Journal of Theoretical and Applied Information Technology
30th June 2016. Vol.88. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HEALTH CARE ANALYSIS FOR PROCESS DEVIATION USING ALPHA-FITNESS ALGORITHM IN PROCESS MINING

[1]GANESHA K, [2]RASHMI NAGARAJ, [3]NAYANA MD

[1]Lecturer, Department of Computer Science, Amrita Vishwa Vidyapeetham, Mysuru Campus ,Amrita University, India

[2,3] P G Scholar, Department of Computer Science, Amrita Vishwa Vidyapeetham, Mysuru Campus, Amrita University, India

E-mail: [1]asasganesha@gmail.com, [2]rashmi.nagraj93@gmail.com , [3]nayanam.md@gmail.com

ABSTRACT

Health care sectors are continuously exploring new and innovative way to improve operational efficiencies. This research study investigate a way to find potential efficiency gains in healthcare sectors by observing how they are carried out in the past and then investigating better ways of implementing them by considering the factors like time, cost and resource utilization. To achieve competitive advantage, healthcare centers try and contour their processes. Process mining can be enforced to extract data from recorded event. The aim of the system is to propose effective process models by applying dataset for each model which indeed identifies the deviation from the actual process with help the of analytical tool ProM. In this paper several blood tests are considered as the baseline scenario wherein effective process models are generated and checked for the efficiency using alpha-fitness algorithm. One of the major parts involved in process improvement is process modeling which can be optimized and analyzed.

Keywords: *Process Deviation, Event Log, Information Management Systems, Prom, Alpha*.

1. INTRODUCTION

With the increasing demand for health, health care centers are striving hard to optimize their processes in order to enhance the efficiency[1], while guaranteeing the quality of the care. Health care centers consists of inter-disciplinary domains such as hospitals, diagnostic centers, information management system, human resource management systems, business processes etc. The main aim of health care system is to provide an efficient and improvised environment for the target population which helps in better decision making. This can be achieved by information management system through continuous monitoring and recording frequent activities from the knowledge base. The information management system[2] mainly deals with data integration, processing for improving health care service for better utilization of available resources. The paper mainly focuses on providing a better information management system for the diagnostic centre which reduces the complexity of the process of diagnosis.

This can be achieved by process mining approach. Process mining aims to discover, monitor and improve genuine processes by extracting facts from event data[3] readily available in current information systems. To be able to apply process mining techniques it is necessary to extract event logs from data sources (e.g., databases, business logs, audit trails, etc.). Event logs mainly consist of CaseID which is unique for different entry, event which consists of activity, timestamp with start and end time and resources which include human resource. Figure 1 describes attributes of an event logs[4]. Event logs contain numerous sets of traces where similar traces may appear multiple times. Traces relates to the activity name.

CaseID	Event	Timestamp	Resource
---	---	---	---

Figure 1: Attributes Of An Event Log

Journal of Theoretical and Applied Information Technology
30th June 2016. Vol.88. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Process mining comprises of three phases namely discovery, conformance analysis and enhancement.

The Discovery technique considers an event log and produces the respected process model. *The Conformance analysis*: where an existing process model is correlated with an event log of the equivalent process. Conformance checking can be used to check for the reality, as listed within the log. It checks for the variation in the process. *The Enhancement* where the belief is to broaden or enhance an existing process model using information about the absolute process recorded in some event log.

Three importatnt events of process mining are: *Play-in:* Generates the process models using the event-logs. *Play-out*: Event logs and process models are checked for the complement with the existing process model. It helps to analyze the performance. *Replay*: It helps in the investigation of frequent paths, bottle-necks, interval of the process etc [3] [4].

Figure2 depicts the role of the process mining with the world .Software systems controls and supports business processes and also generates event logs based on recorded events, transactions and messages. Based on Event logs generated discovery, conformance and enhancement of process models are organized and which in turn analyzes the business processes.

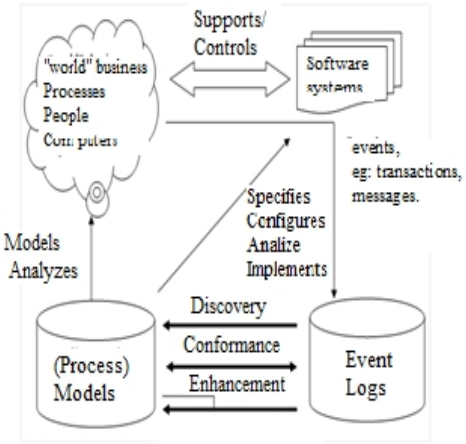

Figure 2: Three Phases Of Process Mining

There is a huge potential for process mining in healthcare as it allows hospitals to analyze and restructure their processes. Event log data may be derived from various data sources in a hospital like intensive care unit, pathology, radiology etc [5]. The log will be analyzed and good process models are generated using alpha algorithm. Using the models generated, the health care centers can compare and improve their process which is beneficial for both healthcare centers as well as target population which reduces the time as well as cost being incurred for unnecessary tests. Further various types of algorithms depicted by various authors are discussed in section2, the actual flow of the process of the proposed method is studied in section3 then the proposed algorithm is discussed in section4 whereas the experimental analysis of the proposed algorithm has been depicted in section5.

2. RELATED WORK

In this section, various contributions of process mining and various existing algorithm contribution are discussed.

[6] The study aimed to find out the probable of a process mining technique to determine an outpatient care process that can be utilized for further improvements. Use of heuristic miner, fuzzy miner, and comp miner of ProM process mining techniques can be used to discover outpatient care process and confirmed that process mining techniques can be useful even in a healthcare environment where a variety of equipments and systems are used.

[7] Provides various process mining techniques to obtain significant knowledge about flows, e.g., to discover usual paths followed by particular groups of patients. This is a significant task given the active nature of healthcare processes. This paper demonstrates the workflow of Dutch hospital for a real case of a gynaecological oncology process with the help of process mining. It uses heuristic miner algorithm which can deal with noise and exceptions so that the user can deal with the actual process.

[8]Describes a case study to illustrate a process mining method for healthcare processes management and improvement. It uses alpha algorithm, fuzzy miner and simulation model of ProM to provide a basic overview of workflow discovery. Finally it tries to exhibit the potential of process mining in health care systems.

[9]Current trends in health management improvement demand the consistency of care protocols to achieve better quality and efficiency. The problems inherent to the design of current Clinical Pathways in this way requires new specific devise tools to help making the system useful. A pattern recognition algorithm (PALIA Algorithm) based in an evolution of the Process Mining classical paradigm is presented and evaluated as a solution to this scenario. The proposed algorithm is

Journal of Theoretical and Applied Information Technology
30th June 2016. Vol.88. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

able to infer Clinical Pathways from functioning logs to support the design of Clinical Pathways.

[10]Process mining was applied with success to determine, characterize and analyze user activities recorded from VEPR. Worth noting the completion of tasks observed after log out, which reveals security inconvenience, and which is not an mistake at log level as were first thought. It uses Fuzzy miner algorithm to determine the most frequent paths followed by the user.

3. SYSTEM MODEL

Process mining is the bridge between model-based process analysis and data-oriented analysis techniques. Through concrete data sets and easy to use software, data science knowledge can be applied directly to analyze and improve processes in a variety of domains. Process discovery corresponds to Play-In wherein it concentrates on four attributes: fitness test, precision, simplicity, generalization. To improve the processes, we are going to discover good process models which help to increase the efficiency of the health care operation. The input for the system is event logs which is being collected from the information system of health care.

To generate process models , the use of analytical tool called ProM and Disco , where event logs being input and its corresponding process models being the output which can be optimized. The first step Play-In, where the event logs are used to generate process models with the help of analytical tool ProM as shown in Figure 3.

Figure 3: The Play-In Process

The data which is being recorded in the log files will be represented as model. In the second step, consider the model generated in previous step and verify if reality is matched as recorded in the log file. Using Play-out the history of process models can be generated. If the event recorded is not matching with the pre defined model, conformance checking is done for the data which shows the deviations from the actual process as shown in Figure 4. Conformance checking is done with fitness test which reveals the worst and the best cases.

Figure 4: The Play-Out Process

Figure 5(a) depicts the actual model and Figure 5(b) shows the deviation that has occurred from the actual model. The model suggests some set of predefined rules that has to be followed in each case. When the rule has been violated, deviation from the actual process occurs which leads to inefficiency.

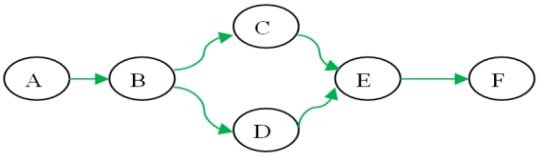

Figure 5(A): The Actual Process Model

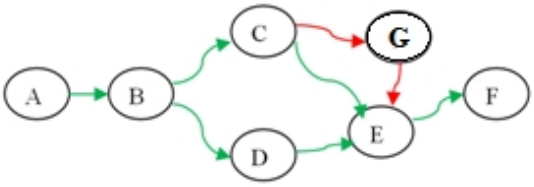

Figure 5(B): The Model With Deviation

4. ALPHA FITNESS ALGORITHM

Process mining techniques can be applied to various domains for streamlining the organizational processes. Process discovery and conformance checking plays a vital role as the workflow process contains concurrency, loops and choices that has to be matched with model behaviour and observed behaviour. Through conformance checking we would like to diagnose the difference from the actual process through token based fitness test.

Algorithm: To depict the best and the worst fit.

Let L be an event log over T. α(L) is defined as follows.

$TL = \{ t \in T | \exists \sigma \in L\ p \in \sigma \}$,
$T_I = \{ t \in T | \exists \sigma \in L\ p = first(\sigma) \}$,
$T_0 = \{ t \in T | \exists \sigma \in L\ p = last(\sigma) \}$,

α(L) =(PL,TL,FL).
Calculate token K for M,C,R,P
Calculate fitness(σ)
If f(σ)=1,
Best fit
Else
If f(σ)= 0,
worst fit(deviation)
end if
end

Journal of Theoretical and Applied Information Technology
30th June 2016. Vol.88. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

First four steps of the above algorithm generates a process models for the logs recorded using α-algorithm. Further to the model that is generated, logs are provided as input and checked for synchronization using the fitness formula(1). The fitness test contains 4 attributes namely P=produced tokens, C=consumed tokens, M=missing tokens, R=remaining tokens and T being the transitions recorded in the log, t is the traces from the log T_1 is the first entry in the log file and T_0 is the last entry in the log file

Fitness of a particular case will be calculated using the equation (1) which reveals the worst case or best case where worst case being 0 and best case being 1.

$$\text{fitness } (\sigma) = \frac{1}{2}\left(1 - \frac{M}{C}\right) + \frac{1}{2}\left(1 - \frac{R}{P}\right) \quad (1)$$

The above formula can be lifted to the entire log file and can check for the fitness as given in (2). Tokens for each instance will be provided. Token will be given during initialization and finalization. In the beginning of the process a token is produced for source place p=1. While at the end a token is consumed from the sink place. Later if there is no token remaining at the end, it declares that the trace has been terminated.

$$f(\sigma) = \frac{1}{2}\left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times MN,\sigma}{\sum_{\sigma \in L} L(\sigma) \times CN,\sigma}\right) + \frac{1}{2}\left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times RN,\sigma}{\sum_{\sigma \in L} L(\sigma) \times PN,\sigma}\right) \quad (2)$$

5. EXPERIMENTAL RESULTS

In this section the implementation of the above algorithm is discussed in order to discover the process and the deviations that occur in the processes. Event logs being the baseline for the model generation. The main finding is to check for the deviation from the actual process.

Case ID	Resource	Activty	Starting Time	Ending Time
1	mdc_1_anemia	bl-sam	0:00:00	0:15:00
1	mdc_1_anemia	HB	0:15:00	0:25:00
1	mdc_1_anemia	Peripheral smear	0:15:00	0:25:00
1	mdc_1_anemia	RBC count	0:15:00	0:25:00
1	mdc_1_anemia	Reticulocyte count	0:15:00	0:35:00
1	mdc_1_anemia	PCV	0:15:00	0:25:00
2	mdc_2_Fever	TC	0:15:00	0:50:00
2	mdc_2_Fever	DC	0:15:00	0:55:00
2	mdc_2_Fever	ESR	0:15:00	1:45:00
2	mdc_2_Fever	CBC	0:15:00	1:15:00
2	mdc_2_Fever	Widal test	0:15:00	0:45:00
2	mdc_2_Fever	Dengue IGM(card)	0:15:00	0:49:00
2	mdc_2_Fever	Chikungunya IGM(c	0:15:00	0:48:00
2	mdc_2_Fever	Brucella Agglutinati	0:15:00	0:45:00

Figure 6: The Event Logs Of Blood Test

Considering few diseases and blood tests that should be conducted for that particular disease is shown in figure 6. The model for case1 and 2 from the event log has been generated as shown in Figure 7(a), Figure 7(b), while the combination of these cases is shown in Figure 7(c) using ProM. Further considering other cases, models for each disease is generated. The models are being compared with the recorded log and checked for synchronization. A prior model will be proposed for the system and fitness will be checked for it. A set of predefined steps which determines succession, choices, causality will be proposed. The log should follow these set of rules which gives the best fit for that particular case. If the log fails to follow the predefined rules it leads to worst fit which means it contains deviation from the actual process.

Finally the proposed algorithm will help in discovering good process model wherein the health care centers can follow. The proposed model specifies the test that has to be conducted for that particular disease in the specific order. If there are any unnecessary test that has been conducted will be highlighted. Thus by following this model, the healthcare centers can improve their efficiency while focusing on timeliness as well.

Consider Table 1 where for a particular disease A, B, C, D and E is the actual test process. Case1 satisfies the test and hence a best fit.

Table 1:The Tests Conducted For Particular Disease

	A	B	C	D	E
Case1	X	X	X	X	X

Table 2: The Tests Conducted For Particular Disease With Some Deviation

	A	B	C	D	G	E	F
Case1	X	X	X	X		X	
Case2	X	X	X	X	**X**	X	
Case3	X	X	X	X	X	X	**X**

Consider table 2 where Case 1 fails to perform adding one more test G i.e. A, B, C, D, G and E which leads to worst fit in turn reduces the efficiency of the process and in third case an extra test has been conducted which is not there in the actual process which is redundant. Hence the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

proposed model highlights the deviations from the process wherein the healthcare centers can restructure the process accordingly.

5. CONCLUSION

The paper, focused on how process mining techniques can be applied in the healthcare domain, by generating process model and check for the fitness. It also checked for the process with recorded log for correctness of the process. The fitness test highlights the deviation in the process. Thus by implementing the proposed algorithm, the efficiency of process can be improved which in turn reduces the time and cost for the tests which are being conducted unnecessarily in health care centers. Future work can be an extension of implementing alpha-fitness algorithm for diagnosing all the tests in the health care sectors like x-ray, scanning, ECG and etc.

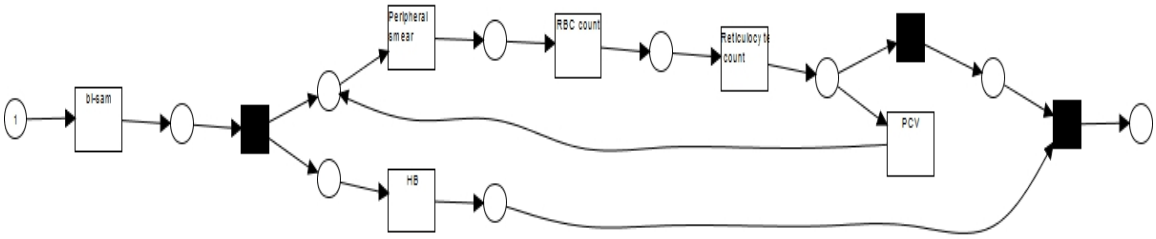

Figure 7(a):The model obtained for log for case 1.

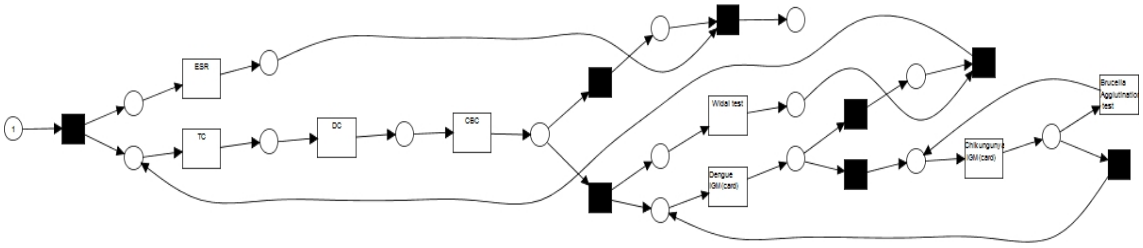

Figure 7(b): The Model Obtained For Log For Case 2.

Figure 7(C): The Model Obtained For Log For Combination Of Case 1 And Case 2.

REFRENCES:

[1] Kaymak U, Mans R, van de Steeg T, Dierks M. On process mining in health care. InSystems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on 2012 Oct 14 (pp. 1859-1864). IEEE.

[2] Homayounfar P. Process mining challenges in hospital information systems. InComputer Science and Information Systems (FedCSIS), 2012 Federated Conference on 2012 Sep 9 (pp. 1135-1140). IEEE.

[3] Van Der Aalst W. Process mining: overview and opportunities. ACM Transactions on Management Information Systems (TMIS). 2012 Jul 1;3(2):7.

[4] Sonawane SB, Patki RP. Process Mining by using Event Logs. International Journal of Computer Applications. 2015 Jan 1;116(19).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[5] Mans RS, van der Aalst WM, Vanwersch RJ, Moleman AJ. Process mining in healthcare: Data challenges when answering frequently posed questions. InProcess Support and Knowledge Representation in Health Care 2013 (pp. 140-153). Springer Berlin Heidelberg.

[6] Kim E, Kim S, Song M, Kim S, Yoo D, Hwang H, Yoo S. Discovery of outpatient care process of a tertiary university hospital using process mining. Healthcare informatics research. 2013 Mar 1;19(1):42-9.

[7] Mans RS, Schonenberg MH, Song M, van der Aalst WM, Bakker PJ. Application of process mining in healthcare–a case study in a dutch hospital. InBiomedical Engineering Systems and Technologies 2008 Jan 28 (pp. 425-438). Springer Berlin Heidelberg.

[8] Zhou Z, Wang Y, Li L. Process mining based modeling and analysis of workflows in clinical care-A case study in a chicago outpatient clinic. InNetworking, Sensing and Control (ICNSC), 2014 IEEE 11th International Conference on 2014 Apr 7 (pp. 590-595). IEEE.

[9] Fernández-Llatas C, Meneu T, Benedí JM, Traver V. Activity-based process mining for clinical pathways computer aided design. InEngineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE 2010 Aug 31 (pp. 6178-6181). IEEE.

[10] Rebuge Á, Velez Lapao L, Freitas A, Cruz-Correia R. A process mining analysis on a Virtual Electronic Patient Record system. InComputer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on 2013 Jun 20 (pp. 554-555). IEEE.

[11] Yang WS, Hwang SY. A process-mining framework for the detection of healthcare fraud and abuse. Expert Systems with Applications. 2006 Jul 31;31(1):56-68.

[12] van der Heijden TH. Process mining project methodology: Developing a general approach to apply process mining in practice. Master of Science in Operations Management and Logistics. Netherlands: TUE. School of Industrial Engineering. 2012.

[13] Bozkaya M, Gabriels J, Werf JM. Process diagnostics: a method based on process mining. InInformation, Process, and Knowledge Management, 2009. eKNOW'09. International Conference on 2009 Feb 1 (pp. 22-27). IEEE.

[14] Zhang X, Chen S. Pathway identification via process mining for patients with multiple conditions. InIndustrial Engineering and Engineering Management (IEEM), 2012 IEEE International Conference on 2012 Dec 10 (pp. 1754-1758). IEEE.

[15] Rebuge Á, Ferreira DR. Business process analysis in healthcare environments: A methodology based on process mining. Information Systems. 2012 Apr 30;37(2):99-116.