

---

# Contents

---

<b>Acknowledgments</b> . . . . .	xi
<b>Introduction</b> . . . . .	xiii
<b>Chapter 1. Optimization and Big Data</b> . . . . .	1
1.1. Context of Big Data . . . . .	1
1.1.1. Examples of situations . . . . .	2
1.1.2. Definitions . . . . .	3
1.1.3. Big Data challenges . . . . .	5
1.1.4. Metaheuristics and Big Data . . . . .	8
1.2. Knowledge discovery in Big Data . . . . .	10
1.2.1. Data mining versus knowledge discovery . . . . .	10
1.2.2. Main data mining tasks . . . . .	12
1.2.3. Data mining tasks as optimization problems . . . . .	16
1.3. Performance analysis of data mining algorithms . . . . .	17
1.3.1. Context . . . . .	17
1.3.2. Evaluation among one or several dataset(s) . . . . .	18
1.3.3. Repositories and datasets . . . . .	20
1.4. Conclusion . . . . .	21
<b>Chapter 2. Metaheuristics – A Short Introduction</b> . . . . .	23
2.1. Introduction . . . . .	24
2.1.1. Combinatorial optimization problems . . . . .	24
2.1.2. Solving a combinatorial optimization problem . . . . .	25
2.1.3. Main types of optimization methods . . . . .	25

---

2.2. Common concepts of metaheuristics . . . . .	26
2.2.1. Representation/encoding . . . . .	27
2.2.2. Constraint satisfaction . . . . .	28
2.2.3. Optimization criterion/objective function . . . . .	28
2.2.4. Performance analysis . . . . .	29
2.3. Single solution-based/local search methods . . . . .	31
2.3.1. Neighborhood of a solution . . . . .	31
2.3.2. Hill climbing algorithm . . . . .	33
2.3.3. Tabu Search . . . . .	34
2.3.4. Simulated annealing and threshold acceptance approach . . . . .	35
2.3.5. Combining local search approaches . . . . .	36
2.4. Population-based metaheuristics . . . . .	38
2.4.1. Evolutionary computation . . . . .	38
2.4.2. Swarm intelligence . . . . .	41
2.5. Multi-objective metaheuristics . . . . .	43
2.5.1. Basic notions in multi-objective optimization . . . . .	44
2.5.2. Multi-objective optimization using metaheuristics . . . . .	47
2.5.3. Performance assessment in multi-objective optimization . . . . .	51
2.6. Conclusion . . . . .	52
<b>Chapter 3. Metaheuristics and Parallel Optimization . . . . .</b>	<b>53</b>
3.1. Parallelism . . . . .	53
3.1.1. Bit-level . . . . .	53
3.1.2. Instruction-level parallelism . . . . .	54
3.1.3. Task and data parallelism . . . . .	54
3.2. Parallel metaheuristics . . . . .	55
3.2.1. General concepts . . . . .	55
3.2.2. Parallel single solution-based metaheuristics . . . . .	55
3.2.3. Parallel population-based metaheuristics . . . . .	57
3.3. Infrastructure and technologies for parallel metaheuristics . . . . .	57
3.3.1. Distributed model . . . . .	57
3.3.2. Hardware model . . . . .	58
3.4. Quality measures . . . . .	60
3.4.1. Speedup . . . . .	60

---

3.4.2. Efficiency . . . . .	61
3.4.3. Serial fraction . . . . .	61
3.5. Conclusion . . . . .	61
<b>Chapter 4. Metaheuristics and Clustering . . . . .</b>	<b>63</b>
4.1. Task description . . . . .	63
4.1.1. Partitioning methods . . . . .	65
4.1.2. Hierarchical methods . . . . .	66
4.1.3. Grid-based methods . . . . .	67
4.1.4. Density-based methods . . . . .	67
4.2. Big Data and clustering . . . . .	68
4.3. Optimization model . . . . .	68
4.3.1. A combinatorial problem . . . . .	69
4.3.2. Quality measures . . . . .	69
4.3.3. Representation . . . . .	76
4.4. Overview of methods . . . . .	81
4.5. Validation . . . . .	82
4.5.1. Internal validation . . . . .	84
4.5.2. External validation . . . . .	84
4.6. Conclusion . . . . .	86
<b>Chapter 5. Metaheuristics and Association Rules . . . . .</b>	<b>87</b>
5.1. Task description and classical approaches . . . . .	88
5.1.1. Initial problem . . . . .	88
5.1.2. <i>A priori</i> algorithm . . . . .	89
5.2. Optimization model . . . . .	90
5.2.1. A combinatorial problem . . . . .	90
5.2.2. Quality measures . . . . .	90
5.2.3. A mono- or a multi-objective problem? . . . . .	91
5.3. Overview of metaheuristics for the association rules mining problem . . . . .	93
5.3.1. Generalities . . . . .	93
5.3.2. Metaheuristics for categorical association rules . . . . .	94
5.3.3. Evolutionary algorithms for quantitative association rules . . . . .	99
5.3.4. Metaheuristics for fuzzy association rules . . . . .	102
5.4. General table . . . . .	105
5.5. Conclusion . . . . .	107

---

<b>Chapter 6. Metaheuristics and (Supervised) Classification</b> . . . . .	109
6.1. Task description and standard approaches . . . . .	110
6.1.1. Problem description . . . . .	110
6.1.2. K-nearest neighbor . . . . .	110
6.1.3. Decision trees . . . . .	111
6.1.4. Naive Bayes . . . . .	112
6.1.5. Artificial neural networks . . . . .	113
6.1.6. Support vector machines . . . . .	114
6.2. Optimization model . . . . .	114
6.2.1. A combinatorial problem . . . . .	114
6.2.2. Quality measures . . . . .	114
6.2.3. Methodology of performance evaluation in supervised classification . . . . .	117
6.3. Metaheuristics to build standard classifiers . . . . .	118
6.3.1. Optimization of <i>K-NN</i> . . . . .	118
6.3.2. Decision tree . . . . .	119
6.3.3. Optimization of <i>ANN</i> . . . . .	122
6.3.4. Optimization of <i>SVM</i> . . . . .	124
6.4. Metaheuristics for classification rules . . . . .	126
6.4.1. Modeling . . . . .	126
6.4.2. Objective function(s) . . . . .	127
6.4.3. Operators . . . . .	129
6.4.4. Algorithms . . . . .	130
6.5. Conclusion . . . . .	132
<b>Chapter 7. On the Use of Metaheuristics for Feature Selection in Classification</b> . . . . .	135
7.1. Task description . . . . .	136
7.1.1. Filter models . . . . .	136
7.1.2. Wrapper models . . . . .	137
7.1.3. Embedded models . . . . .	137
7.2. Optimization model . . . . .	138
7.2.1. A combinatorial optimization problem . . . . .	138
7.2.2. Representation . . . . .	139
7.2.3. Operators . . . . .	140
7.2.4. Quality measures . . . . .	140
7.2.5. Validation . . . . .	143

---

7.3. Overview of methods . . . . .	143
7.4. Conclusion . . . . .	144
<b>Chapter 8. Frameworks . . . . .</b>	<b>147</b>
8.1. Frameworks for designing metaheuristics . . . . .	147
8.1.1. Easylocal++ . . . . .	148
8.1.2. HeuristicLab . . . . .	148
8.1.3. jMetal . . . . .	149
8.1.4. Mallba . . . . .	149
8.1.5. ParadisEO . . . . .	150
8.1.6. ECJ . . . . .	150
8.1.7. OpenBeagle . . . . .	151
8.1.8. JCLEC . . . . .	151
8.2. Framework for data mining . . . . .	151
8.2.1. Orange . . . . .	152
8.2.2. R and Rattle GUI . . . . .	153
8.3. Framework for data mining with metaheuristics . . . . .	153
8.3.1. RapidMiner . . . . .	154
8.3.2. WEKA . . . . .	154
8.3.3. KEEL . . . . .	155
8.3.4. MO-Mine . . . . .	157
8.4. Conclusion . . . . .	157
<b>Conclusion . . . . .</b>	<b>159</b>
<b>Bibliography . . . . .</b>	<b>161</b>
<b>Index . . . . .</b>	<b>187</b>