Data Analytics for Drilling Operational States Classifications

Galina V. Veres and Zoheir A. Sabeur *

University of Southampton IT Innovation Center - Electronics and Computer Science, Faculty of Physical Sciences and Engineering,
Gamma House, Enterprise Road, Southampton, SO16 7NS, UK

Abstract. This paper provides benchmarks for the identification of best performance classifiers for the detection of operational states in industrial drilling operations. Multiple scenarios for the detection of the operational states are tested on a rig with various drilling wells. Drilling data are extremely challenging due to their non-linear and stochastic natures, notwithstanding the embedded noise in them and unbalancing. Nevertheless, there is a possibility to deploy robust classifiers to overcome such challenges and achieve good automated detection of states. Three classifiers with best classification rates of drilling operational states were identified in this study.

1 Introduction

Offshore industrial engineering involves the management of highly complex operations in drilling rigs. Critical situations such as "*Kick*s", "*Fluid loss*" or "*Stuck pipe*" may occur during drilling operations. Such conditions are gradually reached following various stages of criticalities in time. Therefore, it is important that those stages of operations are detected and controlled during drilling processes. One way of achieving it is to automate the detection of drilling Operational States (OS). It involves the breaking of a drilling process into ten well-defined and exclusive drilling OS [1]: *1) Drilling Rotary (DrlRot); 2) Drilling sliding (DrlSld); 3) Clean Downwards (CleanDN); 4) Clean Upwards (CleanUP); 5) Wash Upwards (WashUP); 6) Wash Downwards (WashDN); 7) Move in hole (MoveDN); 8) Move out of hole (MoveUP); 9) Circulation on (CirclHL); and 10) Make Connection (MakeCN)*. The OS have been successfully detected on a drilling run using machine learning techniques with five additional principal states [1]. Further, Echo State Networks were adjusted to cope with unbalanced datasets in order to perform well in the classification of OS at a given well [2]. However, knowledge of labeled data for training was assumed to be available during the drilling process. Therefore, the challenge is to consider a real operational scenario which considers a drilling plan at multiple wells when labelled data becomes available after drilling, at least in one well on the rig.

---

* This work was partly supported by the European Commission FP7 IST Programme under contract TRIDEC IP FP7-258723.

A framework for the selection of the best performing OS classifiers is proposed in this study. The classifiers are trained by using a portion of available labeled data of OS, i.e. the training is done for a given well, while testing is performed on other wells for the detection of unseen OS.

2 Data analytics

Two types of drilling data are generated: Sensor *measurements* data (time series); and data created by drilling experts as *observations*, so-called OS labels. The analyses of both observations and measurements data are addressed in this section.

2.1 Measurement data analyses (Time series)

Measurement data are generated by sensing devices. The data exhibit *complex behaviour* (Fig. 1.) which led to further data analysis for selecting suitable classifiers. The complexity of the drilling time series (data behaviour) required two tests in the classification process: *Linearity* and *Normality* tests.

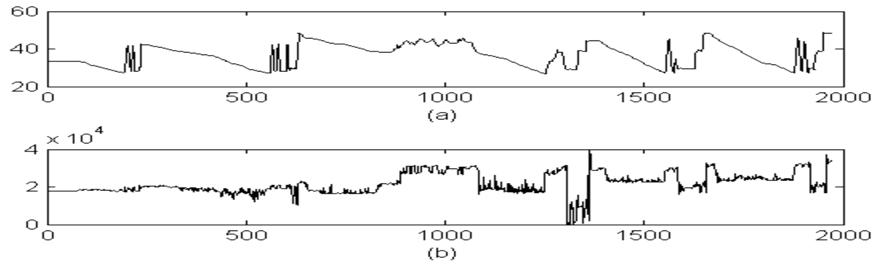

Fig. 1. The complexity of data dynamics: (a) Block Position and (b) Hook Load

Scatter plots for each time series data have been produced in order to check on data linearity. From the ten available measurement time series data only two showed linearity trends. These include Bit Depth and Hole Depth measurements. The test for data *Normality* was performed using Mardia's goodness-of-fit test for multivariate normality [3]. The results have shown that the drilling measurement time series data are non-Gaussian.

2.2 Observation data analysis (OS labels)

OS data are generated by drilling engineers as real-time observations, using expert knowledge assessments and *Morning Reports*. The latter are filled up when phases of drilling operations are complete and passed to the next operating drilling teams. The OS labels are consequently noisy and subjective. The statistical analysis of the 9 wells showed that 15% to 25% of OS labels were missing for each well. Also, the generated labeled OS occurred at different durations and frequencies, i.e. they are statistically imbalanced. Table 1 illustrates such issue (Well140).

Data complexity measures [4] were recently proposed to quantify the characteristics of data which affect accuracy of classification such as 1) *Overlaps of* *classes in feature space*; 2) *Separability of classes;* and 3) *Class density in overlap region*.

CircHL	CleanDN	CleanUP	DrlRot	DrlSld	MakeCN	MoveDN	MoveUP	WashDN	WashUP
9.2%	5.8%	1.3%	33.8%	21.3%	16.7%	3.5%	3.8%	2.4%	2.2%

Table 1: Distribution of OS label, Well140

Fisher's discriminant ratio (F1), the ratio of average intra/inter class nearest-neighbour distance (N2), and class density in overlapped regions (D3) represent a useful set of indicators for the good classification characteristics of the dataset. The generalization for L classes which also considers all feature dimensions was suggested in [5] for F1 and can be calculated as follows:

$$F1 = \frac{\sum_{i=1}^{L} n_i \cdot \delta(\boldsymbol{\mu}, \boldsymbol{\mu}_i)}{\sum_{i=1}^{L} \sum_{j=1}^{n_i} \delta\left(\mathbf{x}_j^i, \boldsymbol{\mu}_i\right)}, \tag{1}$$

where n_i is the number of samples in class i, δ is a similarity metric, $\boldsymbol{\mu}$ is the overall mean, $\boldsymbol{\mu}_i$ is the mean of class i, and $\mathbf{x}_j^i$ corresponds to the sample j of class i. When $F1 = 0$, a complete overlap exists between classes, while $F1 > L - 1$ means that there is no overlap. The intermediate values of $F1$ show the level of overlap between some classes. In this study, $F1 = 1.97$ for training data set of 10 classes. This shows that although there is no complete overlap between all classes, some classes may still overlap.

N2 measures class separability in the following way:

$$N2 = \frac{\sum_{i=1}^{N} \mathrm{intra}(\mathbf{x}_i)}{\sum_{i=1}^{N} \mathrm{inter}(\mathbf{x}_i)}, \tag{2}$$

N is the number of data samples, $\mathrm{intra}(\mathbf{x}_i)$ is the distance to the nearest neighbour within a class for a sample i; and $\mathrm{inter}(\mathbf{x}_i)$ is the distance to the nearest neighbour of any other class. Low values of N2 suggest that samples of the same class are well separated from other classes, whereas large values of N2 indicate that they are dispersed. Table 2 shows N2 values, calculated for each OS (class) for drilling data. The DrlRot class is the best separated class from the rest of classes. Hence one expects that this class can be easily classified. CleanDN and MakeCN classes exhibit good separability, while the MoveDN has the worst separability measure, followed by the DrlSld and MoveUP classes. These last three states mentioned may consequently present some confusion during the classification process.

The aim of the class density D3, as introduced in [5], is to determine the relative density of each class within an overlapping region. The lower the values of D3, the less number of samples lie within the overlapping region. Table 2 shows D3 values of samples in overlapping regions for all OS. The OS with the smallest D3s include DrlRot, MakeCN and CircHL. However, CleanUP, DrlSld and WashUp have shown D3s exceeding 60%. The rest of the OSs has shown significant high proportion of the overlapping regions. This analyses shows the type of challenges exists in this classification problem with overlapping classes and high class densities overall. As a result, the selected classification algorithms need to overcome the multi-class imbalance and complexity of the drilling data.

OS	N2	D3 (%)
CircHL	1.6	9.5
CleanDN	0.7	34.3
CleanUP	1.5	66.6
DrlRot	0.1	4.1
DrlSld	4.7	62
MakeCN	0.2	6.1
MoveDN	8.3	48
MoveUP	4.2	28.9
WashDN	1.5	49.0
WashUP	1.3	66.0

Table 2: Measures of data complexity N2 and D3

3 Selection of classifiers for the best performance

Following the above data analysis, reference to one of the most comprehensive review in [6] and the authors' experience with complex data classification, eight machine learning algorithms were selected. These include: *1) k-Nearest-Neighbour (kNN), 2) Support Vector Machines (SVM), 3) Linear Discriminant Analysis (LDA) 4) Echo State Network (ESN), 5) Random Forest (RF), 6) AdaBoostM2, 7) RusBoost and 8) Subspace*. Each of the algorithms were evaluated using micro-averaged and macro-averaged F-measures [2]; together with the Matthews Correlation Coefficient (MCC) [7] for their respective overall performances. Correct Classification Rates (CCR) were adopted for the assessment of individual OSs. The larger the F-measure is the higher the classification rate. Micro-average F-measure gives equal weights to each label and tends to be dominated by the classifier's performance on common classes. Macro-average F-measure gives equal weights to each class regardless of its frequency. It is influenced more by the classifier's performance on rare classes. Both measurement scores are used to analyze how well classifiers perform under common and rare classes. MCC summarizes the confusion matrix into a single value and is regarded as a good measure for problems with unbalanced classes. It returns a value between -1 and 1, where 1 is a perfect prediction, 0 no better than a random prediction and -1 indicates a total disagreement between prediction and observation. The selected eight classifiers were trained using sensor measurements with given OS at **Well140**. Their testing was subsequently performed on two other Wells of the same Rig (**Well80** and **Well85**). Only **Well80** is presented in this instance.

Ten sensor measurement were considered: *1) Block Position; 2) Bit Depth; 3) Hole Depth; 4) Weight on Bit; 5) Mud Flow; 6) Pump Pressure; 7) Rate of Penetration; 8) Rotary Torque; 9) Hook Load and;10) Rotary Speed*. Six additional features were also considered: *1) Hole Depth - Bit Depth; 2) Hole Depth + Block Position; 3) Bit Depth + Block Position; 4) Rotary Torque * Rotary Speed; 5) Pump Pressure * Mud Flow and; 6) Rate of Penetration * Weight on Bit*. Three experimental scenarios were designed according to various utilizations of the amount of labeled data from **Well140** for classifiers training: 1) 100% of data are considered for training (All); 2) 30% of data are considered for training using uniform sampling without replacement (30_UWR) and; 3) 30% of data are considered for training using a hybrid sampling (30_HS). These scenarios were considered to assess the possibility of reducing training sets without losing in the classification accuracy on the testing sets. The investigation on the sensitivity of the classifiers to various sub-sampled sets and the comparison of confidence intervals under scenarios 2) and 3) were performed using ten different Monte Carlo samples which were respectively drawn for each sampling scheme. Each algorithm was fine-tuned in order to achieve best performance. Table 3 shows the algorithms performance for **Well80**. Three algorithms such as RF, AdaBoostM2 and RUSBoost show best performance for the different training sets. They achieved similar performance according to all three assessment criteria: F_1 and F_2 measures reached values above 80% and 55% respectively; while MCC was above 0.7 for all these algorithms. The kNN and SubSpace algorithms consistently performed poorly Table 3 shows that the volume of the training datasets can be reduced by a third without significantly reducing the classifiers performances.

Method	ALL (%F_1,%F_2,MCC)	30_UWR (%F_1,%F_2,MCC)	30_HS (%F_1,%F_2,MCC)
kNN	(56,35.2,0.42)	(66.3±3.2,36.8±1.1, 0.53±0.05)	(50±2.6,33.6±0.7, 0.37±0.02)
SVM	(71.2,47.2,0.58)	(73.7±0.6,46.9±0.4, 0.62±0.008)	(70.6±2.4,45.8±1.0, 0.58±0.03)
LDA	(70.4,37.2,0.58)	(71.2±0.5,38.4±1.2, 0.59±0.006)	(69.9±0.2,37±0.4, 0.56±0.002)
ESN	(67.4,34.5,0.56)	(58.8±7,33±1.7, 0.48±0.06)	(66.2±2.3,36.3±2, 0.53±0.03)
RF	(84.1,57.1,0.77)	(85.2±0.3,57.4±0.5, 0.79±0.006)	(83.5±0.5,57.5±1.6, 0.74±0.01)
AdaBoostM2	(85.2,61.4,0.77)	(85.3±0.6,60.2±1.9, 0.77±0.01)	(85±0.4,60.5±0.6, 0.76±0.005)
RUSBoost	(85.3,61.6,0.76)	(84.2±0.5,59.5±1.1, 0.75±0.005)	(83.1±1.7,57.9±2.3, 0.73±0.03)
SubSpace	(65.1,22.3,0.61)	(64.6±1.4,21.1±2.7, 0.59±0.02)	(56.4±0.7,20.8±1.1, 0.44±0.01)

Table 3: Comparison of algorithms, Well 80

Uniform sampling Without Replacements (30_UWR) led to good overall performances for **Well80**. Though HS produced more balanced classes for classification, it did not preserve data structure. DrlRot, DrlSld and MakeCN operational states should not be misclassified, since they are critical for decision-making during normal operations. However, the accurate classification of WashUP/WashDN, CleanUP/CleanDN or MoveUP/MoveDN could become more important, when critical situations. As shown in Figure 2 below, three classifiers fulfill best results. These are **RF**, **AdaBoostM2** and **RUSBoost**. These nominated classifiers achieved high CCRs which are greater than 90% in the majority of cases of the critically important states under normal conditions such as DrlRot, DrlSld and MakeCN.

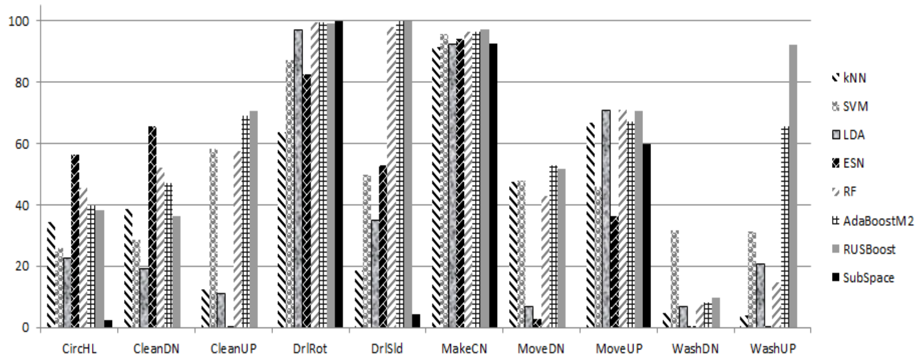

Fig. 2: Classification results for algorithms selection

4 Conclusions

A thorough benchmarking study has been achieved for the selection of the most performing classifiers for the detection of operational states in drilling operations. Strategies were put in place to filter out the less performing classifiers and maintain those which efficiently coped with complex drilling operation data and multiple states classification. Prior knowledge on the geophysical strata of the operating rig can potentially assist on further inferences for improving the identified performing classifiers. RF, AdaBoostM2 and RUSBoost were found highly reliable for achieving real-time automated detection of operational drilling states. They are proposed as the best classifiers for building the next generation decision-support information systems for achieving safer drilling operations in industrial rigs.

References

[1] B. Esmael, R. Fruhwirth, A. Arnaout, and G. Thonhauser. "Operations Recognition at Drill-Rigs," *EGU General Assembly 2012,* Vienna, Austria, April, 2012.

[2] G.V.Veres and Z. A. Sabeur. "Automated operational states detection for drilling system control in critical conditions". In Proceedings of 21st European Symposium on *Artificial* Neural *Networks, Computational Intelligence and Machine Learning (ESANN2013)*, Burges, Belgium, 24-26 April, 2013. i6doc.com publ., ISBN 978-2-87419-081-0.

[3] K. Mardia. "Measures of multivariate skewness and kurtosis with applications". *Biometrika*, vol. 57, pp. 519-530, 1970

[4] T.K. Ho, M. Basu, "Complexity measures of supervised Classification Problems". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 289-300, 2002

[5] J.S. Sanchez, R.A. Mollineda, J.M. Sotoca, "An analysis of how training data complexity affects the nearest neighbor classifiers". *Pattern Analysis and Applications*, vol. 10, no. 3, pp. 189–201, 2007

[6] X. Wu , V. Kumar , J. R. Quinlan , J. Ghosh , Q. Yang , H. Motoda , G. J. McLachlan , A. F. M. Ng , B. Liu , P. S. Yu , Z.-H. Zhou , M. Steinbach , D. J. Hand and D. Steinberg "Top 10 Algorithms in Data Mining", *Knowledge and Information Systems*, vol. 14, no. 1, pp.1 -37 2008

[7] J. Gorodkin. "Comparing two K-category assignments by a K-category correlation coefficient"**.** *Comp Biol and Chem*, vol. 28, pp. 367-374, 2004.