

A Hybrid Multiple Classifier System for Recognizing Usual and Unusual Drilling Events

Bilal Esmael
University of Leoben
Leoben, Austria
Bilal@stud.unileoben.ac.at

Arghad Arnaout
TDE GmbH
Leoben, Austria
Arghad.Arnaout@tde.at

Rudolf K. Fruhwirth
TDE GmbH
Leoben, Austria
Rudolf.Fruhwirth@tde.at

Gerhard Thonhauser
University of Leoben
Leoben, Austria
Gerhard.Thonhauser@unileoben.ac.at

Abstract- Up to very recently, the applications of machine learning in the oil & gas industry were limited to using a single machine learning technique to solve problems in-hand. As the complexity of the demanded tasks being increased, the single techniques proved insufficient. This gave rise to intelligent systems that are hybrids of several machine learning techniques to solve the most challenging problems.

In this paper we propose a hybrid multiple classifier approach for recognizing usual and unusual drilling events. We suggest using two different information sources namely: (1) real time data collected by sensors located around the drilling rig, and (2) daily morning reports written by drilling personnel to describe the drilling process. Text mining techniques were used to analysis the daily morning reports and to extract textual features that include keywords and phrases, whereas data mining techniques were used to analysis the sensors data and extracting statistical features. Three base classifiers were trained and combined in one ensemble to obtain better predictive performance.

Experimental evaluation with real data and reports shows that the ensemble outperforms the base classifiers in every experiment, and the average classification accuracy is about 90% for usual events, and about 75% for unusual events.

Index Terms – Machine learning, Classification, Multiple Classifiers, Unusual Events Detection.

I. INTRODUCTION

During drilling oil wells all mechanical parameters, such as torque, hook load and block position, are continuously measured by rig sensors and stored in real time databases. In addition to these data, drilling personnel write daily morning reports to describe the drilling process from their point of view. Although sensors data represent the main source of information, daily morning reports are helpful enough to offset the lack of understanding of unusual events.

Nowadays, many different types of learning techniques are available; therefore it has become possible for the researchers to use the off-the-shelf technique to classify their databases. But on the other hand, researchers are facing difficulties in selecting the best technique that can apply to their tasks and datasets. The "No Free Lunch" Theorem states that there is no single learning algorithm that in any domain always induces the most accurate learner [2]. The common method is to test many techniques and select the one which achieves the best accuracy on new test dataset.

In recent years, hybrid intelligent systems have become more acceptable in the oil and gas industry [1]. These systems employ different machine learning techniques to solve real-world complex problems [5]. Murillo in [8] proposes a hybrid approach using fuzzy logic and neural network to predict pipe sticking. In addition, the authors of [3] created a hybrid system based on fuzzy logic to solve lost circulation problem.

To recognize complex patterns and make intelligent decisions, we propose using different techniques and creating multiple classifiers to learn from sensors data and daily morning reports. The objective is to utilize the strengths of one technique to complement the weaknesses of another [4, 6].

Multiple classifier systems, known as ensemble systems, are more promising than those of single classifier systems for a wide range of applications [2, 4, 6]. It is not easy, if not impossible, to find a single-classifier system that performs as well as a good ensemble of classifiers [9]. Multiple classifier systems weigh the particular opinions of classifiers, and combine them to reach a final decision that is likely the most informed one [4]. Many ensemble techniques were investigated in [2, 6 and 9].

Indeed, there are many reasons to use multiple classifier technique. Polikar in [4] mentioned four reasons: existing large volume of data, existing different types of features, solving complex problems and statistical reasons.

Training one classifier with large amount of data or with different types of features, like in our case, is usually not practical. In our case, the natures of features are different because we have complementary information from different sources, and a single classifier cannot be used to extract the knowledge contained in all of the data. In this case, data from each source will be used to train a different classifier.

Like all ensemble methods, our approach has a weakness which is increase computation because in order to classify an input, all base classifiers must be processed.

The remainder of the paper is organized as follows: Section II presents the general framework of our approach. Section III shows the details of features extraction. Section IV discuss the details of base classifiers training, and finally section V shows the experimental results of the proposed approach.

II. THE GENERAL FRAMEWORK

The most important peculiarity of our approach is its capability to exploit data from two different data sources, and to make an intelligent decision based on combining multiple classifiers.

The general framework of the approach is sketched in Fig. 1 and further described below.

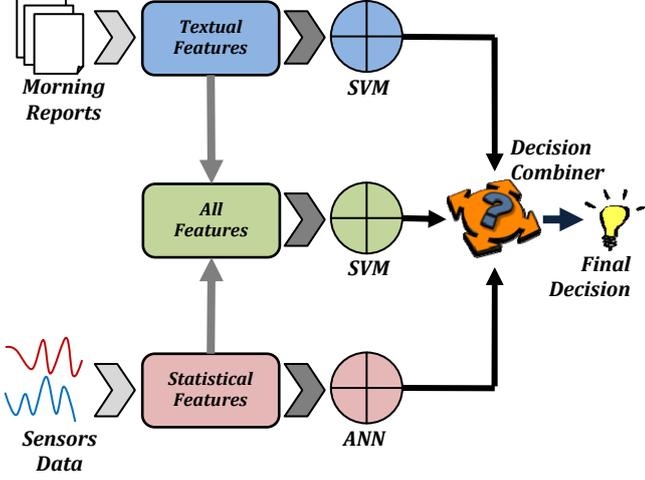


Figure 1. The general framework

The ensemble contains three base classifiers (learner) l_1 , l_2 and l_3 , each one of these classifiers deals with different types of features.

The first classifier l_1 , which is a support vector machine, will be trained only with textual features extracted from daily reports. The second classifier l_2 , which is a neural network, will be trained only with statistical features extracted from sensors data. The third classifier l_3 , which is a support vector machine, will be trained with the whole features set.

To train the ensemble, a training dataset D and a training report set R with correct labels for usual and unusual events are needed, then the feature spaces will be constructed and the base classifiers will be learned.

For each new instance x , three decisions will be made using the three base classifiers, whereas the output of each base classifier l_m is a class label $C_m \in \{C_1, \dots, C_k\}$ where k is number of classes, as well as confidence values or probability information:

$$P_{1,m} \dots P_{k,m} \text{ where } P_{i,m} \in [0,1] \forall i = 1, \dots, k.$$

Finally, the decision combiner will be used to predict the final class C of the instance x using the following equation:

$$C = \underset{c=1, \dots, k}{\operatorname{argmax}} \prod_{m=1}^3 P_{c,m} \quad (1)$$

The argument of the maximum function in Equation 1 means that the final class C is the class of the given argument for which the product function \prod attains its maximum value.

Fig. 2 shows the prediction algorithm used by the approach.

Prediction Algorithm:

Input:

- An instance consist of short text t and 10 channels of sensors data

Output:

- A class label $C_k, k = 1, \dots, K$

Do

- Convert the text t to a vector of textual features v_t
- Claculate the statistical features for each channel and create a vector of statistical features v_s
- Combine v_t and v_s in one comprehensible vector v_{ts}
- Use the three classifiers l_1, l_2 and l_3 to get three class labels $\{C_{1,k}, C_{2,k}, C_{3,k}\}, k = 1, \dots, K$ and probability information $P_{1,m} \dots P_{k,m}$ where k is number of classes
- Use Equation 1 to claculate the final class label C
- Return C

End

Figure 2. The prediction algorithm

III. FEATURE SPACES CONSTRUCTION

Specifying appropriate feature spaces is an important aspect for achieving good performance when designing classification systems. In our approach, we suggest using three feature spaces as shown in Fig. 1:

- Statistical feature space,
- Textual feature space,
- Complete feature space.

A. Statistical Feature Space

The first feature space contains statistical measures extracted only from sensors data. To construct this space, we used the standard sensors data (channels) described in the following table1.

For each channel in Table 1 overall 22 statistical measures were calculated. The calculated measures are: mean, median, mode, variance, standard deviation, interquartile range IRQ, range, skewness, kurtosis, second moment, p10, p25, p50, p75, p90, count, min, max, sum, first, last and entropy. Thereby the total number of statistical features is 220 which can be considered as a compacted representation of sensors data in a given time range.

TABLE 1. STANDARD DATA CHANNELS

Channels	Description
flowinav	Average mud flow-rate
hkldav	Average hook load
mdbit	Measured depth of the bit
mdhole	Measured depth of the hole
posblock	Block position
prespumpav	Average pump pressure
ropav	average rate of penetration
rpmav	Average drill string revolutions
tqav	Average torque
wobav	Average weight on bit

High dimensional space can contain high degree of irrelevant and redundant information which may greatly degrade the performance of classification algorithms. Therefore, features selection methods were used to reduce the dimensionality of the space. We proposed in [7] a method to reduce the dimensionality of statistical feature space.

B. Textual feature space

The second feature space contains textual features extracted only from daily morning reports. Actually, drilling reports mining represents a challenge because drilling personnel predominantly write short technical texts to describe the events in specific time ranges (e.g. “String stuck at 2670 m. attempted to jar down and upwards but no movement”). Therefore the “bag of word” representation used in text mining methods was extended to represent each description (short text) as features vector contains two groups of features namely: keyword features and phrase features. Extracting these types of features has been done by exploiting the internal semantics from the original text and external knowledge bases such as WordNet¹.

For keyword features extraction the traditional syntactic technologies were used. Full parsing to analyze the original text is done, then a stemmer to return each word to its basic form (root) is used. Finally the WordNet dictionary was used to select the type of each word.

Three categories of part of speech were created: verbs, nouns and adjectives, and all extracted keywords were stored under the appropriate category to be used later.

After analysing a huge amount of morning reports, and extracting all verbs, nouns and adjectives, we got 300 adjectives, 700 verbs and 1200 nouns.

Since the keyword features are too general to generate meaningful features, Link Grammar Parser² was used to analyze the text syntactically and extract the key phrases.

The most frequent keywords and phrases were selected to generate the textual feature space.

C. Complete feature space

The complete feature space contains two different types of features because it was generated by grouping the statistical features and the textual features together.

Although the complete feature space is a comprehensible space, experimental results shown that training only one classifier with this space will not produce accurate predictions.

IV. BASE CLASSIFIERS TRAINING

As we mentioned in section II, three base classifiers will be trained and combined to obtain better predictive performance.

¹ <http://wordnet.princeton.edu/>

² <http://www.link.cs.cmu.edu/link/>

Support vector machine (SVM), which is a state-of-the-art classification method, was used with textual features due to its high accuracy and ability to deal with high-dimensional data like we have in textual feature space [11].

Although classical SVM predicts only class label without probability information, but there are some research that extend SVM to give probability estimates, paper [12] introduce one of these research.

SVMs belong to the general category of kernel methods. The most important point should be taken into account when training SVM is selecting an appropriate kernel, and determining the best parameters.

Most people randomly try a few kernels and parameters, and in most cases they cannot build an accurate classifier. The authors of paper [10] propose the following procedure when using SVM tools:

- Transform data to the format of an SVM package
- Conduct simple scaling on the data
- Consider the RBF kernel $K(x,y)=\text{Exp}(-|x-y|^2)$
- Use cross-validation to find the best parameter C and γ .
- Apply the best C and γ to train the whole training set
- Test the classifier

We applied this procedure and used different values for C and γ . We got different accuracy each time. In addition a linear kernel instead of RBF kernel was tested and found that for text classification task, using linear kernel is better than using RBF kernel

For statistical features, four classification techniques were constructed and employed. These techniques are: artificial neural network (ANN), rule induction (RI), decision tree (DT) and naïve Bayes (NB). Each one of these classifiers contains number of parameters that can be tuned to improve the output of these classifiers. Many values and options of these parameters were tested to get the best results.

The performance of the classifiers was evaluated by using the cross-validation method. We found that the worst classifier –in most cases – is Naïve Bayes, and the best on is rule induction and neural network. We used neural network because it provides probabilities information that can be used in the ensemble to combine the results. Fig. 3 shows the structure of the used neural network.

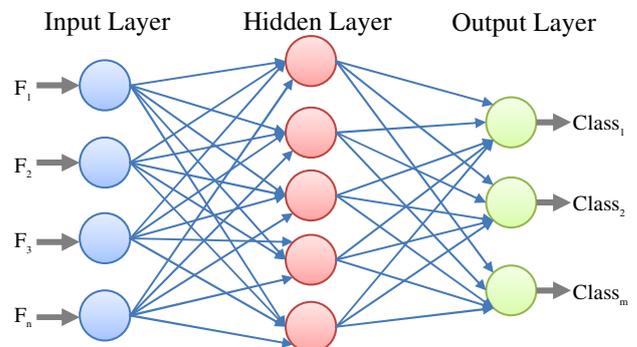


Figure 3. The structure of the neural network

We trained a feed-forward neural network by a backpropagation algorithm (multi-layer perceptron). The structure of the neural network has three layers: input, hidden and output. The input layer consists of input neurons which receive the input (statistical features). The output layer consists of output neurons which represented the classes (drilling operations).

V. EXPERIMENTAL RESULTS

To evaluate our approach, two types of experiments were executed. In the first type we used the approach to recognize only usual drilling events. We collected real time data and daily morning reports from two different drilling scenarios described in table 2. The following Fig. 4 shows the time versus depth (TxD) curve of the scenario1, and Fig. 5 shows the histogram of the classes included in Scenario#1.

TABLE 2. DRILLING SCENARIOS WITHOUT UNUSUAL EVENTS

Scenario	Instances	Duration [day]	Depth [m]	Classes
#1	991	95	7825	5
#2	770	87	4863	7

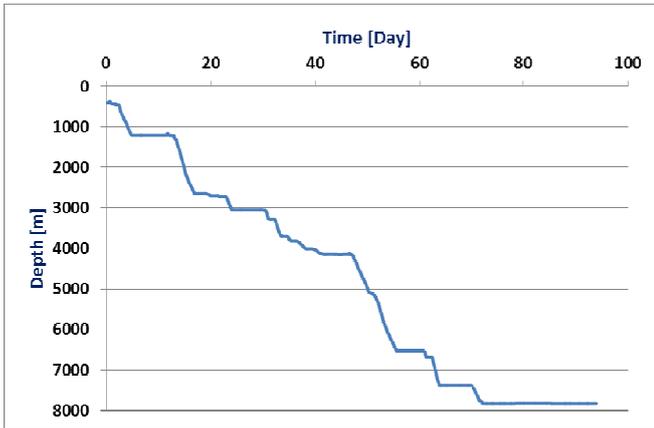


Figure 4. TxD curve of scenario#1

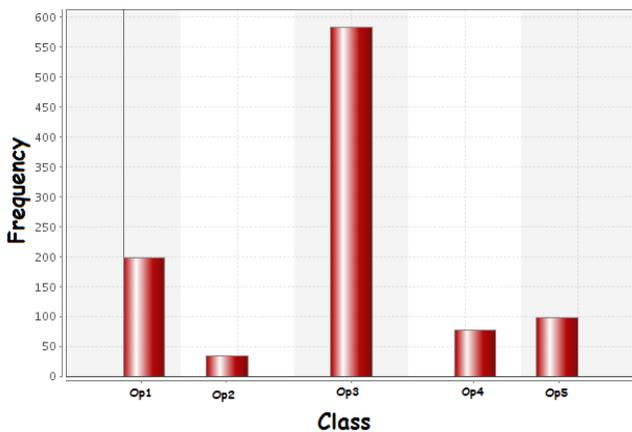


Figure 5. Histogram – operations of scenario#1

We divided the whole dataset into training dataset and test dataset. We used the training dataset to train the base classifiers, and then tested the accuracy using the test dataset. Table 3 shows the results.

TABLE 3. MULTIPLE CLASSIFIERS ACCURACY

Scenario	Accuracy [%]			
	L_1 (SVM)	L_2 (NN)	L_3 (SVM)	Ensemble
#1	81.12	82.90	72.95	88.26
#2	80.2	85.08	79.00	92.55

In the second type of experiments we used the ensemble to recognize usual and unusual drilling events. We collected two datasets that have usual and unusual events. Table 4 describes the collected datasets.

TABLE 4. DRILLING SCENARIOS WITH UNUSUAL EVENTS

dataset	Examples	Events		Classes
		Usual	Unusual	
#1	500	450	50	6
#2	1000	900	100	9

Each one of the collected dataset contains two types of unusual events (problems) which are: “stuck pipe” and “over pull”.

The following tables 5 and 6 show the results of classification.

TABLE 5. SINGLE CLASSIFIERS ACCURACY

N.	L_1 (SVM)-Textual Features			L_2 (NN)-Statistical Features		
	Usual Events	Unusual Events	Total	Usual Events	Unusual Events	Total
#1	80 %	55 %	77.5 %	82 %	62 %	80 %
#2	81 %	57 %	78.6 %	83 %	60 %	80.7 %

TABLE 6. MULTIPLE CLASSIFIERS ACCURACY

N.	L_3 (SVM)-Whole Features			Ensemble		
	Usual Events	Unusual Events	Total	Usual Events	Unusual Events	Total
#1	75 %	60 %	73.5 %	90 %	69 %	87.9 %
#2	77 %	40 %	73.3 %	88 %	70 %	86.2 %

The ensemble outperforms the base classifiers in every experiment.

VI. FURTHER WORK

Our aim for future work is to extend this approach by adding a new group of features. The new features will be extracted from sensors data using Symbolic Aggregate appRoXimation SAX representation. A new classifier based on Hidden Markov Models will be added to improve the results and add the ability to classify more drilling problems. In addition, we are planning to extend this approach to a learning system that can be used not only for usual and unusual drilling events detection, but also for extracting the steps that have been taken to solve unusual events.

ACKNOWLEDGMENT

We thank TDE Thonhauser Data Engineering GmbH for the support of this work and the permission to publish this paper.

REFERENCES

- [1] S. D. Mohaghegh, "E&P Gets Smart With Virtual Intelligence", American Oil & Gas Reporter Kansas, July 2001. pp 77-83.
- [2] E. Alpaydin, "Introduction to Machine Learning", The MIT Press, February 2010.
- [3] L. Sheremetov, I. Batyrshin, J. Martinez, H. Rodriguez and D. Filatov "Fuzzy Expert System for Solving Lost Circulation Problem", IEEE, Fifth International Conference on Hybrid Intelligent Systems (His05), 2006.
- [4] R. Polikar, "Ensemble Based Systems in Decision Making", IEEE Circuits and systems Magazine, Third Quarter 2006.
- [5] S. D. Mohaghegh, "Essential Components of an Integrated Data Mining Tool for the Oil & Gas Industry, With an Example Application in the DJ Basin", SPE 84441, 2002.
- [6] Lei Xu, Adam K, Ching Y. Suen, "Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition", IEEE Transactions On Systems, Vol, 22, No3, May/June 1992.
- [7] E. Esmael, A. Arnaout, R.K. Fruhwirth and G. Thonhauser, "Automated System for Drilling Operations Classification Using Statistical Features", IEEE, Eleventh International Conference on Hybrid Intelligent Systems (HIS11), MALAYSIA, 2011.
- [8] A. Murillo, J. Neuman and R. Samuel, "Pipe Sticking Prediction and Avoidance Using Adaptive Fuzzy Logic and Neural Network Modeling", SPE production and Operations Symposium held in Oklahoma, USA, SPE 120128, 2009.
- [9] S. B. Kotsiantis, "Bagged voting ensembles", In Artificial Intelligence: Methodology, Systems, Applications, 2004.
- [10] C.-C. Chang and C.-J. Lin. LIBSVM: "a library for support vector machines", ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011.
- [11] K. Karpouzis, "Emerging artificial intelligence applications in computer engineering", ISO Press, 2007.
- [12] T.-F. Wu, C.-J. Lin, and R. C. Weng. "Probability Estimates for Multi-class Classification by Pairwise Coupling". Journal of Machine Learning Research, 5(2004), 975-1005. A short version appears in NIPS 2003.