RESEARCH ARTICLE OPEN ACCESS

Extract Business Process Performance using Data Mining

Faisal M. Nafie *, Mergani.A.Eltahir **
**(Department of Computer Science, Majmaah University, Saudi Arabia*
*** (Department of Information Technology, Nile Valley University, Sudan*

ABSTRACT
This paper aimed to analyze the performance of the business process using process mining. The performance is very important especially in large systems .the process of repairing devices was used as case study and the Fuzzy Miner Algorithm used to analyze the process model performance
Keywords: Data mining, Workflow Mining, Process Mining, Business Process.

I. INTRODUCTION

Nowadays there are a lot of Information Systems in companies, organizations, educational institutes, and etc. This means the data already found but this data not used in the ideal way to be more useful, because the normal way that used to get data not highly enough to give amore and intelligent information. Business Process may contain a lot data because its use a process covers many business steps to achieve the goal. Data mining is one of the useful ways to extract and data and obtains more valuable information.

Data Mining:

Data mining is an essential step in the knowledge discovery in databases (KDD) process that produces useful patterns or models from data [1].Data Mining becomes more considerable nowadays especially in the industry[2] .
KDD process consists of iterative sequence methods as follows[x]:

1. Selection: Selecting data relevant to the analysis task from the database.
2. Pre-processing: make the data more consist by Removing noise and inconsistent data; and then combining multiple data sources.
3. Transformation: Transforming data into appropriate forms to perform data mining.
4. Data mining: Choosing a data mining algorithm which is suitable to pattern in the data and perform the extraction to the data patterns.
5. Interpretation/Evaluation: Interpreting the patterns into knowledge by removing Redundant or irrelevant patterns; translating the useful patterns into terms that human-understandable. Data Mining can be used many fields to achieve the many purposes as follows:

1- Business: Data Mining can be applied in retail, banking, and insurance to fraud detection, and marketing analysis.
2- Education: Data Mining can be applied in predicting student's performance, understand student behaviour, and course scheduling.
3- Science and Engineering: Data Mining can be applied in bioinformatics, medicines, and genetics.
4- Telecommunications: Data Mining can be applied to facilitate many other beneficial communication services including fax, pager, cellular phone, Internet messenger, images, e-mail, computer and Web data transmission, and other data traffic [3][4].

Business Process:

A business process is the combination of a set of activities within an enterprise with a structure describing their logical order and dependence whose objective is to produce a desired result. The activities that together create a customer value, Business process modeling enables a common understanding and analysis of a business process [5] [6].

Process Mining:

The goal of process mining is to extract information about processes from transaaction logs .Process mining aims at the automatic construction of models explaining the behavior observed in the event log [7]. Process mining encompasses "techniques, tools and methods to discover, monitor and improve real processes by extracting knowledge from event logs. It is an innovative approach and builds a bridge between data mining and business process management [8]. Process Mining helps in view the interaction between the parts of the whole organization and also helps in [8]:

1. **Find problems:** Locate anomalies even before they become problems.
2. **Understand root causes:** Identify the underlying problems instead of immediately visible symptoms.

3. **Achieve compliance**: Uncover non-compliant situations and uncover incidental non-compliance that would otherwise be impossible to locate.
4. **Compare performance**: Compare departments or similar organizations based on verifiable data, not hunches.
5. **Validate process changes**: Establish baselines for the current situation and use them to determine whether specific changes are effective or not.
6. **Understand the entire process:** Departments often focus on completing their own subtask, unaware of the complete chain of actions that bring a product or service to the customer.

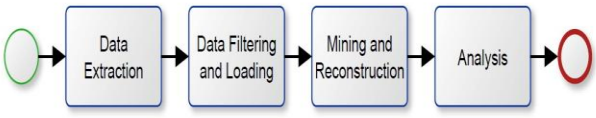
Gain transparency: Since causes of problems can be identified and seen in the context of the organization's overall workflow, employees understand where and why change is needed

Process Models and Event Logs

The aim of process mining is the construction of process models based on available logging data. In the context of information system science a model is an immaterial representation of its real world counterpart used for a specific purpose. Models can be used to reduce complexity by representing characteristics of interest and by omitting other characteristics [11]. A process model is a graphical representation of a business process that describes the dependencies between activities that need to be executed collectively for realizing a specific business objective. It consists of a set of activity models and constraints between them [12]. Process models can be represented in different process modelling language for example using the Business Process Model and Notation (BPMN. BPMN provides more intuitive semantics that are easier to understand for recipients that do not possess a theoretical background in informatics. We therefore rely on BPMN models for illustration in this paper.

Mining Procedure

Fig 1 provides an overview of the different process mining activities. Before being able to apply any process mining technique it is necessary to have access to the data. It needs to be extracted from the relevant information systems. This step is far from trivial. Depending on the type of source system the relevant data can be distributed over different database tables. Data entries might need to be composed in a meaningful manner for the extraction. Another obstacle is the amount of data.Depending on the objective of the process mining up to millions of data entries might need to be extracted which



requires efficient extraction methods. A further important aspect is confidentiality. Extracted data might include personalized information and depending on legal requirements anonymization or pseudonymization might be necessary.

Fig 1 Illustration of Mining Procedure

Before the extracted event log can be used it needs to be filtered and loaded into the process mining software. There are different reasons why filtering is necessary. Information systems are not free of errors. Data may be recorded that does not reflect real activities. Errors can result from malfunctioning programs but also from user disruption or hardware failures that leads to erroneous records in the event log. Other errors can occur without incorrect processing. A specific process is normally analyzed for a certain time frame. When the data is extracted from the source system process instances can get truncated that were executed over the boundaries of the selected time frame. They need to be deleted from the event log or extracted completely. Otherwise they lead to erroneous results in the reconstructed process models. Event logs commonly do not exclusively contain data for a single process. Filtering is necessary to curtail the event log so that it only contains events that belong to the scrutinized process. Such a filtering needs to be conducted carefully because it can lead to truncated process instances as well. A common criterion is the selection of activities that are known to belong to the same process. Data filtering and loading is commonly supported by software tools and performed in a single step. But it can also be done separately. Once the data is loaded into the process mining software the actual mining and Mining Algorithms.

Mining Algorithms

The main component in process mining is the mining algorithm. It determines how the process models are created. A broad variety of mining algorithms does exist. The following three categories will be discussed in more detail. Various advanced mining algorithms do exist that can be used for different purposes. Fig 2 illustrates the mined model using the heuristic Fuzzy Miner algorithm. The model does not follow the BPMN notation but instead uses a dependency graph representation. It does not contain any gateway operators but shows

the dependencies between different activities. The dependency graph illustrates for example that A was followed three times by B and two times by C [9] [10].

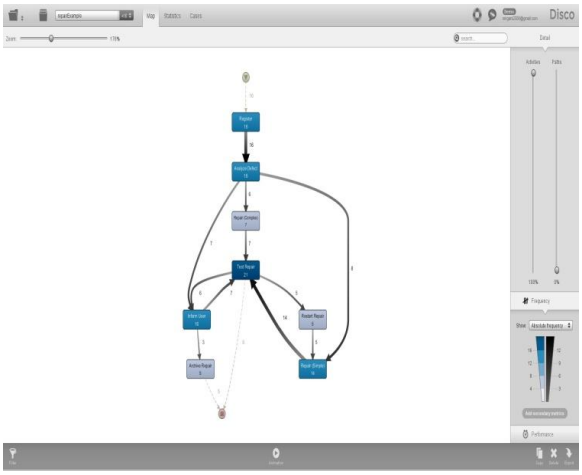


Figure 2 Mined Process Model Using the Fuzzy Miner Algorithm3

It is important to identify which requirements need to be considered for achieving the intended objectives for each individual process mining project. The appropriateness of an algorithm should be evaluated depending on the area of application

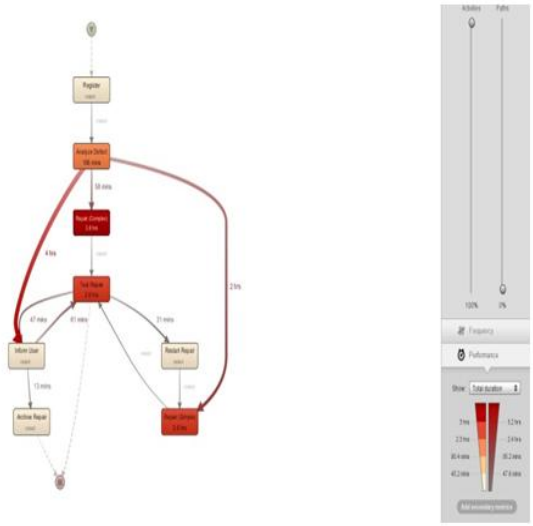


Fig 3 Show Process Mining Performances

II. CASE STUDY - REPAIR SERVICE PROCESS

To optimize a process, one has to first understand the as-Is process. And this is usually far from simple, because business processes are performed by a number of people, often across different organizational units or even companies. Everybody only sees a part of the process. The manual discovery through classical workshops and interviews is costly and time-consuming, remains incomplete and subjective. With Process Mining tools it is possible to leverage existing IT data from operational systems to quickly and Objectively visualize the as-Is processes as they are really taking place.

In workshops with process stakeholders one can then focus on the root cause analysis and the value-adding process improvement activities. In one of our projects we have analyzed a Repair Service Process of a big electronics manufacturer. The following process description has been slightly changed to protect the identity of the manufacturer. The starting point for the project was the feeling of the project manager that the process had severe problems. Customer complaints and the inspection of individual cases indicated that there were inefficiencies and too long throughput times in the process. The project was performed in the following phases: First, the concrete questions and problems were collected, and the IT logs of all cases from the running business year were extracted from the corresponding service platform. The log data were then analyzed together with the process managers in an interactive workshop.

Fig 4 shows a business process model of a simple Repair Service Process. It starts with the registration of the device defect information then the defect analyzed, there are three situations, first one is defect may be complex then repair and test repair after that inform user or restart the repair, if user informed and need more repair then test repair, or restart repair loop between inform user, test repair, restart repair and repair and at the end inform user and archive repair. In the second situation after analyzed defect info user then archive defect or test repair if test repair then, either inform user or test repair simple till end the case by archiving or test repair. In the third situation if simple then repair and test the repair then either test repair or inform user the case ended by archiving repair or optionally by test repair.

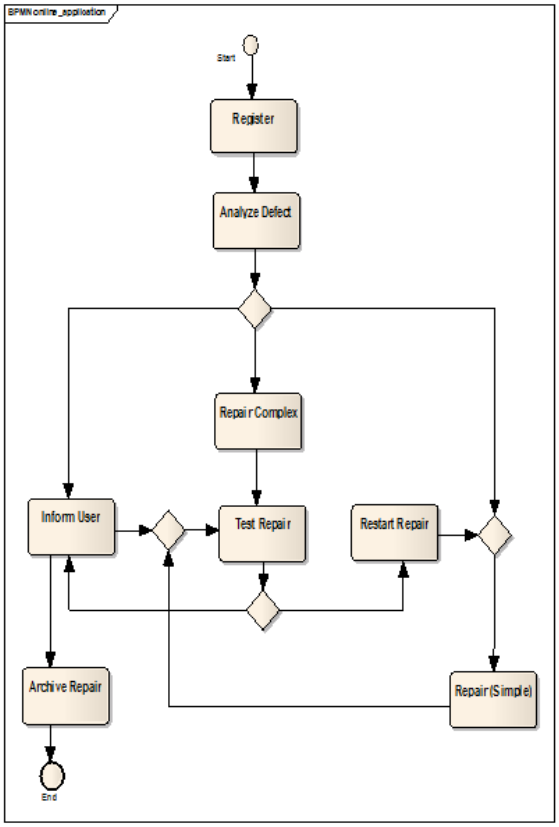


Fig 4 show process mining

III. THE RESEARCH METHOD

The approach used in process mining for answering this question bases on the exploitation of data stored in information systems that is created during the Processing of business transactions. An information system stores data in log files or database tables when processing transactions. In the case of issuing an order data about the type and quantity of ordered goods, preferred suppliers, time of ordering etc. gets recorded. The stored data can be extracted from the information system and be made available in so called event logs. They constitute the data basis for process mining algorithms.

An event log is basically a table. It contains all recorded events that relate to executed business activities. Each event is mapped to a case. A process model is an abstraction of the real world execution of a business process. A single execution of a business process is called process instance. They are reflected in the event log as a set of events that are mapped to the same case. The sequence of recorded events in a case is called trace. The model that describes the execution of a single process instance is called process instance model. A process model abstracts from the single behaviour of process instances and provides a model that reflects the behaviour of all instances that belong to the same process. Cases and events are characterized by classifiers and attributes. Classifiers ensure the distinctness of cases and events by mapping unique names to each case and event. Attributes store additional information that can be used for analysis purposes. An example of an event log is given in Table 1.

Table (1): Event Log

	A	B	C	D	E	F	G	H	I	J	K
1	Case IC	Activity	Resource	Start Ti	Comple	(case) description	defectFix	defectTyp	lifecycle:transi	numberRepairs	phoneType
2	1	Register	System	23:00.0	23:00.0	Simulated process instance			complete		
3	1	Analyze Defe	Tester3	23:00.0	30:00.0	Simulated process instance			6 complete		T2
4	1	Repair (Comp	SolverC1	31:00.0	49:00.0	Simulated process instance			complete		
5	1	Test Repair	Tester3	49:00.0	55:00.0	Simulated process instance	TRUE		complete		0
6	1	Inform User	System	10:00.0	10:00.0	Simulated process instance			complete		
7	1	Archive Repa	System	10:00.0	10:00.0	Simulated process instance	TRUE		complete		0
8	1	End	End	10:00.0	10:00.0	End	End	End	End	End	End
9	10	Register	System	09:00.0	09:00.0	Simulated process instance			complete		
10	10	Analyze Defe	Tester2	09:00.0	15:00.0	Simulated process instance			3 complete		T1
11	10	Repair (Simp	SolverS1	35:00.0	42:00.0	Simulated process instance			complete		
12	10	Test Repair	Tester6	42:00.0	48:00.0	Simulated process instance	FALSE		complete		1
13	10	Restart Repai	System	54:00.0	54:00.0	Simulated process instance			complete		
14	10	Repair (Simp	SolverS2	54:00.0	03:00.0	Simulated process instance			complete		
15	10	Inform User	System	55:00.0	55:00.0	Simulated process instance			complete		
16	10	Test Repair	Tester4	03:00.0	09:00.0	Simulated process instance	TRUE		complete		2
17	10	Archive Repa	System	14:00.0	14:00.0	Simulated process instance	TRUE		complete		2
18	10	End	End	14:00.0	14:00.0	End	End	End	End	End	End
19	100	Register	System	28:00.0	28:00.0	Simulated process instance			complete		
20	100	Analyze Defe	Tester4	28:00.0	36:00.0	Simulated process instance			8 complete		T2
21	100	Repair (Comp	SolverC1	52:00.0	09:00.0	Simulated process instance			complete		
22	100	Test Repair	Tester1	09:00.0	18:00.0	Simulated process instance	TRUE		complete		0
23	100	Inform User	System	20:00.0	20:00.0	Simulated process instance			complete		
24	100	Archive Repa	System	28:00.0	28:00.0	Simulated process instance	TRUE		complete		0
25	100	End	End	28:00.0	28:00.0	End	End	End	End	End	End
26	1000	Register	System	23:00.0	23:00.0	Simulated process instance			complete		
27	1000	Analyze Defe	Tester6	23:00.0	28:00.0	Simulated process instance			3 complete		T3
28	1000	Repair (Simp	SolverS3	35:00.0	49:00.0	Simulated process instance			complete		
29	1000	Test Repair	Tester6	49:00.0	58:00.0	Simulated process instance	TRUE		complete		1
30	1000	Inform User	System	03:00.0	03:00.0	Simulated process instance			complete		
31	1000	Archive Repa	System	08:00.0	08:00.0	Simulated process instance	TRUE		complete		1
32	1000	End	End	08:00.0	08:00.0	End	End	End	End	End	End
33	1001	Register	System	51:00.0	51:00.0	Simulated process instance			complete		
34	1001	Analyze Defe	Tester3	51:00.0	56:00.0	Simulated process instance			8 complete		T1
35	1001	Repair (Comp	SolverC3	06:00.0	54:00.0	Simulated process instance			complete		
36	1001	Inform User	System	35:00.0	35:00.0	Simulated process instance			complete		
37	1001	Test Repair	Tester5	54:00.0	00:00.0	Simulated process instance	TRUE		complete		0
38	1001	End	End	00:00.0	00:00.0	End	End	End	End	End	End
39	1002	Register	System	40:00.0	40:00.0	Simulated process instance			complete		
40	1002	Analyze Defe	Tester2	40:00.0	50:00.0	Simulated process instance			3 complete		T3
41	1002	Repair (Simp	SolverS3	12:00.0	32:00.0	Simulated process instance			complete		
42	1002	Inform User	System	26:00.0	26:00.0	Simulated process instance			complete		
43	1002	Test Repair	Tester4	32:00.0	40:00.0	Simulated process instance	TRUE		complete		1
44	1002	End	End	40:00.0	40:00.0	End	End	End	End	End	End
45	1003	Register	System	43:00.0	43:00.0	Simulated process instance			complete		
46	1003	Analyze Defe	Tester5	43:00.0	49:00.0	Simulated process instance			1 complete		T3
47	1003	Repair (Simp	SolverS1	08:00.0	14:00.0	Simulated process instance			complete		
48	1003	Test Repair	Tester6	14:00.0	20:00.0	Simulated process instance	FALSE		complete		1
49	1003	Inform User	System	18:00.0	18:00.0	Simulated process instance			complete		
50	1003	Restart Repai	System	27:00.0	27:00.0	Simulated process instance			complete		
51	1003	Repair (Simp	SolverS3	27:00.0	41:00.0	Simulated process instance			complete		
52	1003	Test Repair	Tester6	41:00.0	51:00.0	Simulated process instance	TRUE		complete		2
53	1003	End	End	51:00.0	51:00.0	End	End	End	End	End	End

IV. DISCUSSION

Organizations have been focusing more and more on their business processes in order to improve results and reduce costs. A business process can be defined as a related group of tasks that together create value for a (internal or external) customer. Examples of business processes are the processing of orders in a factory or the handling of insurance claims at an insurance company.

Throughout this paper, the emphasis of our research has been on the analysis of the performance of business processes.The Performance Analysis process of repairing devices was used as case study and the Fuzzy Miner Algorithm used to analyze the process model performance. The success of process mining will depend on whether it is able to achieve balance between these conflicting goals sensibly. Fuzzy Mining is a first step in that direction.

V. CONCLUSION

Process mining builds the bridge between data mining and business process management. The increasing integration of information systems for supporting and automating the execution of business transactions provides the basis for novel types of data analysis. The data that is stored in the information systems can be used to mine and reconstruct business process models. These models are the foundation for a variety of application areas including process analysis and optimization or conformance and compliance checking. The basic constructs for process mining are event logs, process models and mining algorithms. We have summarized essential concepts of process mining in this article, illustrating the main application areas and available tool support Process mining is still a young research discipline and limitations concerning noise, adequate representation and competing quality criteria should be taken into account when using process mining. Although some areas like the labelling of events, complexity reduction in mined models and phenomena like concept drift need to be addressed by further research the available set of methods and tools provide a rich and innovative resource for effective and efficient business process management.

REFERENCES

[1]. Boehm B., A spiral model of software development and enhancement, *IEEE Computers, vol. 21, no. 5*, pp. 61-72, 1988.
[2]. Booch G., Rumbaugh J., and Jacobson I., *the Unified Modeling Language*,(Addison-Wesley, 1999).
[3]. Cutts Q., Fostering engagement, covering content: reflecting on large, mixed ability, first year computer programming module, in Walker M. (Ed), *Reconstructing Professionalism in University Teaching*,(Open University Press, pp. 168-189, 2001).
[4]. Favela J. and Pena-Mora F., An Experience in Collaborative Software Engineering Education, *IEEE Software*, pp. 47-52, March/April 2001.
[5]. Griffiths M., *Educational Research for Social Justice*, Open University Press, pp. 38-39, 1998.
[6]. Kermmis S., Action Research, in Hammersley M. (Ed), *Educational Research – Current Issues, vol. 1*, Open University Press, 1993.
[7]. Krutchen P., *The Rational Unified Process Model: an Introduction*, 2nd ed,(Addison-Wesley, 2000).
[8]. Krutchen P., *The 4+1 View Model of Architecture, IEEE Software, vol. 12, no. 6*, 1995.
[9]. Günther, C., and Van der Aalst, W. 2007. *Fuzzy mining–adaptive process simplification based on multi-perspective metrics*, Business Process Management, pp. 328–343.
[10]. Van der Aalst, W. M. P., Weijters, A. J. M. M., and Maruster, L. 2002. *Workflow Mining: Which Processes can be Rediscovered, Vol. 2480*, Presented at the Proc. Int'l Conf. Eng. and Deployment of Cooperative Information Systems (EDCIS 2002), pp. 45–63.
[11]. Becker, J., Delfmann, P., Eggert, M., and Schwittay, S. 2012a. Generalizability and Applicability of Model-Based Business Process Compliance-Checking Approaches – A State-of-the-Art Analysis and Research Roadmap, *BuR -Business Research* (5:2), pp. 221–247.
[12]. Weske, M. 2012. *Business process management concepts, languages, architectures*,) Berlin, New York: Springer .)