

# **Cognitive Intelligence and Robotics**

## **Series Editors**

Amit Konar, Department of Electronics and Telecommunication Engineering,  
Jadavpur University, Kolkata, India

Witold Pedrycz, Department of Electrical and Computer Engineering, University of  
Alberta, Edmonton, AB, Canada

Cognitive Intelligence refers to the natural intelligence of humans and animals, it is considered that the brain performs intelligent activities. While establishing a hard boundary that distinguishes intelligent activities from others remains controversial, most common behaviors and activities of living organisms that cannot be fully synthesized using artificial means are regarded as intelligent. Thus the acts of sensing and perception, understanding the environment, and voluntary control of muscles, which can be performed by lower-level mammals, are indeed intelligent. Besides the above, advanced mammals can perform more sophisticated cognitive tasks, including logical reasoning, learning, recognition, and complex planning and coordination, none of which can yet be realized artificially to the level of a baby, and thus are regarded as cognitively intelligent.

This book series covers two important aspects of brain science. First, it attempts to uncover the mystery behind the biological basis of cognition, with a special emphasis on the decoding of stimulated brain signals or images. Topics in this area include the neural basis of sensory perception, motor control, sensory-motor coordination, and understanding the biological basis of higher-level cognition, including memory, learning, reasoning, and complex planning. The second objective of the series is to publish consolidated research on brain-inspired models of learning, perception, memory, and coordination, including results that can be realized on robots, enabling them to mimic the cognitive activities performed by living creatures. These brain-inspired models of machine intelligence complement the behavioral counterparts studied in traditional artificial intelligence.

The series publishes textbooks, monographs, and contributed volumes.

More information about this series at <http://www.springer.com/series/15488>

Nandita Sengupta · Jaya Sil

# Intrusion Detection

A Data Mining Approach

 Springer

Nandita Sengupta  
Department of Information Technology  
University College of Bahrain  
Manama, Bahrain

Jaya Sil  
Department of Computer Science  
and Technology  
Indian Institute of Engineering Science  
and Technology (IEST), Shibpur  
Howrah, West Bengal, India

ISSN 2520-1956                      ISSN 2520-1964 (electronic)  
Cognitive Intelligence and Robotics  
ISBN 978-981-15-2715-9              ISBN 978-981-15-2716-6 (eBook)  
<https://doi.org/10.1007/978-981-15-2716-6>

© Springer Nature Singapore Pte Ltd. 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, 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.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

# Preface

Data mining is an integrated process of data cleaning, data integration, data selection, data transformation, data extraction, pattern evaluation, and knowledge presentation. The exponential growth of data opens up new challenges to extracting knowledge from large repositories consisting of vague, incomplete, and hidden information. Data mining research attracted many people working in different disciplines for quite a long period of time. However, the methods lack a comprehensive and systematic approach to tackle several problems in data mining techniques, which are interrelated.

The phrase *intrusion detection* refers to the detection of traffic anomaly in computer networks/systems with an aim to secure data resources from possible attacks. Several approaches to intrusion detection mechanisms are available in the literature. Most of these techniques utilize principles of machine learning/pattern recognition. Unfortunately, the existing techniques fail to incrementally learn network behavior. The book fills this void. It examines the scope of reinforcement learning and rough sets in handling the intrusion detection problem.

The book is primarily meant for graduate students of electrical, electronics, computer science and technology. It is equally useful to doctoral students pursuing their research on intrusion detection and practitioners interested in network security and administration.

The book includes five chapters. Starting from the foundations of the subject, it gradually explores more sophisticated techniques on intrusion detection, including Fuzzy Sets, Genetic Algorithm, Rough Sets, and Hierarchical Reinforcement Learning. The book serves a wide range of applications, covering general computer security to server, network and cloud security.

Chapter 1 provides an overview of intrusion detection. Two distinct types of Intrusion Detection Systems have been examined. They are referred to as misuse detection and anomaly detection systems. Next, the chapter outlines the types of possible attacks. It then emphasizes the main steps usually undertaken in an Intrusion Detection System. The steps include data preprocessing, discretization, dimensionality reduction, and classification. The data preprocessing includes data cleaning, such as missing value prediction, filtering of noisy data, and management of

inconsistent data. Subsequent major steps in data preprocessing are data integration, data transformation, and data reduction. Data reduction has been examined in two ways: attribute reduction and object reduction. The next main step in intrusion detection is discretization, i.e., the transformation of continuous data into quantized intervals. The discretization techniques covered are equal width and equal frequency discretization, bottom-up margin, ChiMerge, entropy-based, and nonparameterized supervised discretization. The rest of Chap. 1 covers the classification of network traffic data. Finally, the chapter comes to an end with a list of concluding remarks.

Chapter 2 is concerned with the well-known discretization procedure of network traffic data. The discretization begins with preprocessing of NSL-KDD data set. Two specific discretization techniques have been examined. The former one, called cut generation method, focuses at the center of a data range dividing the range into two halves. The latter one deals with machine learning techniques.

Chapter 3 introduces the principles and techniques of data reduction. Data reduction refers to either dimension reduction or instance reduction. In this chapter, dimension reduction is achieved by two ways: Rough Sets and Fuzzy-Rough Sets. Instance reduction is performed using clustering algorithms. The rest of the chapter deals with experiments and reporting of results to demonstrate the relative performance of different techniques introduced therein. The metrics used include a confusion matrix.

Chapter 4 provides a novel approach to Q-learning induced classifier to classify the traffic data. In classical Q-learning, we develop a Q-table to store the Q-values at given state space. The Q-table is indexed by states as rows and actions by columns. After the Q-learning algorithm converges, the Q-table is used for the planning application, where the optimal action at a state is determined by the highest Q-value at the state. Here, the authors employ cuts of the continuous traffic attribute to represent the states, and the attributes represent the action set.

The Q-table contains immediate reward/penalty at a given cut for selecting an action (attribute). The Q-table adaptation is undertaken by classical Q-learning. To improve the performance of the Q-learning algorithm, we used rough sets to select a fewer alternatives from a long list so as to improve the classification accuracy. Thus, the attributes used in the Q-table are minimized. This chapter also aims at improving the speed of classification of intrusion traffic data using novel hierarchical learning. Here, the hierarchy is required to determine the attributes in coarse level at the higher level of the hierarchy and at a relatively finer level at lower level in the hierarchy. Generally, the reducts (important attributes obtained by the rough set algorithm) are used to represent the row indices of the Q-table. After an attribute at a given reduct is selected by Q-learning in a top hierarchy, the sub-tasks involved in the selected attribute are determined at the next level of the hierarchy. Thus, multiple levels of hierarchy are used to determine tasks as well as sub-tasks at higher speed of completion.

Chapter 5 concludes the book and provides future research path.

# List of Publication by the Authors Relevant to the Book

## List of Publication by Dr. Nandita Sengupta Relevant to the Book

1. Nandita Sengupta, “Designing Encryption and IDS for Cloud Security”, published in The second International Conference on Internet of Things, Data and Cloud Computing (ICC 2017) held in Cambridge city, Churchill College. University of Cambridge, UK. 22–23 March 2017 (<http://icc-conference.org/index.php/conference-program>).
2. Nandita Sengupta, “Designing of Intrusion Detection System using Efficient Classifier”, Eighth International Conference on Communication networks (ICCN 2014), Elsevier, 25–27 July 2014, Bangalore, India. (<http://www.elsevierst.com/ConferenceBookdetail.php?pid=80>).
3. Nandita Sengupta, “Designing Intrusion Detection System and Hybrid Encryption for e-Government System of Bahrain”, Middle East and North Africa Conference for Public Administration Research, Bahrain, 23–24 April 2014.
4. Nandita Sengupta, Jeffrey Holmes and Jaya Sil, “Detecting Intrusion and Designing of CCRX Dynamic Encryption Algorithm of a Network System”, International Journal of Information Processing, 8(2), 37–46, 2014, ISSN: 0973-8215. Each of the IJIP articles are archived and indexed with prestigious academic indexes including Google Scholar, arXiv, getCITED.
5. Nandita Sengupta, Jaydeep Sen, Jaya Sil and Moumita Saha “Designing of On Line Intrusion Detection System Using Rough Set Theory and Q Learning Algorithm”, Elsevier Journal, Neurocomputing, volume 111, pages 161–168, July 2013.
6. Rahul Mitra, Sahisnu Mazumder, Tuhin Sharma, Nandita Sengupta and Jaya Sil, “Dynamic Network Traffic Data Classification for Intrusion Detection using Genetic Algorithm”, Springer, SEMCCO 2012, pp. 509–518, [https://doi.org/10.1007/978-3-642-35380-2\\_60](https://doi.org/10.1007/978-3-642-35380-2_60), Series Volume 7677, Print ISBN:

- 978-3-642-35379-6, Online ISBN: 978-3-642-35380-2, 20–22 December 2012, Bhubaneswar, India ([http://link.springer.com/chapter/10.1007%2F978-3-642-35380-2\\_60](http://link.springer.com/chapter/10.1007%2F978-3-642-35380-2_60)).
7. Sahisnu Mazumder, Tuhin Sharma, Rahul Mitra, Nandita Sengupta and Jaya Sil, “Generation of Sufficient Cut Points to Discretize Network Traffic Data Sets”, Springer, SEMCCO 2012, pp. 528–539, [https://doi.org/10.1007/978-3-642-35380-2\\_62](https://doi.org/10.1007/978-3-642-35380-2_62), Series Volume 7677, Print ISBN: 978-3-642-35379-6, Online ISBN: 978-3-642-35380-2, 20–22 December 2012, Bhubaneswar, India ([http://link.springer.com/chapter/10.1007%2F978-3-642-35380-2\\_62](http://link.springer.com/chapter/10.1007%2F978-3-642-35380-2_62)).
  8. Nandita Sengupta and Jaya Sil, “Comparison of Supervised Learning and Reinforcement Learning in Intrusion Domain”, International Journal of Information Processing, 7(1), 51–56, 2013.
  9. Nandita Sengupta, Amit Srivastava and Jaya Sil, “Reduction of data Size in Intrusion Domain using Modified Simulated Annealing Fuzzy Clustering Algorithm”, Yogesh Chabba (Ed.): AIM 2012, pp. 97–102, Series Volume 296, [https://doi.org/10.1007/978-3-642-35864-7\\_14](https://doi.org/10.1007/978-3-642-35864-7_14), Print ISBN: 978-3-642-35863-0 2012, Online ISBN: 978-3-642-35864-7, Springer-Verlag Berlin Heidelberg 2012 ([http://link.springer.com/chapter/10.1007%2F978-3-642-35864-7\\_14](http://link.springer.com/chapter/10.1007%2F978-3-642-35864-7_14)).
  10. Moumita Saha, Jaya Sil and Nandita Sengupta, “Designing of an Efficient Classifier using Hierarchical Reinforcement Learning”, ICGST Conference on Computer Science and Engineering, Dubai, UAE, 16–19 July, 2012 (<http://www.icgst.com/paper.aspx?pid=P1121213150>).
  11. Nandita Sengupta and Jaya Sil, “Comparison of Different Rule Calculation Method for Rough Set Theory”, International Journal of Information and Electronics Engineering, Vol. 2, No. 3, May 2012.
  12. Nandita Sengupta and Jaya Sil, “Evaluation of Rough Set Theory Based Network Traffic Data Classifier Using Different Discretization Method”, International Journal of Information and Electronics Engineering, Vol. 2, No. 3, May 2012.
  13. Nandita Sengupta and Jaya Sil, “Comparison of Performance for Intrusion Detection System using Different Rules of Classification”, ISBN 978-3-642-22785-1, ICIP 2011, Volume 157, pp. 87–92, [https://doi.org/10.1007/978-3-642-22786-8\\_11](https://doi.org/10.1007/978-3-642-22786-8_11), Print ISBN: 978-3-642-22785-1, Online ISBN: 978-3-642-22786-8, 5–7 August 2011, Springer Berlin Heidelberg, Bangalore, India ([http://link.springer.com/chapter/10.1007/978-3-642-22786-8\\_11](http://link.springer.com/chapter/10.1007/978-3-642-22786-8_11)).
  14. Nandita Sengupta and Jaya Sil, “Decision Making System for Network Traffic”, KBIE 2011, 8–10 January 2011, Bahrain.
  15. Nandita Sengupta and Jaya Sil, “Information Retrieval Techniques in Intrusion Detection”, International Journal of Information Processing Volume 4, Number 4, 2010, pp. 1–6.
  16. Nandita Sengupta and Jaya Sil, “Intelligent Control of Intrusion Detection System using Soft Computing Techniques”, One Day-Control Engineering Symposium New Directions in Automatic Control: Theories and Applications, 26th April 2010, Bahrain.



17. Nandita Sengupta and Jaya Sil, "Dimension Reduction using Rough Set Theory for Intrusion Detection System" Proceedings of 4th National Conference INDIACom 2010, ISSN: 0973-7529, pp. 251–256, New Delhi, India.
18. Nandita Sengupta and Jaya Sil, "Network Intrusion Detection using RST, k means and Fuzzy c means Clustering", International Journal of Information Processing Volume 4, Number 2, 2010, pp. 8–14.
19. Nandita Sengupta and Jaya Sil, "An Integrated Approach to Information Retrieval using RST, FS and SOM" ICIS 2008, Bahrain, December 2008.

### **List of Publication by Prof. Jaya Sil Relevant to the Book**

1. Asit Kumar Das, Jaya Sil, An efficient classifier design integrating rough set and set oriented database operations, Applied Soft Computing, vol. 11 pp. 2279–2285, 2011.
2. Santi P. Maity, Seba Maity, Jaya Sil and Claude Delpha, Optimized Spread spectrum watermarking for fading-like collusion attack with improved detection, Special Issue on Wireless Personal Communications Journal, Springer Verlag, vol. 69, no. 4, March, 2013.
3. Santi P. Maity, Seba Maity, Jaya Sil, Claude Delpha, Collusion resilient spread spectrum watermarking in M-band wavelets using GA-fuzzy Hybridization, The Journal of Systems and Software, Elsevier Science Direct, vol. 86, pp. 47–59, January, 2013.
4. Santanu Phadikar, Jaya Sil, Asit Kumar Das, Rice diseases classification using feature selection and rule generation techniques, Computers and Electronics in Agriculture, Elsevier Science Direct, pp. 76–85, vol. 90, January, 2013.
5. Nandita Sengupta, Jaydeep Sen, Jaya Sil and Moumita Saha, Designing of On Line Intrusion Detection System Using Rough Set Theory and Q Learning Algorithm, Elsevier Neurocomputing Journal, vol. 111, 161–168, July, 2013.
6. Indrajit De and Jaya Sil, No-reference image quality assessment using interval type 2 fuzzy sets, Applied Soft Computing 30 (2015) 441–453.
7. Nanda Dulal Jana and Jaya Sil, Levy distributed parameter control in differential evolution for numerical optimization, Springer Natural Computing (2015), pp. 1–14.
8. Pratyay Konar, Jaya Sil, Paramita Chattopadhyay, Knowledge extraction using data mining for multi-class fault diagnosis of induction motor, Elsevier Neurocomputing Journal, 166 (2015) 14–25.
9. Nanda Dulal Jana, Jaya Sil and Swagatam Das, Continuous fitness landscape analysis using a chaos-based random walk algorithm, Soft Computing, The Springer journal, pp. 1–28, 2016.
10. Nanda Dulal Jana and Jaya Sil, Interleaving of particle swarm optimization and differential evolution algorithm for global optimization, International Journal of Computers and Applications, Taylor and Francis, Volume 0, - Issue 0, pp-1-18, 2016.

11. Amit Paula, Jaya Sil, Chitrangada Das Mukhopadhyay, Gene selection for designing optimal fuzzy rule base classifier by estimating missing value, *Applied Soft Computing*, 55 (2017) 276–288.
12. Nanda Dulal Jana, Jaya Sil, Swagatam Das, Selection of Appropriate Algorithm for Protein Structure Minimization in AB off-Lattice Model using Fitness Landscape Analysis, *Information Sciences* (2017), <https://doi.org/10.1016/j.ins.2017.01.020>.
13. D. Dutta, P. Dutta and J. Sil, Simultaneous feature selection and clustering with mixed features by multi objective genetic algorithm, *International Journal of Hybrid Intelligent Systems*, 11 (2014) 41–54 41, <https://doi.org/10.3233/his-130182> IOS Press.
14. D. Dutta, P. Dutta and J. Sil, Categorical Feature Reduction Using Multi Objective Genetic Algorithm in Cluster Analysis, *Transactions on Computational Science XXI, Lecture Notes in Computer Science Volume, 8160, 2013*, pp. 164–189.
15. Jaya Sil and Asit K Das, Variable Length Reduct Vs. Minimum Length Reduct - A Comparative study, *Procedia Technology*, Elsevier, 00 (2011) 1–10.
16. P. Dey, S. Dey, S. Datta and J. Sil, Dynamic Discredation Using Rough Sets, *Applied Soft Computing*, Elsevier Science Direct, 11 (2011), 3887–3897.
17. Asit Das and Jaya Sil, An efficient classifier design integrating Rough Set and Dempster-Shafer Theory, *Int. J. Artificial Intelligence and Soft Computing*, Vol. 2, No. 3, 2010, 245–261.
18. Asit K. Das and Jaya Sil, Cluster Validation Method for Stable Cluster Formation, *Canadian Journal on Artificial Intelligence, Machine Learning and Pattern Recognition*, Vol. 1, No. 3, July 2010.
19. Amit Paul and Jaya Sil, Dimension Reduction of Gene Expression data Based on Rough Set Theory, *Serial Publications IJCSIT*, Vol. 1 no. 2 (2008), ISSN: 0974-8385.
20. Amit Paul, Anupam Ghosh and Jaya Sil, Dimension Reduction of Gene Expression data using Redundancy Removal Algorithm—Data Compression Approach, *International Journal of Bioinformatics*, Serial Publications, January-June (2008), vol 1, No. 1, 19–30.
21. N. D. Jana, J. Sil and S. Das, “Particle Swarm Optimization with Population Adaptation”, *IEEE Congress on Evolutionary Computation (CEC’14)*, Beijing, 2014.
22. Amit Paul and Jaya Sil, Dimension Reduction of Gene Expression Data for Designing Optimized Rule Base Classifier, *Recent Advances in Information Technology (RAIT-2014)*, Springer, Dhanbad, India, 2014, 133–140.
23. Amit Paul and Jaya Sil, Feature Filtering of Amino acid sequences Using Rough Set Theory, *International Conference on Computational Intelligence in Data Mining (ICCIDM-2014)*, Springer, Sambalpur, India, 2014 (accepted).
24. Zenefa Rahaman and Jaya Sil, DE Based Q-learning Algorithm to Improve Speed of Convergence In Large Search Space Applications, *2014 International Conference on Electronic Systems, Signal Processing and Computing Technologies*, pp. 408–412, 2014.

25. Amit Paul and Jaya Sil, "Gene Selection for Classifying Patients using Fuzzy Importance Factor", IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, <https://doi.org/10.1109/fuzz-ieee.2013.6622383>, India, 2013, pp. 1–7.
26. Pratyay Konar, Moumita Saha, Dr. Jaya Sil, Dr. Paramita Chattopadhyay, Fault Diagnosis of Induction Motor Using CWT and Rough-Set Theory, 2013 IEEE Symposium on Computational Intelligence in Control and Automation (CICA), pp. 9–15, 2013.
27. Nanda Dulal Jana and Jaya Sil, "Particle Swarm Optimization with Levy Flight and Adaptive Polynomial Mutation in gbest Particle", In 2nd International Symposium on Intelligent Informatics (ISI'13) Mysore, August 2013.
28. Nanda Dulal Jana and Jaya Sil, "Particle Swarm Optimization with Exploratory Move", (PREMI'13), Proceedings. Springer 2013 Lecture Notes in Computer Science ISBN 978-3-642-45061-7, Kolkata, December 2013, pp. 614–621.
29. Nanda Dulal Jana and Jaya Sil, "Hybrid Particle Swarm Optimization Techniques for Protein Structure Prediction using 2D Off-lattice Model", In SEMCCO 2013, Tamilnadu, December 2013.
30. D. Dutta, P. Dutta and J. Sil Feature Weighted Clustering of Mixed Numeric and Categorical datasets by Hybrid Evolutionary Algorithm, 2013 IEEE INDICON to be held at Victor Menezes Convention Centre, Indian Institute of Technology (IIT) Bombay, Mumbai, India from 13–15 December, 2013.
31. D. Dutta, P. Dutta and J. Sil, Simultaneous continuous feature selection and K clustering by multi objective genetic algorithm, in: *Proceeding of 3rd IEEE International Advance Computing Conference* (2013), 937–942.
32. Monidipa Das and Jaya Sil, Query Selection using Fuzzy Measures to Diagnose Diseases, B. K. Kaushik and Vinu V. Das (Eds.): AIM 2013, LNCS pp. 19–30, 2013. © Communications in Computer and Information Science 2013.
33. D. Dutta, P. Dutta and J. Sil, Clustering by multi objective genetic algorithm, in: *Proceeding of 1st IEEE International Conference on Recent Advances in Information Technology* (2012), 548–553.
34. D. Dutta, P. Dutta and J. Sil, Clustering data set with categorical feature using multi objective genetic algorithm, in: *Proceeding of IEEE International Conference on Data Science and Engineering* (2012), 103–108.
35. D. Dutta, P. Dutta and J. Sil, Data clustering with mixed features by multi objective genetic algorithm, in: *Proceeding of 12th IEEE International Conference on Hybrid Intelligent Systems* (2012), 336–341.
36. D. Dutta, P. Dutta and J. Sil, Simultaneous feature selection and clustering for categorical features using multi objective genetic algorithm, in: *Proceeding of 12th IEEE International Conference on Hybrid Intelligent Systems* (2012), 191–196.
37. Nandita Sengupta, Amit Srivastava and Jaya Sil, Reduction of data Size in Intrusion Domain using Modified Simulated Annealing Fuzzy Clustering Algorithm, Yogesh Chabba (Ed.): AIM 2012, pp. 99–104, 2012 © Springer-Verlag Berlin Heidelberg 2012.

38. Moumita Saha and Jaya Sil, Dimensionality Reduction Using Genetic Algorithm And Fuzzy-Rough Concepts, 978-1-4673-0125-1 c\_2011 IEEE, pp. 385–390, 2011.
39. Amit Paul and Jaya Sil, Missing Value Estimation in Microarray Data using Coregulation and Similarity of Genes, 978-1-4673-0125-1 c\_2011 IEEE, pp. 709–714, 2011.
40. Tapas Si, N.D. Jana and Jaya Sil, Particle Swarm Optimization with Adaptive Polynomial Mutation, 978-1-4673-0125-1 c\_2011 IEEE, pp. 143–147, 2011.
41. Amit Paul and Jaya Sil, Estimating Missing value in Microarray Gene Expression Data, FUZZ-IEEE 2011.
42. Nandita Sengupta and Jaya Sil, Information Retrieval Techniques in Intrusion Detection, ICIP-2010, pp. 447–455, 2010.
43. Nandita Sengupta and Dr. Jaya Sil, Network Intrusion Detection using RST, k-means and Fuzzy c means Clustering, Third International Conference on Information Processing, 2009 ICIP 2009 pp. 401–409, ISBN:978-93-80026-72-5, 2009.
44. Amit Paul and Jaya Sil, “Sample Selection of Microarray data using Rough-Fuzzy based Approach”, World Congress on Nature and Biologically Inspired Computing (NABIC 2009), IEEE Computer Society Press, ISBN: 978-1-4244-5612-3 pp. 379–384, 2009.

# Acknowledgements

We sincerely thank Shaikh Khaled M Al Khalifa, Founder member and Chairman of Board of Trustee, University College of Bahrain, for continuous support for research work by providing conducive environment, infrastructure, and facilities and for her passionate encouragement, and Prof. Shala Emara, President, Prof. Geoffrey Elliott, Vice President and colleagues of University College of Bahrain for their moral support. We are also sincerely thankful to Prof. Sekhar Mandal, Head of Computer Science and Technology Department, Indian Institute of Engineering Science and Technology (IEST), Shibpur, India, for his support and providing infrastructure of IEST. We are deeply indebted to Prof. Amit Konar, Department of Electronics and Telecommunication, Jadavpur University, India, for his valuable time and guidance for completion of this work.

Nandita Sengupta  
Jaya Sil

# Contents

<b>1</b>	<b>Introduction</b>	1
1.1	Intrusion Detection Systems	3
1.1.1	Types of IDS	3
1.1.2	Types of Attacks	5
1.2	Data Preprocessing	6
1.2.1	Cleaning of Data	6
1.2.2	Integration of Data	8
1.2.3	Transformation of Data	8
1.2.4	Data Reduction	8
1.3	Discretization	11
1.3.1	Classification of Discretization Methods	11
1.3.2	Methods of Discretization	12
1.4	Learning Classifier	14
1.4.1	Dynamic Learning	14
1.4.2	Dynamic Classification	15
1.5	The Work	15
1.5.1	Contributions	16
1.6	Summary	18
	References	19
<b>2</b>	<b>Discretization</b>	27
2.1	Preprocessing	27
2.2	Cut Generation Method	28
2.2.1	Algorithm for Generation of Cut	29
2.2.2	Encoding Method of Center-Spread Technique	32
2.2.3	Discrete Value Mapping	32
2.3	Cut Generation Using Machine Learning Technique	34
2.3.1	Optimized Equal Width Interval (OEWI)	34
2.3.2	Split and Merge Interval (SMI)	37

2.4	Discussions on Results	39
2.5	Summary	40
	References	44
<b>3</b>	<b>Data Reduction</b>	<b>47</b>
3.1	Dimension Reduction Using RST	48
3.1.1	Preliminaries of RST	48
3.1.2	Reduct Using Discernibility Matrix	51
3.1.3	Reduct Using Attribute Dependency	55
3.2	Dimension Reduction Using Fuzzy-Rough Set	58
3.2.1	Fuzzy-Rough Sets	58
3.2.2	Rule-Base	60
3.2.3	Fuzzy-Rough-GA	63
3.3	Instance Reduction	67
3.3.1	Simulated Annealing-Based Clustering Algorithm	68
3.3.2	Modified_S AFC Algorithm	69
3.3.3	Most Significant Cluster	71
3.4	Results and Discussions	72
3.4.1	Results of Dimension Reduction on Discrete Domain	72
3.4.2	Confusion Matrix	75
3.4.3	Results of Dimension Reduction on Continuous Domain	76
3.4.4	Accuracy After Instance Reduction	77
3.5	Summary	79
	References	79
<b>4</b>	<b>Q-Learning Classifier</b>	<b>83</b>
4.1	Q-Learning	83
4.1.1	Extended-Q-Learning Algorithm for Optimized Cut Generation	85
4.2	Hierarchical-Q-Learning Approach	95
4.2.1	Definition of Semi-Markov Decision Process (SMDP)	96
4.2.2	Optimization of Linguistic Labels	96
4.3	Results and Comparisons	98
4.3.1	Result of Extended-Q-Learning Algorithm	98
4.3.2	Experiments Using Synthetic Data Set	101
4.3.3	Results of the Proposed Hierarchical-Q-Learning Algorithm	104
4.4	Summary	106
	References	109

- 5 Conclusions and Future Research** . . . . . 113
  - 5.1 Essence of the Proposed Methods . . . . . 113
  - 5.2 Outstanding Issues . . . . . 114
  - 5.3 Future Research Directions . . . . . 116
  - References . . . . . 117
  
- Annexure** . . . . . 119
  
- References** . . . . . 129
  
- Subject Index** . . . . . 131



## About the Authors



**Dr. Nandita Sengupta** did her Bachelor of Engineering from Indian Institute of Engineering Science and Technology (IEST), Shibpur, India (formerly known as Bengal Engineering College, Shibpur, Calcutta University). She completed her Post-Graduate Course of Management in Information Technology from IMT. She did M.Tech. (Information Technology) and subsequently obtained her Ph.D. in Engineering (Computer Science and Technology) from IEST, Shibpur, India. She has 29 years of working experience out of which 11 years was in industry and 18 years in academics teaching various subjects of IT. Presently she is working as Associate Professor at University College of Bahrain, Bahrain. Her areas of interest are Analysis of Algorithm, Theory of Computation, Soft Computing Techniques, Network Computing. She achieved “Amity Best Young Faculty Award” on 9th International Business Horizon INBUSH 2007 by Amity International Business School, Noida in February 2007. She has around 35 publications in reputed conferences and journals.



**Jaya Sil** is associated with the Department of Computer Science and Technology in the Indian Institute of Engineering Science and Technology, Shibpur as a professor since 2003. She passed out B.E. in Electronics and Telecommunication Engineering from B.E. College under Calcutta University, India on 1984 and M.E. (Tele) from Jadavpur University, Kolkata, India on 1986. Professor Jaya Sil obtained her Ph.D. (Engineering) degree from Jadavpur University, Kolkata on 1996 in the topic Artificial Intelligence. She started her teaching career from B.E. College, Howrah, India in the department of Computer Science and Technology as a lecturer on 1987. Professor Sil worked as Postdoctoral Fellow in Nanyang Technological University, Singapore during 2002–2003. Professor Sil visited Bioinformatics Lab in Husar, Heidelberg, Germany for collaborative research. INSA Senior scientist fellowship has been awarded to her and she visited Wraclaw University of Technology, Poland in 2012. Professor Sil also delivered tutorial, invited talk, presenting papers and chairing sessions in different International conferences in abroad and India. Professor Sil has more than 200 research papers in the field of Bioinformatics, Machine learning and Image Processing along with applications in different Engineering fields. She has published many books and several book chapters and acted as reviewers in IEEE, Elsevier and Springer Journals.