Healthcare Network Modeling and Analysis

Dario Antonelli and 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. A number of initiatives are nowadays in act to reorganize health systems and increase the interconnection of services offered to citizens to build a collaborative care network that maximizes service quality and minimizes costs. To achieve these purposes, a non standard quality assessment must be performed, as models for the analysis and evaluation of healthcare systems considered from the point of view of the entire network instead of single nodes have not been proposed yet. This paper aims at describing how to obtain a rigorous modeling and a meaningful analysis of the mass of data produced by a healthcare network. This analysis can be useful to inform the healthcare managers about the status of services under their responsibility and the degree of collaboration among entities of the network. The final outcome of these activities will be an improvement in the quality of service perceived by patients.

Keywords: Healthcare network, collaboration, performance indicators, UML.

1   Introduction

A promising trend in healthcare pushes the interconnection of the offered services to reduce diseconomies of scale. It implies the need for new analysis methods. Healthcare systems consist of a variety of providers (e.g., medical centers, hospital departments, emergency rooms) and prescribers (e.g., general and specialist physician). The entire network must be assessed, considering the relationships among these entities. Several works address the problem of analyzing single units by applying data/process mining and simulation [1,2]. However, the analysis of a single entity is limitative, because it cannot take into consideration the previous clinical history of the patient. For example, if a patient is addressed by the general practitioner to a specialist physician, and these in turn directs the patient to a specialist physician of another branch, it can be seen as an inefficiency of the system, because the general practitioner could have sent the patient directly to the second specialist. These anomalous behaviors are not detected by the analysis of individual service centers, but will be detected by analyzing patients' pathways through the network of services.

A first reason for which these analyses are not performed is the lack of a formal model to integrate information from different sources. Healthcare agencies maintain large administrative databases, which contain demographic information as well as clinical information such as diagnoses and medical procedures. While these data are

collected for administrative data processing, taking advantage of them for other purposes would be highly desirable [3]. Furthermore, another reason of the incomplete analysis of data is the limited application to the healthcare of some ICT tools developed to perform data analysis in industrial environments. This paper aims at providing a methodology to fruitfully exploit data collected by healthcare networks. The first contribution is the definition of a metamodel of the healthcare network, then the analyses useful to extract information for the network evaluation are discussed. To make the methodology more concrete, real data collected by an Italian Healthcare Territorial Agency (HTA) are exploited as case study.

2   Related Works

Healthcare systems evolved from systems with few hospitals into integrated care delivery systems [4]. An integrated care system includes multiple structural functional aspects of healthcare organizations. The recent trend towards healthcare networks and integrated care even increases the need to effectively support interdisciplinary cooperation along with the patient treatment process [5]. In literature, significant attention has been devoted to the evaluation of the level of integration of healthcare services [6,7] and to the mining of the potential benefits of such integrated systems [8,9], such as a better quality of care, better services, more accessibility, more efficient operations, and reduced unit costs. On the other hand, severe problems were reported in applying advances in information technology to improve administrative and clinical processes [10], not only because of technical issues, but also due to the lack of cooperation among the healthcare network actors [11].

   Despite its limitations, IT can provide positive effects to two kinds of processes: organizational processes (e.g., medical order entry and result reporting) and medical process (e.g., diagnostic and therapeutic procedures to be carried out for a particular patient) [5]. Since the effect on the first kind of process has been widely investigated (i.e., by reducing patient waiting times and costs), this work focuses on describing how IT solutions can be exploited to evaluate the medical processes in a healthcare network. Among data analysis methods, the use of data mining has taken great attention due to the large amount of generated data [12]. In [13], authors worked to identify the risk factors of a disease by using open source data mining tools, particularly association rule mining and classification techniques. A similar technique was proposed in [14] for mining association rules to identify possible side-effects of using multiple drugs during pregnancy. An approach of mining time dependency patterns is discussed in [15], where hospital paths are predicted for new patients. In [16], the authors proposed a temporal pattern mining framework to find predictive patterns for detecting and diagnosing adverse medical conditions associated with a disease. Pattern extraction techniques were already proposed in [17,18,19] to detect frequent medical treatments.

   Present study advances previous works along two main dimensions: (i) it proposes a metamodel of the healthcare network to organize healthcare data in view of the following analyses, and (ii) since it considers the healthcare system as a network

instead of single nodes, it proposes to extend the analysis also considering the interactions among nodes.

3   Methodology

The proposed approach to manage healthcare systems considered as a whole network instead of individual nodes is organized into three main phases, as shown in Figure 1. In "Data modeling" inputs come from the data collected by the healthcare control center. In "Data analysis", two kinds of analyses are proposed: a vertical analysis, i.e., fixed a node, analyzing the frequency of its children, and a horizontal analysis, i.e., fixed a level, extract the sequential pathways involving multiple nodes of that level. By combining the two kinds of analyses, an overview of the behavior of the network in terms of services accessed and pathways frequently followed by patients is provided. In the "Data evaluation" phase, the results emerged from the previous phase are evaluated with the help of domain experts.
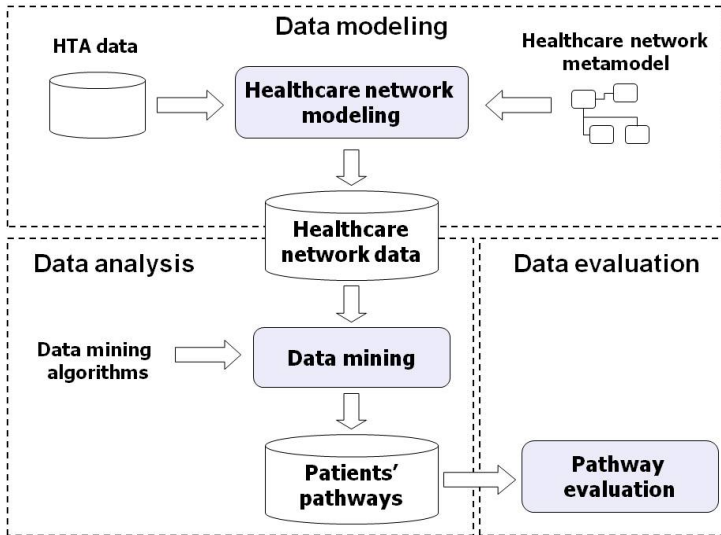

Fig. 1. Proposed approach to model, analyze and assess a healthcare network

4   Data Modeling

The metamodel of healthcare network was defined as a UML class diagram, as shown in Figure 2. In a healthcare system there are four main entities: Service, Prescriber, Provider, and User. A service is any kind of healthcare service provided to a citizen, from examinations to drugs to hospitalizations, while a user is any person who access the 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. Two other entities are needed to represent the interaction and collaboration between these four elements, i.e., Prescription and Provision. A prescription contains the information of the specific services that a prescriber prescribes to a user, and a provision store the information of services provided by providers to the users. A user can be connected with a provision without any prescription, i.e. accessing an emergency room.

Once the entities and relationships in the metamodel have been defined, the model of a specific network is created as an instance of the metamodel. To this aim the case of a Italian healthcare territorial agency (HTA) is considered, which includes all users, providers and prescribers in a certain geographical area [20]. In a HTA, three types of providers can be identified, i.e., the medical structures without hospitalization capacities, the hospitalization structures and the drugstores. There are three services: (i) examination, (ii) the hospitalizations and (iii) the drugs distribution. Data regarding the prescriptions and provisions are routinely collected by the HTA.
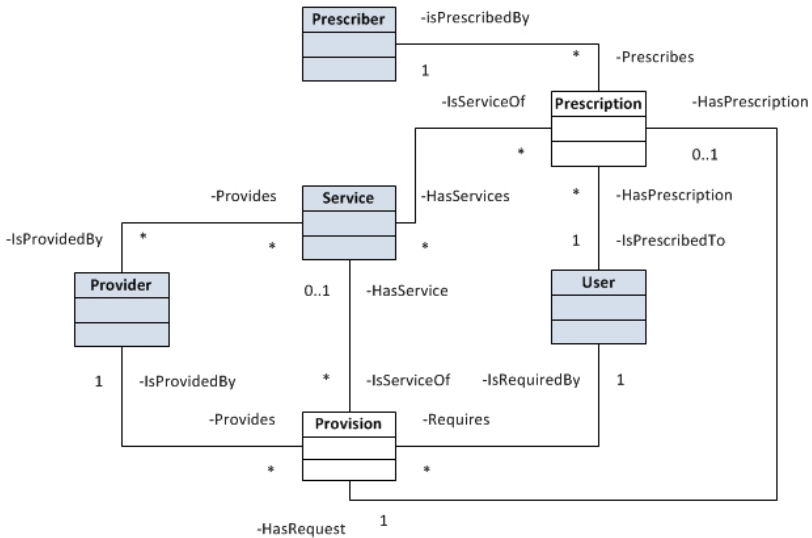

Fig. 2. Metamodel of the healthcare network

5   Data Analysis

Figure 3 shows an example of a hierarchy of service concepts. Starting from the most general concept of service, the first specialization is among the three types of services previously described, i.e., the examinations, the hospitalizations and the drugs. Then for each service type, other specializations can be done, e.g., the one based on the medical branch to which the examinations belong. In this way, a hierarchy of N levels is produced, where the N-th level is composed by the instances, i.e., the concrete services.

With respect to this data representation, two different analyses can be done. The first can be defined as a "vertical" analysis, i.e., fixed a node of the hierarchy,

analyzing the frequency on which the patients access the services of its children. For example, if the healthcare manager is interested in analyzing the services most accessed in the network, he/she can perform a vertical analysis directly on the Service node. If he/she is interested in seeing how patients accessed the different medical branches of examinations, the vertical analysis can be performed on the Examination node. As a result, the number of patients that accessed each medical branch (e.g., Laboratory, Ophthalmology and Radiology in our example) is computed. This analysis can be useful to compare the behavior of different patients' groups in terms of services accessed, or to compare the accessibility of one service with respect to the others. The second kind of analysis, which is generally not performed by healthcare organizations, is a "horizontal" analysis, because it considers simultaneously different nodes at the same hierarchical level (e.g., level L1 in Figure 4). This analysis aims at computing the frequency of the movements of patients among the network entities through the extraction of sequential patterns. To accomplish this task, the data have to be transformed in a sequence dataset containing the temporally ordered activities of patients in the network.
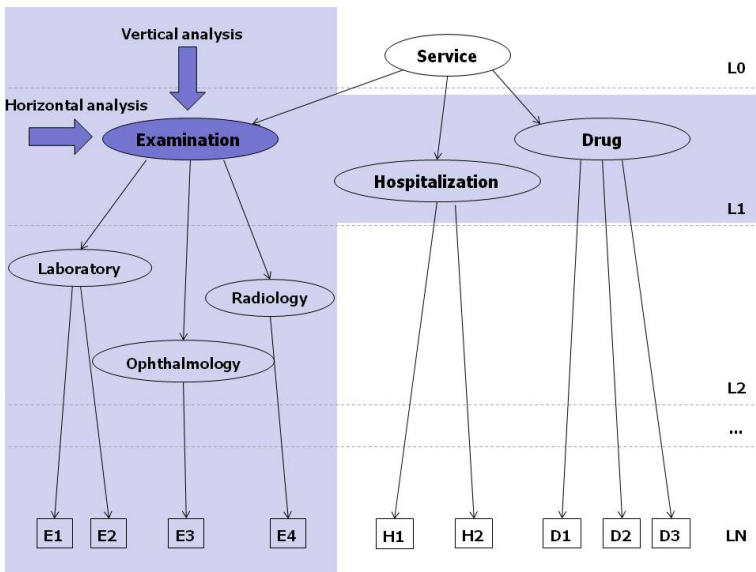

Fig. 3. Horizontal and vertical analysis on healthcare network data

More formally, a sequence dataset is a collection of tuples (Pid, S), where Pid is the patient identifier and S is the temporal list of sets of activities done by the patient. An activity set is a set of activities done in the same date, and it is indicated by curly brackets. Once the data were structured according to this format, data mining algorithms can be used to reconstruct the frequent pathways of patients in the network. The percentage of tuples (i.e., patients) that contain a sequence is called the support of the sequence and is indicated as sup(S). A sequence S is called frequent if its support is above a specified support threshold. To avoid the generation of

redundant information and compactly represent the solution set, only the frequent sequences that are not subsequences of other sequences with the same frequency are extracted. Such sequences are named closed sequences [21]. The frequent closed sequences are extracted by means of the BIDE algorithm.

6   Data Evaluation

The dataset considered as a case study contains data collected by an Italian HTA all along one year. In the database, there is a number of 204,138 patients and a total of 341,373 provisions of services. Particularly, the percentage of services accessed are the following: 62.0% examinations, 30.3% drugs and 7.7% hospitalizations. The dataset was received in an anonymized form directly from the health territorial agency, and was stored and analysed on a dedicated machine not accessible from outside exclusively by authorized personnel.

The application of the vertical analysis to the Examination node in the L1 level identifies the number of patients which accessed examinations in each medical branch. This analysis can be useful to compare the behaviors of patients with different pathology. For example, by applying this analysis separately to patients who suffer from diabetes and who suffer from asthma, results show that the two groups of patients accessed different branches from each others. Patients who suffer from diabetes, access more frequently the branches of Laboratory, Ophthalmology, Cardiology, Radiology, Dietetics, Neurology and Vascular surgery. Instead, patients who suffered from asthma, access more frequently the Pneumology, the Laboratory, the Radiology, the Dermatology and the Allergology branches. The vertical analysis can also be done at a lower hierarchical level. For example, the application of the vertical analysis to the Laboratory branch (L2) shows the specific examinations most accessed by patients.

The horizontal analysis can reveal more interesting information because it considers the chronological order of activities performed by patients. If applied to nodes of level L1, it reveals the frequent patterns among types of services. From this analysis emerges that 100% of the diabetic patients accessed a medical center in the considered year, 79% accessed two times and 61% three times. The frequency of accesses to drugstores is a little bit low, in fact only 76% of patients accessed one time, 68% two times and 60% three times. The analysis at this level is useful to estimate the accessibility of patients to the different types of services, but it gives not details on treatments performed. Thus, the horizontal analysis was repeated at a lower level (i.e., LN). The analysis of extracted frequent patterns, reported in Table 1 for diabetic patients, reveals that most frequent examinations are the base tests repeated to monitor the concentration of sugar in the blood, such as Glucose, Venous blood sample, Capillary blood sample, and Urinalysis. After these, the most frequent examinations are the ones concerning cardiovascular complications and the ones for determining the liver health. Among sequences of length greater than 1, the most frequent sequences show that Glucose is repeated two times (58% of patients), three times (32%), or four times (15%) during the considered period. These patterns are

coherent with the medical knowledge, but the frequency with which they are performed is lower than expected. The fact that around 15% of diabetic patients did not control the blood glucose level along a year can reveal some problem in patient management and need a further investigation of causes.

Table 1. Results of the horizontal analysis on the LN level of the Examination node (for diabetic patients)

Frequent sequences - length 1	*Support*
{Glucose}	85%
{Venous blood sample}	79%
{Capillary blood sample}	75%
{Urinalysis}	75%
…	
{Fundus oculi}	27%
Frequent sequences - length 2	*Support*
{Glucose}{Glucose}	58%
{Capillary blood sample}{Glucose}	54%
Frequent sequences - length 3	*Support*
{Glucose}{Glucose}{Glucose}	32%

7   Conclusion

The evaluation of a health network is a complex task because the patients have autonomy in the choice of which nodes (services) to visit. 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 extracted knowledge represents a picture of the actual processes occurring in the 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 different entities of the network

References

1. 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. In: Fred, A., Filipe, J., Gamboa, H. (eds.) Biomedical Engineering Systems and Technologies. CCIS, vol. 25, pp. 425–438. Springer, Heidelberg (2008)
2. Di Leva, A., Femiano, S.: The BP-M* Methodology for Process Analysis in the Health Sector. Intelligent Information Management 3, 56–63 (2011)

3. Tsechansky, M.S., Pliskin, N., Rabinowitz, G., Porath, A.: Mining relational patterns from multiple relational tables. Decision Support Systems 27, 177–195 (1999)
4. Wan, T.T.H., Wang, B.B.L.: Integrated Healthcare Networks Performance: A Growth Curve Modeling Approach. Health Care Management Science 6, 117–124 (2003)
5. Lenz, R., Reichert, M.: IT support for healthcare processes - premises, challenges, perspectives. Data & Knowledge Engineering 61(1), 39–58 (2007)
6. Ahgren, B., Axelsson, R.: Evaluating integrated health care: a model for measurement. International Journal of Integrated Care 5 (2005)
7. Tjerbo, T., Kjekshus, L.: Coordinating health care: lessons from Norway. International Journal of Integrated Care 5 (2005)
8. Axelsson, R., Bihari, A.S.: Integration and collaboration in public health: a conceptual framework. International Journal of Health Planning and Management 21(1), 75–88 (2006)
9. Bazzoli, B.J., Chan, B., Shortell, S., D'Aunno, T.: The financial performance of hospitals belonging to health networks and systems. Inquiry 37(3), 234–252 (2000)
10. Hurtado, M.P., Swift, E.K., Corrigan, J.M.: Crossing the quality chasm: a new health system for the 21st century. National Academy Press (2001)
11. Cesarini, M., Mezzanzanica, M., Cavenago, D.: ICT Management Issues in Healthcare Coopetitive Scenarios. In: Information Resources Management Association International Conference (2007)
12. Hardin, J.M., Chhieng, C.: Data mining and clinical decision support systems. In: Health Informatics. Clinical decision support systems, pp. 44–63 (2007)
13. Nuvangi, S.M., Oruthotaarachchi, C.R., Tilakaratna, J.M.P.P., Caldera, H.A.: Usage of Association rules and Classification Techniques in Knowledge Extraction of Diabetes. In: International Conference on Advanced Information Management and Service, pp. 372–377 (2010)
14. Chen, J., He, H., Li, J., Jin, H., McAullay, D., Williams, G., Sparks, R., Kelman, C.: Representing Association Classification Rules Mined from Health Data. In: Khosla, R., Howlett, R.J., Jain, L.C. (eds.) KES 2005. LNCS (LNAI), vol. 3683, pp. 1225–1231. Springer, Heidelberg (2005)
15. Lin, F., Chou, S., Pan, S., Chen, Y.: Mining Time Dependency Patterns in Clinical Pathways. International Journal of Medical Informatics 62, 11–25 (2001)
16. Batal, I., Fradkin, D., Harrison, J., Moerchen, F., Hauskrecht, M.: Mining Recent Temporal Patterns for Event Detection in Multivariate Time Series Data. In: ACM SIGKDD Conference on Knowledge Discovery and Data Mining (2012)
17. Baralis, E., Bruno, G., Chiusano, S., Domenici, V.C., Mahoto, N.A., Petrigni, C.: Analysis of medical pathways by means of frequent closed sequences. In: Setchi, R., Jordanov, I., Howlett, R.J., Jain, L.C. (eds.) KES 2010, Part III. LNCS, vol. 6278, pp. 418–425. Springer, Heidelberg (2010)
18. 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), 379–399 (2011)
19. 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 40(11), 4672–4678 (2013)
20. Antonelli, D., Bellomo, D., Bruno, G., Villa, A.: Evaluating Collaboration Effectiveness of Patient-to-Doctor Interaction in a Healthcare Territorial Network. In: Camarinha-Matos, L.M., Xu, L., Afsarmanesh, H. (eds.) PRO-VE 2012. IFIP AICT, vol. 380, pp. 128–136. Springer, Heidelberg (2012)
21. Wang, J., Han, J.: BIDE: efficient mining of frequent closed sequences. In: International Conference on Data Engineering, pp. 79–90 (2004)