



Creation of an Event Log from a Low-Level Machinery Monitoring System for Process Mining Purposes

Edyta Brzychczy^(✉) and Agnieszka Trzcionkowska

AGH University of Science and Technology, Kraków, Poland
{brzych3, toga}@agh.edu.pl

Abstract. Industrial event logs, especially from low-level monitoring systems, very often have no suitable structure for process-oriented analysis techniques (i.e. process mining). Such a structure should contain three main elements for process analysis, namely: timestamp of activity, activity name and case id.

In this paper we present example data from a low-level machinery monitoring system used in underground mine, which can be used for the modelling and analysis of the mining process carried out in a longwall face. Raw data from the mentioned machinery monitoring system needs significant pre-processing due to the creation of a suitable event log for process mining purposes, because case id and activities are not given directly in the data.

In our previous works we presented a mixture of supervised and unsupervised data mining techniques as well as domain knowledge as methods for the activity/process stages discovery in the raw data. In this paper we focus on case id identification with an heuristic approach. We summarize our experiences in this area showing the problems of real industrial data sets.

Keywords: Event logs · Process mining · Low-level monitoring system
Underground mining · Longwall face

1 Introduction

The Internet of Things and Industry 4.0 in the mining industry have become a fact. The great step in underground industrial advancement which completed automatization in the field is the convergence of industrial systems with the power of advanced computing, analytics, low-cost sensing and new levels of connectivity. Smart sensor technologies and advanced technics of analysis play an important role in mining process monitoring and improvement [11].

Automation has enabled access to very detailed data characterizing the operation of machines and devices (stored in monitoring systems). Longwall automation and monitoring systems allows a closer look at the ongoing processes underground. A vast amount of data is generated that should be used and handled more efficiently in a modern mining operation [4]. Nowadays, the analytics of collected data is mainly based on data-oriented techniques: BI techniques for operational report creation as well as on more advanced analytic data mining and machine learning techniques for predictive

maintenance purposes [8]. Thus, for acquiring new knowledge about ongoing processes underground, we proposed process-oriented analysis of the gathered data [2].

In the paper we present example data from a low-level machinery monitoring system used in underground mine, which can be used for the modelling and analysis of the mining process carried out in a longwall face.

Our work aims proposal of original extension of data analysis from low-level longwall machinery monitoring system with process mining techniques and according to the authors' best knowledge, it is the first attempt of process mining usage in the mining domain [6].

The basic challenge raising from the proposed analysis extension is the creation of a suitable event log for process mining purposes containing [1]: timestamp of activity, activity name and case id. This is not a trivial task since case id and activities (process stages) are not given directly in the raw data from the low-level longwall machinery monitoring system. Moreover, there is no procedure to identify case id in the raw data from the longwall monitoring system, since no one has applied process mining in the mining domain.

To address the mentioned challenges we prepared two procedures for data processing:

1. For activities' name definition we proposed the mixture of supervised and unsupervised data mining techniques as well as domain knowledge, presented in more detail in [2]. That proposal contains among others: data cleaning, clustering and classification for labelling the process stages in the raw data.
2. For case id identification, we propose the heuristic approach presented in this paper. Our solution is an example of how we can handle with raw data related to the cyclic process in a specific production domain without clear marking of the start and the beginning of the case.

Our both procedures are written in R, mainly using libraries: `dplyr`, `arules`, `cluster`, `forecast`, `CHAID` and `rpart`.

The paper is structured as follows: Sect. 2 includes mining process description. In Sect. 3 identification of case id in raw data is presented. An example of created event log is described in Sect. 4. Conclusions are presented in Sect. 5.

2 Process Description

The mining process can be defined as a collection of mining, logistics and transport operations. One of the most complex and difficult examples of its realisation is underground mining characterized by changeable geological and mining conditions as well as natural hazards not occurring on the surface. Very interesting is the nature of the mining process in the longwall system that is performed by machines and devices moving in a workspace and also in relation to each other.

Main longwall equipment includes (Fig. 1): a shearer (A), an armoured longwall conveyor (B) and mechanized supports (C).

Each of the mentioned machines realises its own operation process, consequently the mining process in a longwall face can be seen as collection of machinery processes.



Fig. 1. Longwall machinery (https://famur.com/upload/2016/09/FAMUR_01-1.jpg)

The mining process includes even up to a hundred processes (depending on the dimensions of a mining excavation and number of mechanized supports).

In the paper we focus on the operation process of main longwall machinery, namely the shearer. The operation of the shearer indicates the cycle of a whole mining process [12], therefore it is the most intuitive choice for case id in an event log. The theoretical shearer operation process is presented in Fig. 2.

In general a shearer operation cycle consists of several characteristic phases. Firstly, the shearer starts cutting from the driver unit side (1). The next phase is indentation where a shearer is cutting into the turning station direction for a distance of 30–40 m. Together with the movement of the shearer a longwall conveyor is shifting (2). The third step is cutting into the driver unit side – longwall cleaning (3). Along with the shearer the powered roof support is moving. In the next phase a shearer is cutting without loading for a distance of 30–40 m (4), after that it is cutting throughout the longwall till the turning station. Along with the movement of the shearer, the conveyor and powered roof support are moved (5, 6, 7).

The basic indicator of the shearer’s movement is the value of the “Location in the longwall” variable. The ideal model of the shearer operation with activity names is presented in Fig. 3.

The real location of a shearer in a raw data is presented in Fig. 4.

Two main challenges in modelling the mining process based on real data are illustrated well on the picture: data quality and cycles variability. The first challenge has a major source in technical problems in data transfer from the machinery to the surface, especially in the case of power off events and data retrieving from the machinery local data containers. The second challenge is strictly related to the mining and geological conditions of process realisation.

It should be also mentioned that raw data contain various quantitative and qualitative (mostly binary) variables that in some way describe the process stages, not directly as activity names. It needs a lot of analytic efforts to build event logs on top of it (activity recognition, abstraction level choice etc.). Our contribution in this area is presented in the following sections.

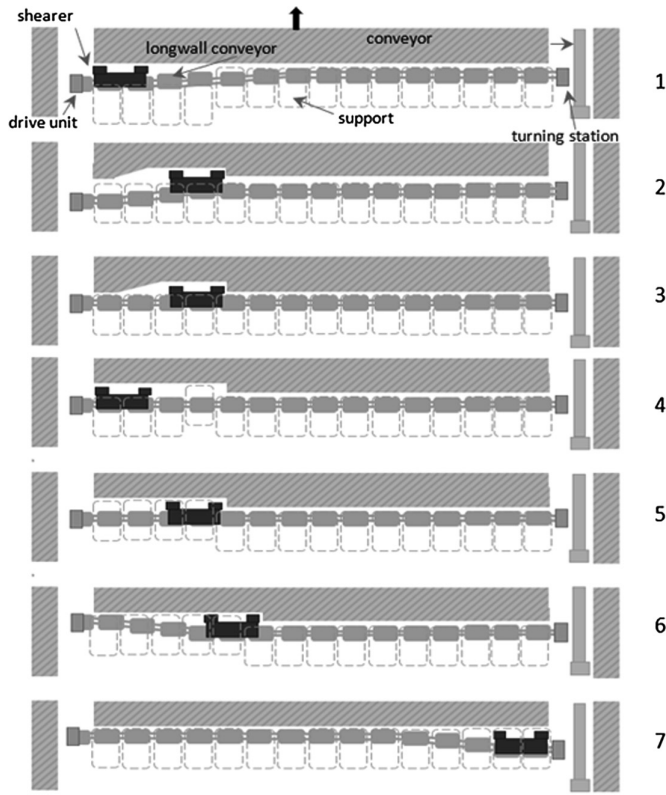


Fig. 2. Example cycle of shearer operation. Source: based on [10]

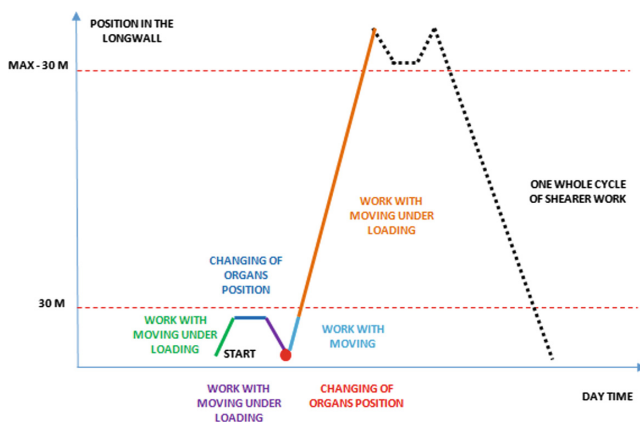


Fig. 3. Example cycles of shearer operation in time dimension

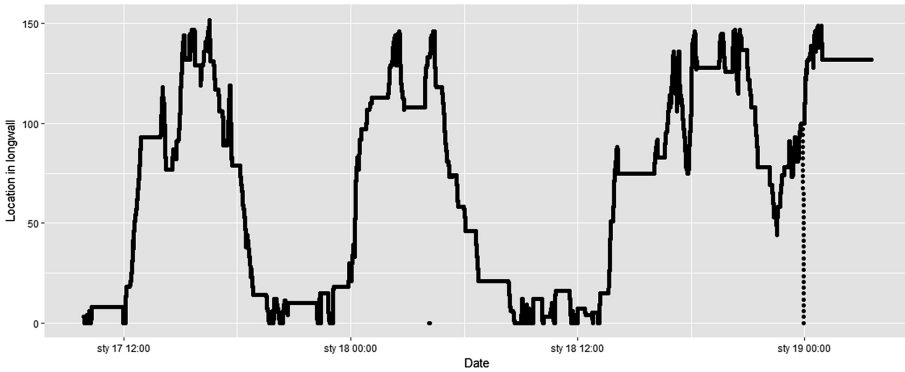


Fig. 4. Example cycles of the shearer operation

3 Identification of Case ID in a Raw Data

In this section we present issues related to case id identification for the purpose of event log creation based on the shearer operation data from the selected hard coal mine. Raw data related to the mentioned process include 2.5 million records from a monthly period obtained from one of the Polish mining companies.

In Table 1 the selected variables characterizing the shearer operation are presented.

Table 1. Selected variables characterizing the shearer operation

Variable	Type	Range
Location in the longwall	Numerical	0–200 [m]
Shearer speed	Numerical	0–20 [obr/min]
Arm left up/down	Binary	0/1
Arm right up/down	Binary	0/1
Move in the right	Binary	0/1
Move in the left	Binary	0/1
Current on the left organ	Numerical	0–613 [A]
Current on the right organ	Numerical	0–680 [A]
Current on the left tractor	Numerical	0–153 [A]
Current on the right tractor	Numerical	0–153 [A]
Security DMP left organ	Binary	0/1
Security DMP right organ	Binary	0/1
Security DMP left tractor	Binary	0/1
Security DMP right tractor	Binary	0/1

The identification of the shearer’s work cycles (case id) was mainly based on the analysis of the attributes “Location in the longwall” (distance below 5 m and over 135 m) and “Shearer speed” (equal to 0 m/s).

The first approach of the cycle start and finish identification was based on the classic analysis of local minimum and maximum. This approach did not yield satisfactory results. The main problem was related to large local process variability.

Therefore, the heuristic approach with the following steps was proposed.

1. The shearer's position in the longwall face was split into three ranges (Fig. 5) according to the technological conditions and theoretical model of the cycle:

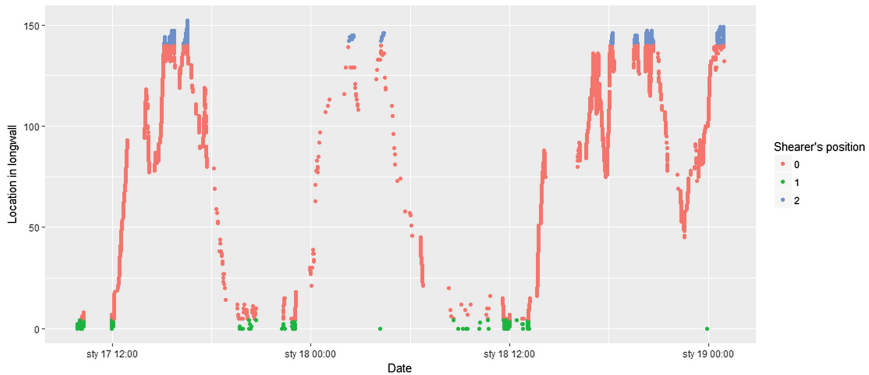


Fig. 5. Ranges of the shearer's position

- the beginning of the longwall face - distance below 5 m (was marked with 2),
 - the end of the longwall face - over 135 m (was marked with 1),
 - and in the middle of the longwall face (marked with 0).
2. In the range sets, below 5 m and over 135 m, the local minimum (1) and maximum (2) were detected accordingly.
 3. Characteristic peaks (start and end of the cycle) were identified by the selection of sequences only with specific range (1) and (2) order (Fig. 6).

In the analysed dataset 75 cycles were identified (9 cycles are presented in Fig. 7)

In the most cases the proposed heuristic enabled the identification of the cycle start and end correctly. The errors in identification were caused mainly by data quality. The red arrows in the Fig. 7 points out one of the main issues: incorrect state of location. Thus, the presented approach is sensitive on data quality and further works will be focused on improving data cleaning at the early stages to avoid the mentioned issue.

In the case of lack of data in a shearer location variable extrapolation between the nearest two points existing in the data can be done. We know how the theoretical and technological cycle looks like, so extrapolation, based also on other variables values, could be verified.

The shearer cycle is crucial for all machinery working in the longwall face, because the rest of the machines and devices are just adjusting to the shearer position in the cycle. Therefore, for process modelling purposes, it is very important to find the way how in raw data a start and end of the cycle can be identified. Especially, when real cycles are varied very much from theoretical models.

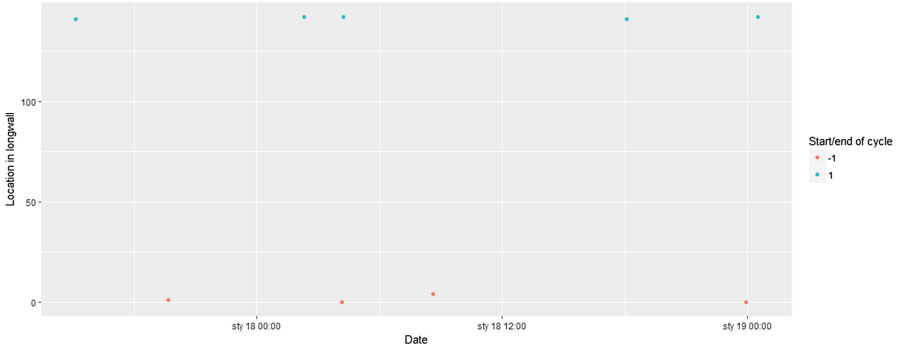


Fig. 6. Visualization of the beginning and the end of the cycles

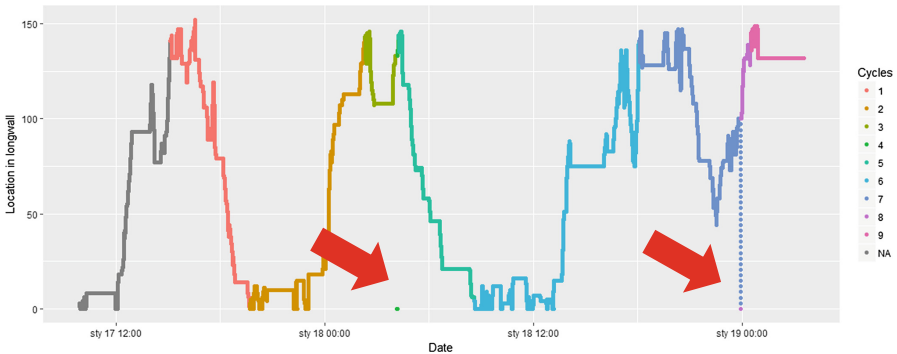


Fig. 7. Example of identified cycles

Although our approach is for a very specific domain, we think that it can be helpful for the creation of event logs for similar problems and processes.

4 Creation of an Event Log

The creation of an event log based on sensor data, beside a case id identification requires the recognition and identification of activities (process stages). In these cases supervised and unsupervised techniques of data mining can be applied [3, 5, 7, 9, 13].

Selected variables (Table 1) were used for distinguishing the unique states of the shearer operation, according to the procedure, described in [2]. The following stages were performed:

1. *Data preprocessing.* In this stage exploratory data analysis was conducted. Subsequently an analysis of correlation for the numerical variables and cross tables for the logical variables were performed to exclude the depended variables. Then the discretization of all continuous variables into a categorical variables was carried out. Furthermore, in the final data set, containing discretized and logical variables, duplicate rows were removed.

2. *Data clustering.* For the final data set dissimilarity matrix with Gower’s distance was created. Then hierarchical clustering was carried out using the Ward’s minimum variance method. Finally, selected clusters have been labelled with activity names, based on the statistical analysis results and an expert knowledge.
3. *Classification for labelling activity names in the raw data.* In this stage instances with a labeled activity name (process stage) have been used as a learning sample in the CHAID tree algorithm. For each label, according to the CHAID tree model, unique rules have been generated and, on this base, activity labelling in the raw data was done.

The identification of case id and activity definition enable the creation of an event log presented in Table 2. The process stages labelled on the example traces are shown in Fig. 8.

Table 2. Fragment of an event log (selected in the Fig. 8)

Case id/trace	Timestamp	Activity/process stage
1	17.01.2018 15:23:11	Mining – Type II
1	17.01.2018 15:24:14	Start mining
1	17.01.2018 15:25:20	Stope change – Type IX
1	17.01.2018 15:25:35	Start mining
1	17.01.2018 15:26:19	Stope change – Type IX
1	17.01.2018 15:26:21	Start mining
1	17.01.2018 15:27:15	Stope change – Type IX
1	17.01.2018 15:28:17	Start mining
1	17.01.2018 15:28:37	Shearer stoppage
1	17.01.2018 15:29:18	Start mining
1	17.01.2018 15:32:39	Shearer stoppage
1	17.01.2018 15:41:29	Start mining

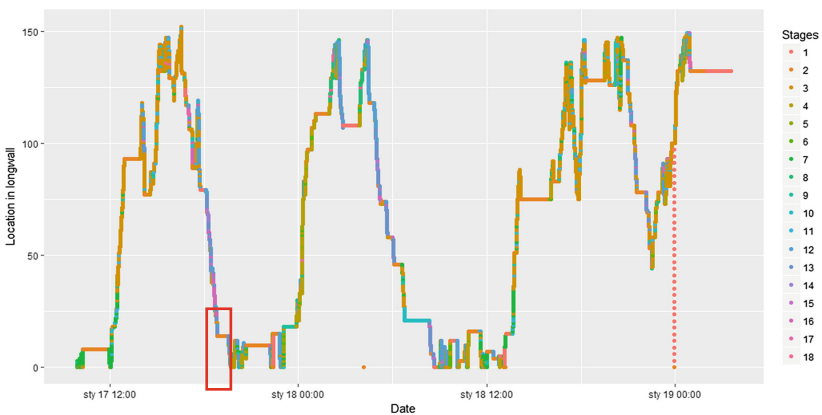


Fig. 8. Labelled process stages on example traces

A created event log enables the performance of process modelling with selected techniques and formalisms [1] and further works in this scope are carried out.

5 Conclusions

Current underground machinery monitoring systems can contain streaming data from hundreds of sensors of various types. The efficient processing of such an amount of data (Big Data) for process improvements is possible only with the specific techniques of advanced analysis from data mining and process mining fields.

Process mining techniques require a specific structure of an event log with activity names and case id, that very often are not present in raw industrial sensor data. The challenges related to activity recognition and case identification are strongly connected to the data quality and nature of an analyzed process. Therefore, cleaning and pre-processing activities are needed and adequate analytic approaches should be found.

In the paper we presented case id identification problems on a selected example from the longwall monitoring system in an underground mine. The classic approach in this case has not yielded correct results due to the high variability of the process and the existence of many local optima, thus the heuristic approach was developed.

We have contributed original solutions (procedures) for an event log creation from a low-level machinery monitoring system in underground mining for process mining purposes. Future challenges will be related to process modelling based on prepared event logs in the case of high process variability.

Acknowledgements. This paper presents the results of research conducted at AGH University of Science and Technology – contract no. 15.11.100.181.

References

1. van der Aalst, W.M.P.: Data science in action. In: van der Aalst, W.M.P. (ed.) *Process Mining*, pp. 3–23. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49851-4_1
2. Brzychczy, E., Trzcionkowska, A.: Process-oriented approach for analysis of sensor data from longwall monitoring system. In: Burduk, A., Chlebus, E., Nowakowski, T., Tubis, A. (eds.) *ISPEM 2018. AISC*, vol. 835, pp. 611–621. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-97490-3_58
3. Cook, D.J., Krishnan, N.C., Rashidi, P.: Activity discovery and activity recognition: a new partnership. *IEEE Trans. Cybern.* **43**(3), 820–828 (2013). <https://doi.org/10.1109/tsmcb.2012.2216873>
4. Erkayaoğlu, M., Dessureault, S.: Using integrated process data of longwall shearers in data warehouses for performance measurement. *Int. J. Oil Gas Coal Technol.* **16**(3), 298–310 (2017). <https://doi.org/10.1504/ijogct.2017.10007433>
5. van Eck, M.L., Sidorova, N., van der Aalst, W.M.P.: Enabling process mining on sensor data from smart products. In: *IEEE RCIS*, pp. 1–12. IEEE Computer Society Press, Brussels (2016). <https://doi.org/10.1109/rcis.2016.7549355>
6. Gonella, P., Castellano, M., Riccardi, P., Carbone, R.: *Process mining: a database of applications*. Technical report, HSPI SpA - Management Consulting (2017)

7. Guenther, C.W., van der Aalst, W.M.P.: Mining activity clusters from low-level event logs. BETA Working Paper Series, WP 165, Eindhoven University of Technology, Eindhoven (2006)
8. Korbicz, J., Koscielny, J.M., Kowalczyk, Z., Cholewa, W. (eds.): Fault Diagnosis: Models, Artificial Intelligence, Applications. Springer, Heidelberg (2004). <https://doi.org/10.1007/978-3-642-18615-8>
9. Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P., Toussaint, P.J.: From low-level events to activities - a pattern-based approach. In: La Rosa, M., Loos, P., Pastor, O. (eds.) BPM 2016. LNCS, vol. 9850, pp. 125–141. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45348-4_8
10. Napieraj, A.: The method of probabilistic modelling for the time operations during the productive cycle in longwalls of the coal mines (in Polish). Wydawnictwa AGH, Cracow (2012)
11. Ralston, J.C., Reid, D.C., Dunn, M.T., Hainsworth, D.W.: Longwall automation: delivering enabling technology to achieve safer and more productive underground mining. *Int. J. Mining Sci. Technol.* **25**(6), 865–876 (2015). <https://doi.org/10.1016/j.ijmst.2015.09.001>
12. Snopkowski, R., Napieraj, A., Sukiennik, M.: Method of the assessment of the influence of longwall effective working time onto obtained mining output. *Archives Mining Sci.* **61**(4), 967–977 (2016). <https://doi.org/10.1515/amsc-2016-0064>
13. Tax, N., Sidorova, N., Haakma, R., van der Aalst, W.M.P.: Event abstraction for process mining using supervised learning techniques. In: Bi, Y., Kapoor, S., Bhatia, R. (eds.) *IntelliSys 2016*. LNNS, vol. 15, pp. 251–269. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-56994-9_18