number of different centers visited by patients and also the longest pathways involving different centers. The process extracted from the senior segment include less centers and shortest paths. This can be due to the low mobility of elder patients with respect to the others. The number of accesses to medical centers of each of these three segments is reported in Fig. 9.

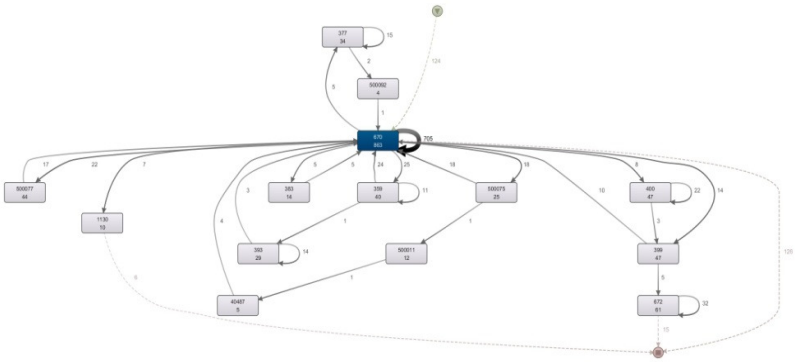

Fig. 6. Process flow extracted by Disco on the young segment for all the genders.

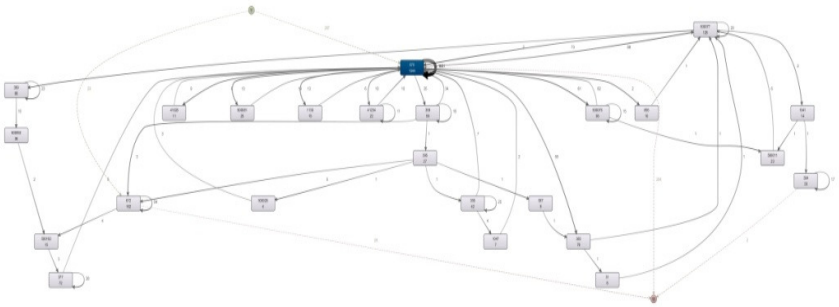

Fig. 7. Process extracted by Disco on the adult segment for all the genders.

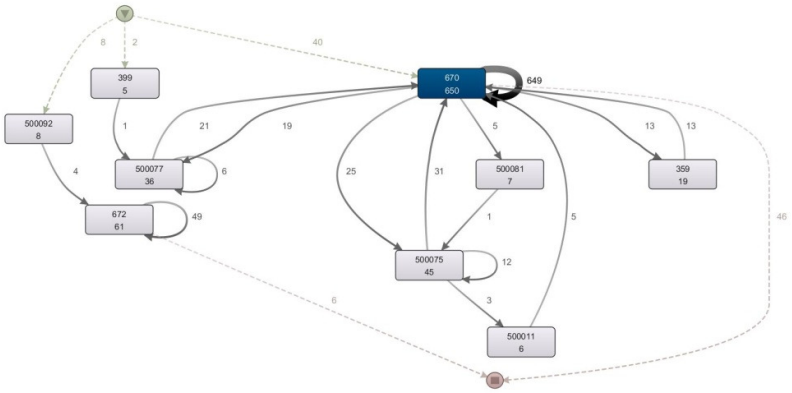

Fig. 8. Process flow extracted by Disco on the senior segment for all the genders.

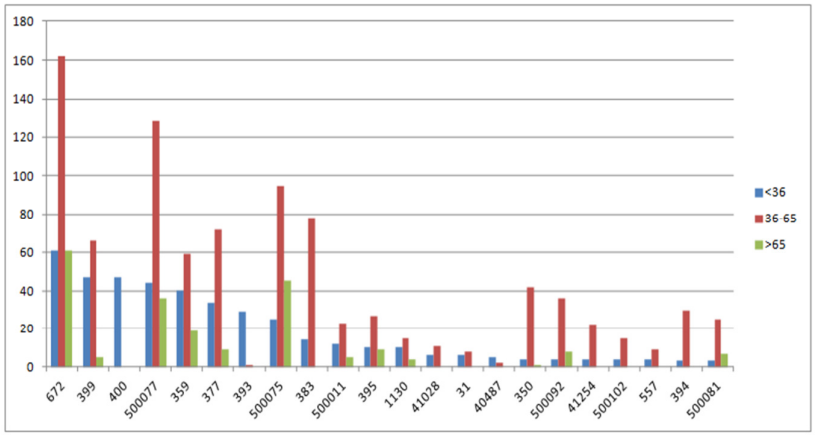

Fig. 9. Different numbers of accesses to the medical centers by young patients (first bar), adult patients (second bar) and senior patients (third bar).

6 Conclusion

The evaluation of a health network is a complex task because the management of patients is done by entities not controlled by a central unit, with the result that guidelines for the evaluation of the diagnostic and therapeutic pathways through the network are not available.

This paper aims at adapting tools and technologies derived from other research fields and using them firstly to obtain a model of a healthcare network, and secondly to perform a meaningful analysis of the mass of data produced by a healthcare network. The obtained results can be useful for healthcare managers to inform them clearly about the status of services under their responsibility, and to suggest improvements to system inefficiencies. It is also useful to evaluate the degree of collaboration among the different entities of the network.

We are currently working on extending the analysis to involve different kind of entities, and to explore different mining algorithms to refine the extracted model.

References

1. Hillestad R., Bigelow J., Bower A., Girosi F., Meili R., Scoville R., Taylor R.: Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. Health Aff (Millwood), 24(5), pp. 1103-17 (2005)
2. Mans R.S., van der Aalst W.M.P., Vanwersch R.J.B., Moleman A.J.: Process Mining in Healthcare: Data Challenges when Answering Frequently Posed Questions. Lecture Notes in Computer Science, 7738, Springer Berlin Heidelberg, pp. 140-153 (2013)

3. Sun J., Reddy C. K. Big data analytics for healthcare, Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1525-1525 (2013)
4. Mans R.S., Schonenberg M.H., Song M., van der Aalst W.M.P., Bakker P.J.M.. Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital, BIOSTEC 2008, CCIS 25, pp. 425–438, (2008)
5. Coelli, F. C. , Ferreira, R. B., Almeida, R. M.V.R., Pereira, W. C. A.: Computer simulation and discrete-event models in the analysis of a mammography clinic patient flow, Computer Methods and Programs in Biomedicine, Vol.87 (3), pp.201--207 (2007)
6. Di Leva, A., Femiano, S.: The BP-M* Methodology for Process Analysis in the Health Sector, Intelligent Information Management, 3, pp.56--63 (2011)
7. Cardoso, E.C.S, Guizzardi, R.S.S., Almeida, J.P.A.: Aligning goal analysis and business process modelling: a case study in healthcare, International Journal of Business Process Integration and Management, Vol.5(2), pp.144 -- 158 (2011)
8. Wan, T.T.H., Wang, B.B.L. Integrated Healthcare Networks Performance: A Growth Curve Modeling Approach, Health Care Management Science, Vol. 6, pp.117--124 (2003)
9. Lenz, R., Reichert, M.: IT support for healthcare processes - premises, challenges, perspectives, Data & Knowledge Engineering, Vol.61(1), pp.39--58 (2007)
10. Scott, W.R.: The organization of medical care services: Toward an integrated theoretical model, Medical Care Review, Vol.50(3), pp.271--302 (1993)
11. Ahgren, B., Axelsson, R.: Evaluating integrated health care: a model for measurement, International Journal of Integrated Care, 5 (2005)
12. Tjerbo T., Kjekshus L.. Coordinating health care: lessons from Norway, International Journal of Integrated Care, Vol.5 (2005)
13. Axelsson R., Bihari A. S.: Integration and collaboration in public health: a conceptual framework, International Journal of Health Planning and Management, Vol.21(1), pp.75--88 (2006)
14. Bazzoli, B.J., Chan, B., Shortell, S., D'Aunno, T.: The financial performance of hospitals belonging to health networks and systems, Inquiry, Vol.37(3), pp.234--252 (2000)
15. Provan K.G., Milward H.B.: Do networks really work? A framework for evaluating public-sector organizational networks, Public Administration Review, Vol.61(4), pp.414--423 (2001)
16. Hurtado M.P., Swift, E.K., Corrigan, J.M.: Crossing the quality chasm: a new health system for the 21st century, National Academy Pres (2001)
17. Cesarini, M. Mezzanzanica, M. Cavenago, D.: ICT Management Issues in Healthcare Coopetitive Scenarios, Information Resources Management Association International Conference (2007)
18. Lin, F., Chou, S., Pan S., Chen, Y. Mining Time Dependency Patterns in Clinical Pathways. International Journal of Medical Informatics, Vol.62, pp.11–25 (2001)
19. Batal, I., Fradkin, D., Harrison, J. Moerchen, F., Hauskrecht, M.: Mining Recent Temporal Patterns for Event Detection in Multivariate Time Series Data. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (2012)
20. Baralis, E., Bruno, G., Chiusano, S., Domenici, V.C., Mahoto, N.A., and Petrigni, C.: Analysis of medical pathways by means of frequent closed sequences, Lecture Notes in Computer Science, pp.418--425 (2010)
21. Antonelli, D. Baralis, E., Bruno, G., Chiusano, S., Mahoto, N.A., Petrigni, C.: Analysis of diagnostic pathways for colon cancer. Flexible Services and Manufacturing Journal, 24(4), pp. 379--399 (2011)
22. Antonelli, D., Baralis, E., Bruno, G., Cerquitelli, T., Chiusano, S., Mahoto, N.A.: Analysis of diabetic patients through their examination history, Expert Systems with Applications, Vol. 40(11), pp:4672--4678 (2013)

23. Partington A, Wynn M, Suriadi S, Ouyang C, Karnon J, Process Mining for Clinical Processes: A Comparative Analysis of Four Australian Hospitals, ACM Trans. Manage. Inf. Syst., 5(4), 1-18, 2015
24. Rebuge A., Ferreira D.R.: Business Process Analysis in Healthcare Environments: A Methodology Based on Process Mining. Information Systems, 37(2), (2012)
25. Poelmans J., Dedene G., Verheyden G., Van der Mussele H., Viaene S., Peters E.: Combining business process and data discovery techniques for analyzing and improving integrated care pathways. In Proceedings of the International Conference on Data Mining (ICDM'10), Vol. 6171. Springer, 505–517 (2010)
26. ICD10: http://www.who.int/classifications/icd
27. MDC: http://health.utah.gov/opha/IBIShelp/codes/MDC.htm,
28. DRG: http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MedicareFeeforSvcPartsAB/downloads/DRGDesc05.pdf,
29. ATC: http://www.whocc.no/atc_ddd_index
30. Antonelli D., Bruno G.: Healthcare network modeling and analysis, Collaborative Systems for Smart Networked Environments, IFIP Advances in Information and Communication Technology, Springer Berlin Heidelberg, Vol. 434, pp. 691-698, (2014)
31. Antonelli, D., Bruno, G., Bellomo, D., Villa, A.: Evaluating Collaboration Effectiveness of Patient-to-Doctor Interaction in a Healthcare Territorial Network, Collaborative Networks in the Internet of Services, Springer Berlin Heidelberg, pp. 128--136 (2012)
32. Günther, C.W., van der Aalst, W. M. P.: Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics, Lecture Notes in Computer Science, 4714, pp 328-343 (2007)