Available online at http://www.journalcra.com

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

International Journal of Current Research
Vol.11, pp.215-221, December, 2010

ISSN: 0975-833X

RESEARCH ARTICLE

PROCESS MINING IN HEALTHCARE USING CONTROL FLOW PERSPECTIVE: A CASE STUDY

Saravanan, M.S[1] and Rama Sree, R.J[2].

[1]R & D Centre, Bharathiar University, Coimbatore, Tamil Nadu, INDIA.
[1]VEL TECH Dr. RR & Dr. SR Technical University Avadi, Chennai, INDIA.
[2]Department of Computer Science Rashtriya Sanskrit University, Tirupati, Andhra Pradesh, INDIA.

ARTICLE INFO

Article History:
Received 12th September, 2010
Received in revised form
15th October, 2010
Accepted 13th November, 2010
Published online 28th December, 2010

Key words:
Process mining,
Healthcare,
Event log,
Prom framework.

ABSTRACT

The main operations in hospitals are trying to maintain their process in the progressive manner that is healthcare sectors are try to streamline their processes. In order to do so, it is essential to have an accurate view of the care flows under consideration. In this paper, we apply process mining techniques to obtain meaningful knowledge about these flows, e.g., to discover typical paths followed by particular groups of symptoms that create a disease of a patient. This is a non-trivial task given the dynamic nature of healthcare processes. The paper demonstrates the applicability of process mining using a real case of a patient affected by a disease in a hospital. Using a variety of process mining techniques, we analyzed the healthcare process using three different algorithms (1) α-algorithm (2) Heuristics Miner algorithm and (3) Damped Working Set (DWS) algorithm. In order to do so we extracted relevant event logs from the hospital information system and analyzed these logs using the ProM framework. This paper only deals about control flow perspective of the healthcare processes. Therefore the results show that process mining can be used to provide new insights that facilitate the improvement of existing treatment system.

INTRODUCTION

The hospitals are now searching for best technology or software to streamline their daily patient care flow process. So, it is a need to find the best tool or software to provide better service to attend the various cases called patients.

***Corresponding author:** *saranenadu@yahoo.co.in*

So the healthcare sector, normally hospitals have to focus on ways to streamline their processes in order to deliver high quality care while at the same time reducing costs (Anyanwu *et al.,* 2003). Furthermore, a lot of private and public organizations are putting more pleasure on the hospital to work in efficient way as much as possible. Therefore, more demand in the future to maximize the service is expected. The common factor that require in the present treatment system in the hospitals, require highly *simple* and extremely *flexible* healthcare processes. It also referred to as "control flows". Moreover, in healthcare there are many disciplines are involved for which it is found that they are working in isolation and hardly have any idea about what happens within other disciplines. Another issue is that within industry or healthcare sector many autonomous, independently developed applications are found (Lenz *et al.,* 2002). A consequence of this all is that *it is not known what happens in a healthcare process for a group of patients with the same process or a group of same symptoms for various diseases*. The concept of process mining provides an interesting opportunity for providing a solution to this problem. Process mining (van der Aalst *et al.,* 2003) aims at extracting process knowledge from so called "event logs" which may originate from all kinds of systems, like enterprise information systems or hospital information systems. Typically, these event logs contain information about the start or completion of process steps together with related context data that is actors and resources. Furthermore, process mining is a very broad area both in terms of (1) applications for example from banks to embedded systems and (2) techniques.

The main idea of this paper focuses on the applicability of process mining in the healthcare domain. Process mining has already been successfully applied in the service industry (van der Aalst *et al.,* 2007). In this paper, the applicability of process mining using control flow perspective algorithms are used to create a best model for the healthcare processes. We will show how process mining can be used for obtaining insights related to control flows, e.g., the identification of control paths and strong comparison between different symptoms for the same diseases to minimize the processing time. So, there are several process mining techniques, which will also show the diversity of process mining techniques but in this paper we will discuss about control flow discovery.

In view of the better care flows, the raw data were collected from S.M. Hospital in Ranipet, Tamil nadu, INDIA, a multispeciality hospital in south India. This raw data contains data about a symptoms of four diseases of a patient, This is treated and for which all deep analysis and treatment activities have been recorded to analyze the flow time, waiting time, processing time and synchronization time for better process. Note that we did not use any apriori knowledge about the control process of this group of diseases and that we also did not have any process model at hand.

Today's Business Intelligence (BI) tools (Snezana *et al.,* 2010) used in the healthcare domain, like Cognos, Business Objects, or SAP BI, typically look at aggregate data seen from an external perspective that is frequencies, averages, utilization, service levels, etc,. These BI tools focus on performance indicators such as the number of tasks or operations of a disease, the length of waiting list and the success rate of operations. Process mining looks "inside the process" at different abstraction levels. So, in the context of a healthcare, unlike BI tools, we are more concerned with the control paths followed by individual disease and whether certain procedures are followed or not.

This paper is structured as follows: Section 2 provides an overview of process mining. In Section 3 we will show the applicability of process mining in the healthcare unit domain using data obtained for symptoms of four diseases of patient. Section 4 discusses about the control flow perspective of process mining and section 5 concludes the paper.

Process mining

The concept or the division of process mining is applicable to a wide range of systems. These systems may be pure information systems for example ERP systems or systems where the hardware plays a more prominent role for example embedded systems. The only requirement is that the system produces *event logs*, thus recording that parts of the actual behavior. The information which is an interesting class of information systems that produce event logs, so called Process-Aware Information Systems (PAISs) (Dumas *et al.,* 2005). Examples are classical workflow management systems, example; Staffware, ERP systems, example; SAP, case handling systems, example; FLOWer, PDM systems example; Windchill, CRM systems example; Microsoft Dynamics CRM, middleware example; IBM's WebSphere, hospital information systems example; Chipsoft, etc. These systems provide very detailed information about the activities that have been executed.

However, not only PAISs are recording events. Also, in a typical healthcare there is a wide variety of systems that record events. For example, in a treatment process, a system can record the treatment of a patient undergoes and also it can record occurring symptoms, which differ from other diseases. For a treatment department the whole process depends on a patient, because the testing the symptoms different from each disease, at the same time it matches few cases for the same symptoms for different diseases, also can be recorded. However, frequently these systems are maintained in the healthcare department. In healthcare process, the process is started from admission to discharge of a patient, so that the different activity of treatment process has distributed in and around the departments of every hospital. Hence, information needs to be processed at each stage without any delay and loss. Therefore the data from different systems needs to be identified and processed at earliest using best software tools. The data fed in to these type of systems need to be fine tuned or refined data or relative data are discovered using some modern system. The one of the leading study says that process mining is one of the best technologies to fulfill the needs of the knowledge discovery. In this way, these systems within the healthcare unit can contain information about processes within one department but also across departments. This information can be used for improving processes within departments itself or improving the services offered to patients for various diseases. The goal of process mining is to extract information for example process models from these logs that is process mining describes 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 that is a well defined step in the process and is related to a particular case that is a process instance. Furthermore, some mining techniques use additional information such as the performer or originator of the event that is the person or resource executing or initiating the activity, the timestamp of the event, or data elements recorded with the event for example the size of an order. Process mining addresses the problem that most "process or system owners" have limited information about what is actually happening. In practice, there is often a significant gap between what is prescribed or supposed to happen, and what *actually* happens. Only a concise assessment of reality, which process mining strives to deliver, can help in verifying process models, and ultimately be used in system or process redesign efforts. The idea of process mining is to discover, monitor and improve real processes that are not assumed processes by extracting knowledge from event logs. We consider three basic types of process mining (Figure 1): (1) *discovery*, (2) *conformance*, and (3) *extension*.

Discovery: Traditionally, process mining has been focusing on *discovery*, i.e., deriving information about the original process model, the organizational context, and execution properties from enactment logs. An example of a technique addressing the control flow perspective is the α-algorithm (van der Aalst *et al.*, 2004) which constructs a Petri net model describing the behavior observed in the event log. It is important to mention that there is no apriori model, i.e., based on an event log some model is constructed. However, process mining is not limited to process models that is control flow and recent process mining techniques are more and more focusing on other perspectives, for example the organizational perspective, performance perspective or the data perspective. For example, there are approaches to extract social networks from event logs and analyze them using social network analysis (van der Aalst *et al.,* 2008). This allows organizations to monitor how people, groups, or software or system components are working together. Also, there are approaches to visualize performance related information, e.g. there is an approach which graphically shows the bottlenecks and all kinds of performance indicators, e.g., average or variance of the total flow time or the time spent between two activities.

Conformance: There is an apriori model. This model is used to check if reality conforms to the model. For example, there may be a process model indicating that purchase orders of more than one million Euro require two checks. Another example is the checking of the so called "four eyes" principle. Conformance checking may be used to detect deviations, to locate and explain these deviations, and to measure the severity of these deviations.

Extension: There is an apriori model. This model is extended with a new aspect or perspective that is the goal is not to check conformance but to enrich the model with the data in the event log. An example is the extension of a process model with performance data, that is some apriori process model is used on which bottlenecks are projected. At this point in time there are mature tools such as the ProM framework (van der Aalst *et al.,* 2007), featuring an extensive set of analysis techniques which can be applied to real life logs.

Patient care process

In every process, there should be input to process, in this case the healthcare system taken as a input raw data to process the patient data. The patient data need to be maintained in every section or division of the treatment of the patient, hence there should be a need or guarantee of data is required to do process mining. To this case study, we use raw data collected from S.M. Hospital of Ranipet, a division of public healthcare by the government of Tamilnadu in India. The raw data contains information about a group of few patients treated for a fever in august 2010 to September 2010 and for which all treatment activities have been recorded. The information is available in different departments for the disease of fever, for example blood test, urine test and several labs, etc. For this data set, we have extracted event logs from the healthcare database, where each event refers to a service delivered to a patient. As the data is coming from a hospital, we have to face the interesting problem that for each service delivered for a patient it is only known on which day the service has been delivered. In other words the information about the actual timestamps of the start and completion of the service delivered are recorded with difficulties. The collected log contains 50 different events, which indicated that we are dealing a non-trivial care flow process. In the remainder of this section we will focus on obtaining, in an explorative way, insights into the treatment process of healthcare process. So, we will only focus on the discovery part of process mining, instead of the conformance and extension part. Furthermore, obtaining these insights should not be limited to one perspective only. Therefore, in section 4, we focus on the discovery of control paths followed by the doctors in the hospital. This also demonstrates the diversity of process mining techniques available. However discuss about the different control flow perspective mining such as respectively alpha, heuristics and DWS algorithm in detail using the healthcare data to discover the better knowledge.

Control flow perspective

In this section, we present some results obtained through a detailed analysis of the healthcare event log for the treatment process. We concentrate on the discovery part to show actual situations for example control flows in the hospital, healthcare process. Hence, in this section, we more specifically elaborate on mining results based on the control flow perspective in process mining.

***Converting the raw data into MXML format*:** In healthcare process the patient first diagnosed with different testing then the actual treatment is given by the doctors. Therefore the testing decides the division of treatment and medicine. So, the testing process is recorded and fed into the system as a data. These data are stored in the form of database file format. Then this database information is converted into Mining Extensible Markup Language (MXML) file format using the ProM Import Framework. Finally this converted MXML file is sent as an input to ProM Framework to do different types of mining. In this paper we concentrated to deal with process or control flow of activities or tasks. Hence the view of control flow perspective deal about the processing time, flow time, waiting time and synchronization time to identify and minimize for better process for lot of patients at a time, after the better knowledge getting from the sampling process. The following sections deals about the different application of mining algorithm on healthcare data.

Alpha Mining Algorithm: One of the most promising mining techniques is control flow mining which automatically derives process models from process logs. The generated process model reflects the actual process as observed through real process executions. The generated process models from healthcare unit process were give insight into control paths for different treatment process. Until now, there are several process mining algorithms such as the α-mining algorithm, heuristic mining algorithm, Damped Working Set (DWS) mining algorithm, etc., (Greco *et al.,* 2006). Hence the α-mining algorithm enables users to focus on the main process flow instead of on every detail of the behavior appearing in the process log (Weijters *et al.,* 2003). Figure 1 shows the process model for all cases obtained using the α-mining algorithm. Despite its ability to focus on the most frequent paths, the process, depicted in figure 1, is still spaghetti-like and too complex to understand easily. The "patient start" event log start the process execution and end with the "patient complete" event log, between these two process all other event logs are mined for new knowledge, not like human thinking. After the "admission complete" the patient number, testing of blood and urine process are started for treatment. The treatment process is decided after the different testing process, hence the process of medicine is decided, so here the synchronous testing of different symptoms is discovered for right treatment. Hence, the doctors are need for this information to process the further treatment. Therefore the parallel and serial processes are identified thru this algorithm and also the exceptions such as treatment failure are not managed by this algorithm. Hence, this method of process, which is used for real time applications, such as healthcare process are very helpful to the doctors, hospitals and patients for quick treatment with respect to processing time, waiting time, etc.

Heuristics Miner Algorithm: Heuristics mining algorithm can deal with noise and exceptions, and enables users to focus on the main process flow instead of on every detail of the behavior appearing in the process log (Weijters *et al.,* 2003).

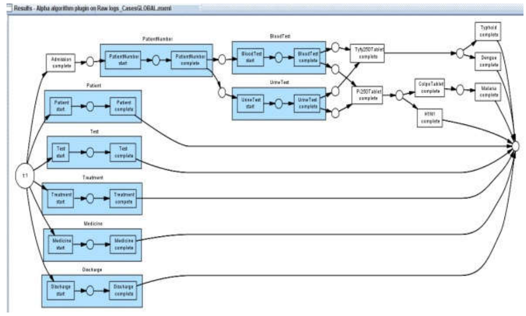


Fig 1. α-mining algorithm process model for all event logs

Heuristics Miner Algorithm

Heuristics mining algorithm can deal with noise and exceptions, and enables users to focus on the main process flow instead of on every detail of the behavior appearing in the process log (Weijters *et al.,* 2003). The figure 2 shows the process model for all cases obtained using the Heuristics Miner. It has the most frequent paths such as treatment, testing and medicine etc. Hence the model shows in figure 2 has the clear picture, that simple and understandable without more direction arrows, also it has less spaghetti like output. The Heuristics Miner algorithm described in this paper is more or less a replacement for the older heuristics mining tool like Little Thumb, etc. The most important characteristic of the Heuristics Miner is the robustness for noise and exceptions. Hence this model handles exceptions; it can be automated using the meta data from database information using various automated tool ProM, etc.

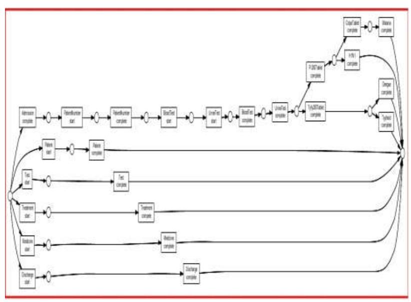


Fig 2. Heuristics mining algorithm process model for all event logs

DWS Algorithm

This algorithm describes the implementation of DWS algorithm (Greco *et al.,* 2006). The mining is carried out through a top down hierarchical clustering process, where the log is recursively split into homogeneous clusters from a behavioral viewpoint. All discovered clusters are then equipped with a specific workflow model by exploiting algorithm Heuristics Mining provided with ProM. Any partitioning step hence produces a refinement of the workflow model being discovered. Specific behavioral patterns, named discriminate rules, are used as features for clustering log instances by means of classical k-means algorithm. The figure 3 shows the flowchart like approach to understand the process in vertical phase. This type of representation help to the healthcare domain for identifying the same diagnosis undergoes for a patient. In figure 3, it shows that the different testing for the same patient to take a treatment. In this algorithm, we can come to know that the parallel processing is represented as parallel columns.

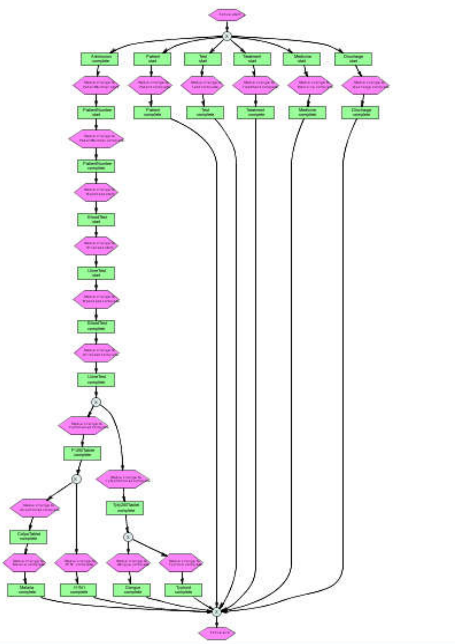
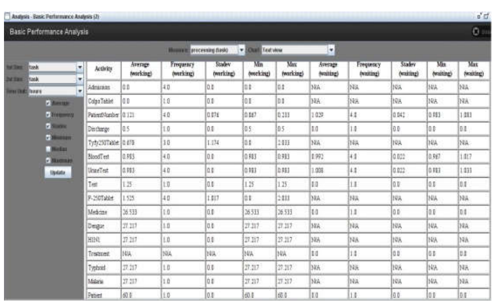


Fig. 3. DWS algorithm process model for all event logs

Since, processes in the healthcare domain do not have a single kind of flow but a lot of variants based on different diseases of the patients. Therefore using the figure 1, 2 and 3, it can understand that the similar diseases require different diagnosis and treatment for each patient by comparing each testing of log events. It is surprising that the derived process model is spaghetti-like and convoluted. One of the methods for handling this problem is breaking down a log into two or more sub logs until these become simple enough to be analyzed clearly. We apply clustering techniques to divide a process log into several groups that is clusters, where the cases in the same cluster have similar properties. Clustering is a very useful technique for logs which contain many cases following different procedures, as is the usual case in healthcare unit process systems. Figure 4 shows the Average, Minimum and Maximum of processing time and waiting time for all events or activities. Hence from this figure, it is clear that the time taken for processing that is processing time of each event or activity, this helps to reduce the processing time for few activities. The waiting time (Elhuizen *et al.,* 2007) or idle time of activity is used to reduce the flow time for effective control over each activity. The synchronization time is used to process some activities as parallel to minimize the overall processing time. The time taken for processing the every event is managed efficiently for every end user in the real time applications. So, this healthcare case study deals the main problem of patient processing time with better analysis of mining.



Activity	Average (working)	Frequency (working)	Stadev (working)	Min (working)	Max (working)	Average (waiting)	Frequency (waiting)	Stadev (waiting)	Min (waiting)	Max (waiting)
Admission	0.0	4.0	0.0	0.0	0.0	N/A	N/A	N/A	N/A	N/A
ColpaTablet	0.0	1.0	0.0	0.0	0.0	N/A	N/A	N/A	N/A	N/A
PatientNumber	0.121	4.0	0.076	0.067	0.233	1.029	4.0	0.042	0.983	1.083
Discharge	0.5	1.0	0.0	0.5	0.5	0.0	1.0	0.0	0.0	0.0
Tyfy250Tablet	0.678	3.0	1.174	0.0	2.033	N/A	N/A	N/A	N/A	N/A
BloodTest	0.983	4.0	0.0	0.983	0.983	0.992	4.0	0.022	0.967	1.017
UrineTest	0.983	4.0	0.0	0.983	0.983	1.008	4.0	0.022	0.983	1.033
Test	1.25	1.0	0.0	1.25	1.25	0.0	1.0	0.0	0.0	0.0
P-250Tablet	1.525	4.0	1.017	0.0	2.033	N/A	N/A	N/A	N/A	N/A
Medicine	26.533	1.0	0.0	26.533	26.533	0.0	1.0	0.0	0.0	0.0
Dengue	27.217	1.0	0.0	27.217	27.217	N/A	N/A	N/A	N/A	N/A
H1N1	27.217	1.0	0.0	27.217	27.217	N/A	N/A	N/A	N/A	N/A
Treatment	N/A	N/A	N/A	N/A	N/A	0.0	1.0	0.0	0.0	0.0
Typhoid	27.217	1.0	0.0	27.217	27.217	N/A	N/A	N/A	N/A	N/A
Malaria	27.217	1.0	0.0	27.217	27.217	N/A	N/A	N/A	N/A	N/A
Patient	60.0	1.0	0.0	60.0	60.0	0.0	1.0	0.0	0.0	0.0

Fig. 4. Average, Minimum and Maximum of processing time and waiting time for all Events or Activities.

Conclusion

In this paper, we have focused on the applicability of process mining in the healthcare unit domain. In this case study, we have used data coming from non-trivial healthcare process of the S.M. Hospital. We focused on obtaining insights into the care flow that is control flow by looking at the control flow perspective. For this perspective, it has some initial results. In which it has been shown that, it is possible to mine complex healthcare unit processes giving insights into the process. In addition, with existing techniques, it is able to derive understandable models for large groups of treatments to identify the flow time and synchronization time and to minimize the processing time and waiting time. Furthermore, the results compared with the flowchart of the treatment process, more or less the automated system developed model has detailed mining result for better process management. However a lot of effort was required for creating the flowchart and obtaining the logistical data, where with the process mining there is the opportunity to obtain these kinds of data in a semi automatic way. Unfortunately, traditional process mining approaches have problems dealing with unstructured processes as for example can be found in a hospital environment. Future work will focus on both developing new mining techniques and on using existing techniques in an innovative way to obtain understandable, high level information instead of "spaghetti like" models showing all details. Obviously, we plan to evaluate these results in healthcare organizations such as the S.M. Hospital. The results are not derived by human thinking, it goes as per the recorded information and hence the automated process of process mining gives the further necessary measures to the healthcare experts called doctors, well in advance after some experiments or treatments, so it is sufficient and very helpful for the better healthcare process.

REFERENCES

Anyanwu, K., Sheth, A., Cardoso, J.,Miller, J., and Kochut, K. 2003. Healthcare Enterprise Process Development and Integration. *Journal of Research and Practice in Information Technology*, 35 (2): 85-98.

Lenz, R., Elstner, T., Siegele, H., and Kuhn, K. A. 2002. Practical Approach to Process Support in Health Information Systems. *Journal of the American Medical Informatics Association*, 9(6): 571–585.

van der Aalst, W., van Dongen, B., Herbst, J., Maruster, L., Schimm, G., and Weijters, A. 2003. Workflow Mining: A survey of Issues and Approaches. *IEEE Journal of the Data and Knowledge Engineering*, 47(2).

van der Aalst, W., Reijers, H., Weijters, A., van Dongen, B., de Medeiros, A. A., Song, M., and Verbeek, H. 2007. Business process mining : an industrial application. *J. of the Information Systems*, 32(5).

Snezana, S. and Violeta, M. 2010. Business Intelligence tools of Statistical Analysis. 32nd International Conference on Information Technology Interfaces (ITI) by the University of St. Kliment Ohridski Bitola, Bitola, Macedonia and published in IEEE Journal,: 21-24: 199-204.

Dumas, M., van der Aalst, W., and ter Hofstede, A. 2005. Process-Aware Information Systems: Bridging People and Software through Process Technology. *Published by Wiley & Sons*.

van der Aalst, W., Weijters, A., and Maruster, L. 2004. Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9):1128–1142.

van der Aalst, W., Reijers, H., and Song, M. 2005. Discovering Social Networks from Event Logs. *Published in Computer Supported Cooperative Work*, 14(6):549–593.

van der Aalst, W., van Dongen, B., Gunther, C., Mans, R., de Medeiros, A. A., Rozinat, A., Rubin, V., Song, M., Verbeek, H., and Weijters, A. 2007. ProM 4.0: Comprehensive Support for Real Process Analysis. In Kleijn, J. and Yakovlev, A., editors, Application and Theory of Petri Nets and Other Models of Concurrency (ICATPN 2007), Published by Springer-Verlag, Berlin, 2007, Vol 4546 of Lecture Notes in Computer Science, pp. 484–494.

Weijters, A. and van der Aalst, W. 2003. Rediscovering Workflow Models from Event-Based Data using Little Thumb. *Published in Integrated Computer-Aided Engineering*, 10(2): 151–162.

Greco, G., Guzzo, A., Pontieri, L. and Sacca, D. 2006. Discovering Expressive Process Models by Clustering Log Traces. *IEEE Transactions on Knowledge and Data Engineering (TKDE),* 18(8):1-15.

Elhuizen, S., Burger, M., Jonkers, R., Limburg, M., Klazinga, N., and Bakker, P. 2007. Using Business Process Redesign to Reduce Wait Times at a University Hospital in the Netherlands. *The Joint*

Commission Journal on Quality and Patient Safety, 33(6): 332–341.