APPLICATION OF BICLUSTERING TECHNIQUE IN MACHINE MONITORING

Marcin Michalak

Institute of Informatics, Silesian University of Technology, Gliwice, Poland
`Marcin.Michalak@polsl.pl`

ABSTRACT

Nowadays we can observe the change of the structure of energy resources, which leads to the increasing fraction of a renewable energy sources. Traditional underground coal mining loses its significance in a total but there are countries, including Poland, which economy is still coal based. A decreasing coal resources imply an exploitation a becoming harder accessible coal beds what is connected with the increase of the safety of the operation. One of the most important technical factor of the safety of underground coal mining is the diagnostic state of a longwall powered roof support. It consists of dozen (or hundreds) of units working in a row. The diagnostic state of a powered roof supports depends on the diagnostic state of all units. This paper describes the possibility of unit diagnostic state analysis based on the biclustering methods.

KEYWORDS

Biclustering, Machine Monitoring, Machine Diagnosis

1. INTRODUCTION

In a coal mining industry – similarly as in the case of other industry branches – the growth of monitoring systems application. Initially, monitoring systems were designed just for the purpose of data acquisition and presentation is being observed. Over time, their abilities were extended in the direction of simple dangerous situations recognition and finally – to the advanced machine diagnostic state analysis and its prediction for the nearest future.

Longwall systems are the basis of the coal mining, because the longwall is the place in the process of mining from which we can say about the output. Mechanised longwall systems consist of longwall shearer (which tears off the output from the rock), longwall conveyor (transports the output from the longwall to the heading) and units of powered roof support (prop the roof after mining the output).

Longwall systems are very interesting objects from the collected data point of view. Its most important part is a power roof support. Its primary task is to protect the other elements of the longwall system, especially the coal shearer which is an essential part as of a coal mining process, not to mention the protection the workers from the falling rocks. Power roof support consists of units. In a particular moments of time – after shearer takes the another part of the

longwall output – each unit has to move in the direction of the whole longwall face advance (treading), protecting the rock material, exposed by the shearer, from collapsing.

Unequal propping can be caused by leaks in the hydraulic system (pipes, valves) or leaks in legs of the unit. To long times of treading can point the wrong diagnostic state of the unit or be caused by the wrong usage (so called: moving with the contact of roof-bar with the roof). It is also dangerous to perform the treading too long as the roof is not propped. So it can be stated that the safety of coal mining is determined by the diagnostic state of all parts of the longwall mining system, including the diagnostic state of all units of power roof support.

In this article the ability of adaptation of biclustering method for the purpose of the analysis the data from longwall monitoring systems is presented. The paper is organized as follows: it starts from the brief description of monitoring systems with a special consideration of underground coal mining monitoring system is presented. Than the description of a construction, the role and a working cycle of a powered roof support is presented together with the proposed monitored (and extracted) parameters. Next part presents a wide group om continuous and binary data biclustering methods. The paper ends with application of an OPSM biclustering algorithm for the real data.

2. MONITORING SYSTEMS

Nowadays, software producers and monitoring systems users point the need of analysis of the data, collected in repositories of these systems. In particular, the definition of diagnostic models of monitored devices can be a goal of this analysis [8]. The process of a diagnostic model identification can be carried out by planned experiments or on the basis of a data from the past device operation. In a specific situation, when there are no data describing an improper state of the machine it is possible to develop a model describing only the proper state of the machine and treat the deviation from the model as the possibility of improper machine state [11].

The problem of a monitoring and diagnosing of a coal mine industry devices was raised recently in [1][4][7][11][12][17]. These topics are presented widely and review in [23]. In these works also new methods of extraction and processing of new diagnostic features in new diagnostic relations discovering are presented. Especially in [1] the diagnostic of conveyor belts is described. In [18] a current consumption and a temperature of roadheader cutting head. On the basis of these parameters three roadheader working states were defined. Two of them described different but correct underground mining conditions. In this paper also the parameter reflecting the roadheader cooling system efficiency was defined.

Longwall conveyors diagnostic was an aim of the following works [4][7][12]. In [4]the way of conveyor chute failures detection on the bottom side of a conveyor was presented. On the basis of the conveyor engines power consumption analysis the failure was detected with an accuracy to the one unit. In [7] the complex subassemblies management system was proposed, which allows to generate operational and analytical reports as well as summary statements. In [12] a harmonic analysis was used for the prediction of a diagnostic state of longwall conveyor chain.

3. BICLUSTERING

Biclustering is the problem of unsupervised data analysis, where we are grouping scalars from the two-dimensional matrix. It called also as co-clustering, two-dimensional clustering or twomode clustering. This approach has been started in 70's in the last century [6] and is successfully applied in bioinformatics [3][15][21]. The idea of biclustering is to find in a matrix a subset of colulmns and subset of rows, which intersection gives a submatrix of cells with similar (or the same) values.

In the literature there are a lot of algorithms of biclusters induction. In [3] authors define bicluster as a subset of rows under subset of columns, for which calculated parameter (mean squared residue score) is below threshold defined by the user. The minimum value of the considered parameter is 0. The algorithm consists of two steps. Initially the rows and columns are removed from input dataset, until the value of mean squared residue score is below assumed level. Then rows and columns, which were removed during the first step, are added to obtained in the previous step submatrix until its score fulfils the criterion of being bicluster. After each iteration, the founded bicluster has been hidden with random values. The extension of this algorithm proposed in [22] allows to avoid noise among input dataset, which was a consequence of masking discovered biclusters.

The Order Preserving Submatrix Algorithm was presented in [2]. The bicluster was defined as subset of rows, which preserves linear ordering across subset of conditions. The set of valid biclusters is identifying by algorithm based on stochastic model. This idea was also evolved in [9].

The algorithm X-Motif is dedicated to the extraction of conserved gene expression motifs from gene expression data and has been proposed in [14]. Bicluster is defined as subset of genes, which expression level belong to the same state across subset of conditions. The states are assigned to genes during preprocessing step. In order to find multiple biclusters an algorithm is running in an iterative way. Each iteration starts from different initial sets.

There exists also methods of biclustering dedicated for matrices with the binary values. Bimax [16] uses a simple divide-and-conquer approach for finding all inclusive maximal biclusters for a given minimal number of rows and columns. Bicluster, which is maximal in the sense of inclusion is defined as not entirely contained in any other bicluster. Such assumption allows to exclude from analysis individual cells equal to one, which can be considered as a single biclusters, however they provide no important information.

BicBin [20] is an algorithm dedicated for binary sparse matrices. It consists of three main components: the score function to evaluate a submatrix, the search algorithm to restrict the search space of all possible submatrices and an algorithm used to extract all biclusters in an ordered way from a dataset. BicBin is dedicated for finding inexact biclusters. Each run of BicBin may give different results, because algorithm finds set of random biclusters, which fulfil its restrictions and cover all ones in dataset.

A novel approach of the binary matrix biclustering is based on the rough sets theory [13] where non-exact biclusters are defined as the ordered pair of biclusters called a lower and an upper approximation. The lower approximation is the exact submatrix of the given one and the upper

approximation is non-exact matrix that is the superset of a given one. The algorithm is hierarchic similarly as the Ward clustering algorithm and it starts from the previously generated set of exact biclusters [19]. In every step two rough biclusters can be joined if the intersection of their lower approximations is nonempty. The generalisation of the data description allows to limit the number of final biclusters assuring the assumed level of the description accuracy.

The analogical hierarchical strategy can be also applied for classical biclusters (not considered as the rough bicluster) and was presented in [10].

4. POWERED ROOF SUPPORT

The final safety of operations in a coal mine is dependent on several components: a human factor, natural influence, a technical reasons and – of course – from the unexpected circumstances. A technical operating conditions are an object of interest of multiple monitoring systems [7]. The most common monitored signal is the pressure in the leg (legs) of the section, which reflects the real strength of propping the roof.

A single section of a powered roof support consists of one or more hydraulic prop (legs), which holds up an upper part of the section (roof-bar). A section has also and hydraulic shifting system, which is responsible for shifting the unit with the longwall advance simultaneously. Most of the time the unit props the roof, assuring the safety of mining, but after each shearer run it moves to prop the newly bared roof. A single unit is presented on the Fig. 1.

Fig. 1 Single unit of powered roof support (www.joy.com).

A typical unit working cycle will be presented on the real – 6000 seconds long – observation of the pressure in one of the two-leg unit, shown on the Fig. 2. As it can be observed, it is common situation when two legs are propping the roof-bar with different force (they have different pressure levels in legs).
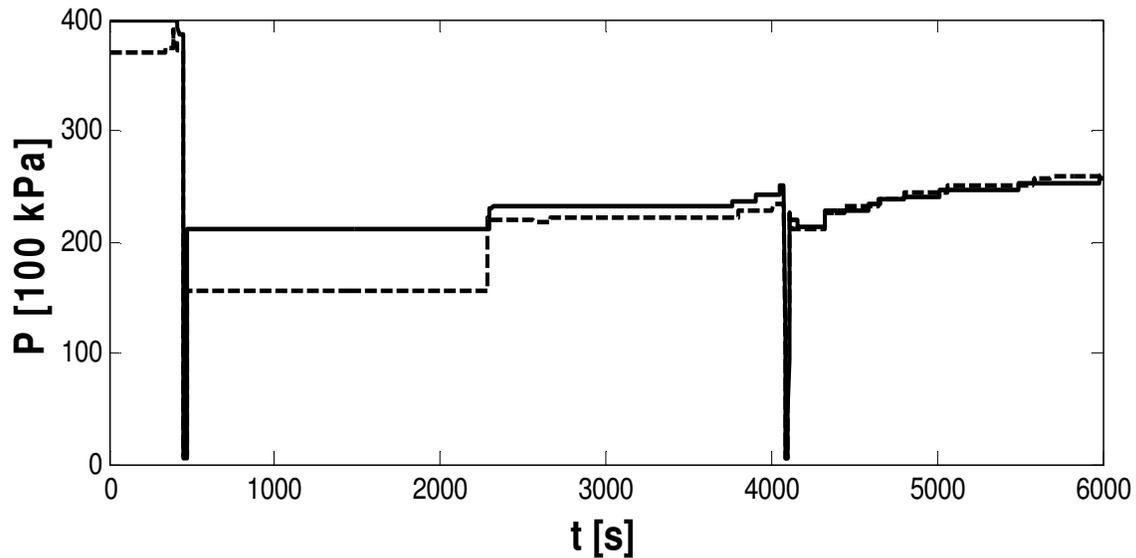
Fig. 2 A real time series of pressure in two legs of the unit.

A typical powered roof support unit working cycle can be divided into several phases. Starting from the moment of the beginning of unit shifting the following phases are named and described as follows:

- treading,
- spragging,
- overbuilding,
- pre-treading,
- pressure lowering.

For each of these phases some similar and some specific variables can be defined, for example: phase duration (for all phases), pressure increase speed (spragging, overbuilding), pre-treading dynamic type, pressure level equability.

The analysis of the mentioned phase statistics requires a diagnostic information, or a diagnostic algorithm, that classified the moment of time (a value of a pressure) into a proper working cycle phase. Otherwise, it is not possible to analyse the whole data – each column representing a single second of powered roof support section work. Then some aggregation of the data must be performed.

The time series, describing a single powered roof support unit that contains several legs, can be an average pressure level in legs or a difference between leg pressure levels. As an aggregate a multiple statistics can be taken into considerations, starting from a mean, maximal and minimal value. The way of data aggregation and a time of aggregation should be dependent on the goal of analysis. For example: using a minimum value in the aggregation time will help to detect a time slices when the section was in a treading phase.

5. EXPERIMENTS

In this paper an application of biclustering methods for pressure level equability in legs of one powered roof support unit will be presented. Application of these methods will help to find a subset of sections which behave similarly in a selected period (or periods – the time interval does not have to be connected). On the basis of a real data, a procedure of diagnostic procedure will be presented. The real data describes the over 8 hours of over 100 sections of powered roof support work. Each section was a double-leg and a pressure level measurement was taken once in a second. On the Fig. 3 an average pressure of two legs is presented. Small values of the absolute difference are marked as white points. As the absolute difference increases the point color becomes more black.
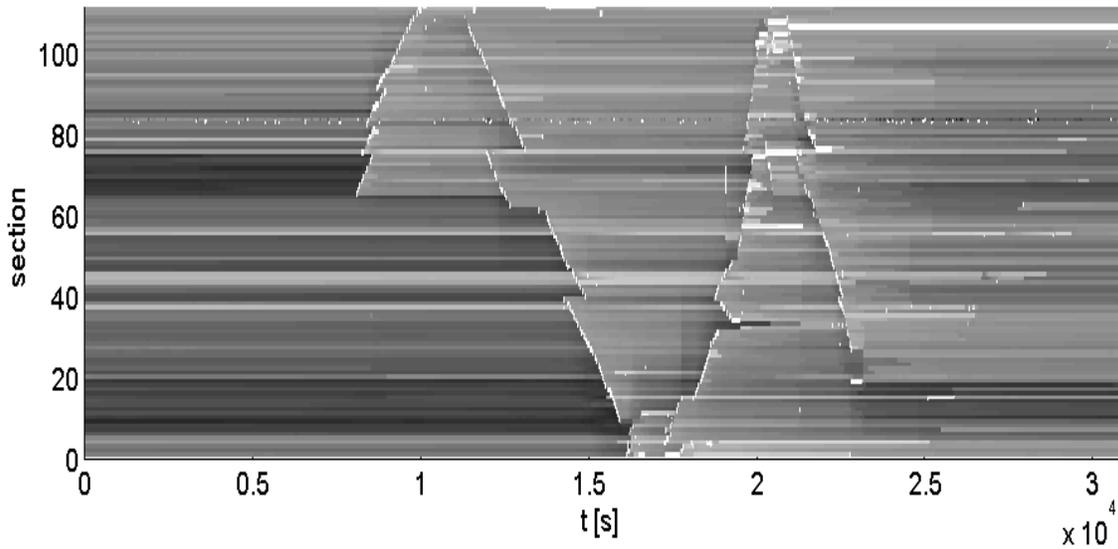

Fig. 3 Average pressure in two legs.

The horizontal zigzag, starting at the 150th minute and ending at the 450th minute refers to the shearer passage and the sections treading. As it can be seen, each section was moved several times. Such small number of treadings does not allow to analyse phases durations. Instead, the pressure level inequality will become the goal of the analysis.

The relative difference of a pressure levels was calculated with the following formula:

$$avg = \frac{|leg_1 - leg_2|}{\min\{leg_1, leg_2\}}$$

where leg_1 and leg_2 are pressure levels in according legs.

Due to the large amount of the data and a goal of data generalisation the aggregation time is set to 60 seconds. As the point of the interest is a difference of a pressure levels the aggregation measure is maximal value of differences in one minute.

The definition of a difference for the pressure level allows to have infinities, as the division by zero is not forbidden. This situation occurs when one of the legs is not propping or there was an error in a measurement device or in a data transmission. The infinity can be easily replaced by 750, which is higher than the highest absolute difference between legs pressure (equal to 70.4 [MPa]). The interpretation of this replacement can be as follows: as only one leg props there is a maximal inequality.

For the further analysis only a subset of sections and a period of time is selected: it will contain 51 middle sections monitored during the middle 200 minutes (Fig. 4).
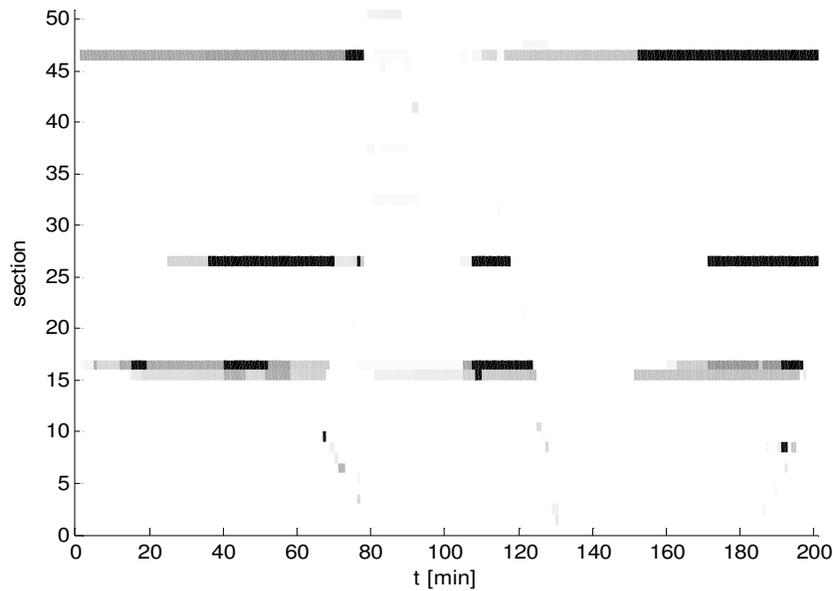

Fig. 4 An aggregated maximal differences of pressure levels in units.

A statistical description of the subset values is presented in the Table 1. The bicluster 0 means the whole data while the following ones are the selected biclusters generated with OPSM algorithm [5]. With the standard algorithm settings it generated 42 biclusters. Four of them are briefly described in the Table 1 and also presented on the Fig. 5. We can observe that the algorithm detected biclusters with the very high pressure level inequality.

Table 1. Statistical description of real data and selected of generated biclusters.

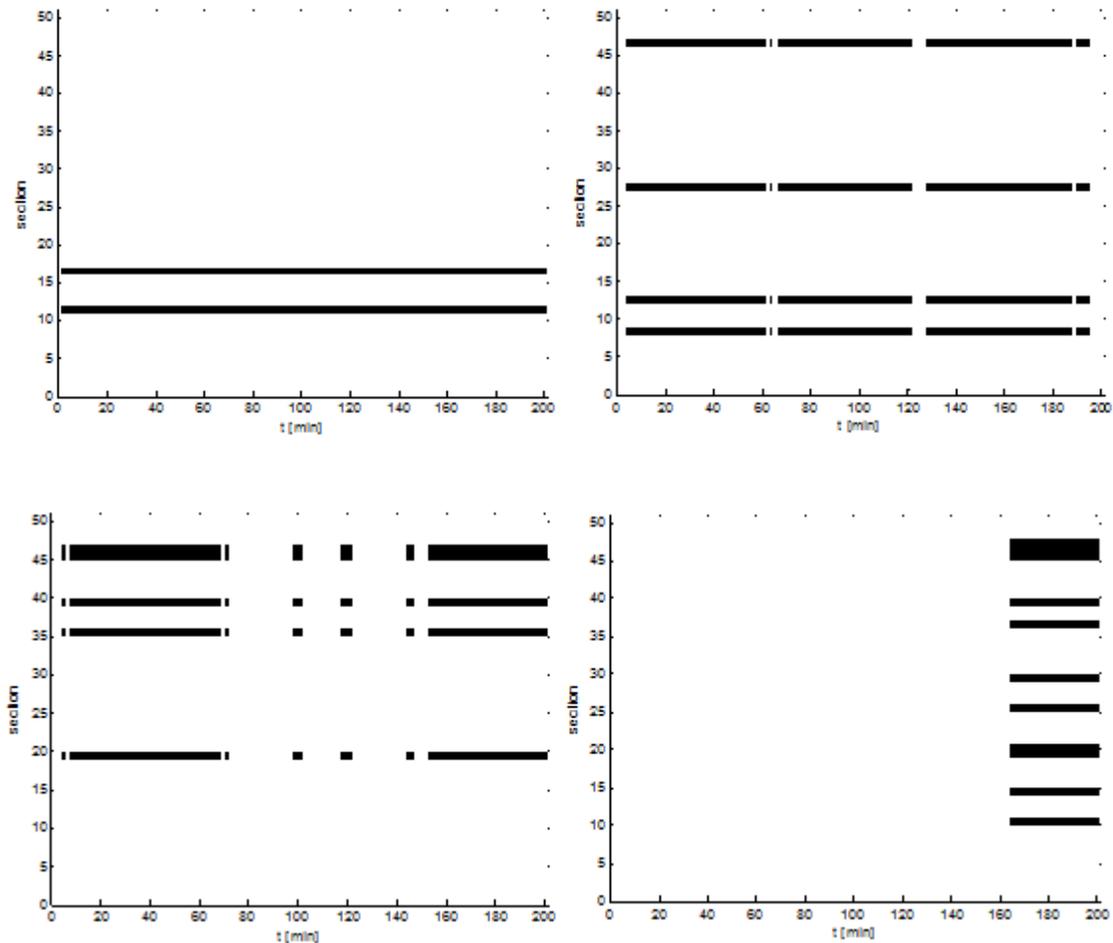| #	min	Q1	Q2	Q3	mean	max	std	duration [min]	sections
0	0.000	0.036	0.085	0.200	20.278	750	103.823	201	51
1	0.000	0.045	0.683	122.286	159.000	750	228.400	201	2
3	0.000	0.053	0.372	85.428	9.667	750	197.493	182	4
5	0.004	0.091	0.152	74.451	2.225	750	194.031	127	6
11	0.020	0.100	0.157	62.837	0.386	750	207.416	38	12

Fig. 5 Four biclusters: 1st (upper left), 3rd (upper right), 5th (lower left) and 11th (lower right).

5. CONCLUSIONS

In the paper the application of biclustering technique for the off-line machine monitoring was presented. This simple example shows a possibility of building procedures for the purpose of a variety aspects of powered roof support monitoring and diagnosing. The presented approach does not exhaust the subject of finding biclusters in a longwall complex data. Future works in this area will focus on joining biclustering results with real improperities in the real data.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Bartelmus W.: Condition Monitoring of Open Cast Mining Machinery. Wroclaw University of Technology Press, Wroclaw 2006.

[2]   Ben-Dor A., Chor B., Karp R., Yakhini Z.: Discovering Local Structure in Gene Expression Data: The Order-Preserving Sub-Matrix Problem, Proceedings of the 6th Annual International Conference on Computational Biology, 2002

[3]   Cheng Y., Church G.: Biclustering of Expression Data, Proceedings of 8th Interantional Conference on Intelligent Systems for Mollecular Biology, 93-103, 2000.

[4]   Gąsior S.: Diagnosis of longwall chain conveyor, Przegląd Górniczy, 57(7–8):33–36, 2001.

[5]   Gupta J., Singh S., Verma N.:MTBA: MATLAB Toolbox for Biclustering Analysis, IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions, IIT Kanpur, India, pp. 94-97, July 2013.

[6]   Hartigan J.A.: Direct Clustering of a Data Matrix. Journal of American Statistical Association, 67(337):123–129, 1972

[7]   Kacprzak M., Kulinowski P., Wędrychowicz D.: Computerized information system used for management of mining belt conveyors operation. Eksploatacja i Niezawodnosc - Maintenance and Reliability, 13(2):81-93, 2011.

[8]   Korbicz J., Kościelny J.M., Kowalczuk Z., Cholewa W.: Fault Diagnosis: Models, Artificial Intelligence, Applications, Springer, 2004.

[9]   Liu J.,Wang W.: OP-Clusters: Clustering by Tendency in High Dimensional Space, Proceedings of the 3rd IEEE Int. Conf. on Data Mining, 187-194, 2003.

[10] Michalak M., Lachor M., Polański A.: HiBi - The Algorithm of Biclustering the Discrete Data. Lecture Notes in Computer Science 8468:760–771, 2014.

[11] Michalak M., Sikora B., Sobczyk J.: Diagnostic Model for Longwall Conveyor Engines, Advances in Intelligence and Soft Computing , 391:437-447.

[12] Michalak M., Sikora M., Sobczyk J.: Analysis of the longwall conveyor chain based on a harmonic analysis. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2013; 15 (4): 332–336.

[13] Michalak M., Stawarz M.: HRoBi - The Algorithm for Hierarchical Rough Biclustering, Lecture Notes in Computer Science, 7895:194–205, 2013.

[14] Murali T.M., Kasif S.: Extracting Conserved Gene Expression Motifs from Gene Expression Data, Pacific Symposium on Biocomputing, 77-88, 2003.

[15] Pensa R., Boulicaut J.F.: Constrained Co-clustering of Gene Expression Data, Proc. SIAM International Conference on Data Mining, SDM 2008, 25-36, 2008

[16] Prelić, A., Bleuler, S., Zimmermann, P.,Wille A., Buhlmann P., GruissemW., Hennig L., Thiele L., Zitzler E.: A Systematic Comparison and Evaluation of Biclustering Methods for Gene Expression Data, Bioinformatics 22(9):1122-1129, 2006.

[17] Sikora M.: Induction and pruning of classification rules for prediction of microseismic hazards in coal mines. Expert Systems with Applications 38(6):6748-6758, 2011.

[18] Sikora M, Michalak M.: Eksploracja baz danych systemów monitorowania na przykładzie obserwacji pracy kombajnu chodnikowego, Bazy Danych: Rozwój metod i technologii. Tom I, (in Polish), WK L, Warsaw 2008:429-437.

[19] Stawarz M., Michalak M.: eBi — The Algorithm for Exact Biclustering. Lecture Notes in Computer Science 7268:327-334, 2012.

[20] van Uitert M., Meuleman W., Wessels L.: Biclustering Sparse Binary Genomic Data, Journal of Computational Biology, 15(10):1329-1345, 2008.

[21] Yang E. Foteinou P.T,. King K.R., Yarmush M.L., Androulakis I.P.: A Novel Nonoverlapping bislutering Algorithm for Network Generation Using Living Cell Array Data, Bioinformatics 17(23) 2306-2313, 2007.

[22] Yang J., Wang H., Wang W., Yu P.: Enhanced biclustering on expression data, Third IEEE Symposium on Bioinformatics and Bioengineering, 321-327, 2003.

[23] Zimroz R.: Metody adaptacyjne w diagnostyce układów napędowych maszyn górniczych (in Polish). Wroclaw University of Technology Press, Wroclaw 2010.

AUTHOR

Marcin Michalak Marcin Michalak was born in Poland in 1981. He received his M.Sc. Eng. In computer science from the Silesian University of Technology in 2005 and Ph.D. degree in 2009 from the same university. His scientific interests are in machine learning, data mining, rough sets and biclustering. He is an author and coauthor of over 60 scientific papers.