PRACTICAL ASPECTS OF EVENT LOGS CREATION FOR INDUSTRIAL PROCESS MODELLING

MSc., Eng. Agnieszka Trzcionkowska
DSc., PhD. Edyta Brzychczy
AGH University of Science and Technology**, Poland**

Abstract: In the paper we address the challenge of applying process mining techniques for discovering models of underground mining operations based on a sensor data. The paper presents practical approach of creation an event log based on industrial sensors data gathered in an underground mine monitoring systems. The proposed approach enables to generate event logs at different generalization levels based on several numbers of discovered stages of devices performance. For discovering process stages data mining techniques such as exploratory data analysis, clustering and classification have been applied. Created event log has been used in one of the process mining tasks – process model discovery.

Keywords: event logs, sensor data, process mining, longwall face, underground mine

1. INTRODUCTION

Modern diagnostic systems applied in an underground mine machinery provide full and integrated in one place access to information from whole spectrum of sensors monitoring of working conditions of machines and devices (Brzychczy and Trzcionkowska, 2017). This access enables comprehensive and ongoing supervision of the operations of industrial facilities. The constant monitoring and diagnosis of the mining process in real time require the collection of large amounts of historical data (Big Data). Hidden knowledge in that data, through modern processing and visualization can be useful for improving the operations and give the possibility of effective management of the equipment and industrial infrastructure.

Having technology enabling integration of heterogeneous data from many sources and its multidimensional analysis is one of the main pillars of the Industry 4.0 concept and development of the Industrial Internet of Things.

Looking for new analytic possibilities for industrial data, process mining should be taken into consideration as process–oriented approach for data analysis. Process mining aims to discover, monitor and improve real processes by extracting knowledge from dedicated data structures, namely event logs.

Nowadays, in the industrial practice there is a gap between sensor data and the event logs. In the monitoring systems generally low–level events are recorded and there is no one approach for process activity definition and creation of high-level event logs. Various methods has been proposed for activity discovery and recognition of event abstraction. In the literature two general approaches can be found: the first, more traditional, is based on experts knowledge (Mannhardt et al., 2016) and the second takes advantage of machine learning techniques (Guenther and van der Aalst, 2006; Bose and van der Aalst, 2009; Cook et al., 2013; van Eck et al., 2016; Tax et al., 2018).

In the paper we propose approach for log creation based on low-level event data from monitoring system of underground mine machinery.

There is no conventional event log describing the operations in a longwall face. Theoretical pre-defined stages (activities) have been traditionally created by domain experts based on

longwall working technology manual for the reporting needs, but they are abbreviated and do not cover real diversity of a working conditions. Preparation of relevant event log requires suitable database structure and several steps for transformation of raw data into event logs adapted to the requirements of the process mining techniques. Many challenges are related to that task: data quality problems, discovering and describing of unique stages in machinery and devices operation and, finally, problems with precise determination of cases in the longwall process realization. Therefore, the necessary step is to find an mathematical solution to generate real stages of the mining process in a longwall face based on the collected sensor data.

In Figure 1 we present the main steps of proposed approach for event log creation based on industrial sensor data.
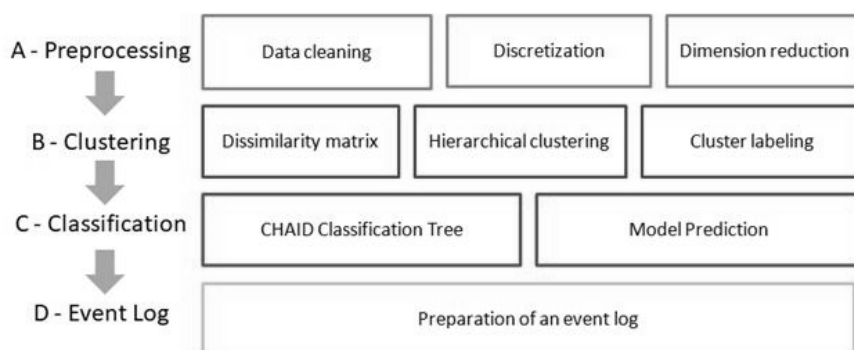
Fig. 1. Workflow of an event log creation from industrial sensor data

Our approach consists of the four main stages: preprocessing, clustering, classification and preparation of an event log. These stages are described in detail in section 3. In addition to this part, the paper is structured as follows: in section 2 brief introduction to process mining is presented, in section 4 preliminary results of process modeling based on created event log are shown, finally, in section 5 conclusions and summary is presented.

2. PROCESS MINING

Over last ten years process mining has grown as a new scientific discipline linking process science to data science (van der Aalst, 2016). From a company perspective it becomes as one of the most important innovations in the field of business process management. Process mining provides several methods and tools which allow: discovery of real process models, conformance checking by monitoring deviations, identification of bottlenecks, prediction and automated construction of simulation models.

Main core of process mining are proper, high quality event logs where each event refers to a case, an activity, and a point in time (van der Aalst, 2013). Typically, they contain information recorded by the information systems. Real challenge appears when such event logs have to be created and described for the purpose of reconstruction what has actually happened in the process. This problem is common to a large number of industry processes.

Four types of process mining could be conducted (van der Aalst, 2018):
1. Process discovery,
2. Conformance checking,
3. Process reengineering,
4. Operational support.

Process discovery is related to learning of process models from event data. This technique without using additional information produces a real process model based on an event log.

Conformance checking brings ability to find differences and commonalities between existing process model and an event log. Reality captured in the event log is verified with process model which can be made by hand or discovered from the event data.

Process reengineering provides improving and extending of the existing model based on an event log. In this task enriching of the model with additional perspectives is possible due to bottleneck detection, resource usage analysis or case mining. As a result process model is updated and can be used for actual process improvement.

Operational support providing warnings, predictions or recommendations during process running based on the model and event data.

3. CREATION OF AN EVENT LOG FROM INDUSTRIAL SENSOR DATA

The main challenge of our work is to create event logs from available sensor data which will enable further analysis of underground mining processes with selected process mining techniques.

Currently, data from longwall monitoring system are mainly used for the purpose of real time visualization and daily reports. Such analysis contains only time summary of simplified pre-defined stages of devices which were described by domain experts and other existing stages are disregarded.

In this section we present complete overview of processing raw sensors data into an event log using supervised and unsupervised data mining techniques.

Sensor data received from one of the Polish mining company include information about enormous numbers of working condition parameters of many devices.

Process mining comprises many different techniques which can easily handle with analysis of complex and extensive processes. However, for the purpose of more effective process analysis it is better to split one complex process into several subprocesses (Hompes et al. 2016). In this paper we are focus on one of the processes realised in a longwall face – operation of a longwall shearer.

3.1. Preprocessing

The raw data set contains 2,5 million records and 176 variables (continuous and logical) related to the shearer operation process from monthly period.

The main objective of the *Preprocessing* stage was to reduce amount of variable only to the most accurate which describe work of the shearer in the pertinent way. The first step of variables reduction was made during the *Data Cleaning*. For that purpose exploratory data analysis was conducted. During analysis variables with more than 40% of missing values or variables with only one logical value were excluded. Subsequently analysis of correlation for numerical data and cross tables for logical data were performed to exclude depended variables. Received list of variables was compared with domain knowledge of experts and final list of variables regarding to shearer operations was created including basic parameters of operation (i.e. speed, currents on organs and tractors, position of organs) and parameters related to position in a longwall face (i.e. location, direction of movement).

In the *Discretization* stage we converted all continuous variables into a categorical variables. Variables related to currents were divided into: idle running, small load, medium load, high load and overload. The remaining variables, i.e. position in the longwall and speed, were divided into 6 intervals.

In the *Dimension Reduction* step in the final data set containing discretized and logical variables duplicate rows were removed. As a result of this step data frame used for more general process stages definition was created.

3.2. Clustering

Discovering more general process stages has contained initial step *Dissimilarity Matrix* creation. The challenge in our case was to create dissimilarity matrix for qualitative variables. For this purpose, the Gower's distance was used. This measure is taking into account distances between pairs of variables over all dataset and then combines those distances into a single value per record-pair. The distance is always a number from range: 0 (identical) to 1 (maximally dissimilar).

The next step *Hierarchical clustering* was carried out using the Ward's minimum variance method. The result of hierarchical clustering is visualized in Figure 2 as dendrogram with marked clusters.
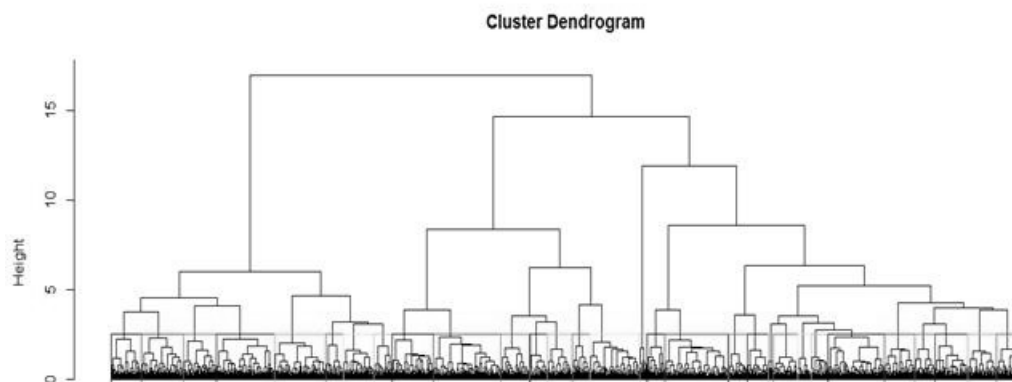

Fig. 2. Cluster dendrogram of unique parameters of shearer work

Hierarchical clustering is one of the unsupervised data mining techniques that requires a predetermined number of clusters. The optimal number of clusters was determined based on Silhouette information (Rousseeuw and Silhouettes, 1987) according to a specified number of clusters and final statistical description. Initially 30, 18, 10 and 7 numbers of clusters have been investigated.

In the *Cluster labelling* step 18 groups for further analysis were selected. The following process activities based on expert knowledge and practical aspects of the selected abstraction level have been defined: (1) Traction difficulty, (2) Shearer stoppage, (3) Start mining, (4) Mining – Type I, from (5) to (12) Stope change – from Type I to Type VII, (13) Mining – Type II, (14) Mining – Type III, (15) Stope change – Type VIII, (16) Security – Type I, (17) Stope change – Type IX and (18) Security – Type II. The results of *Clustering* step were used as labeled data for automatic detection of activities in the raw data set.

3.3. Classification

Detection of activities in the raw data requires a set of rules describing unique stages of the analyzed process. Previous steps enabled labeling of instances which can be used as *predictor variables* in supervised learning. In the *Classification* step firstly classification tree algorithm – *Chi-square Automatic Interaction Detection* has been used. This algorithm was selected due to qualitative character of the data and its accuracy in the validation stage. Events with labeled stages have been treated as a learning sample. For each label unique rules have been generated in a form of CHAID model. Then *Model Prediction* based on obtained rules was done and its examples results are presented in Figure 3.
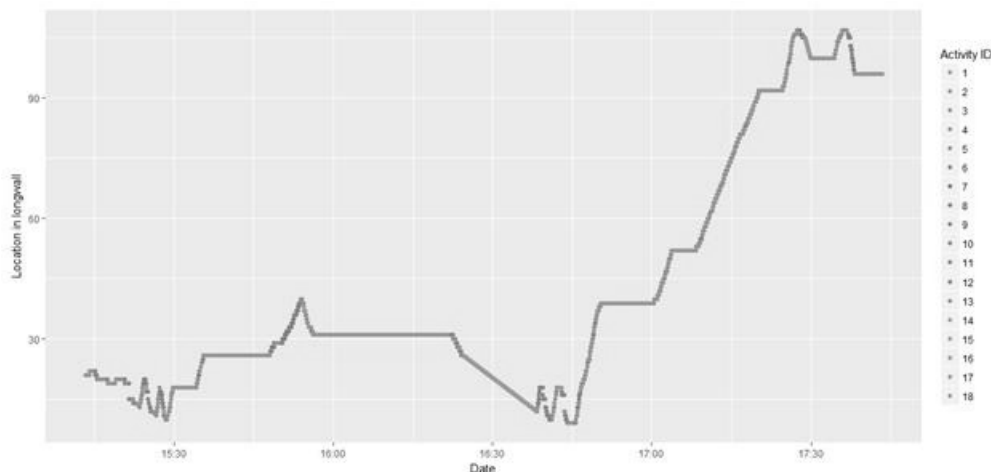
Fig. 3. Labeled process stages on an example trace

On the example trace 18 defined process stages are labeled. This assignment has enabled creation of eligible event log.

3.4. Creation of an event log

In Table 1 an example of event log based on discovered activities related to one trace – cycle of shearer operation is presented. Created event log contains main information required for process mining: *Case* – attribute based on location of shearer in a longwall face, *Timestamp* – precise time of an activity occurrence, *Activity ID* – number of defined activity and *Activity/process stage* – name of an activity. Additionally, more detail *Description* of process stages is provided.

Table 1
The fragment of created event log

Case	Timestamp	Activity ID	Activity/process stage	Description
1	29.01.2018 15:29:49	3	Start mining	Start mining process after stoppage
1	29.01.2018 15:29:51	11	Stope change – Type IV	Change the stope position during stoppage, arm left up
1	29.01.2018 15:29:53	3	Start mining	Start mining process after stoppage
1	29.01.2018 15:29:54	2	Shearer stoppage	Roof problems, cleaning or non-technological breaks
1	29.01.2018 15:34:02	3	Start mining	Start mining process after stoppage
1	29.01.2018 15:34:04	4	Mining – Type I	Normal mining, move in the right
1	29.01.2018 15:34:24	6	Stope change – Type II	Change the stope position during mining, move in the right, arm left up
1	29.01.2018 15:34:25	4	Mining – Type I	Normal mining, move in the right
1	29.01.2018 15:34:44	8	Stope change – Type IV	Change the stope position during mining, move in the right, arm right up

The created event log enabled to perform shearer operation process modelling, briefly presented in the next section.

4. PROCESS MODELLING

In process modelling various process mining techniques and formalisms can be used (van der Aalst 2016). In Figure 4 process model discovered from obtained event log in a form of Petri net is presented.
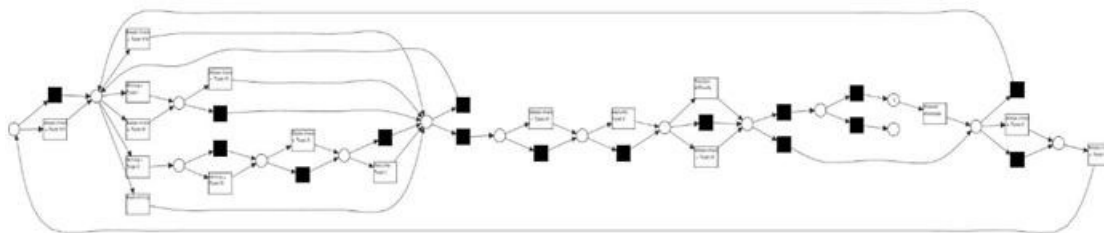
Fig. 4. Discovered process model in a form of Petri net

The process model was created with *Inductive Miner* algorithm in the most powerful open-source tool for process mining – ProM (2016).
The model shows general behavior which was observed and the most frequent activity flows on the specific level of abstraction (18 unique stages).
Model is quite complex and several differences can be found in comparison to ideal process model of a longwall shearer operation presented in Figure 5.
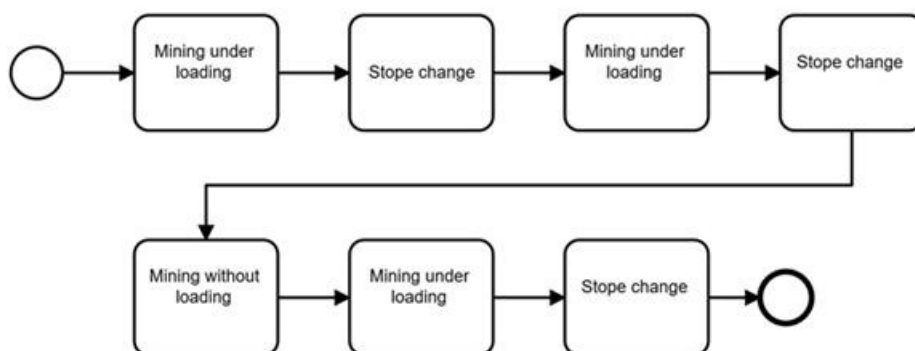
Fig. 5. The ideal process model of a longwall shearer operation

The most important is that those two models are not at the same level of abstraction and theoretical model is more simple. Real-life model contains much more detailed stages which describe never defined random activities.
Discovered model provides valuable insights about real process performance. Firstly it shows diversity of possible process instances. Model enabled also identification of discrepancies with ideal model e.g. activity repetitions. More detail comparison can be done with conformance checking techniques and further works in this scope will be carried out.

5. CONCLUSIONS

In the paper we presented practical aspects of event logs creation based on industrial sensors data gathered in an underground mine monitoring systems. We provide an approach to transform sensor data into an event log with supervised and unsupervised data mining techniques. An event log enables process modelling and its analysis with various process mining techniques.
Our approach for an event log creation consists of four main stages, namely preprocessing, clustering, classification and final preparation of an event log. During analysis various levels of event log abstraction were investigated (30, 18, 10 and 7 activities respectively). Created event log (with 18 unique defined activities) was used in automatic process model discovery in ProM software. Discovered process model in a form of Petri net has shown the real process performance recorded in the event log.

According to presented approach a new event log can be easily created on the lower level of abstraction i.e. with 30 unique stages or the higher level of abstraction i.e. with 7 stages.

The paper presents one part of research work on new possibilities for analysis of underground mining process with process mining. Process-oriented nature of process mining techniques can bring valuable insights for improvement of underground mining process performance increasing its safety, productivity and efficiency. In our opinion process mining can be innovative solution for sensor data analysis in various industry branches.

ACKNOWLEDGEMENTS

REFERENCES

van der Aalst, W.M.P. (2018). Process Discovery from Event Data: Relating Models and Logs Through Abstractions. WIREs Data Mining and Knowledge Discovery, to appear. http://wwwis.win.tue.nl/~wvdaalst/publications/z6.pdf [Accessed 30 May. 2018].

van der Aalst, W.M.P. (2016). Process Mining: Data Science in Action. Berlin: Springer–Verlag.

van der Aalst, W.M.P. (2013). Process Mining in the Large: A Tutorial. . Lecture Notes in Business Information Processing,172, pp. 33–76.

Bose, R.P.J.C., van der Aalst, W.M.P. (2009). Abstractions in process mining: A taxonomy of patterns. Germany. 7 th International BPM Conference Proceedings, pp. 159–175.

Brzychczy, E., Trzcionkowska, A. (2017). New possibilities for process analysis in an underground mine. In: Management in mining production, economic, social and technical perspectives and experiences. Zeszyty Naukowe. Organizacja i Zarządzanie, 111, pp. 13–25.

Cook, D.J., Krishnan, N.C., Rashidi, P. (2013). Activity discovery and activity recognition: A new partnership. IEE T. Cybernetics. 43(3), pp. 820–828.

van Eck, M.L., Sidorova, N., van der Aalst. W.M.P. (2016). Enabling Process Mining on Sensor Data from Smart Products. Brussels. IEEE RCIS. IEEE Computer Society Press, pp. 1–12 .

Guenther, C.W., van der Aalst, W.M.P. (2006). Mining Activity Clusters from Low–Level Event Logs. Eindhoven. BETA Working Paper Series, WP 165.

Hompes, B.F.A., Verbeek, van der Aalst, H.M.W. (2014). Finding Suitable Activity Clusters for Decomposed Process Discovery. Proceedings of the 4th International Symposium on Data–driven Process Discovery and Analysis (SIMPDA 2014), 1293, pp. 16–30.

Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P., Toussaint, P.J. (2016). From low–level events to activities – a pattern–based approach. International Conference on Business Process Management, pp.125–141.

ProM (2016). Process Mining Group. Eindhoven Technical University. Available at: http://www.promtools.org.

Rousseeuw, P., Silhouettes J. (1987). A graphical aid to the interpretation and validation of cluster analysis,. Journal of Computational and Applied Mathematics, 20, pp. 53–65.

Tax, N., Sidorova, N., Haakma, R., van der Aalst, W.M.P. (2018). Event Abstraction for Process Mining Using Supervised Learning Techniques. In: Bi. Y. Cham, S. Kapoor, R. Bhatia, eds., Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016. Lecture Notes in Networks and Systems,15, pp. 51–269.