Mining and Retrieving Medical Processes to Assess the Quality of Care

Stefania Montani[1], Giorgio Leonardi[1,2], Silvana Quaglini[2],
Anna Cavallini[3], and Giuseppe Micieli[3]

[1] DISIT, Computer Science Institute, Università del Piemonte Orientale, Alessandria, Italy
[2] Dipartimento di Informatica e Sistemistica, Università di Pavia, Italy
[3] IRCCS Fondazione "C. Mondino", Pavia, Italy - on behalf of the Stroke Unit Network (SUN) collaborating centers

Abstract. In a competitive healthcare market, hospitals have to focus on ways to deliver high quality care while at the same time reducing costs. To accomplish this goal, hospital managers need a thorough understanding of the actual processes. Process mining can be used to extract process related information (e.g., process models) from data. This process information can be exploited to understand and redesign processes to become efficient high quality processes. Process analysis and redesign can take advantage of Case Based Reasoning techniques.

In this paper, we present a framework that applies *process mining* and *case retrieval* techniques, relying on a novel distance measure, to stroke management processes. Specifically, the goal of the framework is the one of analyzing the quality of stroke management processes, in order to verify: (i) whether different patient categories are differently treated (as expected), and (ii) whether hospitals of different levels (defined by the absence/presence of specific resources) actually implement different processes (as they auto-declare). Some first experimental results are presented and discussed.

1 Introduction

Healthcare institutions are increasingly facing pressure to reduce costs, while at the same time improving the quality of care. In order to reach such a goal, healthcare administrators and expert physicians need to evaluate the services the institution provides. Service evaluation requires to analyze medical processes, which are often automated and logged by means of the workflow technology.

Process analysis (PA) covers functions of simulation and diagnosis of processes. While simulation can support performance issues evaluation, diagnosis can highlight e.g., similarities, differences, and adaptation/redesign needs. Indeed, the existence of different patients categories, or of local resource constraints, can make differences between process instances necessary, and process adaptation compulsory (even when the medical process implements a well-accepted clinical guideline). Proper PA techniques are strongly needed when

S.J. Delany and S. Ontañón (Eds.): ICCBR 2013, LNAI 7969, pp. 233–240, 2013.

a given process model does not exist, e.g., because a full clinical guideline has not been provided, and only some recommendations are implemented. In this case, *process mining* techniques [4] can be exploited, to extract process related information (e.g., process models) from log data. It is worth noting, however, that the mined process can also be compared to the existing guideline (if any), e.g., to check conformance, or to understand the required level of adaptation to local constraints. Thus, the mined process information can always be used to understand, adapt and redesign processes to become efficient high quality processes.

The *agile workflow* technology [15] is the technical solution which has been invoked to deal with process adaptation/redesign. In order to provide an effective and quick adaptation support, many agile workflow systems share the idea of recalling and reusing concrete *examples of changes* adopted in the past. To this end, *Case Based Reasoning (CBR)* [1] has been proposed as a natural methodological solution (see e.g, [10,11,7]). In particular, the *case retrieval* step has been extensively studied in PA applications, since the nature of processes can make distance calculation and retrieval optimization non-trivial [12,13,2,8].

In this paper, we propose a framework for medical process analysis and adaptation, which relies on **process mining** and **case retrieval** techniques.

Specifically, our goal is the one of analyzing the quality of stroke management processes, in order to verify: (i) whether different patient categories are differently treated (as expected), and (ii) whether hospitals of different levels (defined by the absence/presence of specific resources for stroke management) actually implement different processes (as they auto-declare).

First, our system extracts process models from a database of real world process logs. In particular, we learn different models for every patient category, and/or for every hospital. Given one of the models as an input, we then retrieve and order the most similar models we have learned. An examination of the distance among the models, to be conducted by a medical expert, can provide information about the quality of the processes, by verifying and quantifying issues (i) and (ii) above. To this end, we have introduced a proper *distance definition*, that extends previous literature contributions [5,3,2] by considering the available information, learned through process mining.

Experimental results (related to issue (ii)) and future research directions are discussed in the paper as well.

2 Methods

2.1 Process Mining and the ProM Tool

Process mining describes a family of a-posteriori analysis techniques exploiting the information recorded in logs, to extract process related information (e.g., process models).

Traditionally, process mining has been focusing on discovery, i.e., deriving process models and execution properties from enactment logs. It is important to mention that there is no a-priori model, but, based on process logs, some model, e.g., a Petri net, is constructed. However, process mining is not limited to process models (i.e., control flow), and recent process mining techniques are more and more focusing on other perspectives, e.g., the organization perspective, the performance perspective or the data perspective. Moreover, as well stated in [6], process mining also supports conformance analysis and process enhancement.

To be able to understand whether the healthcare organizations under study achieve their goals of providing timely and high quality medical services, we conducted several experiments (see also [9]) using the process mining tool called ProM, extensively described in [14]. ProM is a platform independent open source framework which supports a wide variety of process mining and data mining techniques, and can be extended by adding new functionalities in the form of plug-ins.

In particular, we relied on ProM's Heuristic miner [16] for mining the process models, and on a performance analysis plug-in which projects information of the mined process on places and transitions in a Petri net.

2.2 Distance Definition for Case Retrieval

In order to retrieve process models and order them on the basis of their distance with respect to a given query model, we have introduced a distance definition that extends previous literature contributions [5,3,2] by properly considering the available information, learned through process mining.

In particular, since mined process models are represented in the form of graphs (where nodes represent activities and edges provide information about the control flow), we define a distance based on the notion of graph edit distance [3]. Such a notion calculates the minimal cost of transforming one graph into another by applying insertions/deletions and substitutions of nodes, and insertions/deletions of edges.

As in [5], we provide a normalized version of the approach in [3], and as in [5,2], we calculate a *mapping* between the two graphs to be compared, so that edit operations only refer to mapped nodes (and to the edges connecting them).

Moreover, with respect to all the previous approaches, we introduce two novel contributions:

1. we calculate the cost of node substitution $fsubn$ (see Definition 2 below) by applying **taxonomic distance** [13,12] (see Definition 1), and not string edit distance on node names as in [5]. Indeed, we organize the various activities executable in our domain in a taxonomy, where activities of the same type (e.g., Computer Assisted Tomography (CAT) *with* or *without* contrast) are connected as close relatives. The use of this definition allows us to explicitly take into account this form of domain knowledge: the closer two activities are in the taxonomy, the less penalty has to be introduced for substitution;
2. we add a cost contributions related to edge substitution ($fsube$ in Definition 2 below), that incorporates information learned through process mining, namely (i) the percentage of patients that have followed a given edge, and

(ii) the reliability of a given edge, i.e., of the control flow relationship between two activities. The percentage of patients that followed an edge is calculated as the fraction over all the traces in the database in which the activities connected by the edge at hand take place in sequence. The reliability of a relationship (e.g., activity x follows activity y) is not only influenced by the number of occurrences of this pattern in the logs, but is also (negatively) determined by the number of occurrences of the opposite pattern (y follows x). Both items (i) and (ii) are outputs of Heuristic miner [16].

Formally, the following definitions apply:

Definition 1: Taxonomic Distance
Let α and β be two activities in the taxonomy t, and let γ be the closest common ancestor of α and β. The *Taxonomic Distance* $dt(\alpha, \beta)$ between α and β is defined as:

$$dt(\alpha, \beta) = \frac{N_1 + N_2}{N_1 + N_2 + 2 * N_3}$$

where N_1 is the number of arcs in the path from α and γ in t, N_2 is the number of arcs in the path from β and γ, and N_3 is the number of arcs in the path from the taxonomy root and γ.

Definition 2: Extended Graph Edit Distance. Let $G1 = (N1, E1)$ and $G2 = (N2, E2)$ be two graphs, where Ei and Ni represent the sets of edges and nodes of graph Gi. Let M be a partial injective mapping [5] that maps nodes in $N1$ to nodes in $N2$ and let *subn*, *sube*, *skipn* and *skipe* be the sets of substituted nodes, substituted edges, inserted or deleted nodes and inserted or deleted edges with respect to M. In particular, a substituted edge connects a pair of substituted nodes in M. The fraction of inserted or deleted nodes, denoted $fskipn$, the fraction of inserted or deleted edges, denoted $fskipe$, and the average distance of substituted nodes, denoted $fsubn$, are defined as follows:

$$fskipn = \frac{|skipn|}{|N1| + |N2|}$$

$$fskipe = \frac{|skipe|}{|E1| + |E2|}$$

$$fsubn = \frac{\sum_{n,m \in M} dt(n, m)}{|subn|}$$

where n and m are two mapped nodes in M.

The average distance of substituted edges $fsube$ is defined as follows:

$$fsube = \frac{\sum_{(n1,n2),(m1,m2) \in M}(|rel(e1) - rel(e2)| + |pat(e1) - pat(e2)|)}{|2 * sube|}$$

where edge $e1$ (connecting node $n1$ to node $m1$) and edge $e2$ (connecting node $n2$ to node $m2$) are two substituted edges in M, $rel(ei)$ is the reliability $\in [0, 1]$ of edge ei as extracted by Heuristic miner [16], and $pat(ei)$ is the percentage of patients that crossed edge ei.

The extended graph edit distance induced by the mapping M is:

$$ext_{edit} = \frac{wskipn * fskipn + wskipe * fskipe + wsubn * fsubn + wsube * fsube}{wskipn + wskipe + wsubn + wsube}$$

where $wsubn$, $wsube$, $wskipn$ and $wskipe$ are proper weights $\in [0, 1]$.

The extended graph edit distance of two graphs is the minimal possible distance induced by a mapping between these graphs. To find the mapping that leads to the minimal distance we resort to the greedy algorithm described in [5].

3 Experimental Results

In clinical practice, no support is available to physicians/administrators to verify whether hospitals of different levels actually implement different processes when caring a specific pathology (see issue (ii) described in the Introduction). In a previous version of this work [9], process mining was relied upon to provide physicians with a graphical view of the mined processes. A visual inspection of those figures was a first help towards the fulfillment of the tasks related to issue (ii). However, mined processes can be huge and very complex, so that an automated comparison among them, like the one we are providing in this framework, can truly be an added value for quality evaluation.

In the rest of this section, we discuss our experimental results, related to issue (ii). In particular, we wished to test whether the level of 37 hospitals located in the Lombardia Region (Northern Italy) could be verified (or corrected) through our framework, when referring to stroke care. Hospital levels (i.e., 1, 2, 3) have to be defined in Lombardia Region according to the available human and instrumental resources. Every hospital auto-declares its own level. Specifically, we mined the stroke management processes implemented in all 37 hospitals. We then chose one level-2 hospital as a query, and we retrieved and ordered the mined processes of the 36 others (21 of which were declared as level-2 hospitals as well). We performed retrieval and ordering both resorting to the distance defined in [5], and to the novel one introduced in section 2.2. Results are reported in figure 1.

First, we can observe that our distance is able to discriminate among every single mined processes, while the one in [5] only identifies some macro-classes, composed by several processes, whose distance from the query does not change

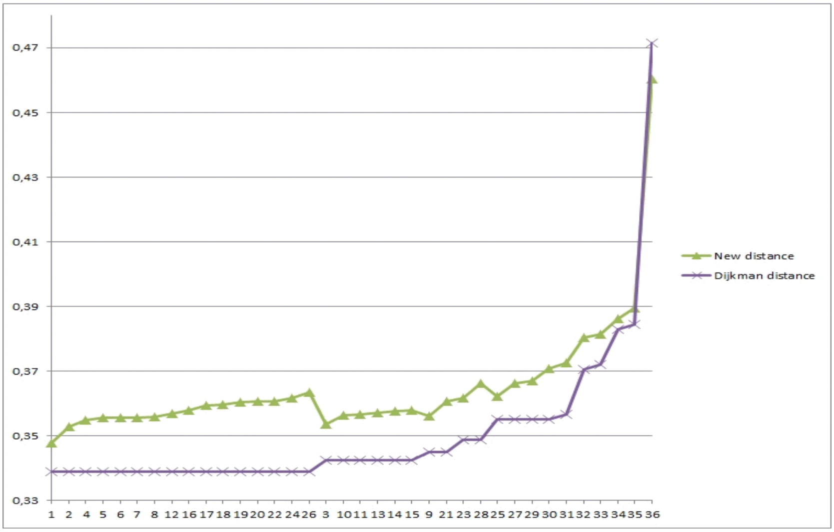

Fig. 1. Retrieval and ordering of 36 mined processes, implemented in 36 different hospitals in the Lombardia region, with respect to the selected query process (on the x-axis: process number; on the y-axis: distance value from the query). Results are shown in two different framework settings: when relying on the metric in [5] (Dijkman distance), and when relying on the metric defined in section 2.2

(see horizontal segments in figure 1). We believe that the finer distinction we could obtain is due to the use of taxonomic distance, and of edge information, which are disregarded by [5]. This additional information can be very significant from a medical viewpoint. For instance, hospitals 2 and 20 are not distinguishable according to [5], but in hospital 20 more than 70% of the patients undergo ECG immediately after CAT, while in hospital 2 this occurs for only 10% of the patients. Almost all patients undergo these tests in the two hospitals indeed, but within different control flow patterns. In hospital 20 there seems to be a behavioral rule pushing for the pattern CAT *immediately followed by* ECG, while in the other hospital this direct sequential pattern does not exist. This is an edge-related information extracted by Heuristic miner, and properly used by our metric for providing its finer ordering.

As for the declared hospital levels, we considered the 22 closest processes (i.e., hospitals) with respect to the query. This number was chosen because it is the sum of the number of processes in the two closest macro-classes when resorting to [5] (16 processes belong to the first macro-class, 6 to the second), and with [5] it is not possible to further refine the ordering among these examples. If the auto-declared level of these examples was correct (and confirmed by the mined processes), we should find 21 level-2 hospitals in this set. However, this did not happen. When resorting to [5], we found only 13 level-2 hospitals in these nearest neighbors. Of them, only 9 were listed in the closest 16 (i.e., the first macro-class). When exploiting our distance, we still found 13 level-2 hospitals in the first 22, but 11 of them were in the first 16. Our results were thus closer to the expected ones.

We analyzed the situation of the remaining 8 level-2 hospitals, that were not found in the nearest neighbors. Very interestingly, 7 of these missing examples are the very same when resorting to the two different metrics. Indeed, the visual examination of the graphs highlight important differences with respect to the query hospital. For example, one of them does not perform the thrombolisys treatment, even if typical of level 2 stroke units. We have to say that some local conditions (e.g., specific resources availability) may have recently changed, altering the real level of some hospitals with respect to the originally declared one. This conclusion thus supports the quality of the implemented metrics, and of our novel contribution in particular.

As a final consideration, we can quickly comment on 4 cases, that were differently ordered by the two metrics. According to the auto-declared levels, our ordering is closer to reality in 3 of them (no. 9, 22 and 24), while in the fourth case (no. 26) our metric overestimates the distance between the hospital and the query. Despite the overall positive outcome, this motivates further improvements, like the ones we will discuss in section 4.

4 Discussion, Conclusions and Future Work

This work showed that process mining and case retrieval techniques can be applied successfully to clinical data to gain a better understanding of different medical processes adopted by different hospitals (and for different groups of patients). It is interesting to analyze the differences, to establish whether they concern only the scheduling of the various tasks or also the tasks themselves. In this way, not only different practices may be discovered that are used to treat similar patients, but also unexpected behavior may be highlighted.

In this paper we have shown some first experimental results. More tests are obviously needed, including leave-one-out style experiments and comparisons with other metrics, and are planned for the next months.

In the future we also wish to extend our contribution, by including the treatment of time in $fsube$ (see Definition 2 in section 2.2). Indeed, by projecting the mined process on a Petri Net (see section 2.1), we can obtain information about delays between activities, possible overlaps and synchronizations . We would like to explicitly compare this information between mapped processes. We believe that, since in emergency medicine the role of time is clearly central, this enhancement could represent a relevant added value in our framework, and make it even more reliable and useful in practice.

Acknowledgments. This research is partially supported by the GINSENG Project, Compagnia di San Paolo.

References

1. Aamodt, A., Plaza, E.: Case-based reasoning: foundational issues, methodological variations and systems approaches. AI Communications 7, 39–59 (1994)
2. Bergmann, R., Gil, Y.: Retrieval of semantic workflows with knowledge intensive similarity measures. In: Ram, A., Wiratunga, N. (eds.) ICCBR 2011. LNCS, vol. 6880, pp. 17–31. Springer, Heidelberg (2011)
3. Bunke, H.: On a relation between graph edit distance and maximum common subgraph. Pattern Recognition Letters 18(8), 689–694 (1997)
4. Van der Aalst, W., van Dongen, B., Herbst, J., Maruster, L., Schimm, G., Weijters, A.: Workflow mining: a survey of issues and approaches. Data and Knowledge Engineering 47, 237–267 (2003)
5. Dijkman, R., Dumas, M., Garca-Banuelos, R.: Graph matching algorithms for business process model similarity search. In: Proc. International Conference on Business Process Management, pp. 48–63 (2009)
6. IEEE Taskforce on Process Mining: Process Mining Manifesto, `http://www.win.tue.nl/ieeetfpm`
7. Kapetanakis, S., Petridis, M., Knight, B., Ma, J., Bacon, L.: A case based reasoning approach for the monitoring of business workflows. In: Bichindaritz, I., Montani, S. (eds.) ICCBR 2010. LNCS, vol. 6176, pp. 390–405. Springer, Heidelberg (2010)
8. Kendall-Morwick, J., Leake, D.: On tuning two-phase retrieval for structured cases. In: Lamontagne, L., Recio-García, J.A. (eds.) Proc. ICCBR 2012 Workshops, pp. 25–334 (2012)
9. Mans, R., Schonenberg, H., Leonardi, G., Panzarasa, S., Cavallini, A., Quaglini, S., Van der Aalst, W.: Aprocess mining techniques: an application to stroke care. In: Proc. Medical Informatics Europe (MIE), pp. 573–578 (2008)
10. Minor, M., Tartakovski, A., Schmalen, D., Bergmann, R.: Agile workflow technology and case-based change reuse for long-term processes. International Journal of Intelligent Information Technologies 4(1), 80–98 (2008)
11. Montani, S.: Prototype-based management of business process exception cases. Applied Intelligence 33, 278–290 (2010)
12. Montani, S., Leonardi, G.: Retrieval and clustering for supporting business process adjustment and analysis. Information Systems, doi: `http://dx.doi.org/10.1016/j.is.2012.11.006`
13. Montani, S., Leonardi, G.: Retrieval and clustering for business process monitoring: results and improvements. In: Agudo, B.D., Watson, I. (eds.) ICCBR 2012. LNCS, vol. 7466, pp. 269–283. Springer, Heidelberg (2012)
14. van Dongen, B.F., de Medeiros, A.K.A., Verbeek, H.M.W(E.), Weijters, A.J.M.M.T., van der Aalst, W.M.P.: The proM framework: a new era in process mining tool support. In: Ciardo, G., Darondeau, P. (eds.) ICATPN 2005. LNCS, vol. 3536, pp. 444–454. Springer, Heidelberg (2005)
15. Weber, B., Wild, W.: Towards the agile management of business processes. In: Althoff, K.-D., Dengel, A.R., Bergmann, R., Nick, M., Roth-Berghofer, T.R. (eds.) WM 2005. LNCS (LNAI), vol. 3782, pp. 409–419. Springer, Heidelberg (2005)
16. Weijters, A., Van der Aalst, W., Alves de Medeiros, A.: Process Mining with the Heuristic Miner Algorithm, BETA Working Paper Series, WP 166. Eindhoven University of Technology, Eindhoven (2006)