Analytical treatment of transport logistics business processes by the Process Mining technology

A A Popov¹, S N Masaev², D A Edimichev² and O V Pomolotova²

¹ Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, prospect “Krasnoyarsk Rabochy”, 31, Krasnoyarsk, Russia
² Siberian Federal University, Svobodny av., 79, Krasnoyarsk, 660041, Russia

E-mail: tolynbms@yandex.ru

Abstract. The purpose of the research is to analyze the business processes associated with goods movement through the warehouse of a transport company by the Process Mining method based on ProM software. This article describes the analytical treatment of the event log of goods movement through the warehouse of a transport company. A model of a business process based on an event log was built and analyzed. The result confirms the possibility of business processes analytical treatment based on the transport company's event logs. Using this technology, you can analyze various logistics companies, identify errors in the processes and look for ways to eliminate them. The results of the research contain conclusions, recommendations, and options for further research of the transport company.

1. Introduction

Currently, information analysis systems market, including business process modeling, is developing dynamically. Various technologies are emerging that enhance analysis capabilities [1,2,3].

Modern organizations have process models that describe how they should be handled (ISO 9000 family of standards) [4]. Leaders of organizations are puzzled by the documentation of their processes. These processes are often not handled by information systems. In [4], the authors proposed an approach for measuring the process model accuracy in relation to the event log. In contrast to previous approaches, models were first built and recorded, which made the data approach more reliable even in case of deviations.

In [5], a method of replay analysis was presented that can measure the compliance of the event log for the given process model. The approach quantifies compliance and provides intuitive diagnostics (omitted and inserted actions). This technique was implemented in ProM6.

The article [6] summarized the applied approach and how it can be used to study a wide range of e-commerce information systems. Authors noted the problems and difficulties in the application of process analysis methods in the journals of e-commerce systems. The software environment ProM6 was used.

Process detection algorithms are usually aimed at discovering process models from event logs that best describe the recorded behavior [7].

The review of works showed that the Process mining methodology based on the actual of business processes is the most effective for solving the research problem. The main idea of process analysis is to identify, track, improve and eliminate errors in real processes in the enterprise by extracting data from event logs. Event logs represent the real execution of business processes through the interaction of users
with information systems. The company's business processes or information system settings are adjusted based on the analysis of the event log.

The proposed approach is known all over the world. It is used and developed actively by a group of researchers at the University of Eindhoven under the leadership of Wil van der Aalst. The main statements of the approach are highlighted, for example, in the works [8], [9].

At the same time, the analysis of past publications did not reveal the application of the Process Mining approach to the analysis of business processes in transport logistics.

A large transport company was used as an object of research to solve the problem. The event log of the goods movement in the warehouse of the transport company was used for this task. This process was chosen because it is quite laborious and critical. And the erroneous execution of the process can lead to an error in the shipment of the goods and fines from the superior organization.

2. Methods

The event log must have four properties to automatically build a model by the Process mining method:

- activity - some event or action (authorization, page view, task setting).
- time stamp - the start time of the event.
- case id - defining the sequence of actions.
- recourse - the executor of the action, it can be a user or some kind of information system.

The process mining toolkit will automatically group events by activity and time ID after collecting all the data.

By analyzing the process model in this way, you can identify unnecessary cycles, regular delays in the execution time of tasks, unnecessary actions in processes, a task queue and you can also track executor errors. Filtering by time, you can see the behavior of processes in dynamics and the results after changes in the processes.

The analysis tool was ProM software. It builds human-readable event log-based models with plugins. This statement is supported by a review of famous works and personal experience.

Before applying any analysis methods and plugins to the event log, you need to get an idea of the general information in it.

In this work, to analyze business processes using the Process mining method, the event log was used, which was collected by the employees of the transport company.

The important idea is that all data is confidential. Client data is not included in the event log. Each cargo has its own ID and cargo code.

This event log is well processed, events are clearly formulated, understandable to humans and easy to use.

The event log includes 151 private groups and 2170 actions. Each action has 2 types of actions (event start and event end).

Next, business process model was built. A heuristic algorithm was used to construct a model of event dependency in this work. The plugin "Mine for a Heuristics using Heuristics miner" was used for the implementation.

3. Results

The constructed model contains three starting events: cargo acceptance, cargo redirection, arrival at the warehouse. Start events fully correspond to the logical model by the heuristic model.

The heuristic model is fairly simple, since the sequence of events for the movement of goods in the warehouse is basically the same.

In the logical model, the "Flight movement" event should be the last one because after this event, the car is sent from the warehouse to another city. This model contains 2 input events and 2 outputs about the "Flight movement" process.
After all the events were analyzed, several flaws were found in the model. The constructed heuristic model doesn’t show the connection between the transition of the "Cargo execution" event to the "Discharge zone" event. Although in the logical model, the cargo can move after the registration of the cargo immediately to the dispatching area. Also, this data is contained in the event log.

The most common event is the "Start of loading". At the input there are 4 events and 2 output. In further analysis of the sequences of events, we will try to find out for what reason one of the input events is the event "Flight movement".

Next, we looked at the most common sequences of events. We visualize the sequence of events by the "Explorer events log" plugin. The plugin allows you to fetch event streams for a sequence system. The results of the work of this plugin are built routes of the sequence of transitions through the event log so it is passed to this plugin as input data. Transitions of the constructed system correspond to events in the log. The total number of sequences is 151. Figure 1 shows the most frequently repeated chain of sequences.

![Figure 1. The most frequently repeated chain of sequences.](image)

This sequence is repeated 63 times. It is 42.72% of all sequences of events. The chain of sequences is the “ideal” life cycle of the cargo movement process. Chain contains the event "Regular Slotted Container zone (RSC zone)". The client orders a rigid packaging service so protects himself and the transport company from damage to the cargo during transportation. Also, this service is paid so generates income for the company.

Figure 2 shows the second common chain of sequences of events.

![Figure 2. Chain of sequences.](image)

This sequence is obtained when clients refuse from the rigid packaging service. This sequence is repeated 35 times, which in percentage was 23.18% of the total number of sequences.

Figure 3 shows a stray signal sequence chain. This sequence is repeated 6 times - 3.97% of all sequences.

![Figure 3. Found stray signal in sequence chains.](image)
The chain is not built in the correct sequence. After analyzing the chain data so it was revealed that some events have a date, but no execution time.

In Figure 4, you can see that all these 6 sequences were entered into the database one after another when sending the cargo. So, it follows that there is stray signal in the event log. Now we aren't understanding for these errors, but most likely it was a breakdown of the radio data collection terminal, a human factor or an employee's inattention.

Figure 4. Chain of Sequences.

Figure 5 shows the sequence "cargo acceptance> cargo execution> RSC zone> discharge zone" is repeated 2 times. At first it is impossible to understand that there was an error in sequence since the cargo could still be sent at the time of data extraction.

Figure 5. Sequence chain with stray signal.

The cargo was placed in the dispatching area on 12/14/2019 so we can conclude that there is no error in the first sequence.

In the second sequence the cargo was placed in the dispatching area on 11/30/2019 so the cargo has been in the warehouse for 15 days (data taken before 12/15/2019) already. Therefore, we conclude that the cargo is either lost or sent in fact without a mark in the database or the client has suspended
dispatching. Only the dispatching suspension is not an error in the business process of the transport company. Loss of cargo and dispatch in fact is a mistake of warehouse workers and can lead to penalties.

Next, we mapped the business process model to the event log. We have checked the correspondence of the routes by the plugin "Fuzzy miner", in which you can analyze the "conformance checking" sequences.

Verification is necessary to identify two types of nonconformities:

- Inappropriate log behavior: behavior observed in the log that is not permitted by the model;
- Optional Model Behavior: behavior allowed in the model but never seen in the log.

There are generally three families of methods for detecting inappropriate log behavior: replay, trace alignment, and behavior alignment.

In replay methods [10], each trace is replayed in the process model one event at a time. A common limitation of conversion methods is that error handling is done locally. These methods cannot identify the minimum number of errors that might explain the inappropriate log behavior. This problem is eliminated by trace alignment methods [11]. These methods determine the closest matching trace for each trace in the log. The trace is analyzed by the model.

Figure 6 shows sets of fuzzy sequences. Each sequence refers to the life cycle of the goods movement through the warehouse.

![Figure 6. Sequence of fuzzy events.](image)

In our example, the model is descriptive. You can find inconsistencies in order to improve the model in reality. A 100% fit route is the ideal route for model analysis, but no one was found.

4. Discussion
By the results of the research of business processes, recommendations were developed based on the event log of the transport company.
It should be noted that some sequences have stray signal. So, the information system periodically fails due to breakdown of the radio data collection terminal. Since in some routes the events do not have the exact time of their execution. Failures destroy the data written to the event log.

It was noted that there are goods that are in the warehouse for a long time, which may violate the client's deadlines. Management needs to analyze why such situations occur. Since the delay in the client's deadlines leads to penalties for TOP management.

The most frequently repeated chain of sequences occurs only 63 times. Since it is the "ideal" sequence of this business process than we would like to advise storekeepers to more persistently offer the service of rigid packaging. It protects the company and the client from damage to property.

5. Conclusion
The completed research work is proof that the Process mining methodology can be applied for analytical treatment of transport logistics business processes. The tasks of building a model, identifying sequences and their correspondence to the event log, identifying and causing errors and noise were completed. After analyzing, it is clear that a deeper processes analytical treatment of the transport company is possible: total labor intensity, time spent on the route, analysis of the detection of cargo delays at a particular event (exit from the client's deadlines), determination of popular geographic directions, the average time of loading the car at different points followings.

We have data to form a model of the life cycle of the goods movement in a warehouse by the formalism of Markov chains. Scientific work is planned for development in this direction.

Also, an interesting development of the work is the organization of work execution by an alternative probabilistic method of network planning - GERT networks.

References
[1] Barsegyan A A, Kupriyanov M S, Holod I I, Tess M D and Elizarov S I 2009 Data and process analysis 3 (Petersburg) p 512
[2] Repin V V et al. 2014 “Business processes. Modeling, implementation, management”
[3] 2005 Business Processes: Regulation and Management 2 319
[4] Adriansyah A, Munoz-Gama J, Carmona J, Van Dongen B F and Van der Aalst W 2012 Alignment Based Precision Checking Business Process Management Workshops Springer 137-49
[5] Adriansyah A, Sidorova N and van Dongen B F 2011 Cost-Based Fitness in Conformance Checking International Conference on Application of Concurrency to System Design pp 57-66
[6] Mitsyuk A, Kalenkova A, Shershakov S and Van der Aalst W 2014 Using process mining for the analysis of an e-trade system: a easy study Business Informatics 3(29) 21-4
[7] Buijs J, Van Dongen B F and Van der Aalst W 2012 On the Role of Fitness Precision Generalization and Simplicity in Process Discovery 20th International Conference on Cooperative Information Systems pp 112-79
[8] Van der Aalst W 2011 Process Mining Manifesto
[9] Van der Aalst W 2011 Process Mining: Discovery Conformance and Enhancement of Business Processes (Berlin) 370
[10] Adriansyah, A 2014 Aligning observed and modeled behavior Technische Universiteit Eindhoven https://doi.org/10.6100/IR770080
[11] Van Dongen B F, Alves de Medeiros A K, Verbeek H M, Weijters A J and Van der Aalst W 2005 The ProM Framework: A New Era in Process Mining Tool Support Lecture Notes in Computer Science Springer pp 444-54