PROBABILISTIC MODELLING OF OIL RIG DRILLING OPERATIONS FOR BUSINESS DECISION SUPPORT: A REAL WORLD APPLICATION OF BAYESIAN NETWORKS AND COMPUTATIONAL INTELLIGENCE.

FRANÇOIS A. FOURNIER

A thesis submitted in partial fulfilment of the
requirements of the
Robert Gordon University
for the degree of Doctor of Philosophy.

This research programme was carried out in
collaboration with ODS-Petrodata Ltd. and IHS Inc.,
under the Knowledge Transfer Partnership (006922),
with funding from the Technology Strategy Board, UK.

March 2013

Abstract

This work investigates the use of evolved *Bayesian networks* learning algorithms based on *computational intelligence meta-heuristic* algorithms. These algorithms are applied to a new domain provided by the exclusive data, available to this project from an industry partnership with ODS-Petrodata, a business intelligence company in Aberdeen, Scotland. This research proposes statistical models that serve as a foundation for building a novel operational tool for forecasting the performance of rig drilling operations. A prototype for a tool able to forecast the future performance of a drilling operation is created using the obtained data, the statistical model and the experts' domain knowledge. This work made the following contributions: applied K2GA and *Bayesian networks* to a real-world industry problem; developed a well-performing and adaptive solution to forecast oil drilling rig performance; used the knowledge from industry experts to guide the creation of competitive models; created models able to forecast oil drilling rig performance consistently with nearly 80% forecast accuracy using either Logistic regression or *Bayesian network* learning using *genetic algorithms*; introduced the node juxtaposition analysis graph which allows the visualisation of the frequency of nodes links appearing in a set of orderings, providing new insights when analysing node ordering landscapes; explored the correlation factors between model score and model predictive accuracy and showed that the model score does not correlate with the predictive accuracy of the model; explored a method for feature selection using multiple algorithms and drastically reduced the modelling time by multiple factors; proposed new fixed structure *Bayesian network* learning algorithms for node ordering search-space exploration. Finally, this work proposed real-world applications for the models, based on current industry needs, such as *recommender systems*, an oil drilling rig selection tool, a user-ready rig performance forecasting software and rig scheduling tools.

Keywords

Bayesian Networks; Industry application; Oil and Gas; Metaheuristic; Business Intelligence; Computational Intelligence; Weka; Recommender Systems; Scheduling; Forecasting

Table of Contents

Abstract 1

Keywords 1

Table of Contents 2

List of Figures, Tables and Equations 5

Acknowledgements 8

Chapter 1: Introduction 10

1.1 Objective and Motivation 12

1.2 Ethical Considerations of Technological Developments 13

1.3 Publications and Presentations 14

1.4 Thesis Organisation 14

Chapter 2: Offshore Oil Drilling Rig: Commercial Background 16

2.1 Background - Offshore Drilling 16

2.2 Background - The Rig Tendering Process 18

2.3 The Problems at Hand 19

2.4 Gulf of Mexico Dataset 20

2.4.1 Review of Available Data 21

2.4.2 Expert Guided Selection 23

2.4.3 Data Linking 23

Chapter 3: Data Modelling: Technology Review 26

3.1 Bayesian Networks 26

3.1.1 Other Applications of Bayesian Networks 28

3.1.2 Limitations of Bayesian Networks 29

3.2 Search and Score, Learning Bayesian Networks Using K2 29

3.3 Evolutionary Computation for Bayesian Network Learning 31

3.4 Nature-inspired Metaheuristic Algorithms 31

3.4.1 Genetic Algorithms 31

3.4.2 Ant Colony Optimisation 33

3.5 Logistic Regression 34

3.6 Metrics of Quality 35

3.6.1 Building the Models 36

3.6.2 Models Evaluation 36

Chapter 4: Evolved Bayesian Network Models of Rig Drilling Operations in the Gulf of Mexico 40

4.1 Dataset - WRD1 40

4.2 K2 and Genetic Algorithms 41

4.3 K2 and Ant Colony Optimisation 43

4.4 Experimental Results 46

4.4.1 Structures Performances 46

4.4.2 Expert Evaluation of the Model 49

4.4.3 Node Juxtaposition Analysis 50

4.4.4 Algorithm Analysis 51

Chapter 5: Drilling Performance Models of Rig Operations in the Gulf of Mexico 54

5.1 Selection of Data for Model Building – WRD2 54

5.1.1 The AveragePerformanceFootagePerDay Field: dataset variations and category optimisation of the forecasted performance 57

5.1.2 Categorising the Data 59

5.2 Experimental Exploration 68

5.2.1 K2 Parameter Search 68

5.2.2 K2GA Runs 71

5.2.3 Node Juxtaposition 75

Chapter 6: Result Validation, Scoring Variations and Feature Selection 76

6.1 Lower Scored Ordering Cross-validation Forecast Ability 76

6.2 Alternative Modelling Techniques 78

6.3 Comparing Results with a Simulated ‘Manual’ Approach 79

6.3.1 NoModel Classification 80

6.3.2 Average and Majority Vote Forecast Validation 80

6.3.3 Model Validation Conclusion 81

6.4 Study of Fitness Landscape Covariance from Random Orderings Using Fixed Structures and CH-Score 81

6.5 Feature Selection 85

6.5.1 Weka Feature Selection 85

6.5.2 Pearson Correlation 88

6.5.3 Feature Selection Methods Combination 90

6.5.4 Accuracy and Model Learning Time 90

Chapter 7: Value Creation and Commercial Applications 94

7.1 Creating Value from Computational Intelligence 94

7.2 Rig Performance Forecasting: Demonstration Interface 96

7.3 Recommender System for Oil Drilling Rig Selection 100

7.3.1 Recommender Systems: Technology Review 100

7.3.2 Recommender Systems and Bayesian Networks 102

7.4 Rig Scheduling 103

7.4.1 Automatic Scheduling in the Industry 103

7.4.2 Scheduling Technologies 104

Chapter 8: Conclusions 106

8.1 Summary of Chapter Conclusions 106

8.2 Future Work 108

8.3 Conclusion and Summary of Contributions 110

References 114

Appendix 128

Review of Extracts from Kordon's Work in View of This Research 132

Glossary 134

List of Figures, Tables and Equations

Figure 1: An oil drilling rig (semi-submersible): J.W.Mclean, stacked (parked) in Invergordon, Scotland........ 11
Figure 2: Rig21sWE database extract with the four main tables used for data selection and the link between Well and Deployment........ 22
Figure 3: Data Linking Algorithm........ 24
Figure 4: An example of Bayesian Network [50]........ 28
Figure 5: An example of data points with a standard Logistic function fitted to them........ 35
Figure 6: K2GA [43]........ 42
Figure 7: ChainGA [24]........ 43
Figure 8: ChainACO [26]........ 44
Figure 9: ChainACO pseudo-code [26]........ 44
Figure 10: K2ACO [26]........ 45
Figure 11: K2ACO pseudo-code [26]........ 45
Figure 12: Network representations for K2GA, ChainGA, K2ACO and ChainACO [129]........ 47
Figure 13: Grayscale representation of node juxtapositions for Genetic Algorithms/Ant Colony Optimisation and K2/Chain algorithms on WRD1........ 50
Figure 14: *Expectation-Maximisation* clustering algorithm [136]........ 61
Figure 15: (a) Pre-discretisation data distribution graph as visualised in Weka........ 62
Figure 16: (a) Post-discretisation data distribution graph as visualised in Weka........ 64
Figure 17: Overview of WRD2.0 experiment 1 best network score results........ 73
Figure 18: Link details of WRD2.0 experiment 1 best network score results........ 73
Figure 19: Overview of WRD2.0 experiment 2 best network score results........ 74
Figure 20: Link details of WRD2.0 experiment 2 best network score results........ 74
Figure 21: WRD2.0, experiment 1 & 2 best ordering node juxtaposition graphs........ 75
Figure 22: CH-Score and model accuracy correlation analysis (matching instances sorted by descending CH-score)........ 77
Figure 23: Shape of best network based on % correctly classified instances from a 10-fold cross-validation........ 77
Figure 24: Link details of best network based on % correctly classified instances from a 10-fold cross-validation........ 78
Figure 25: Base fields used by experts to evaluate oil drilling rig performance........ 80
Figure 26: NoModel forecast results........ 80

Figure 27: Average and Majority vote forecast validation 81
Figure 28: An example of pyramid fixed structure 82
Figure 29: An example of vertical block fixed structure 83
Figure 30: An example of horizontal block fixed structure 83
Figure 31: Accuracy for the tests of reduced WRD2.0 dataset 92
Figure 32: Starting screen for the demonstration allowing the user to select one type of forecasting. 97
Figure 33: First step of the demonstration wizard allowing to select a rig and some geographical data 98
Figure 34: One selected rig when using the demo 98
Figure 35: Step two of the demonstration wizard, selecting well information 99
Figure 36: A display of the performance forecast of an oil drilling rig in the demonstration software 99

Table 1: Offshore drilling platform types in the Gulf of Mexico 16
Table 2: Main offshore oil drilling rigs type definition 17
Table 3: Well-Deployment ↔ Rig record matching example 25
Table 4: WRD1 selected fields and variable value count for preliminary experiments 41
Table 5: Means and standard deviations of best individuals K2 scores 48
Table 6: Paired t-test of best individuals K2 score across all runs 48
Table 7: Time statistics per run over all runs 48
Table 8: Data fields selected for WRD2 55
Table 9: Fields not selected and data not available but of potential interest for WRD datasets 56
Table 10: WRD2.x Bayesian and Logistic testing 58
Table 11: An example of WRD2.5 categories possible presentation 58
Table 12: Expert suggested categories for WRD2 selected data 60
Table 13: Discretisation and cluster values to manual category selection 66
Table 14: Clustering seeds and boundary values identified for data categorisation 67
Table 15: Weka K2 parameter search for WRD2.0 70
Table 16: WRD2.0 experimental run score results 71
Table 17: t-Test 2-sample assuming unequal variances for WRD2.0 71
Table 18: WRD2.0 experiments, measures of model quality 71
Table 19: Accuracy on predicting AveragePerformanceFootagePerDay from WRD2.0 with various Weka algorithms and 10-folds cross-validation 79

Table 20: Accuracy on predicting AveragePerformanceFootagePerDay from WRD2.5 with various Weka algorithms 79
Table 21: ASIA10000 random ordering score correlations (40320 random orderings) 84
Table 22: ASIA random ordering scores statistic description........ 84
Table 23: ALARM random ordering score correlations (100 000 random orderings) 84
Table 24: ALARM random ordering scores statistic description 84
Table 25: CAR random ordering score correlations (100 000 random orderings)........ 84
Table 26: CAR random ordering scores statistic description........ 84
Table 27: Weka Feature Selection, variable selected by each algorithm and overall selection counts 87
Table 28: Summary of feature selection methods and assembly of ranks into one value........ 89
Table 29: Test of reduced WRD2.0 dataset (using the feature selected) with the main Weka algorithms used previously 91
Figure 30: Pearson Correlations of all continuous variables for WRD2.0........ 128
Figure 31: Pearson Correlations of all continuous variables for WRD2.0, ordered by correlation with AveragePerformanceFootagePerDay........ 128
Figure 32: Pearson Correlations of all continuous variables for WRD2.5........ 129
Figure 33: Pearson Correlations of all continuous variables for WRD2.5, ordered by correlation with AveragePerformanceFootagePerDay........ 129
Figure 34: Pearson Correlations of all continuous variables for WRD2.0-preD 130
Figure 35: Pearson Correlations of all continuous variables for WRD2.0-preD, ordered by correlation with AveragePerformanceFootagePerDay........ 130
Table 36: Parameter search with WRD2.5........ 131

Equation 1: Application of the Bayesian theorem to Bayesian networks 27
Equation 2: An example of Ant Colony Optimisation state transition rule [107]........ 34
Equation 3: An example of Ant Colony Optimisation pheromone updating rule [107] 34
Equation 4: CH-Score [28]........ 36
Equation 5: Probabilities for Concordance Index 38
Equation 6: Concordance Index calculation........ 38
Equation 7: Pearson product moment correlation........ 88

Acknowledgements

First of all, I would like to thank bullet points for the clarity they provide to any text every time they are used. Then, I would like to thank the producers of my favourite whiskies for the inspiration they gave me during the long hours of writing:

- Bruichladdich Rocks 14 years old
- Clynelish 14 years old
- Cragganmore 12 years old
- Lagavulin 16 years old
- Bunnahubhain Darach ùr
- Glenfiddich 15 distillery edition
- Jura (Prophecy & Elixir)

Then I would like to thank Star Trek in all of its iterations (TOS, TAS, TMP, TWOK, TSFS, TVH, TFF, TUC, TNG, GEN, FC, INS, NEM, DS9, VOY, ENT, ST), from 1965 to 2394, for the inspiration it gave me when I was not writing and for the excuses to procrastinate it also provided.

Human-wise, I would mostly like to thank my partner and friends:

- Alexandra Weber, for being here, for all her enduring support and for the proofreading of every draft but even more for the encouragement to actually complete the writing in the dark hours of slow progress.
- My entire family for supporting me at a distance even though I have been away from home for 9 years and counting.
- Graeme Nesham, for his kindness, support and all the useful details about oil rig specifications. From crane to water depth, he knows it all!
- John Hartley, for his friendship and for teaching me years ago nearly all I know about oil exploration and extraction.
- François Delobel, Serge Kruppa and Eric Werkhoven for showing me what computing was all about and inspiring me to innovate continuously, early in my computing career.

The important people behind this work are also:

- John McCall, Andrei Petrovski and Peter Barclay, my supervisors, for the advice and for all the draft reading they endured.
- Robert Steven and John Hartley for the expertise they provided to the project.

Finally, I would like to thank the following organisations:

- ODS-Petrodata Ltd. and IHS Inc. for employing me before, after and somehow during the length of my research efforts.
- The Technology Strategy Board for the funding of the research project through the Knowledge Transfer Partnership program (partnership 006922).

There are lots of other people I would like to thank for where I am today. Thank you all, all along the road.