SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2018
Series: ORGANIZATION AND MANAGEMENT vol. 116 No. 1995

Edyta BRZYCHCZY, Aneta NAPIERAJ, Marta SUKIENNIK
AGH University of Science and Technology
Faculty of Mining and Geoengineering
brzych3@agh.edu.pl, aneta.napieraj@agh.edu.pl, marta.sukiennik@agh.edu.pl

MODELLING OF PROCESSES WITH USE OF PROCESS MINING TECHNIQUES

Abstract. The main purpose of the paper is presentation of new opportunities for process modelling. In the literature review section, Petri nets as one of the formal modelling notation of processes is highlighted and introduction of relatively young research discipline – process mining – is presented. One of the process mining tasks is process model discovery from event logs gathered in informatics systems in enterprise. In the article practical example of process model discovery with ProM software is given with use of real event log from Volvo IT Belgium. In conclusions further opportunities of process mining techniques in process management are emphasized.

Keywords: Process mining, Petri net, process discovery, process model, ProM software

1. Introduction

Processes represent a core asset of any operating enterprise. They have direct impact on the attractiveness of products and services as perceived by the market [21]. Processes enable realisation of company strategic goals, that is why they should be in the centre of attention.

Single process can be defined as a set of events and activities, which are interconnected by time, resource or other dependencies. Main features of the process are: defined model of the process, involved resources, KPI's and responsible persons (owner and manager).

The basic element necessary for process management is knowledge about its real performance. This knowledge could be expressed in the process model. Creation of the process model could be performed in two ways:

1. model is created on base of employers' knowledge (hand-made model);
2. model is created on base of evidences gathered in the informatics systems of enterprise (discovered automatically).

Hand-made model could be biased by imaginative and wishful thinking point of view and could be different than original process execution. That is why fact-based modelling is more often emphasized as opportunity for more suitable expression of real process into model. For this purpose process mining techniques could be used.

Process mining is one of the ingredients of so called process science. Process science *is an umbrella term which refers to the broader discipline that combines knowledge from information technology and knowledge from management science to improve and run operational processes* [3].

Process science contains inter alia:

- Business process management (BPM) – discipline including techniques for the design, execution, control, measurement and optimisation of business processes,
- Process automation and workflow management – focusing on development of information systems supporting operational business processes,
- Stochastics – techniques which enable analysis of random processes (include Markov models, simulation, queueing systems),
- Operation research and optimisation – analysis of mathematical models of the processes, services or supply chains in order to their design, control and manage, also to provide the best option.
- Process mining – including techniques for discovery process models, confrontation between event data and process models and further analysis of processes.

In the paper more particularly process mining is presented, which could be used for process model discovery, its analysis, further improvements or optimisation of processes in organisations.

2. Process models

Process modelling enables understanding and analysis of a business process [29]. There are many business process modelling techniques. The most well-known include [12, 13]: flow chart technique, data flow diagram (DFD), role activity diagrams (RAD), role interaction diagrams (RID), Gantt chart, The Integrated Definition for Function Modelling (IDEF), Petri nets, Object oriented methods (i.e. Booch's Object Oriented Design (OOD) Technique or UML), Workflow techniques with various notation languages (Graph-Based languages, Net-Based languages (based on Petri nets) and Workflow Programming languages) as well as very popular nowadays – Business Process Modelling Notation (BPMN) [17] or Event – Driven Process Chain (EPC) [31]. Worth of mentioning are also YAWL [10] and process trees [32]. Brief comparisons of mentioned techniques are presented among other in [12, 26, 27].

Process models could be classified according to degree of formality. The first group are informal models (i.e. DFD, Gantt chart, RAD). Their main purpose is to provide insights or support discussion but cannot be used for enactment and rigorous analysis. The second group are formal models (i.e. Petri nets, process algebra). We can define a model to be formal if it is possible to determine whether a particular scenario (i.e., a trace of activities) is possible of not. Formal models typically allow deeper analysis and enactment, however, they may be more difficult to construct than informal models. In practice also semi-formal models exist (e.g., BPMN, UML activity diagrams, EPCs, etc.) [2].

In the paper we would like to present possibility of automatic creation of formal process model from event data (in a form of Petri net) with use of process mining techniques.

Petri nets were introduced by Carl Adam Petri's (in 1962) in a new model of information flow in systems. Originally developed for systems engineering, nowadays they are used for modelling computer software, hardware, control flow [26].

Petri nets are one of the oldest and best investigated formal process models (Fig. 1).

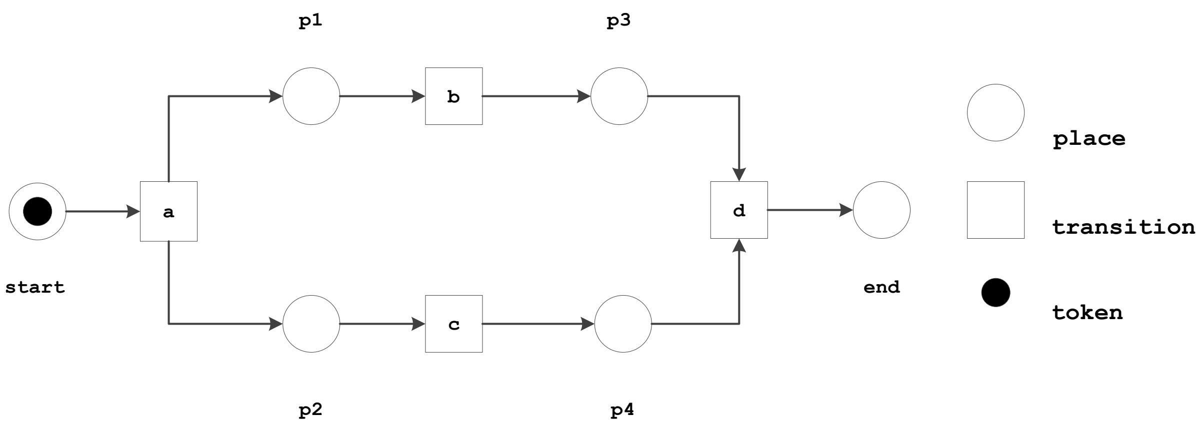

Fig. 1. An example of Petri Net
Source: Author's elaboration.

Petri net is defined as a triplet N = (P,T,F) where P is finite set of places, T is a finite set of transitions such that $P \cap T = \varnothing$, and $F \subseteq (P \times T) \cup (T \times P)$ is a set of directed arcs, called the flow relation [1].

The state of the Petri net is determined by the distribution of tokens over places and is referred to as its marking [3]. A transition is enabled if each of its input places contains a token. An enabled transition consuming one token from each input place and producing one token for each output place [8]. In general transitions in Petri net correspond to observable activities, but there are also so called silent or invisible transitions enable expression of logical dependency in the net.

In the modelling of business processes the sub-class of Petri nets is used, so called Workflow nets (WF-nets) [6, 9, 23]. Such WF-net is a Petri net with source place (process start) and sink place (process end) and all nodes are lying on a path from source to sink place.

Petri Nets are both a graphical notation and a precise mathematical notation, suitable for the analysis and reengineering of business process models. Petri Nets are sometimes considered as non user-oriented technique, difficult in applying by inexperienced users in BPM activities [26]. Nowadays their usage is popularized by very dynamic developing process mining discipline, in which Petri nets can be found as quite simple and intuitive notation for modelling of the processes. In the literature many various techniques for their analysis could be found [15, 16, 25, 28], increasing possibility of their usage in process enhancement.

3. Process mining

Process mining is a relatively young research discipline which can be located between data science on the one hand and process science on the other hand [34]. Process mining enables fact-based insights and can support process improvements [4].

The basic element for process mining is an event log including structured data about process performance. Each event in such a log refers to an activity and is related to a particular case. The events in a case are ordered and can be seen as one iteration of the process. The sequence of activities executed for a case is called a trace and an event log can be viewed as a multiset of traces. Often event logs store additional information about events. For example, resource (i.e. person or device) executing or initiating the activity, the timestamp of the event, or data elements recorded with the event (e.g., the size of an order) [5].

Event log could be expressed in a form of table (Table 1).

Table 1

An example of event log

Case Id	ChangeDate+Time	Status	Sub Status	Product	Activity id*
1	2012-03-31T17:33:07	Accepted	In Progress	PROD 798	a
1	2012-05-08T13:04:27	Completed	Resolved	PROD798	b
1	2012-05-16T00:22:15	Completed	Closed	PROD798	c
2	2012-05-01T02:07:30	Accepted	In Progress	PROD126	a
2	2012-05-01T02:12:29	Completed	Resolved	PROD126	b
2	2012-05-02T00:03:29	Completed	Closed	PROD126	c
3	2012-05-03T08:54:23	Accepted	In Progress	PROD660	a
3	2012-05-03T09:18:57	Completed	In Call	PROD660	d
4	2012-04-24T15:50:53	Accepted	In Progress	PROD514	a
4	2012-04-24T16:07:48	Queued	AwaitingAssignment	PROD514	e
4	2012-04-24T22:37:57	Accepted	In Progress	PROD514	a
4	2012-05-08T18:13:35	Completed	Resolved	PROD514	b
4	2012-05-16T00:21:15	Completed	Closed	PROD514	c

Source: Author's elaboration on the basis of [33], Ghent University. Dataset.

Presented event log can be also presented as a set of traces:

{<a,b,c>, <a,b,c>, <a,d>, <a,e,a,b,c>}

Process mining aims to discover, monitor and improve real processes by extracting knowledge from event logs available in today's information systems [4].

Hence we can identify three main types of process mining [8]:

1. process discovery,
2. conformance checking,
3. enhancement.

Process discovery techniques use an event log and create a model without using any a-priori information [7]. This analysis is based only on informatics systems data.

During conformance checking an existing process model is compared with an event log of the same process to detect and locate deviations between process model and real process execution (measuring the alignment between model and reality).

The process enhancement enables extension of analysis or improvement of the process by use of the additional information recorded in the event log i.e. involved resources or adding other perspectives to the process model (i.e. organizational, time or case perspective).

Process mining projects are executed for needs of various industries and markets (Fig. 3). In the survey performed by HSPI S.p.A [30] the number of ongoing project increases every year, in the HSPI's database above one hundred projects are included.

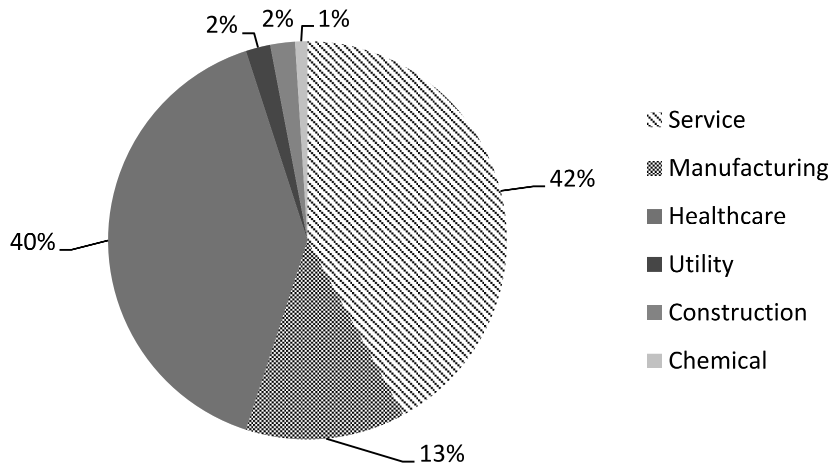

Fig. 3. Structure of process mining projects
Source: Author's elaboration on the basis of [30].

Analysed process mining projects were related to companies operating in services (42%), in healthcare (39%) and manufacturing (14%).

The most actual review of process mining issues is presented in [3] and dynamic development of this discipline can be also observed in publications from BPM Conferences [24].

4. Process discovery

Process discovery can be defined as learning a process model from example behaviour recorded in an event log.

One of the first process discovery algorithms was the α-algorithm [11]. Main assumption of this algorithm is to investigate relations between activities i.e. ($a,b \in A$) in the event log (L). These relations are defined as follows [3]:

$a >_L b$ if and only if there is a trace $\sigma = < t1, t2, t3, \ldots, tn >$ and $i \in \{1, \ldots, n\text{-}1\}$ such that $\sigma \in L$ and $ti = a$ and $ti{+}1 = b$;

$a \rightarrow_L b$ if and only if $a >_L b$ and $b \ngtr_L a$;

$a \#_L b$ if and only if $a \ngtr_L b$ and $b \ngtr_L a$;

$a \,||_L\, b$ if and only if $a >_L b$ and $b >_L a$.

On the basis of these relations the footprint matrix of the event log can be captured (the example of footprint matrix based on the event log example is presented in Tab. 2).

Table 2

Footprint matrix of example event log

activity	a	b	c	d	e
a	#	→	#	→	→
b	←	#	→	II	II
c	#	←	#	#	#
d	←	II	#	#	II
e	←	II	#	II	#

Source: Author's elaboration.

These relations enable to create process model based on discovered patterns (Fig. 4).

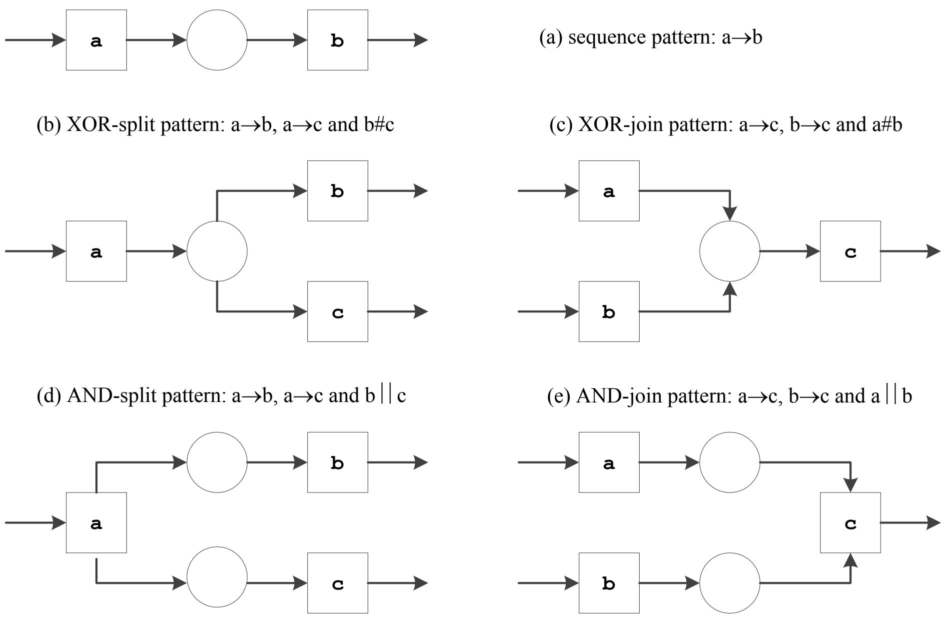

Fig. 4. Typical process patterns based on the chosen footprints
Source: Author's elaboration based on [3].

The α-algorithm is very simple, but because of specific problems with noise, loops discovery, duplicate tasks and advanced routing constructs has not very practical meaning for real event logs analysis. However, some of its ideas have been included in more complex mining techniques which include: heuristic mining [36], inductive mining [3], genetic process mining [14] or region-based mining [18]. A comparison of selected techniques is presented in [19].

The correct process model should be sound. It means that process model have to fulfil the following conditions [3]:

- WF-net should be safe (places cannot hold multiple tokens at the same time),
- proper completion should be available (if sink place is marked all other places should be empty),
- option to complete exists (it is always possible to mark the sink place for any marking),
- there is no dead transitions (all parts of the model are potentially reachable).

Moreover, discovered model should be adequate according to four competing quality criteria: replay fitness, simplicity, precision and generalisation.

Replay fitness of the model expresses the ability of a model to replay all behaviour recorded in the event log. Furthermore, discovered model should be the simplest model that can explain observed behaviour. Precision means that the model does not allow for too much additional behaviour in comparison to the event log (avoiding underfitting). Generalisation enables putting into model additional behaviour (avoiding overfitting) and provides certain abstraction of the process. Creation of good process model is a balancing between these four criteria and model should be the result of a wise compromise.

In process model creation, its analysis or improvement the Process Mining Project Methodology PM^2 could be used [22].

The methodology consists of six stages:

1. planning,
2. extraction,
3. data processing,
4. mining and analysis,
5. evaluation,
6. process improvement and support.

In the first two stages initial research questions are defined and event data are extracted. Next three stages (4-6) are executed iteratively as analytic part of process mining task. In this part answering a specific research questions by applying process mining techniques and evaluation of discovered models is performed. After satisfactory results of previous stages process improvement and support could be executed.

To support process mining, various tools have been developed. Released software includes commercial and open-source software. Among commercial tools one can find Celonis (Celonis GmbH), Disco (Fluxicon), Minit (Gradient ECM), Perceptive (Lexmark), Rialto (Execura) and others [35]. The most powerful and popular tool (especially in academic community) is open-source ProM (can be found at www.promtools.org). It supports many of developed process mining algorithms with over 1500 plug-ins available. ProM enables not only control-flow or time analysis but also organizational mining, decision mining and many others. An example of process mining executed in ProM is presented in the next section.

5. Process model discovery in practice

In analysis an event log from Volvo IT Belgium company is used [33]. This company provides IT services according to terms and conditions regulated in Service Level Agreements (SLAs). The event log was collected in VINST system, which supports incidents and problems handling.

In pre-processing stage the original event log (VINST cases closed problems) was filtered with use of simple heuristic in order to exclude infrequent events and traces with default settings in ProM plug-in.

The first task in process mining is very similar to initial task in data mining as is exploratory data analysis. At the beginning, the analysed data set should be explored and initial findings or hypothesis should be formulated.

The first insights can be done with analysis of event log with Log Vizualizer (Fig. 5) and dotted chart (Fig. 6).

Fig. 5. Window of Log Vizualizer
Source: Author's elaboration.

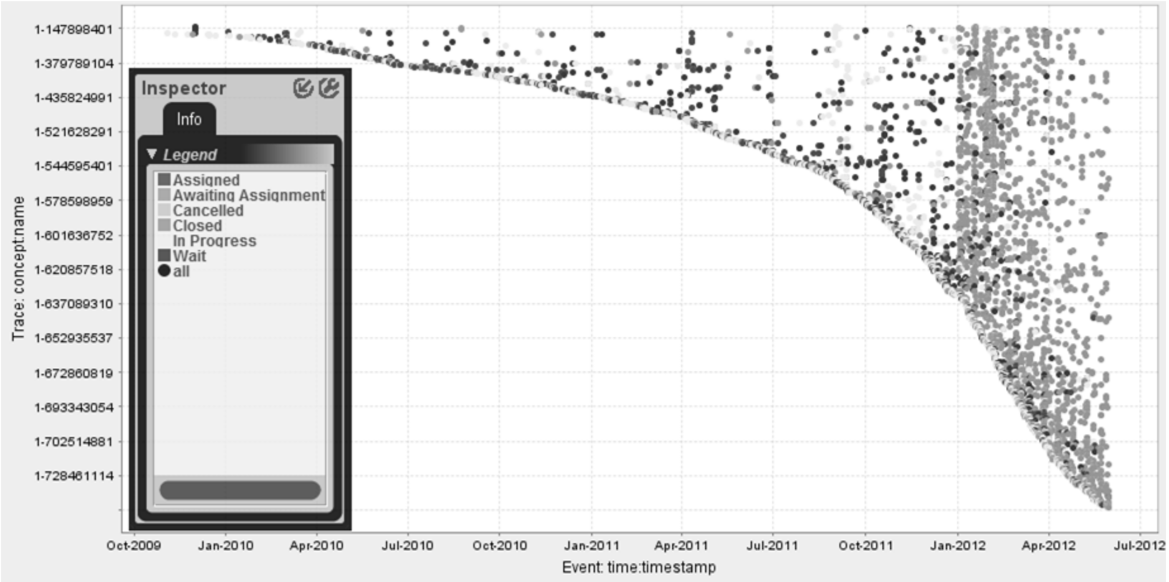

Fig. 6. Dotted chart of event log
Source: Author's elaboration.

The analysed event log includes 6217 events in 1396 cases. The events in log regard time from 4th November 2009 to 31th May 2012. Additional information gathered in log summary includes 6 classes of events ("In progress", "Closed", "Awaiting assignment", "Assigned", "Wait", "Cancelled"). Most of cases start with "In progress" event and all finish with "Closed" event.

Looking at the dotted chart regularity of case coming can be observed. Coloured dots illustrate different event classes. The visible days with completion of cases can be seen, especially in January 2012. It looks like "cleaning" activity of old cases.

The three most popular traces include the following order of events:

1. In progress – Closed (486 traces – 34,81% of event log),
2. In progress – Awaiting assignment – In progress – Assigned – In Progress – Closed (129 traces – 9,24% of event log),
3. In progress – Wait – Closed (110 traces – 7,88% of event log).

In the ProM there are various notations possible to model processes. Discovered Petri net for analysed event log is presented in Fig. 7. To ensure soundness of the Petri net, as plug-in for discovery of net model, the inductive miner was chosen.

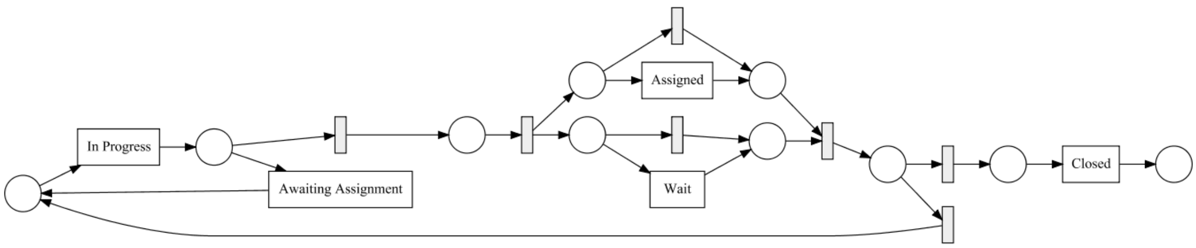

Fig. 7. Discovered Petri net
Source: author's elaboration

Created model is rather simple and is characterised by very high replay fitness (97%). The model does not enable too much behaviour, so model precision is also satisfactory. The model gives the general view of the process realisation.

Such model could be also expressed in the BPMN notation (Fig. 8).

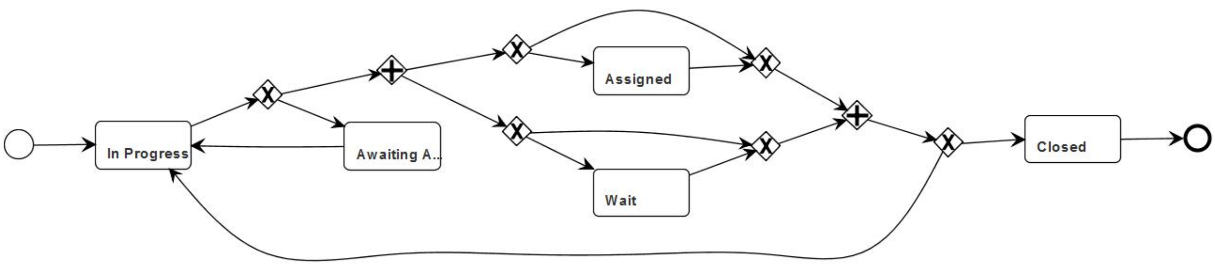

Fig. 8. BPMN model of discovered process
Source: Author's elaboration.

It is worth mentioning that in ProM software as well as Petri net or BPMN models, process tree could be easily created (Fig. 9).

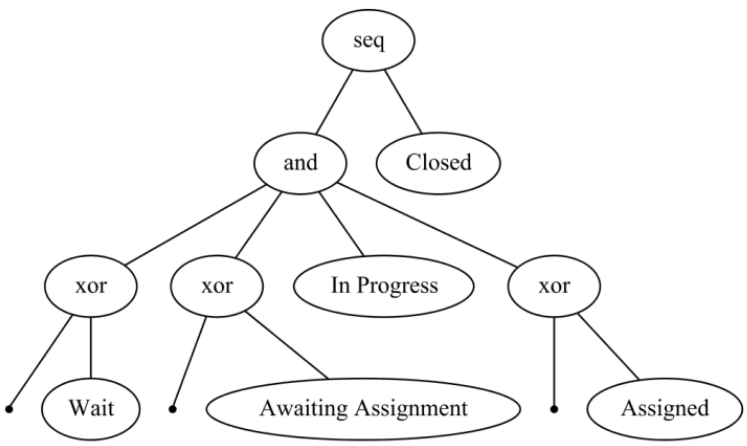

Fig. 9. Process tree for analysed event log
Source: Author's elaboration.

Knowledge of "as-is" process model enable further analysis regarding, first of all, comparison between formal process model (in example formulated in ISO documentation) and event log-based model, in order to identify the deviations (in plus or in minus) in process performance. The same model could be used for deeper analysis of bottlenecks or deeper case analysis (in example to find additional conditions or dependencies in cases' performance).

Conclusions

Nowadays processes have become the most important sources of creating value for company. Managers for decision-making need correct and objective view on processes, especially the core ones. Such view can be made with use of fact-based process models. Opportunity for this type of modelling is given by exploration of event logs gathered in the informatics systems of enterprise. Creation of such "as-is" process models is possible with use of process mining techniques. Hence, discovered process models could help to diagnose the actual processes and could be the base for further improvements.

In the paper discovery of process model was presented with use of real event log data from VINST system (Volvo IT Belgium). In analysis ProM software was used.

At the beginning, using Log Vizualizer option, examples of simple statistics (event class distribution, traces frequency) were presented. Created dotted chart has enabled formulation of first insights about the process (i.e. regularity of case flow, unusual distribution of events in time). As result of analysis, the satisfactory process model in a form of Petri net was obtained. Additionally, BPM model and process tree were presented as well.

Process mining gives much more opportunities to support managers in process modelling and management. Analysis of deviations or identification of bottlenecks in process could be easily performed in ProM and other commercial tools, giving a deep insight into the process structure.

A separate topic is mining of social networks in organization from event log to better understanding resource allocation and employees involvement, also important in terms of process management.

Considering capabilities and success of process mining projects, it has to be stated that era of process mining has just started and its meaning for process management in companies will be growing.

The paper presents results of research conducted in AGH University of Science and Technology no. 11.11.100.693

Bibliography

1. van der Aalst W.M.P.: Decomposing Petri Nets for Process Mining: A Generic Approach. "Distributed and Parallel Databases", Vol. 31(4), 2013, p. 471-507.
2. van der Aalst W.M.P.: Process-Aware Information Systems: Lessons to Be Learned from Process Mining. In Transactions on Petri Nets and Other Models of Concurrency II,

volume 5460 of Lecture Notes in Computer Science. Springer-Verlag, Berlin 2009, p. 1-26.

3. van der Aalst W.M.P.: Process Mining: Data Science in Action. Springer-Verlag, Berlin 2016.
4. van der Aalst W.M.P.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer-Verlag, Berlin 2011.
5. van der Aalst W.M.P.: Process Mining in the Large: A Tutorial, [in:] Zimanyi E. (ed.): Business Intelligence (eBISS 2013). “Lecture Notes in Business Information Processing”, Vol. 172. Springer-Verlag, Berlin 2014, p. 33-76.
6. van der Aalst W.M.P.: Verification of Workflow Nets, [in:] Azéma P., Balbo G. (eds.): Application and Theory of Petri Nets. “Lecture Notes in Computer Science”, Vol. 1248. Springer-Verlag, Berlin 1997, p. 407-426.
7. van der Aalst W.M.P., Bolt A., van Zelst S.J.: RapidProM: Mine Your Processes and Not Just Your Data, [in:] Hofmann M., Klinkenberg R. (eds.): RapidMiner: Data Mining Use Cases and Business Analytics Applications, Chapman and Hall/CRC Press, 2016.
8. van der Aalst W.M.P., van Dongen B.F.: Discovering Petri Nets From Event Logs, [in:] Jensen K., van der Aalst W.M.P., Balbo G., Koutny M., Wolf K. (eds.): Transactions on Petri Nets and Other Models of Concurrency (ToPNoC VII). “Lecture Notes in Computer Science”, Vol. 7480. Springer-Verlag, Berlin 2013, p. 372-422.
9. van der Aalst W.M.P., van Hee K.M., Ter Hofstede A.H.M., Sidorova N., Verbeek H.M.W., Voorhoeve M., Wynn M.T.: Soundness of Workflow Nets: Classification, Decidability and Analysis. “Formal Aspects of Computing”, Vol. 23(3), 2011, p. 333-363.
10. van der Aalst W.M.P., Ter Hofstede A.H.M.: YAWL: Yet Another Workflow Language. “Information Systems”, Vol. 30(4), 2005, p. 245-275.
11. van der Aalst W.M.P, Weijters A.J.M.M., Maruster L.: Workflow Mining: Discovering process models from event logs. “IEEE Transactions on Knowledge and Data Engineering”, Vol. 16, No. 9, 2003, p. 1128-1142.
12. Aguilar-Savén R.S.: Business process modelling: Review and framework. “International Journal of Production Economics”, Vol. 90, Iss. 2, 2004, p. 129-149.
13. Alotaibi Y.: Business process modelling challenges and solutions: a literature review. “Journal of Intelligent Manufacturing”, Vol. 27, Iss. 4, 2016, p. 701-723.
14. Alves de Medeiros A.K.: Genetic Process Mining. PhD thesis. Eindhoven University of Technology, 2006.
15. Best E., Schlachter U.: Analysis of Petri Nets and Transition Systems, [in:] Knight S., Lanese I., LluchLafuente A., Vieira H.T. (eds.), 8th Interaction and Concurrency Experience (ICE 2015), EPTCS 189, 2015, p. 53-67, doi:10.4204/EPTCS.189.6.
16. Biernacki J., Biernacka A., Szpyrka M.: Action-based verification of RTCP-nets with CADP, [in:] Simos T.E., Kalogiratou Z., Monovasilis T. (eds.), ICCMSE 2015:

International Conference of Computational Methods in Sciences and Engineering. AIP Publishing LLC, Athens 2015.

17. Business Process Model and Notation™ (BPMN™) 2.0.2 Object Management Group, 2013, http://www.omg.org/spec/BPMN/index.htm, 21.03.2017.
18. Carmona J., Cortadella J., Kishinevsky M.: A Region-Based Algorithm for Discovering Petri Nets from Event Logs. Business Process Management (BPM2008), 2008, p. 358-373.
19. van Dongen B.F., Alves de Medeiros A.K., Wen L.: Process Mining: Overview and Outlook of Petri Net Discovery Algorithms, [in:] Jensen K., van der Aalst W. (eds.), ToPNoC II. "Lecture Notes in Computer Science", Vol. 5460. Springer-Verlag, Berlin 2009, p. 225-242.
20. Dumas M., La Rosa M., Mendling J., Reijers H.: Fundamentals of Business Process Management. Springer-Verlag, Berlin-Heidelberg 2013, doi: 10.1007/978-3-642-33143-5.
21. Dumas M., Ter Hofstede A.H.M.: UML activity diagrams as a workflow specification language. UML 2001 – The Unified Modeling Language. Modeling Languages, Concepts, and Tools. Springer, Berlin-Heidelberg 2001, p.76-90.
22. van Eck M.L., Lu X., Leemans S.J.J., van der Aalst W.M.P.: PM2: a Process Mining Project Methodology, [in] Zdravkovic J., Kirikova M., Johannesson P. (eds.), International Conference on Advanced Information Systems Engineering (Caise 2015), "Lecture Notes in Computer Science", Vol. 9097. Springer-Verlag, Berlin 2015, p. 297-313.
23. Eshuis R., Dehnert J.: Reactive Petri Nets for Workflow Modeling, [in:] van der Aalst W.M.P., Best E. (eds.): Applications and Theory of Petri Nets 2003. ICATPN 2003. "Lecture Notes in Computer Science", Vol. 2679. Springer-Verlag, Berlin 2003, p. 296-315.
24. http://bpm-conference.org/BpmConference/Conferences, 18.03.2017.
25. Jensen K., Kristensen L.M.: Coloured Petri Nets. Springer-Verlag, Berlin-Heidelberg 2009.
26. Laden A., de Cesare S.: A Comparative Analysis Of Business Process Modelling Techniques. UK Academy for Information Systems Conference Proceedings, Vol. 2, 2009.
27. List B, Korherr B.: An evaluation of conceptual business process modelling languages. Proceedings of the 2006 ACM symposium on Applied computing (SAC '06). ACM, New York, USA 2006, p. 1532-1539.
28. Murata T.: Petri nets: Properties, Analysis and Applications. "Proc. of the IEEE", Vol. 77(4), 1989, p. 541-580.
29. Pietroń R.: Process management. Wrocław University of Technology, PRINTPAP, 2011.

30. Process mining: A database of applications. HSPI SpA – Management Consulting. Bologna, Italy, http://www.win.tue.nl/ieeetfpm/lib/exe/fetch.php?media=casestudies:hspi_process_mining_database_20161207.pdf, 20.03.2017.
31. Scheer A.-W., Thomas O., Adam O.: Process Modeling using Event-Driven Process Chains, [in:] Dumas M., van der Aalst W.M.P., Ter Hofstede A.H.M. (eds): Process-Aware Information Systems: Bridging People and Software through Process Technology. John Wiley & Sons, Inc., Hoboken, NJ, USA 2005, doi: 10.1002/0471741442.
32. Schunselaar D.M.M.: Configurable process trees: elicitation, analysis, and enactment. Technische Universiteit Eindhoven, Eindhoven 2016.
33. Steeman W.: BPI Challenge 2013. Ghent University, Dataset, 2013, http://dx.doi.org/10.4121/uuid:a7ce5c55-03a7-4583-b855-98b86e1a2b07, 18.03.2017.
34. The Process Mining Manifesto by the IEEE Task Force on Process Mining, [in:] Daniel F., Barkaoui K., Dustdar S. (eds.): BPM 2011 Workshops, Part I, LNBIP 99. Springer-Verlag, Berlin 2012, p. 169-194.
35. Turner C.J., Tiwari A., Olaiya R., Xu Y.: Process mining: from theory to practice. "Business Process Management Journal", Vol. 18(3), 2012, p. 493-512.
36. Weijters A., Ribeiro J.T.S.: Flexible heuristics miner (FHM). IEEE Symposium on Computational Intelligence and Data Mining (CIDM), Paris 2011, p. 310-317.