

Contents

Acknowledgement.....	VII
List of Figures	XIII
List of Tables	XVII
Acronyms	XXI
Symbols	XXIII
Abstract	XXIX
Kurzfassung.....	XXXI
1 Introduction.....	1
1.1 Aims	2
1.2 Related work.....	2
1.3 Own publications.....	4
1.4 Outline	5
2 Data foundation	7
2.1 Data sources	7
2.1.1 On-board data: load spectrum data	8
2.1.2 Off-board data: workshop data	13
2.2 Preprocessing of data	15
2.3 Real-world datasets	16
2.4 Conclusion	17
3 Classifying component failures of a vehicle fleet	19
3.1 Fundamentals of classification.....	19
3.1.1 Performance Measures	22

3.1.2 The “class imbalance problem”	23
3.2 Classification methods	25
3.2.1 Support vector machine (SVM)	25
3.2.2 Classification tree	33
3.2.3 Random forest (RF)	38
3.2.4 Oblique random forest (ORF)	41
3.3 Fundamentals of feature selection	47
3.3.1 Variable importance in tree-based classifiers	50
3.3.2 Recursive feature elimination (RFE)	53
3.4 A new RF based classification and feature selection framework	55
3.5 Case study: Classifying component failures of a hybrid car battery	57
3.5.1 Parameter optimization	58
3.5.2 Results	65
3.6 Conclusion	81
4 Visualizing different kinds of vehicle stress and usage	83
4.1 Distance and dissimilarity measures	84
4.1.1 Euclidean distance	84
4.1.2 Random forest dissimilarity	85
4.2 Dimensionality reduction methods	87
4.2.1 Principal Components Analysis	89
4.2.2 Sammon Mapping	90
4.2.3 Locally Linear Embedding	91
4.2.4 Isomap	93
4.2.5 t-Distributed Stochastic Neighbour Embedding	93
4.3 Case study: Dependence of vehicle usage on operating country	97
4.3.1 Preprocessing and parametrization	98
4.3.2 Results	99

<i>Contents</i>	XI
4.4 Case study: Visual distinction of component failures	106
4.4.1 Preprocessing and parametrization	107
4.4.2 Results	108
4.5 Conclusion	113
5 Identifying usage and stress patterns in a vehicle	
 fleet	115
5.1 Fundamentals of rule learning	116
5.1.1 Rule evaluation measures	117
5.2 Rule learning methods	118
5.2.1 RIPPER	119
5.2.2 C5.0rules	120
5.2.3 Random forest based rule learning methods	122
5.3 Case study: Identifying stress patterns for component failures	128
5.3.1 Synthetic data	128
5.3.2 Preprocessing and parametrization	131
5.3.3 Results	133
5.4 Conclusion	145
6 Conclusion	147
6.1 Main contributions	147
6.2 Limitations	148
6.3 Benefits	149
7 Outlook	153