Studies in Computational Intelligence, Volume 190

**Editor-in-Chief**
Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
*E-mail:* kacprzyk@ibspan.waw.pl

---

Further volumes of this series can be found on our homepage: springer.com

Vol. 168. Andreas Tolk and Lakhmi C. Jain (Eds.)
*Complex Systems in Knowledge-based Environments: Theory, Models and Applications,* 2009
ISBN 978-3-540-88074-5

Vol. 169. Nadia Nedjah, Luiza de Macedo Mourelle and Janusz Kacprzyk (Eds.)
*Innovative Applications in Data Mining,* 2009
ISBN 978-3-540-88044-8

Vol. 170. Lakhmi C. Jain and Ngoc Thanh Nguyen (Eds.)
*Knowledge Processing and Decision Making in Agent-Based Systems,* 2009
ISBN 978-3-540-88048-6

Vol. 171. Chi-Keong Goh, Yew-Soon Ong and Kay Chen Tan (Eds.)
*Multi-Objective Memetic Algorithms,* 2009
ISBN 978-3-540-88050-9

Vol. 172. I-Hsien Ting and Hui-Ju Wu (Eds.)
*Web Mining Applications in E-Commerce and E-Services,* 2009
ISBN 978-3-540-88080-6

Vol. 173. Tobias Grosche
*Computational Intelligence in Integrated Airline Scheduling,* 2009
ISBN 978-3-540-89886-3

Vol. 174. Ajith Abraham, Rafael Falcón and Rafael Bello (Eds.)
*Rough Set Theory: A True Landmark in Data Analysis,* 2009
ISBN 978-3-540-89886-3

Vol. 175. Godfrey C. Onwubolu and Donald Davendra (Eds.)
*Differential Evolution: A Handbook for Global Permutation-Based Combinatorial Optimization,* 2009
ISBN 978-3-540-92150-9

Vol. 176. Beniamino Murgante, Giuseppe Borruso and Alessandra Lapucci (Eds.)
*Geocomputation and Urban Planning,* 2009
ISBN 978-3-540-89929-7

Vol. 177. Dikai Liu, Lingfeng Wang and Kay Chen Tan (Eds.)
*Design and Control of Intelligent Robotic Systems,* 2009
ISBN 978-3-540-89932-7

Vol. 178. Swagatam Das, Ajith Abraham and Amit Konar
*Metaheuristic Clustering,* 2009
ISBN 978-3-540-92172-1

Vol. 179. Mircea Gh. Negoita and Sorin Hintea
*Bio-Inspired Technologies for the Hardware of Adaptive Systems,* 2009
ISBN 978-3-540-76994-1

Vol. 180. Wojciech Mitkowski and Janusz Kacprzyk (Eds.)
*Modelling Dynamics in Processes and Systems,* 2009
ISBN 978-3-540-92202-5

Vol. 181. Georgios Miaoulis and Dimitri Plemenos (Eds.)
*Intelligent Scene Modelling Information Systems,* 2009
ISBN 978-3-540-92901-7

Vol. 182. Andrzej Bargiela and Witold Pedrycz (Eds.)
*Human-Centric Information Processing Through Granular Modelling,* 2009
ISBN 978-3-540-92915-4

Vol. 183. Marco A.C. Pacheco and Marley M.B.R. Vellasco (Eds.)
*Intelligent Systems in Oil Field Development under Uncertainty,* 2009
ISBN 978-3-540-92999-4

Vol. 184. Ljupco Kocarev, Zbigniew Galias and Shiguo Lian (Eds.)
*Intelligent Computing Based on Chaos,* 2009
ISBN 978-3-540-95971-7

Vol. 185. Anthony Brabazon and Michael O'Neill (Eds.)
*Natural Computing in Computational Finance,* 2009
ISBN 978-3-540-95973-1

Vol. 186. Chi-Keong Goh and Kay Chen Tan
*Evolutionary Multi-objective Optimization in Uncertain Environments,* 2009
ISBN 978-3-540-95975-5

Vol. 187. Mitsuo Gen, David Green, Osamu Katai, Bob McKay, Akira Namatame, Ruhul A. Sarker and Byoung-Tak Zhang (Eds.)
*Intelligent and Evolutionary Systems,* 2009
ISBN 978-3-540-95977-9

Vol. 188. Agustín Gutiérrez and Santiago Marco (Eds.)
*Biologically Inspired Signal Processing for Chemical Sensing,* 2009
ISBN 978-3-642-00175-8

Vol. 189. Sally McClean, Peter Millard, Elia El-Darzi and Chris Nugent (Eds.)
*Intelligent Patient Management,* 2009
ISBN 978-3-642-00178-9

Vol. 190. K.R. Venugopal, K.G. Srinivasa and L.M. Patnaik
*Soft Computing for Data Mining Applications,* 2009
ISBN 978-3-642-00192-5

K.R. Venugopal
K.G. Srinivasa
L.M. Patnaik

# Soft Computing for Data Mining Applications

Springer

Dr. K.R. Venugopal
Dean, Faculty of Engineering
University Visvesvaraya College of Engineering
Bangalore University
Bangalore 560001
Karnataka
India

Prof. L.M. Patnaik
Professor, Vice Chancellor
Defence Institute of Advanced Technology
Deemed University
Girinagar, Pune 411025
India

Dr. K.G. Srinivasa
Assistant Professor,
Department of Computer Science and Engineering
M.S. Ramaiah Institute of Technology
MSRIT Post,
Bangalore 560054
Karnataka
India

ISBN 978-3-642-00192-5 e-ISBN 978-3-642-00193-2

DOI 10.1007/978-3-642-00193-2

Studies in Computational Intelligence ISSN 1860949X

Library of Congress Control Number: 2008944107

© 2009 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

*Typeset* & *Cover Design:* Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed in acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

*Tejaswi*

# Foreword

The authors have consolidated their research work in this volume titled Soft Computing for Data Mining Applications. The monograph gives an insight into the research in the fields of Data Mining in combination with Soft Computing methodologies. In these days, the data continues to grow exponentially. Much of the data is implicitly or explicitly imprecise. Database discovery seeks to discover noteworthy, unrecognized associations between the data items in the existing database. The potential of discovery comes from the realization that alternate contexts may reveal additional valuable information. The rate at which the data is stored is growing at a phenomenal rate. As a result, traditional ad hoc mixtures of statistical techniques and data management tools are no longer adequate for analyzing this vast collection of data. Several domains where large volumes of data are stored in centralized or distributed databases includes applications like in electronic commerce, bioinformatics, computer security, Web intelligence, intelligent learning database systems, finance, marketing, healthcare, telecommunications, and other fields.

Efficient tools and algorithms for knowledge discovery in large data sets have been devised during the recent years. These methods exploit the capability of computers to search huge amounts of data in a fast and effective manner. However, the data to be analyzed is imprecise and afflicted with uncertainty. In the case of heterogeneous data sources such as text and video, the data might moreover be ambiguous and partly conflicting. Besides, patterns and relationships of interest are usually approximate. Thus, in order to make the information mining process more robust it requires tolerance toward imprecision, uncertainty and exceptions.

With the importance of soft computing applied in data mining applications in recent years, this monograph gives a valuable research directions in the field of specialization. As the authors are well known writers in the field of Computer Science and Engineering, the book presents state of the art technology in data mining. The book is very useful to researchers in the field of data mining.

Bangalore,
November 2008

N.R. Shetty
President, ISTE, India

# Preface

In today's digital age, there is huge amount of data generated everyday. Deriving meaningful information from this data is a huge problem for humans. Therefore, techniques such as data mining whose primary objective is to unearth hithero unknown relationship from data becomes important. The application of such techniques varies from business areas (Stock Market Prediction, Content Based Image Retrieval), Proteomics (Motif Discovery) to Internet (XML Data Mining, Web Personalization). The traditional computational techniques find it difficult to accomplish this task of Knowledge Discovery in Databases (KDD). Soft computing techniques like Genetic Algorithms, Artificial Neural Networks, Fuzzy Logic, Rough Sets and Support Vector Machines when used in combination is found to be more effective. Therefore, soft computing algorithms are used to accomplish data mining across different applications.

Chapter one presents introduction to the book. Chapter two gives details of self adaptive genetic algorithms. An iterative merge based genetic algorithms for data mining applications is given in chapter three. Dynamic association rule mining using genetic algorithms is described in chapter four. An evolutionary approach for XML data mining is presented in chapter five. Chapter six, gives a neural network based relevance feedback algorithm for content based image retrieval. An hybrid algorithm for predicting share values is addressed in chapter seven. The usage of rough sets and genetic algorithms for data mining based query processing is discussed in chapter eight. An effective web access sequencing algorithm using hashing techniques for better web reorganization is presented in chapter nine. An efficient data structure for personalizing the Google search results is mentioned in chapter ten. Classification based clustering algorithms using naive Bayesian probabilistic models are discussed in chapter eleven. The effective usage of simulated annealing and genetic algorithms for mining top-$k$ ranked webpages from Google is presented in chapter twelve. The concept of mining bioXML databases is introduced in chapter thirteen. Chapter fourteen and fifteen discusses algorithms for DNA compression. An efficient algorithm for motif discovery in protein

sequences is presented in chapter sixteen. Finally, matching techniques for genome sequences and genetic algorithms for motif discovery are given in chapter seventeen and eighteen respectively.

The authors appreciate the suggestions from the readers and users of this book. Kindly communicate the errors, if any, to the following email address: venugopalkr@gmail.com.

Bangalore,
November 2008

K.R. Venugopal
K.G. Srinivasa
L.M. Patnaik

# Acknowledgements

We wish to place on record our deep debt of gratitude to Shri M C Jayadeva, who has been a constant source of inspiration. His gentle encouragement have been the key for the growth and success in our career. We are indebted to Prof. K Venkatagiri Gowda for his inspiration, encouragement and guidance throughout our lives. We thank Prof. N R Shetty, President, ISTE and Former Vice Chancellor, Bangalore University, Bangalore for his foreword to this book. We owe debt of gratitude to Sri K Narahari, Sri V Nagaraj, Prof. S Lakshmana Reddy, Prof. K Mallikarjuna Chetty, Prof. H N Shivashankar, Prof. P Sreenivas Kumar, Prof. Kamala Krithivasan, Prof. C Sivarama Murthy, Prof. T Basavaraju, Prof. M Channa Reddy, Prof. N Srinivasan, Prof. M Venkatachalappa for encouraging us to bring out this book in the present form. We sincerely thank Sri K P Jayarama Reddy, T G Girikumar, P Palani, M G Muniyappa for their support in the preparation of this book.

We are grateful to Justice M Rama Jois, Sri N Krishnappa for their encouragement. We express our gratitude to Sri Y K Raghavendra Rao, Sri P R Ananda Rao, Justice T Venkataswamy, Prof. V Y Somayajulu, Sri Sreedhar Sagar, Sri N Nagabhusan, Sri Prabhakar Bhat, Prof. K V Acharya, Prof. Khajampadi Subramanya Bhat, Sri Dinesh Kamath, Sri D M Ravindra, Sri Jagadeesh Karanath, Sri N Thippeswamy, Sri Sudhir, Sri V Manjunath, Sri N Dinesh Hegde, Sri Nagendra Prasad, Sri Sripad, Sri K Thyagaraj, Smt. Savithri Venkatagiri Gowda, Smt. Karthyayini V and Smt. Rukmini T, our well wishers for inspiring us to write this book.

We thank Prof. K S Ramanatha, Prof. K Rajanikanth, V K Ananthashayana and T V Suresh Kumar for their support. We thank Smt. P Deepa Shenoy, Sri K B Raja, Sri K Suresh Babu, Smt. J Triveni, Smt. S H Manjula, Smt. D N Sujatha, Sri Prakash G L, Smt. Vibha Lakshmikantha, Sri K Girish, Smt. Anita Kanavalli, Smt. Alice Abraham, Smt. Shaila K, for their suggestions and support in bringing out this book.

We are indebted to Tejaswi Venugopal, T Shivaprakash, T Krishnaprasad and Lakshmi Priya K for their help. Special thanks to Nalini L and Hemalatha for their invaluable time and neat desktop composition of the book.

# About the Authors

**K.R. Venugopal** is Principal and Dean, Faculty of Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore. He obtained his Bachelor of Technology from University Visvesvaraya College of Engineering in 1979. He received his Masters degree in Computer Science and Automation from Indian Institute of Science Bangalore He was awarded Ph.D. in Economics from Bangalore University and Ph.D. in Computer Science from Indian Institute of Technology, Madras. He has a distinguished academic career and has degrees in Electronics, Economics, Law, Business Finance, Public Relations, Communications, Industrial Relations, Computer Science and Journalism. He has authored and edited twenty seven books on Computer Science and Economics, which include Petrodollar and the World Economy, Programming with Pascal, Programming with FORTRAN, Programming with C, Microprocessor Programming, Mastering C++ etc. He has been serving as the Professor and Chairman, Department of Computer Science and Engineering, UVCE. He has over two hundred research papers in refereed International Journals and Conferences to his credit. His research interests include computer networks, parallel and distributed systems and database systems.

**K.G. Srinivasa** obtained his a Ph.D. in Computer Science and Engineering from Bangalore University. Currently he is working as an Assistant Professor in the Department of Computer Science and Engineering, M S Ramaiah Institute of Technology, Bangalore. He received Bachelors and Masters degree in Computer Science and Engineering from the Bangalore University in the year 2000 and 2002 respectively. He is a member of IEEE, IETE, and ISTE. He has authored more than fifty research papers in refereed International Journals and Conferences. His research interests are Soft Computing, Data Mining and Bioinformatics.

**L.M. Patnaik** is Vice Chancellor of Defence Institute of Advanced Studies, Pune, India. He was the Professor since 1986 with the Department of

Computer Science and Automation, Indian Institute of Science, Bangalore. During the past 35 years of his service at the Institute. He has over 400 research publications in in refereed International Journals and Conference Proceedings. He is a Fellow of all the four leading Science and Engineering Academies in India; Fellow of the IEEE and the Academy of Science for the Developing World. He has received twenty national and international awards; notable among them is the IEEE Technical Achievement Award for his significant contributions to high performance computing and soft computing. His areas of research interest have been parallel and distributed computing, mobile computing, CAD for VLSI circuits, soft computing, and computational neuroscience.

# Contents


**1 Introduction** . . . . . 1
1.1 Data Mining . . . . . 4
1.1.1 Association Rule Mining (ARM) . . . . . 4
1.1.2 Incremental Mining . . . . . 5
1.1.3 Distributed Data Mining . . . . . 6
1.1.4 Sequential Mining . . . . . 6
1.1.5 Clustering . . . . . 6
1.1.6 Classification . . . . . 8
1.1.7 Characterization . . . . . 8
1.1.8 Discrimination . . . . . 9
1.1.9 Deviation Mining . . . . . 9
1.1.10 Evolution Mining . . . . . 9
1.1.11 Prediction . . . . . 10
1.1.12 Web Mining . . . . . 10
1.1.13 Text Mining . . . . . 11
1.1.14 Data Warehouses . . . . . 11
1.2 Soft Computing . . . . . 13
1.2.1 Importance of Soft Computing . . . . . 13
1.2.2 Genetic Algorithms . . . . . 13
1.2.3 Neural Networks . . . . . 14
1.2.4 Support Vector Machines . . . . . 14
1.2.5 Fuzzy Logic . . . . . 15
1.2.6 Rough Sets . . . . . 16
1.3 Data Mining Applications . . . . . 16
References . . . . . 17

**2 Self Adaptive Genetic Algorithms** . . . . . 19
2.1 Introduction . . . . . 19
2.2 Related Work . . . . . 20
2.3 Overview . . . . . 22
2.4 Algorithm . . . . . 23

2.4.1 Problem Definition . . . . . 23
2.4.2 Pseudocode . . . . . 23
2.5 Mathematical Analysis . . . . . 25
2.5.1 Convergence Analysis . . . . . 30
2.6 Experiments . . . . . 32
2.7 Performance Analysis . . . . . 40
2.8 A Heuristic Template Based Adaptive Genetic Algorithms . . . . . 42
2.8.1 Problem Definition . . . . . 42
2.9 Example . . . . . 42
2.10 Performance Analysis of HTAGA . . . . . 44
2.11 Summary . . . . . 48
References . . . . . 49

**3 Characteristic Amplification Based Genetic Algorithms** . . . . . 51
3.1 Introduction . . . . . 51
3.2 Formalizations . . . . . 52
3.3 Design Issues . . . . . 54
3.4 Algorithm . . . . . 55
3.5 Results and Performance Analysis . . . . . 58
3.6 Summary . . . . . 61
References . . . . . 61

**4 Dynamic Association Rule Mining Using Genetic Algorithms** . . . . . 63
4.1 Introduction . . . . . 63
4.1.1 Inter Transaction Association Rule Mining . . . . . 64
4.1.2 Genetic Algorithms . . . . . 65
4.2 Related Work . . . . . 66
4.3 Algorithms . . . . . 67
4.4 Example . . . . . 69
4.5 Performance Analysis . . . . . 74
4.5.1 Experiments on Real Data . . . . . 78
4.6 Summary . . . . . 79
References . . . . . 79

**5 Evolutionary Approach for XML Data Mining** . . . . . 81
5.1 Semantic Search over XML Corpus . . . . . 82
5.2 The Existing Problem . . . . . 83
5.2.1 Motivation . . . . . 84
5.3 XML Data Model and Query Semantics . . . . . 85
5.4 Genetic Learning of Tags . . . . . 86
5.5 Search Algorithm . . . . . 89
5.5.1 Identification Scheme . . . . . 89

5.5.2 Relationship Strength . . . . . 90
5.5.3 Semantic Interconnection . . . . . 91
5.6 Performance Studies . . . . . 93
5.7 Selective Dissemination of XML Documents . . . . . 99
5.8 Genetic Learning of User Interests . . . . . 101
5.9 User Model Construction . . . . . 102
5.9.1 SVM for User Model Construction . . . . . 103
5.10 Selective Dissemination . . . . . 103
5.11 Performance Analysis . . . . . 105
5.12 Categorization Using SVMs . . . . . 108
5.12.1 XML Topic Categorization . . . . . 108
5.12.2 Feature Set Construction . . . . . 109
5.13 SVM for Topic Categorization . . . . . 111
5.14 Experimental Studies . . . . . 113
5.15 Summary . . . . . 116
References . . . . . 117

**6 Soft Computing Based CBIR System** . . . . . 119
6.1 Introduction . . . . . 119
6.2 Related Work . . . . . 120
6.3 Model . . . . . 121
6.3.1 Pre-processing . . . . . 122
6.3.2 Feature Extraction . . . . . 122
6.3.3 Feature Clustering . . . . . 126
6.3.4 Classification . . . . . 126
6.4 The STIRF System . . . . . 128
6.5 Performance Analysis . . . . . 129
6.6 Summary . . . . . 136
References . . . . . 136

**7 Fuzzy Based Neuro - Genetic Algorithm for Stock Market Prediction** . . . . . 139
7.1 Introduction . . . . . 139
7.2 Related Work . . . . . 140
7.3 Model . . . . . 141
7.4 Algorithm . . . . . 146
7.4.1 Algorithm FEASOM . . . . . 146
7.4.2 Modified Kohonen Algorithm . . . . . 146
7.4.3 The Genetic Algorithm . . . . . 148
7.4.4 Fuzzy Inference System . . . . . 149
7.4.5 Backpropagation Algorithm . . . . . 149
7.4.6 Complexity . . . . . 149
7.5 Example . . . . . 150
7.6 Implementation . . . . . 152
7.7 Performance Analysis . . . . . 154

7.8 Summary . . . . . 165
References . . . . . 165

**8 Data Mining Based Query Processing Using Rough Sets and GAs** . . . . . 167
8.1 Introduction . . . . . 167
8.2 Problem Definition . . . . . 169
8.3 Architecture . . . . . 170
8.3.1 Rough Sets . . . . . 171
8.3.2 Information Streaks . . . . . 174
8.4 Modeling of Continuous-Type Data . . . . . 175
8.5 Genetic Algorithms and Query Languages . . . . . 180
8.5.1 Associations . . . . . 181
8.5.2 Concept Hierarchies . . . . . 182
8.5.3 Dealing with Rapidly Changing Data . . . . . 185
8.6 Experimental Results . . . . . 186
8.7 Adaptive Data Mining Using Hybrid Model of Rough Sets and Two-Phase GAs . . . . . 189
8.8 Mathematical Model of Attributes (MMA) . . . . . 190
8.9 Two Phase Genetic Algorithms . . . . . 191
8.10 Summary . . . . . 194
References . . . . . 194

**9 Hashing the Web for Better Reorganization** . . . . . 197
9.1 Introduction . . . . . 197
9.1.1 Frequent Items and Association Rules . . . . . 198
9.2 Related Work . . . . . 200
9.3 Web Usage Mining and Web Reorganization Model . . . . . 200
9.4 Problem Definition . . . . . 202
9.5 Algorithms . . . . . 202
9.5.1 Classification of Pages . . . . . 206
9.6 Pre-processing . . . . . 206
9.7 Example . . . . . 208
9.8 Performance Analysis . . . . . 210
9.9 Summary . . . . . 214
References . . . . . 214

**10 Algorithms for Web Personalization** . . . . . 217
10.1 Introduction . . . . . 217
10.2 Overview . . . . . 219
10.3 Data Structures . . . . . 219
10.4 Algorithm . . . . . 221
10.5 Performance Analysis . . . . . 223
10.6 Summary . . . . . 229
References . . . . . 229

**11 Classifying Clustered Webpages for Effective Personalization** . . . 231
11.1 Introduction . . . 231
11.2 Related Work . . . 232
11.3 Proposed System . . . 233
11.4 Example . . . 237
11.5 Algorithm II: Naïve Bayesian Probabilistic Model . . . 239
11.6 Performance Analysis . . . 241
11.7 Summary . . . 246
References . . . 247

**12 Mining Top - k Ranked Webpages Using SA and GA** . . . 249
12.1 Introduction . . . 249
12.2 Algorithm *TkRSAGA* . . . 252
12.3 Performance Analysis . . . 253
12.4 Summary . . . 258
References . . . 258

**13 A Semantic Approach for Mining Biological Databases** . . . 259
13.1 Introduction . . . 259
13.2 Understanding the Nature of Biological Data . . . 260
13.3 Related Work . . . 262
13.4 Problem Definition . . . 263
13.5 Identifying Indexing Technique . . . 263
13.6 LSI Model . . . 265
13.7 Search Optimization Using GAs . . . 266
13.8 Proposed Algorithm . . . 267
13.9 Performance Analysis . . . 268
13.10 Summary . . . 277
References . . . 277

**14 Probabilistic Approach for DNA Compression** . . . 279
14.1 Introduction . . . 279
14.2 Probability Model . . . 281
14.3 Algorithm . . . 284
14.4 Optimization of $P'$ . . . 285
14.5 An Example . . . 286
14.6 Performance Analysis . . . 287
14.7 Summary . . . 288
References . . . 288

**15 Non-repetitive DNA Compression Using Memoization** ... 291
15.1 Introduction ... 291
15.2 Related Work ... 293
15.3 Algorithm ... 294
15.4 Experimental Results ... 298
15.5 Summary ... 300
References ... 300

**16 Exploring Structurally Similar Protein Sequence Motifs** ... 303
16.1 Introduction ... 303
16.2 Related Work ... 305
16.3 Motifs in Protein Sequences ... 305
16.4 Algorithm ... 307
16.5 Experimental Setup ... 308
16.6 Experimental Results ... 310
16.7 Summary ... 317
References ... 317

**17 Matching Techniques in Genomic Sequences for Motif Searching** ... 319
17.1 Overview ... 319
17.2 Related Work ... 320
17.3 Introduction ... 321
17.4 Alternative Storage and Retrieval Technique ... 323
17.5 Experimental Setup and Results ... 327
17.6 Summary ... 329
References ... 330

**18 Merge Based Genetic Algorithm for Motif Discovery** ... 331
18.1 Introduction ... 331
18.2 Related Work ... 334
18.3 Algorithm ... 334
18.4 Experimental Setup ... 337
18.5 Performance Analysis ... 339
18.6 Summary ... 340
References ... 340

# Acronyms

| | |
|---|---|
| GA | Genetic Algorithms |
| ANN | Artificial Neural Networks |
| AI | Artificial Intelligence |
| SVM | Support Vector Machines |
| KDD | Knowledge Discovery in Databases |
| OLAP | On-Line Analytical Processing |
| MIQ | Machine Intelligence Quotient |
| FL | Fuzzy Logic |
| RS | Rough Sets |
| XML | eXtended Markup Language |
| HTML | Hyper Text Markup Language |
| SQL | Structured Query Language |
| PCA | Principal Component Analysis |
| SDI | Selective Dissemination of Information |
| SOM | Self Organizing Map |
| CBIR | Content Based Image Retrieval |
| WWW | World Wide Web |
| DNA | Deoxyribo Nucleic Acid |
| IGA | Island model Genetic Algorithms |
| SGA | Simple Genetic Algorithms |
| PID | Pima Indian Diabetes |
| Wisc | Wisconsin Breast Cancer Database |
| Hep | Hepatitis Database |
| Ion | Ionosphere Database |
| LVQ | Learning Vector Quantization |
| BPNN | Backpropagation Neural Network |
| RBF | Radial Basis Function |
| ITI | Incremental Decision Tree Induction |
| LMDT | Linear Machine Decision Tree |
| DTD | Document Type Definition |
| MFI | Most Frequently used Index |

| | |
|---|---|
| LFI | Less Frequently used Index |
| hvi | Hierarchical Vector Identification |
| UIC | User Interest Categories |
| KNN | $k$ Nearest Neighborhood |
| DMQL | Data Mining Query Languages |
| TSP | Travelling Salesman Problem |
| MAD | Mean Absolute Deviation |
| SSE | Sum of Squared Error |
| MSE | Mean Squared Error |
| RMSE | Root Mean Squared Error |
| MAPE | Mean Absolute Percentage Error |
| STI | Shape Texture Intensity |
| HIS | Hue, Intensity and Saturation |
| DCT | Discrete Cosine Transform |
| PWM | Position Weight Matrix |
| PSSM | Position Specific Scoring Matrix |
| PRDM | Pairwise Relative Distance Matrix |
| DSSP | Secondary Structure of Proteins |
| LSI | Latent Semantic Indexing |
| GIS | Geographical Information Systems |
| CAD | Computer Aided Design |
| FS | Free Search |
| BGA | Breeder Genetic Algorithm |
| STIRF | Shape, Texture, Intensity-distribution features with Relevance Feedback |