Scientific Journals
of the Maritime University of Szczecin

Zeszyty Naukowe
Akademii Morskiej w Szczecinie

2018, 53 (125), 102–105
ISSN 1733-8670 (Printed)
ISSN 2392-0378 (Online)
DOI: 10.17402/271

Received: 24.10.2017
Accepted: 15.02.2018
Published: 16.03.2018

Guidelines for recording transport event logs

Mariusz Dramski

Maritime University of Szczecin, Institute of Marine Technology
1–2 Wały Chrobrego St., 70-500 Szczecin, Poland
e-mail: m.dramski@am.szczecin.pl

Key words: event log, process mining, XES, CSV, data, method

Abstract
A process is an ordered set of related activities taking place in a given time. Processes are present in all branches of the economy, engineering, science, etc. Due to the huge amount of data produced the rapid development of data mining techniques has been observed. Similar methods are also used in the context of processes and are called process mining. The main task of process mining is to create a process model, which is used to reason about the process and to make decisions inside it. The process model may be used to discuss responsibilities, simulations, predictions, etc. The main data structures in process mining are event logs. It is always very important to have correct data which makes creating a reliable process model possible. In this paper the basic guidelines for recording such event logs have been described and conclusions were drawn. The main focus of this research was transport problems.

Introduction

There is a need to define what an event log is. This definition can be widely found in the literature (Aalst, 2011). In simple words the event log is a data structure that represents the process performance. Let us consider a simple structure like an array; it is only a finite set of elements of the same type. If we change the type of just one element, the data structure cannot be called an array any longer. This leads to the simple conclusion that there is a need to ensure the completeness of the data. Software engineers, researchers, etc. have to know what the data structures they use are. It is obvious that when we model the decision making process, the natural data structure would be trees or graphs. The same problem occurs when we talk about process mining techniques.

Process mining is a brand new branch of data science, focusing mainly on business processes. It allows for the detection of some problems, building process models, makes improvements to the process possible, etc.

Event logs as data structures can be obtained in different ways. The most common is to record the events and activities during the process flow. It is necessary to realize that each activity has its proper place in time.

An example fragment of an event log has been given in Table 1.

The three main parts of the event log are clearly visible; case id, event id, and properties. The most important property is the timestamp. There can be no event log without this element.

Data sources

Process mining is impossible without the event log. The kind of data that is needed was mentioned in the introduction. Unfortunately many organizations don't have good data warehouses. Even if they exist they very often contain only some subsets of the data. Besides, the data is not process oriented. The typical data sources are, e.g.:

- SAP tables – the problem is that the SAP implementation may contain over 10,000 tables;

Table 1. An example of an event log (Aalst, 2011)

Case id	Event id	Properties			
		Timestamp	Activity	Resource	Cost
1	3654423	30-12-2010:11.02	Register request	Pete	50
	3654424	31-12-2010:10.06	Examine thoroughly	Sue	400
	3654425	05-01-2011:15.12	Check ticket	Mike	100
	3654426	06-01-2011:11.18	Decide	Sara	200
	3654427	07-01-2011:14.24	Reject request	Pete	200
2	3654483	30-12-2010:11.32	Register request	Mike	50
	3654484	30-12-2010:12.12	Check ticket	Mike	100
	3654485	30-12-2010:14.16	Examine casually	Pete	400
	3654486	05-01-2011:11.22	Decide	Sara	200
	3654487	08-01-2011:12.05	Pay compensation	Ellen	200
…	…	…	…	…	…

- PDF files – non editable;
- Message exchanges;
- MS Excel files (not process oriented);
- and others…

Each data source is required to be converted into the official process mining data format – Extensible Event Stream (XES).Some tools support also CSV files (Dramski, 2016).

This leads to the conclusion that the event log is going to be created. There is a need to extract and convert the data into the desired format. First of all we have to answer some questions e.g. "Which of the 10,000 SAP tables to convert?" (Aalst, 2011).

Depending on the questions and viewpoint chosen, different event logs may be obtained. In this paper we have considered the event logs which may be applied to transport systems. So there is a need to discover e.g. goods and services flows. Each question requires a different approach to process mining.

In the literature (Aalst, 2011) some challenges of data extraction were mentioned:

- correlation – events need to be related to each other;
- timestamps – events need to be ordered per case (such ordering doesn't require timestamps but are necessary for building the process model);
- snapshots – the recording process can be started and stopped in each moment of the process flow. Therefore sometimes the data doesn't cover the whole process but only a fragment.

So we can say that we have some snapshots of the process:

- scoping – there is a need to decide which part of the data is in our area of interest;
- granularity – events in the event log have different levels of granularity. It can be different even if we consider the event log for one process in one organization.

Examples of transport processes

In order to illustrate the problems with regard to building a correct model of the process, three models were considered (Filipova, Stojadinova & Hadjiatanasova, 2002). The process is very simple; a lorry brings the goods to the port, a ship waits for loading and then heads to the open sea. Although it is not a complicated process, there are several ways to solve this problem. The Petri net was chosen for creating models, where:

- P1 – lorry,
- P2 – ship,
- P3 – port,
- P4 – open sea,
- T1 – movement of the lorry to the port,
- T2 – movement of the ship to the port,
- T3 – processing and leaving the port,
- Pi are places and Ti are transitions.

Figure 1 illustrates the simple model of the process. It can be seen that a lorry enters the port. Then

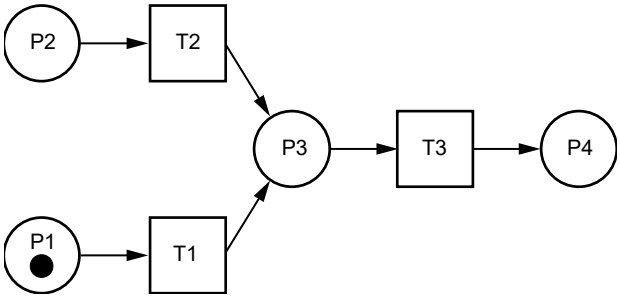

Figure 1. The model of the transport process, (Filipova, Stojadinova & Hadjiatanasova, 2002)

the ship is loaded and may leave the port, but there is a problem. No token in P2 says that there is not a ship in the port. This observation may lead to the example event log such as:

$$L = \{\langle T1, T3\rangle, \{\langle T1, T3\rangle, \ldots\} \quad (1)$$

It does not matter how often the sequence ⟨T1, T3⟩ occurs. If we make an assumption that the initial token is always placed in P1, then it can be observed that this is the only possible path. Anyway, there is a possibility to create the event log. However it won't correspond to reality.

Maybe we can add the other initial token in P2. It won't solve the problem either. Even if both tokens (ship and lorry) meet in P3, only one of them will be able to fire the transition T3. It is not set which of them it would be. So we cannot define which vehicle heads to the open sea. Of course the researcher knows very well that it will be a ship, but the model does not determine it.

Figure 2 illustrates the first proposal for the model's improvement. Now T3 requires two tokens to fire, but it never will fire. The token from P1 will stuck in P3 waiting for the second token (from P1 or P2). Assuming that only one lorry can enter the port at one time, there is still no ship in the port. T3 will wait for the token which will never arrive.

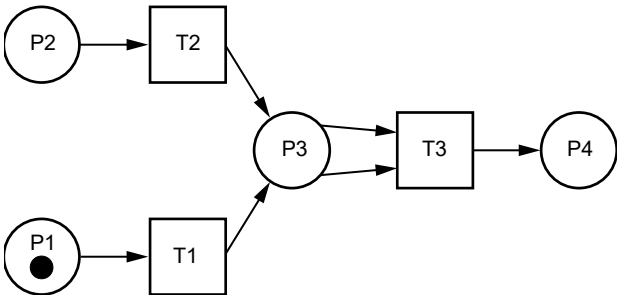

Figure 2. The model of the transport process (Filipova, Stojadinova & Hadjiatanasova, 2002)

Figure 3 illustrates the second improvement made only by adding a token in P2. Now we see that both a lorry and a ship enter the port (place P3). There will be two tokens in P3, so the transition T3 fires. There will be one token in P4. Example event log:

$$L = \{\langle T1, T2, T3\rangle, \{\langle T2, T1, T3\rangle, \ldots\} \quad (2)$$

The order of T1 and T2 is not important because both vehicles can enter the port in different times. Anyway there is still a danger. We can add other lorry and then two tokens will meet in P3, so the T3 transition will fire.

Summing up, these three simple examples show that there is a need to better know the way the process

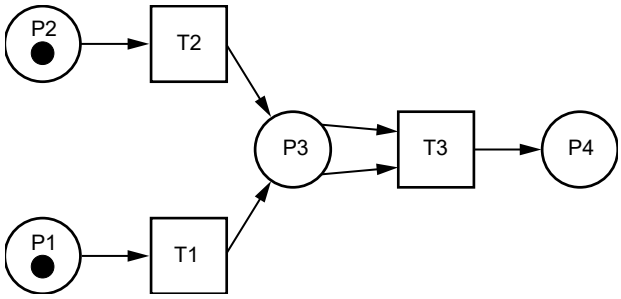

Figure 3. The model of the transport process (authors own research based on (Filipova, Stojadinova & Hadjiatanasova, 2002))

is running. The event logs extracted from the models were different. Some other conclusions have been drawn in the summary of this paper.

Tools used for processes modeling

Processes can be modeled using different approaches. One of the best known are transition systems and Petri nets. These simple tools allow for the modeling of the process flow and for conclusions to be drawn from it. A Petri net (Aalst, 2011, Peterson, 1981) was chosen in the previous section of this paper. Of course this kind of model can be easily converted to other types such as methods like:

- workflow nets (Aalst, 1998);
- causal nets;
- event-driven process chains;
- BPMN diagrams (Business Process Modeling Notation) (OMG, 2010);
- YAWL (Yet Another Workflow Language (Hofstede et al., 2010) ;
- and more…

These approaches allow for the creation and verification of the process model and enable it to be applied to the real events stream. Besides, there is a possibility to convert the model from one notation to the other one without losing any data.

Process mining is not an easy job. It requires a lot of knowledge and experience. It is necessary to know the real process in as much detail as possible. When the model and reality have little in common, model-based analysis doesn't make any sense. The model should satisfy some desirable properties; it should be the simplified description of the real process. On the other hand the model may sometimes describe the idealized version of reality. In each moment of the process flow, some deviations may occur. Very often it is impossible to create a model with perfect alignment with reality. The same problem may be also seen from different points of view focusing on different properties.

Conclusions

In this paper some basic guidelines for recording transport event logs have been given. First the event log data structure was described (a more detailed definition can be found in the literature (Aalst, 2011)), example event logs have also been presented in the literature (Filipova, Stojadinova & Hadjiatanasova, 2002). Three simple models were then presented and some example event logs were shown and some observations were made. According to the above text and the literature there is the possibility to define some guidelines for recording event logs, which are:

- Every process takes place in time, so the timestamps are the most important part of each event log. Besides this the data structure requires timestamps. Process diagnosis or further improvement is impossible when the time is not given;
- It is necessary to define the point of view. We can focus on the entire process or only a fragment. This knowledge is very important, because the model must fit into reality as much as possible;
- There is no ideal model of reality. Perfect alignment is an unachievable utopia;
- There is the need to consider that in each moment of the process flow, some deviations may occur. The most common reason for such errors is the human factor;
- The researcher must know the process. It can be done as a result of experiments, observations, or obtained data but also by cooperating with the experts;
- There are a lot of usable tools in process mining such Petri nets, transition systems, BPMN diagrams, YAWL etc.;
- There are some software systems on the market that support process mining techniques. Some of them are free of charge (such as ProM (Aalst, 2011));
- It is necessary to verify if the event log has recorded properly.

The above guidelines show that process mining techniques require some basic activities which should be taken when building a process model.

Process mining is quite a new area of data science, focusing especially on business processes, but it can also be applied to other branches. In the literature (Aalst, 1998) the application of Petri nets in transport streams was shown. Further examples can be found in the literature.

Acknowledgments

This research outcome has been achieved under research project No. 1/S/ITM/16 financed from a subsidy of the Ministry of Science and Higher Education for statutory activities of the Maritime University of Szczecin.

References

1. Aalst van der, W. (2011) *Process mining – discovery, conformance and enhancement of business processes*. Berlin Heidelberg: Springer-Verlag.
2. Aalst van der, W. (1998) The application of Petri nets to workflow management. *The Journal of Circuits, Systems and Computers* 8(1), pp. 21–66.
3. Dramski, M. (2016) Extensible event stream format for navigational data. *Scientific Journals of the Maritime University of Szczecin, Zeszyty Naukowe Akademii Morskiej w Szczecinie* 47 (119), pp. 61–65.
4. Filipova, K., Stojadinova, T. & Hadjiatanasova, V. (2002) Application of Petri nets for transport streams modeling. *Facta Universitatis, Series: Architecture and Civil Engineering* 2(4), pp. 295–306.
5. Hofstede ter, A.H.M., Aalst van der, W., Adams, M. & Russell, N. (2010) *Modern Business Process Automation: YAWL and its Support Environment*. Berlin: Springer.
6. OMG (2010) *Business Process Model and Notation* (*BPMN*). Object Management Group, dtc/2010-06-05.
7. Peterson, J (1981) *Petri net theory and the modeling of the systems.* Prentice-Hall.