Decision Rule Learning from Stream of Measurements—A Case Study in Methane Hazard Forecasting in Coal Mines

Michał Kozielski[1(✉)], Paweł Matyszok[2], Marek Sikora[2,3], and Łukasz Wróbel[2,3]

[1] Institute of Electronics, Silesian University of Technology, Gliwice, Poland
michal.kozielski@polsl.pl

[2] Institute of Informatics, Silesian University of Technology, Gliwice, Poland
{pawel.matyszok,marek.sikora,lukasz.wrobel}@polsl.pl

[3] Institute of Innovative Technologies EMAG, Katowice, Poland
http://adaa.polsl.pl

Abstract. The approach based on the Very Fast Decision Rules algorithm in application to prediction of alarm state resulting from methane hazard in coal mines is presented in this work. The approach introduces the modification of rule induction process due to application of the Correlation rule quality measure. An evaluation of the introduced method on a real life stream data collected from coal mine sensors is performed. The results show advantages of the introduced method considering both the classification quality and the rule-based knowledge representation.

Keywords: Data stream mining · Rule-based learning · Classification

1 Introduction

Nowadays, the data that have to be analysed are more seldom in a form of a batch data set but increasingly often they are in a form of a data stream. The stream of data is theoretically unrestricted what means that new examples can arrive constantly and it is not possible to store them for later analysis. Machine learning or data mining of such data requires incremental methods that are able to learn with the incoming examples and to apply the created model at any time. There is a number of approaches that meet the restrictions presented above [14].

Within the undertaken task of methane hazard forecasting in coal mines it is important to create a prediction system that uses a comprehensible data model, therefore a rule based approach was taken into consideration. There are several rule based methods dedicated to stream data analysis, starting from the STAGGER approach [16], through the FLORA methods [20], AQ11-PM algorithm [13] and the FACIL method [5].

In recent years a new branch of methods [1,8,10,11] was initiated by application of Hoefding bound to incremental rule induction. This approach was introduced by the Very Fast Decision Rules (VFDR) algorithm [6].

A. Gruca et al. (eds.), *Man-Machine Interactions 5*,
Advances in Intelligent Systems and Computing 659, DOI 10.1007/978-3-319-67792-7_30

Due to the objective of the presented work the application of sequential covering rule induction algorithms seems to be the most sensible solution. The quality of the rule set obtained by the sequential covering algorithm depends on the quality measure [2,4,9,18,19] used in the rule growing and pruning phases. The applied quality measure is one of the factors affecting the prediction accuracy, the number of rules induced and their other characteristics (e.g. the statistical significance). Therefore, possibility of the method quality improvement is a motivation for the presented research.

The presented work aims in introducing the approach based on the VFDR algorithm in application to prediction of alarm state resulting from methane hazard in coal mines. The contribution of this work consists of the QVFDR algorithm resulting from a modification of rule induction process by application of the Correlation rule quality measure. The proposed method is properly evaluated for the considered application taking into account both classification quality and the rule interpretability.

The structure of the paper is as follows. Section 2 presents the VFDR algorithm and the introduced modification of the rule induction method. Section 3 presents a data set and the results of the analysis enabling evaluation of the new approach. The work is summarised in Sect. 4.

2 Incremental Rule Learning

Within this section at first the VFDR method is presented and next the proposed approach named Quality-driven Very Fast Decision Rules (QVFDR) is introduced.

2.1 Very Fast Decision Rules

The VFDR method [6] is a single pass algorithm that learns ordered and/or unordered rules. The algorithm is initiated with an empty rule set (RS) and a default rule $\rightarrow \mathcal{L}$, where $\mathcal{L}$ is initialized to NULL. $\mathcal{L}$ is a data structure containing information used to incoming instances classification, and the sufficient statistics needed to expand the rule. Each rule r, that was learnt, is a conjunction of literals, that are conditions based on attribute values, and an $\mathcal{L}_r$. If all the literals are true for a given example, then the example is said to be covered by the rule. The labelled examples covered by a rule r are used to update $\mathcal{L}_r$. A rule is expanded with the literal that has the highest gain measure (e.g. entropy) among the examples covered by the rule. $\mathcal{L}_r$ accumulates the sufficient statistics to compute the gain measure of all possible literals.

The number of observations, after which a rule can be expanded or a new rule can be induced, is determined by the Hoeffding bound. In order to make the computations more efficient, verification, whether the number of observations is sufficient, is performed after every N_{min} examples instead of after each new example.

The general form of the VFDR method is presented in Algorithm 1, whereas the method of rule expanding is presented in Algorithm 2.

Algorithm 1. VFDR: Rule Learning Algorithm

```
input S: Stream of examples
        N_min: Minimum number of examples
        ordered_set: boolean flag
output RS: Set of Decision Rules
Let RS ← {}
Let default rule L ← ∅
for each example (x, y_k) ∈ S do
   for each rule r ∈ RS do
      if r covers the example then
         Update sufficient statistics of rule r
         if Number of examples in L_r > N_min then
            r ← ExpandRule(r)
         end if
         if ordered_set then
            BREAK
         end if
      end if
   end for
   if none of the rules in RS trigger then
      Update sufficient statistics of the empty rule
      if Number of examples in L > N_min then
         RS ← RS ∪ ExpandRule(default rule)
      end if
   end if
end for
```

2.2 Quality-Driven Very Fast Decision Rules

The method introduced in this work and referred further as QVFDR modifies the VFDR approach presented in the previous section.

In order to present the introduced changes we start with a definition of four numbers: p, n, P and N, which are required during the rule induction. For a classification rule r a value of P is calculated as:

$$P = |Pos(r)|, \tag{1}$$

where $Pos(r)$ is a set of all training examples whose decisions are equal to the rule decision part. A value of N is calculated similarly:

$$N = |Neg(r)|, \tag{2}$$

where $Neg(r)$ is a set of all remaining training examples (that do not belong to $Pos(r)$). Then:

$$p = |Pos(r) \cap [r]|, \tag{3}$$

$$n = |Neg(r) \cap [r]|. \tag{4}$$

Algorithm 2. ExpandRule: Expanding one rule

1: **input** r: One rule
2:         H: Split evaluation function
3:         δ: one minus the desired probability of choosing the correct attribute
4: **output** r: Expanded rule
5: Let h_0 the entropy of the class distribution at $\mathcal{L}_r$
6: Compute $\epsilon = \sqrt{\frac{R^2 ln(1/\delta)}{2n}}$ (Hoeffding bound)
7: **if** $h_0 > \epsilon$ **then**
8:    **for** *each attribute* X_i **do**
9:       Let h_{ij} be the $H()$ of the best split based on attribute X_i and value v_j
10:       **if** $h_{ij} < h_{best}$ **and** $n_{ij} > 0.1 * n$ **then**
11:          Let $h_{best} \leftarrow h_{ij}$
12:       **end if**
13:    **end for**
14:    **if** $h_0 - h_{best} > \epsilon$ **then**
15:       Extend r with a new condition based on the best attribute $X_a = v_j$
16:       Release sufficient statistics of $\mathcal{L}_r$
17:       $r \leftarrow r \cup \{X_a = v_j\}$
18:    **end if**
19: **end if**

The above numbers are required to calculate a quality measure used in the rule growing and pruning phases. Numerous measures were analysed in our previous research concerning classification rule induction [17,18]. In this work the Correlation measure presented in formula 5 was applied.

$$\phi = \frac{pN - Pn}{\sqrt{PN(p+n)(P-p+N-n)}} \tag{5}$$

The Correlation (ϕ) measure computes the correlation coefficient between the predicted and target labels. It was applied to classification rule induction algorithms as well as to subgroup discovery and evaluation of association rules [7,9,21]. Correlation measure values are normalised within $[-1, 1]$ range. It enables comparison of rule quality for different P and N values. It is a particularly valuable property in the context of the presented application, where P and N values change with the new examples.

The presented QVFDR approach modifies the VFDR method within the section presented as Algorithm 2, lines 8–13. The VFDR method uses entropy as a function that evaluates if a new condition based on a given attribute and its value should be introduced into a rule (line 9). Within the QVFDR method this evaluation is performed by means of the Correlation measure (Eq. 5). Additionally, opposite to VFDR the presented method evaluates by means of the Correlation measure not a new condition but a whole rule that is expanded. When the two best candidate conditions are identified in this way, entropy is calculated for them and they are verified against the Hoeffding bound condition (Algorithm 2, line 14).

When a rule is created its quality is evaluated and therefore, the values of p, n, P and N numbers are calculated. Next, these values are modified with the arrival of the new examples. If the rule is modified within the process of incremental learning, the values of p, n, P and N numbers are reset for a new rule form.

Classification by means of the presented method is performed as follows. If a test example is covered by a rule, then it votes for its decision with the weight corresponding to its quality. The weights are summed up and the example is assigned to the class represented by a higher weight. If no rule covers the example, then the default rule is applied.

3 Analysis and Results

This section presents a cases study that was performed in order to evaluate the proposed QVFDR approach.

The analysis was performed utilising the Massive Online Analysis (MOA) tool [3]. The VFDR implementation available in this tool was extended into the QVFDR method.

The case study was performed on a data set containing coal mine sensor measurements. The task was both to predict natural hazard related to methane concentration in a coal mine gallery and to discover a valuable knowledge about the analysed process from the induced rules.

3.1 Data Set

The data set, named *Methane*[1], was collected within the DISESOR project [12]. Originally, the data set consisted of the measurements that were collected each second by the following 9 sensors in a coal mine:

- one anemometer (represented by AN symbol),
- seven methanometers (represented by MM symbol),
- binary indication if the longwall shearer is running.

This data set was aggregated in order to receive a single entry for each 30 s of measurements. The new data set was received applying the following aggregation operators within the given 30 s range:

- minimum operation for anemometer measurements,
- maximum operation for methanometer measurements,
- dominant operation for longwall shearer operation indication.

The resulting data set consists of 100 577 observations. The task defined on this data set was to predict the exceeded level of methane concentration in 3 min horizon. A normal level of methane concentration (class 0) was set to be $\alpha < 1$, whereas an alarm (class 1) takes place when $\alpha \geq 1$. The data set is slightly imbalanced as class 0 contains 58 288 examples and class 1 contains 42 289 examples.

[1] The data set is available at http://adaa.polsl.pl/software.html.

3.2 Quantitative Results

Within the experiments the performance of both the QVFDR and the VFDR methods on *Methane* data set was compared. The results of the analysis are presented in Table 1. They consist of both classification quality and process description conciseness which is expressed by a number of generated rules.

Table 1. Quality of QVFDR and VFDR expressed by means of classification quality and a number of generated rules

	Accuracy [%]	Kappa [%]	Number of rules	% of examples classified by default rule
QVFDR	88.30	75.61	4	12%
VFDR	83.68	66.89	36	33%

The results presented in Table 1 show that QVFDR is characterised by better classification quality and it generated smaller number of rules at the same time. Additionally, as it is shown in the last column of Table 1, classification of QVFDR is rule driven to a greater extent comparing to VFDR, because classification performed by QVFDR is based on the knowledge gained from examples (and expressed in the form of rules) in case of significantly larger number of examples.

3.3 Generated Rules

In this section the rules generated by QVFDR are presented. The descriptions presented in Tables 2 and 3 show how these rules evolved with the new examples

Table 2. Rules generated for majority class 0 (normal operation)

	Rule form	Number of examples by this rule	Comment
Rule 1	$MM532 \leq 1.0$	2 800	First condition
	$MM532 \leq 1.0$ & $AN662 \leq 1.75$	3 035	Second condition
	$MM532 \leq 1.0$ & $AN662 \leq 1.1$	11 065	Second condition modification
Rule 3	$MM534 \leq 0.9$	12 263	First condition
	$MM534 \leq 0.9$ & $AN662 \leq 2.1$	42 870	Second condition
	$MM534 \leq 0.9$ & $AN662 \leq 2.0$	51 801	Second condition modification
Rule 4	$MM534 \leq 0.9$	55 652	First condition

taken into account and what knowledge they represent. The rules are presented in the order in which they were generated.

Table 3. Rules generated for minority class 1 (alarm)

	Rule form	Number of examples by this rule	Comment
Rule 2	$MM534 > 0.9$	5 898	First condition
	$MM534 > 0.9$ & $AN662 \leq 1.9$	6 107	Second condition

The rules presented in Tables 2 and 3 show, consistently with intuition, that the main impact on a decision have the values of the sensor that is a basis of the prediction (methanometer $MM534$). However, registering the high methane concentration level (0.9) by this sensor does not necessarily lead to alarm concentration. Rule 2 expanded with the second condition shows that high methane concentration level ($MM534 > 0.9$) has to be accompanied by poor ventilation ($AN662 \leq 1.9$) to lead to alarm methane concentration (class 1). At the same time, looking at Rule 3 we can learn, that if the methane concentration level is low ($MM534 \leq 0.9$), then even poor ventilation ($AN662 \leq 2.0$) will not lead to alarm methane concentration value. Rule 1 shows the same dependency, however, for the methanometer placed at the beginning of the wall, which is less dangerous and crucial for methane concentration monitoring.

Figures 1, 2, 3 and 4 show how the quality of the rules (expressed by the Kappa measure) evolved with the learning process. It is possible to identify in the presented figures when the rules were expanded and the values p, n, P and N were reset, what resulted in rapid decrease of the rule quality regained with the succeeding examples.

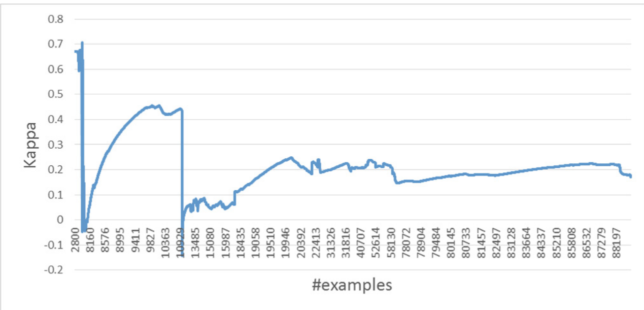

Fig. 1. Quality of Rule 1

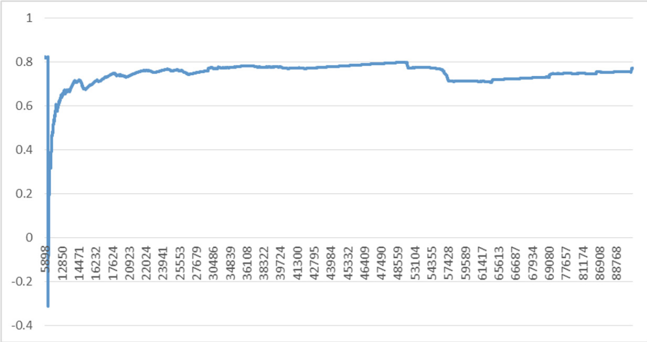

Fig. 2. Quality of Rule 2

Fig. 3. Quality of Rule 3

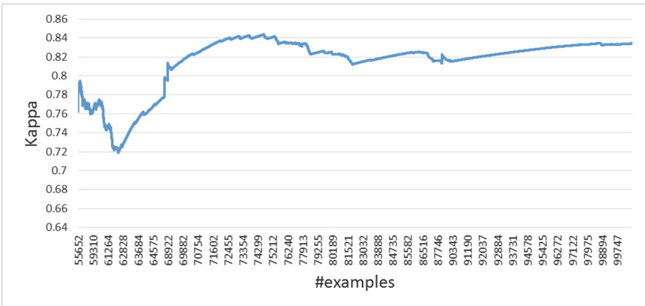

Fig. 4. Quality of Rule 4

4 Conclusions

The work presented introduced a new approach, named Quality-driven Very Fast Decision Rules (QVFDR), to rule-based incremental learning on data stream. The method is based on the VFDR method and modifies the rule generation process by controlling the rule quality by means of Correlation measure.

Within the performed evaluation of the introduced method it was analysed how many examples was classified by the generated rules, and how many examples was classified by a default rule utilising a distribution of the historical examples.

Besides, it was analysed how the rule quality changes taking two different cases into consideration. The first one is the change with the incoming examples and the other one is the case when a rule is expanded (a new condition is added or an existing condition is modified).

Finally, it was shown that the rules that were generated are meaningful and can be interpreted from the domain knowledge perspective. It justifies application of Correlation measure to a rule induction process within the incremental learning method.

The contribution of this work seems to be valuable as, to the best of the authors knowledge, there were no other approaches presenting the analysis of both the classification quality and the rule interpretability. The issue of default rule participation in classification was also not raised before.

Additionally, the presented experiments show that the introduced method performs well for the analysed case study and outperforms the VFDR method.

As future work, removal of the Hoeffding bound from the algorithm is planned as it is expected to be a reason of a large number of examples covered by a default rule instead of generated rules. Additionally, the Hoeffding's inequality was shown [15] to be not appropriate for application to any evaluation function (heuristic measure), e.g., information gain or Gini index. Therefore, it was suggested to be replaced by the McDiarmid's bound in these applications [15]. Besides, the extension of the research on other quality measures (e.g. Lift, C2) and analysis of benchmark data sets accompanied by thorough statistical evaluation is planned. Finally, authors are planing to implement the proposed algorithm as a part of the DISESOR system [12] and include its functionality within the methane concentration prediction task.

Acknowledgements. The work was carried out within the statutory research projects of the Institute of Electronics, Silesian University of Technology (BK_220 /RAu-3/2016 (02/030/BK_16/0017)) and the statutory research fund of the Institute of Innovative Technologies EMAG.

References

1. Almeida, E., Kosina, P., Gama, J.: Random rules from data streams. In: SAC 2013, Coimbra, Portugal, pp. 813–814 (2013)

2. An, A., Cercone, N.: Rule quality measures for rule induction systems: description and evaluation. Comput. Intell. **17**(3), 409–424 (2001)
3. Bifet, A., Holmes, G., Kirkby, R., Pfahringer, B.: MOA: massive online analysis. J. Mach. Learn. Res. **11**, 1601–1604 (2010)
4. Bruha, I., Tkadlec, J.: Rule quality for multiple-rule classifier: empirical expertise and theoretical methodology. Intell. Data Anal. **7**(2), 99–124 (2003)
5. Ferrer-Troyano, F.J., Aguilar-Ruiz, J.S., Santos, J.C.R.: Incremental rule learning and border examples selection from numerical data streams. J. Univ. Comput. Sci. **11**(8), 1426–1439 (2005)
6. Gama, J., Kosina, P.: Learning decision rules from data streams. In: IJCAI 2011, Barcelona, Spain, pp. 1255–1260 (2011)
7. Geng, L., Hamilton, H.J.: Interestingness measures for data mining: a survey. ACM Comput. Surv. **38**(3), 9 (2006)
8. Ikonomovska, E., Gama, J., Džeroski, S.: Learning model trees from evolving data streams. Data Min. Knowl. Disc. **23**(1), 128–168 (2011)
9. Janssen, F., Fürnkranz, J.: On the quest for optimal rule learning heuristics. Mach. Learn. **78**, 343–379 (2010)
10. Kosina, P., Gama, J.: Very fast decision rules for multi-class problems. In: SAC 2012, Trento, Italy, pp. 795–800 (2012)
11. Kosina, P., Gama, J.: Very fast decision rules for classification in data streams. Data Min. Knowl. Disc. **29**(1), 168–202 (2015)
12. Kozielski, M., Sikora, M., Wróbel, Ł.: Decision support and maintenance system for natural hazards, processes and equipment monitoring. Eksploatacja i Niezawodność-Maint. Reliab. **18**(2), 218–228 (2016)
13. Maloof, M.A., Michalski, R.S.: Incremental learning with partial instance memory. Artif. Intell. **154**(1), 95–126 (2004)
14. Nguyen, H.L., Woon, Y.K., Ng, W.K.: A survey on data stream clustering and classification. Knowl. Inf. Syst. **45**(3), 535–569 (2015)
15. Rutkowski, L., Pietruczuk, L., Duda, P., Jaworski, M.: Decision trees for mining data streams based on the McDiarmid's bound. IEEE Trans. Knowl. Data Eng. **25**(6), 1272–1279 (2013)
16. Schlimmer, J.C., Granger, R.H.: Incremental learning from noisy data. Mach. Learn. **1**(3), 317–354 (1986)
17. Sikora, M., Wróbel, Ł.: Data-driven adaptive selection of rules quality measures for improving the rules induction algorithm. In: Kuznetsov, S.O., Ślęzak, D., Hepting, D.H., Mirkin, B.G. (eds.) Rough Sets, Fuzzy Sets, Data Mining and Granular Computing: 13th International Conference, RSFDGrC 2011, Moscow, Russia, 25–27 June 2011. LNCS, vol. 6743, pp. 278–285. Springer, Heidelberg (2011)
18. Sikora, M., Wróbel, L.: Data-driven adaptive selection of rule quality measures for improving rule induction and filtration algorithms. Int. J. Gen Syst **42**(6), 594–613 (2013)
19. Slezak, D., Ziarko, W.: The investigation of the Bayesian rough set model. Int. J. Approx. Reason. **40**(1), 81–91 (2005)
20. Widmer, G., Kubat, M.: Learning in the presence of concept drift and hidden contexts. Mach. Learn. **23**(1), 69–101 (1996)
21. Xiong, H., Shekhar, S., Tan, P.N., Kumar, V.: Exploiting a support-based upper bound of pearson's correlation coefficient for efficiently identifying strongly correlated pairs. In: SIGKDD 2004, Seattle, USA, pp. 334–343 (2004)