

Process Mining Methodology in Industrial Environment: Document Flow Analysis

Paweł Markowski, Michał R. Przybyłek
 Polish-Japanese Academy of Information Technology
 Warsaw, Poland
 Email: {pawel.markowski, mrp}@pjatk.edu.pl

Abstract—A real-world company asked us to solve emerging problem related to the flow of the documents: some of the documents had been lost. We used methods of process mining to solve this problem. We found an error in the information system. Moreover, our analysis of the flow of the documents led to the conclusion that the business processes did not meet the company’s needs, and with our help they were redesigned and reimplemented. This research gave an additional important insight to process mining methodology — in order to extract knowledge about business processes that is useful for the decision makers, it is crucial to create algorithms capable of collaborating with human experts.

I. INTRODUCTION

IN CONTEMPORARY business enterprises the complexity of real-world processes has to be perceived as one of the greatest obstacles in achieving effectiveness. More and more enterprises face with the problem of redesigning their business processes in order to survive in the global economy. The competitive market creates the demand for high quality services at lower costs and with shorter cycle times. Even relatively small companies are frequently confronted — in their daily routines — with processes of very high complexity. In such environment business processes must be identified, described, understood and analyzed to find inefficiencies, which cause financial losses.

One way to achieve it is modeling. Business modeling is the first step towards defining a software system. It enables companies to look afresh at how to improve organization and to discover the processes that can be solved automatically by software that will support the business. However, as it often happens, such a developed model corresponds more to how people think of the processes and how they wish the processes would look like, than to the real processes as they take place.

Another way is by extracting information from a set of events gathered during executions of a process. Process mining [1], [2], [3], [4], [5], [6], [7], [8], [9] is a new and prosperous technique that allows to extract a model of a business process based on information gathered during real executions of the process. The methods of process mining are used when there is no enough information about processes (i.e. there is no a priori model), or there is a need to check whether the current model reflects the real situation (i.e. there is a priori model, but of a dubious quality). One of the crucial advantages of process mining over other methods is its objectiveness — models discovered from real executions of a

process are all about the real situation as it takes place, and not about how people think of the process, and how they wish the process would be like. In this case, the extracted knowledge about a business process may be used to reorganize the process to reduce its time and cost for the enterprise.

Table I shows a typical event-log gathered from a flow of documents in a company. This event-log is a list of entries that contain information about observable actions of a process:

- “Actor” is someone who triggered the event; in our example an “Actor” is described by his full name (we left visible only two first letters of actors’ surname)
- “Time Stamp” is the exact time of the event; in example from Table I “Time Stamp” indicates the exact date and time when the event occurred
- “Event” is an observable action of the event (we shall assume, that we are given only some rough information about the real actions); in example from Table I there are four possible types of events:
 - “Created” — starting point of approval process triggered by new document that was created in the system
 - “GL app” — first level of acceptance ; “Actor” is directly responsible for the area, which document content concerns. Approval confirms that the description coincides with reality
 - “Quality app” — second level of acceptance ; “Actor” is a Quality Engineer and confirms that the change does not have negative impact in quality area
 - “APU app” — 3rd and last approval level ; “Actor” is an Area Supervisor/Manager and confirms that the organization is ready for this kind of described change and costs regarded with it were included in forecast (money, which have to be spent are booked)
 - “Approved” — document has been confirmed by all required employees and is registered in SAP
 - “Rejected” — document has been rejected in approval path ; rejection reason attached to this document
- “Case ID” is an instance of the process that executed the event

The column “Case ID” allows us to divide a list of events on collections of events corresponding to a particular executions of the process (i.e. to a particular instance of the process). The

Table I
AN EVENT LOG GATHERED FROM THE FLOW OF DOCUMENTS.

Actor	Time Stamp	Event	Case ID
Ko* Milosz	2015-11-27 15:28	Created	1
Ko* Milosz	2015-12-08 11:57	Created	4
Gr* Grzegorz	2016-01-04 16:05	Created	2
Gr* Grzegorz	2016-01-04 16:08	GL app	2
Ch* Arkadiusz	2016-01-04 18:30	Created	3
Ch* Arkadiusz	2016-01-04 18:31	GL app	3
Po* Aleksander	2016-01-05 11:15	Quality app	2
Po* Aleksander	2016-01-05 11:18	Quality app	3
Ko* Artur	2016-01-05 11:22	APU app	3
Se* Edyta	2016-01-05 19:38	Approved	3
Si* Aneta	2016-01-11 16:08	Created	5
Si* Aneta	2016-01-11 16:19	GL app	5
Po* Aleksander	2016-01-12 10:23	Quality app	5
Ko* Artur	2016-01-12 10:56	APU app	5
Se* Edyta	2016-01-13 11:05	Approved	5
Ko* Artur	2016-02-01 08:02	APU app	2
Se* Edyta	2016-02-01 18:51	Approved	2
Ja* Iwona	2016-02-15 16:31	Created	6
Ja* Iwona	2016-02-15 16:32	GL app	6
Do* Marek	2016-02-15 17:14	Created	7
Ch* Arkadiusz	2016-02-15 20:57	Created	9
Ch* Arkadiusz	2016-02-15 20:59	GL app	9
Ma* Wiktor	2016-02-15 21:12	Created	10
Ma* Wiktor	2016-02-15 21:13	Created	11
Tu* Lukasz	2016-02-16 10:14	Quality app	8
...

column “Time Stamp” makes it possible to linearly order each of the collections.

Figure 1 shows a model recognized from the log presented on Table I.

The aim of this paper is to show how methods of process mining can be applied to solve difficult problems in a real-world company. A company asked us to solve emerging problem related to the flow of documents: some of the documents had been lost. In the first case an error in the information system was identified and fixed. It also turned out, that the flow of the documents had been broken and after our analysis the whole process was redesigned and reimplemented.

Other examples of process mining usability in real-world companies are described in:

- *Process Mining in Healthcare*[11]
- *Process Mining Methodology for Health Process Tracking Using Real-Time Indoor Location Systems*[12]
- *A general divide and conquer approach for process mining*[17]
- *Process Mining Applied to the Test Process of Wafer Steppers in ASML*[13]

Health care process is modern subject for process mining researchers. One of the best environments for flow tracking are areas where human behaviors influence is the highest possible. Human errors, flow deviations, workarounds can be detected by process mining tools.

Our research gave an additional important insight to process mining methodology — in order to extract knowledge about business processes that is useful for the decision makers, it is crucial to create algorithms capable of collaborating with human experts. Optimization algorithms have been researched in

the area of decision support [10], but much less is known about algorithms for process mining in this context. One of the most significant methods is an explanation and justification of a mined model. This requirement stems from real business cases, where the best model becomes useless if it is not accepted by the decision maker. The lack of acceptance of a business process model can be due to several reasons: undisclosed user preferences; new constraints; a different evaluation; or simply a misunderstanding of the proposed model.

The paper is structured as follows. We checked ProM tools usability in Section II, where present complete document flow analysis. Readers can find there complex problem resolved after data extraction and analysis. We described results after document flow update that are effect of our common work with a company experts in *conformance checking* process [16]. We conclude the paper in Section III.

II. DOCUMENT FLOW ANALYSIS

Flow of the documents, which approvals are required is important due to financial control requirements. Unsupported flow occurrence might be a source of abuses, human errors or problems to meet customer expectations. Corporations have to take care about transparency of financial status required by internal procedures and headquarters policies. Internally developed solution has to ensure document approvals in line with employees permissions. This kind of software is vulnerable to errors. Bug can be source of problems like embezzlement results, communication errors or wrong conclusions. Part of organization that is responsible for programming or management has to ensure that the implemented program works according to the established process.

A. Background

Real-world company has a problem with documents flow. They had mismatch between documents’ flow system results and Financial Department’s reports. Company’s manager contacted us and ordered a complete service of process analysis. The company has implemented Microsoft Sharepoint to ensure proper flow of documents. Each document contains information about produced parts and represents their market value in currency. Developed addon is programmed with optional paths, which are depended on the value of created document. There is additional flow rule that the specialist opinion is needed in particular cases. This specialist is quality engineer. The higher value of a document is, the longer approvals path is required and the people involved in the acceptance process of the document have to be higher in the organization structure. Document creation or approval are triggers for system that generates event logs.

Financial controller and area supervisor have found a problem with budget forecasts. Anomaly was detected by differences in spending calculated by Financial Department and other department manager. Rules in the system looked fine, because each spending value was bigger than specified amount that had to be approved by the department manager. Operation Manager was informed about the situation and commissioned

the analysis. The analysis detected source of anomalies and after that Operation Manager gave instructions how to prevent such anomalies in the future.

Figure 1 presents a diagram mined by ProM Casual Activity Graph. A diagram is consistent with the flow implemented in Sharepoint. It shows that there are 3 levels of approval, but flow depends on specificity of a particular document, which path might be different. Collected logs contains 3 roles in approval system.

B. First level of process analysis CSV to XES

- Events log extraction from Sharepoint application to CSV file.
- CSV text file have to be imported to ProM and transformed to XES
- Analysis of imported data has to be preceded by questions, which may show proper direction of next mining methods

In process mining it is important to understand the data that have been collected. Without this knowledge it may be hard to extract any reasonable conclusion. The analysis of received data should be supported by a person who has theoretical knowledge about the assumptions. Too big deviations in comparison to process draft in organization might warn us about low data quality, which can be source of negative consequences in further analysis.

C. XES log summary

After CSV event log transformation to XES format ProM gives access to the document summary. One of the views is XES log summary where we found information about wrong starting point with 0,06% occurrences value. APU app is 3rd level approval in Figure 1. This is a point, which has to be verified during process exploration. It looks that a process was started from 3rd level acceptance even before its creation.

D. Directly follows graph extracted by ProM

ProM addon generates diagrams where nodes are events and edges are connections between them [15]. This visualization presents process steps flow, which are connected by arrows with labels presenting quantity of occurrences. Figure 1 is a directly follows graph mined from XES log. It transfers event logs into visualization of process steps.

Connection from 3rd role in acceptance process with an event announcing the creation of the document is an anomaly in this flow. The assumption is that the document has to be created before approval. Both in XES summary (Section II-C) and in Figure 1 is visible unusual event.

E. Extracting single case visualization

Process mining helped us with understanding the above anomaly, and we used an inductive visual miner (IvM [14]), which is dedicated ProM tool. It is able to extract information about process for every single instance and to visualize it. IvM allows to track the flow for every process step by graphical user interface. Inductive visual miner includes a functions

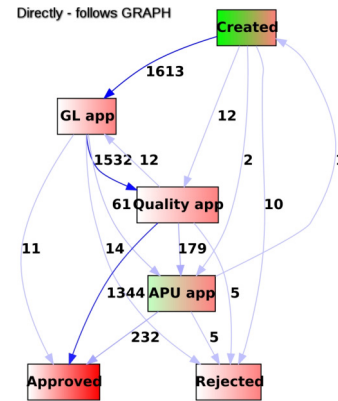


Figure 1. Process discovered by ProM from event log presented in Table I.

based on click event that allow us to extract additional information about process steps or single cases.

Figure 1 APU app(roval) was performed before document creation. Logs visualization shows that the next approval steps were performed even with this anomaly. Analyze of this case led to modifications in path structure. Additional check conditions were added to ensure proper document flow. The source of this problem was in improper database management and software bug. System changed first and second level of approval fields to empty but APU acceptance field did not. Modified document has APU acceptance from previous flow cycle. Trigger for process of documents' approval is document creation. After that instance of process have to be verified by authorized employees. Graph on Figure 1 shows that the created documents are addressed to GL app (first level of approval) in most of cases. This people are able to physical confirmation of compliance created document with the actual situation in the factory. They are working in 3 shift model so verification of documents is possible 24h per working days.

F. System validation after updates

Event logs were imported to ProM two months after system update. In order to confirm that a document approval path is in compliance with procedures. XES event log summary shows single starting point "Created". This summary confirms that the starting point is common and correct for each process instance. Process mining tools help corporation experts in analysis approvals path. Process approval path was redesigned to be in compliance with organization structure. This change is reducing human error risk by change that requires minimum 2 employees approval. Each step in approval path has a possibility to reject a document as like as in previous system version. Obligatory Quality approval step increased description coincidence with reality, because person responsible for that in 1st step knows about further Quality verification.

III. CONCLUSIONS

We investigated how methods of process mining can be applied to solve difficult problems in a real-world company. The paper shows that ProM functionality enables us to take

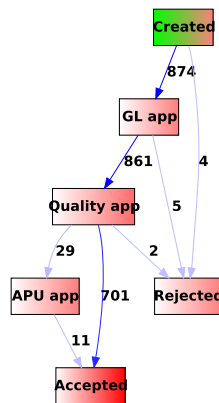


Figure 2. Document flow after system update

a look on processes not only from the designer perspective but also from the perspective of the production environment. IT systems are integral part of companies and sources of important data. Such data should be collected by system administrators, as they are valuable input in process extraction. Event logs gathered during execution of business processes present the reality in an objective way. Therefore, it is crucial to use them in further process optimization. They allow to look at the business process from the production perspective. Process mining has to be supported by people, who understands the relationship between discovered actions. The corporation that we worked with, solved emerging problem related to the flow of documents: some of the documents had been lost. An error in the information system was identified and fixed. It also turned out, that the flow of the documents had been broken and after our analysis the whole process was redesigned and reimplemented. The results obtained in process mining were the most important inputs for the system developers. The developers were able to prepare patches that fixed the problems. Process extraction realized by process mining tools was helpful in problem understanding and fixing. User friendly flow's representation combined with experts knowledge can be the key to successful system patch preparation and implementation.

Our research gave an additional important insight to process mining methodology — in order to extract knowledge about business processes that is useful for the decision makers, it is crucial to create algorithms capable of collaborating with human experts. Optimization algorithms have been researched in the area of decision support, but much less is known about algorithms for process mining in this context. One of the most significant methods is an explanation and justification of a mined model. This requirement stems from real business cases, where the best model becomes useless if it is not accepted

by the decision maker. The lack of acceptance of a business process model can be due to several reasons: undisclosed user preferences; new constraints; a different evaluation; or simply a misunderstanding of the proposed model. In future work we should focus more on human-computer interaction in process mining algorithms and systems.

REFERENCES

- [1] W.M.P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Springer Verlag, 2011.
- [2] A.J.M.M. Weijters, W.M.P. van der Aalst, *Process Mining: Discovering Workflow Models from Event-Based Data*, Proceedings of the 13th Belgium-Netherlands Conference on Artificial Intelligence, Maastricht, pp 283-290, 2001.
- [3] A.K.A de Medeiros, B.F. van Dongen, W.M.P. van der Aalst and A.J.M.M. Weijters, *Process Mining: Extending the alpha-algorithm to Mine Short Loops*, BETA Working Paper Series, WP 113, Eindhoven University of Technology, Eindhoven, 2004
- [4] W.M.P. van der Aalst, A.H.M. ter Hofstede, B. Kiepuszewski, A.P. Barros, *Workflow Patterns*, BPM Center Report BPM-00-02, BPMcenter.org, 2000.
- [5] W.M.P. van der Aalst and A.H.M. ter Hofstede, *Workflow Patterns: On the Expressive Power of (Petri-net-based) Workflow Languages*, BPM Center Report BPM-02-02, BPMcenter.org, 2002.
- [6] W.M.P. van der Aalst, A.J.M.M. Weijters, L. Maruster, *Workflow Mining: Discovering Process Models from Event Logs*, BPM Center Report BPM-04-06, BPMcenter.org, 2004.
- [7] M.T. Wynn, D. Edmond, W.M.P. van der Aalst, A.H.M. ter Hofstede, *Achieving a General, Formal and Decidable Approach to the OR-join in Workflow using Reset nets*, BPM Center Report BPM-04-05, BPMcenter.org, 2004.
- [8] W.M.P. van der Aalst, A.K. Alves de Medeiros, A.J.M.M. Weijters, *Process Equivalence in the Context of Genetic Mining*, BPM Center Report BPM-06-15, BPMcenter.org, 2006.
- [9] W.M.P. van der Aalst, M. Pesic, M. Song, *Beyond Process Mining: From the Past to Present and Future*, BPM Center Report BPM-09-18, BPMcenter.org, 2009.
- [10] A. Wierzbicki, M. Makowski, J. Wessels, *Model-Based Decision Support Methodology with Environmental Applications*, Springer, Series: Mathematical Modelling: Theory and Applications, Vol. 9, Berlin Heidelberg, 2000.
- [11] Mans, R.; van der Aalst, W.; Vanwersch, R. *Process Mining in Healthcare*, Springer Briefs in Business Process Management; Springer International Publishing: Cham, Germany, 2015
- [12] Carlos Fernandez-Llatas, Aroa Lizondo1, Eduardo Monton, Jose-Miguel Benedi, Vicente Traver *Process Mining Methodology for Health Process Tracking Using Real-Time Indoor Location Systems Sensors*, 11/2015; 15(12):29821-29840. DOI: 10.3390/s151229769
- [13] A. Rozinat, I.S.M. de Jong, C.W. Gunther, and W.M.P. van der Aalst *Process Mining Applied to the Test Process of Wafer Steppers in ASML* IEEE Transactions on Systems, Man and Cybernetics - Part C, 39:474–479, 2009.
- [14] Leemans, S.J.J., Fahland, D., Aalst, W.M.P.v.d. *Process and Deviation Exploration with Inductive Visual Miner* Proceedings of the BPM Demo Sessions 2014 Co-located with the 12th International Conference on Business Process Management (BPM 2014), Eindhoven, The Netherlands, September 10, 2014. (2014) 46
- [15] Leemans, S., Fahland, D., van der Aalst, W. *Exploring processes and deviations* Business Process Management Workshops (2014)
- [16] W. M. P. van der Aalst *A general divide and conquer approach for process mining* Federated Conference on Computer Science and Information Systems, 2013, pp. 1-10.
- [17] P. Homayounfar *A general divide and conquer approach for process mining* Proceedings of the Federated Conference on Computer Science and Information Systems, 2012, ISBN 978-83-60810-51-4, pp. 1135–1140