

Approximation of a Coal Mass by an Ultrasonic Sensor Using Regression Rules

Marek Sikora^{1,2}, Marcin Michalak^{1,3}, and Beata Sikora⁴

¹ Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, Poland
{Marek.Sikora, Marcin.Michalak}@polsl.pl

² Institute of Innovative Technologies EMAG, ul. Leopolda 31,
40-189 Katowice, Poland

³ Central Mining Institute, Plac Gwarkow 1, 40-166 Katowice, Poland
Marcin.Michalak@gig.eu

⁴ Silesian University of Technology, ul. Kaszubska 23, 44-100 Gliwice, Poland
Beata.Sikora@polsl.pl

Abstract. A method of approximation the mass of coal moving on a conveyor belt under the ultrasonic sensor that measures a height of coal pile is described in the paper. A process of defining a set of variables that affects the approximated coal mass is presented. A model of multiple regression and an algorithm of regression rules induction based on the M5 algorithm have been exploited to relate momentary values of the coal pile with the mass of moving coal.

Keywords: applications of data mining, regression rules, coal weight approximation.

1 Introduction

A problem of estimation the coal mass is the considerable problem in mining industry. Scales are usually installed at a few locations of the whole transportation system only. The transportation system is a web of joined conveyor belts. In majority of cases, belt scales are mostly installed at the end of transport routes due to high installation costs. However an information (at least approximate) about the mass of currently mined material is desired in many situations during production process. For example, such information is essential for the transportation system in which a sum of productivities of conveyors transporting output to the main conveyor belt is greater then a productivity of the main conveyor. Since a scale is usually installed at the end of the main conveyor, a situation in which the mining process is conducted so intensively that it causes overloading of the main conveyor and consequently its emergency stoppage may happen.

An attempt at substituting a sensor that measures the height of a coal pile shifting below for a scale is described in the paper. Since the correlation between the height of the coal pile and parameters of the conveyor and the shifting coal is unknown, it should be discovered experimentally. The model of multiple regression and the M5 regression rule induction algorithm [6] were applied for that purpose.

The present paper is organized as follows: a problem of defining independent variables is presented in the second section, a problem of regression rules induction is presented concisely in the third section, next two sections contains the description of the conducted experiment and summary of the whole paper.

2 Determination of a Set of Independent Variables

Data coming from a system that consists of a scales, three conveyors and ultrasonic sensor sampling the height of a shifting pile of material with the frequency 10 Hz were put to analysis. The scales frequency was equal to 1 Hz, thus the aggregation of data coming from the height sensor was made. An average value measured by the sensor was determined for each second.

The coal weight was measured at the beginning of a system of three coupled conveyors, whereas the sensor taking the height was placed about two kilometers farther. Therefore making data acquisition for next experiments, ink marking of the coal pile was resorted so as to gain the possibility of unequivocal comparing the weight and height taken by the sensor. Moreover, speed of the conveyor belt travel was known which even more facilitated weight reading at the required time. During the experiment the main conveyor was powered solely by the material from the conveyor on which the height sensor has been installed. After the brief look at the time series it occurs that the distance between the series is about 1000 seconds. Normalized graphs of the coal pile mass and height after the 1000 seconds delay of the weight are presented in Fig. 1. It is visible that the real delay is about 80 seconds shorter. The final delay was calculated as the analysis of the difference between series and was equal to 922 seconds. As it was mentioned before, the research intention was to relate the weight of coal located on a conveyor with both the coal pile height and parameters characterizing a type of the given conveyor. The section area of coal shifted on the conveyor belt (Fig. 2) can be characterized by the following variables [1]:

λ – the angle between conveyor’s surface and a side of the isosceles triangle (A_2 – Fig. 2); β – the slope of the side edge of the conveyor’s section to the horizon; l_s, l_b – widths of horizontal and slant belt of the conveyor; h – the height of coal section. Based on the parameters the following values can be determined:

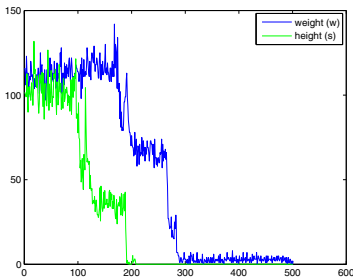


Fig. 1. A part of time series of the coal mass and the height of a coal pile.

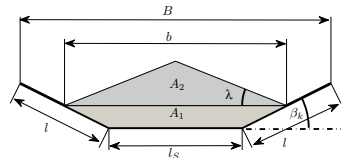


Fig. 2. A section of coal transported on the conveyor belt

h_{li} – the value of height for which the coal section is a triangle while λ is fixed; h_1 – the height of the triangle part of the section (A2 – Fig. 2); h_2 – the height of the quadrangle part (A1 – Fig. 2); $h_{max} = h_1 + h_2$ – the height of coal pile which exceeded means that coal falls down from the conveyor.

Finally, the value of section area depending on the coal pile height can be expressed by the formula (1). For calculations concerning the analyzed data set purposes we assumed that $\lambda = 15^\circ$ [5].

$$S = \begin{cases} h^2/\tan\lambda & 0 \leq h \leq h_{li} \\ h^2/\tan\lambda - 0.25(l_s(h - h_{li})/(h_{li}))^2 \frac{\tan \lambda \tan \beta}{\tan \lambda + \tan \beta} & h_{li} < h \leq h_{max} \\ h_{max}^2/\tan\lambda - 0.25(\frac{h_{max}-h_{li}}{h_{li}} l_s)^2 \frac{\tan \lambda \tan \beta}{\tan \lambda + \tan \beta} & h > h_{max} \end{cases} \quad (1)$$

Approximating the volume of a lump of coal which moved under the conveyor between two aggregated measurements of the height it has been assumed that the volume is the frustum’s volume (2). The frustum has volumes of bases S_1 and S_2 which are sections of the coal pile at succeeding (equal) times $t - 1, t$.

$$V(t) = [S(t) + S(t - 1) + \sqrt{S(t)S(t - 1)}]/3 \quad (2)$$

In the manner presented above it can be generated a data set in which each row (describing the time t_i) contains the following values: areas of the section and the volume of coal lump under the conveyor at times t_{i-1}, t_i ; the weight of coal pile which moved under the conveyor between times t_{i-1} and t_i .

3 Regression Rules Induction

An idea of the M5 algorithm was culled from so-called regression and classification trees (CART) [2] and from the C4.5 algorithm [7]. M5 analyzes the training set Tr and allows to create rules of the form (3).

$$\mathbf{IF} w_1 \wedge w_2 \wedge \dots \wedge w_k \mathbf{THEN} y = f(x) \quad (3)$$

where w_i is the so-called elementary condition which takes the form $a_i \in R_{a_i}$, $R_{a_i} \subset V_{a_i}$, (e.g. pressure small, average) for discrete-valued variables, and the form $a_i \in \langle v_1, v_2 \rangle$ (e.g. $gas_concentration \in \langle 0.4, 1.3 \rangle$ or $gas_concentration \geq 2$) for real-valued attributes. The function f is linear function of the form $s + s_{i1}a_{i1} + s_{i2}a_{i2} + \dots + s_{it}a_{it}$, where $s, s_{i1}, s_{i2}, \dots, s_{it}$ are real numbers (coefficients) and $\{a_{i1}, a_{i2}, \dots, a_{it}\} \subseteq A$. Independent variables contained in a rule conclusion should be real-valued variables.

The M5 algorithm builds a tree which is next transformed into a rule set (nodes that are not leaves create rule premises, and the function f which is rule conclusion is contained in a leaf). The tree is built on the *divide-and-conquer* principle basis. On each stage of the tree construction (in each node that is not a leaf) a procedure checking which attribute $a \in A$ and limitary value $q \in V_a$ will partition an example set P connected with the given node into two subsets $P_{<q}$ and $P_{>q}$ so as to minimize the expected variance of dependent variable, i.e. to maximize the value of expression (4), is invoked.

$$\Delta V = V(P)[(|P_{<q}|/|P|)V(P_{<q}) + (|P_{>q}|/|P|)V(P_{>q})] \quad (4)$$

Additionally, the algorithm applies pruning and smoothing methods that usually lead to decrease of a prognosis error for testing data. Details concerning the algorithm operation can be found, inter alia, in [6] and [8].

4 The Experiment

While conducting the experiment, 252 tons of coal has been transported under the sensor. Results obtained by the multiple linear regression method, the interval multiple linear regression method and results get by the M5 algorithm are contained in Table 1.

A number of independent variables that are delivered to a learning algorithm is of great importance in regression tasks realized with the aid of machine learning methods. In the case described here only three independent variables were available: areas of a coal lump at times $t - 1, t$ (denoted by $s(1), s(2)$) and the lump's volume (denoted by $v(1)$). Applying merely the variables (together or separately) didn't lead to obtain good results neither by the regression methods (rows 4, 5 in Table 1) nor by the M5 algorithm. Therefore values of section's areas and volumes registered earlier were decided to be applied as independent variables, too. For example, the $s(4)$ designation denotes that the variable contains information about the section's area calculated at time $t - 4$. However our implementation of the M5 algorithm didn't bear the enlarged set of independent variables. The algorithm treats succeeding delays (e.g. $s(3), s(4), \dots, s(max)$) of input variables as next independent variables automatically. The only parameter is the maximal delay which the algorithm can reach. The change can be found the useful modification of the M5 algorithm, because it significantly shortens time of data preparing and experiments conducting.

The smallest error, thus the difference between the approximated and real coal mass has been noticed for the M5 algorithm. Obtained rules are presented below. An experiment consisting in calculation the average value from average error and the total average deviation in the coal mass approximation in 10-fold cross validation mode has been also conducted for M5 and the interval regression.

Regression rules obtained by the M5 algorithm:

If $v(t) \leq 90.36$, **then** $w(t) = 2.1 - 19.161v(3) + 11.918v(4) + 10.299s(3) - 8.682v(5) + 6.238s(4) - 5.57v(t) + 3.662s(t) + 3.322s(6) - 2.851s(5) + 1.32 v(6)$

If $v(t) > 90.36$ **and** $v(3) \leq 284.9$, **then** $w(t) = 48.6 - 0.214s(3) + 0.152v(3) + 0.151v(t)$

If $v(t) \leq 425.09$ **and** $v(3) > 284.9$, **then** $w(t) = 87.7 - 1.665v(3) + 0.935s(4) + 0.853s(3) - 0.316v(t) + 0.191s(t)$

If $v(t) > 425.09$, **then** $w(t) = 113.8 + 0.152s(6) + 0.102v(3) - 0.1v(6) - 0.086v(4) - 0.061s(3) - 0.007v(5)$.

Results of the interval multiple regression model for the set of variables enlarged to 5th delay of input variables are presented in last row of the Table 1. One can noticed that the model is subtly worst than the model get by M5, whereas M5 matched a number of delays by itself. It can be observed while analyzing the form of get rules that the M5 algorithm built two regression models (two rules) depends on the volume of the shifting lump of coal. In a premise of the

Table 1. Values of regression error and differences between approximated and real values of coal mass

Method	Average error (%)	$R^2 \cdot 100\%$	Average difference/ Total average (kg)
Rule model – training set	0.11%	99.1	5.0/193
Rule model – cross validation	0.11%	99.0	5.2/194
Regression – volume, weight	7.8%	93.0	8.1/4869
Regression – section areas, weight	7.66%	91.0	8.0/6587
Interval regression – all variables beginning from 5th delay	0.12%	98.8	5.6/196
Interval regression – all variables from 5th delay – crossvalidation	0.13%	98.4	6.1/410

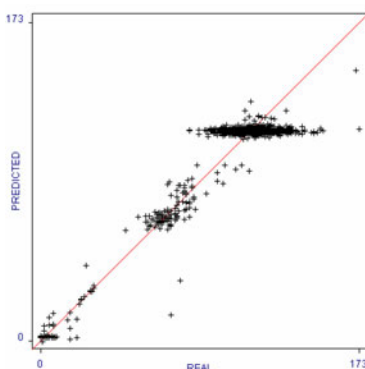


Fig. 3. The graph of estimated and real values for rule based model obtained in the cross validation mode

rules one variable $v(1)$ occurs. A complementation of the principal rules are two rules with additional condition $v(3) > (\leq) 284.9$ in the premise. One may presume that the two rules describe situations in which big, not shattered coal lumps were transported on a conveyor belt between succeeding measurements of transported coal section’s area, which perturbed the model realized by the principal rules. A graph of differences between estimated and real values of the coal mass is presented in Fig. 3.

5 Summary and Conclusions

The innovative method of measuring a mass of material (coal mass has been considered here) moving on a conveyor belt by an ultrasonic sensor has been presented in the paper. The correlation between height of a material pile and its mass has been determined by the M5 regression rules induction algorithm extended by automatic matching of variables reflecting succeeding delays of the time series. Efficiency of the method is high and it exceeds the obtained regression model. The innovative method of measuring a mass of material (coal mass

has been considered here) moving on a conveyor belt by an ultrasonic sensor has been presented in the paper. The correlation between height of a material pile and its mass has been determined by the M5 regression rules induction algorithm extended by automatic matching of variables reflecting succeeding delays of the time series. Efficiency of the method is high and it exceeds the obtained regression model. The data analysis method itself seems to be sufficient for the considered task. Even higher accuracy can be surely get by fuzzy [9] or neuro-fuzzy methods [4]. For example, fuzzy reasoning was applied for approximate and on-line measurement of ash content in coal [3]. However, in conducted works, simplicity and potential of implementation in an electronic device converting measurement signals into mass values directly in mine tunnels should be an important feature of the obtained model. Calibration of such device would consist in transmitting from the mine head to the device values of the model parameters only. In the case of fuzzy or neuro-fuzzy systems, realization of such system is much more difficult in the case of fuzzy-neural networks than for rules of the form (3). Authors plan more measurement experiments, especially for other kinds of coal (a new dependent variable including the λ parameter values that change according to the mined coal kind will occur among measurement variables). Ultimately, the sensor and the method of relating the height of a coal pile with its weight will subject to the procedure for granting patents (for that reason Authors haven't named the sensor type and its specification).

Acknowledgments

Research described in the paper was supported by the Somar Sp. Ltd., Katowice, Poland. The second Author was supported by the European Community from the European Social Fund.

References

1. Antoniuk, J.: Conveyor belts in underground and surface mining (in Polish). Silesian University of Technology Press, Gliwice, Poland (2007)
2. Breiman, L., Friedman, R.A., Olshen, R.S., Stone, C.: Classification and Regression Trees. Wadsworth, Belmont CA (1994)
3. Cierpisz, S., Heyduk, A.: A simulation study of coal blending control using a fuzzy logic ash monitor. *Control Engineering Practice* 10, 449–456 (2002)
4. Czogała, E., Łeski, J.: Fuzzy and Neuro-Fuzzy Systems. *Studies in Fuzziness and Soft Computing*, vol. 47. Physica-Verlag, Heidelberg NewYork (2000)
5. Gładysiewicz, L.: Conveyor belts. Theory and calculations. Wrocław (2003) (in Polish)
6. Quinlan, J.R.: Learning with continuous classes. In: Proc. of the International Conference on Artificial Intelligence (AI 1992), World Scientific, Singapore (1992)
7. Quinlan, J.R.: C4.5 Programs for Machine Learning. Morgan Kaufman Publishers, San Mateo, California (1992)
8. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco (2005)
9. Yager, R.R., Filev, D.P.: Essential of Fuzzy Modelling and Control. John Wiley & Sons, Inc., Chichester (2004)