

Guest Editorial Preface

Special Issue on Advances in Machine Learning and Data Mining

Brojo Kishore Mishra, C.V. Raman College of Engineering, Bhubaneswar, India

V. Santhi, VIT University, Vellore, India

Sanjaya Kumar Panda, V.S.S. University of Technology, Burla, India

Ashish Khanna, Maharaja Agrasen Institute of Technology, New Delhi, India

Vishal Gour, Government Engineering College, Bikaner, India

Vishal Jain, Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi, India

Machine learning and data mining are the growing fields of research in marketing, customer relationship management, trading, advertising, stock analysis and health care. Machine learning enables the computer to learn without any explicitly programming. On the other hand, data mining focuses on finding the interesting patterns in large datasets. Both machine learning and data mining use many of the same set of algorithms, methods and techniques. However, they differ each other in prediction. Machine learning is especially required to reproduce the known patterns and knowledge, whereas data mining is required to uncover early unknown patterns and knowledge.

There are numerous applications in machine learning and data mining such as medical diagnosis, classification, prediction, extraction, image recognition, speech recognition, learning associations, regression, market basket analysis, customer relationship management, fraud and intrusion detection, bioinformatics and many more. The integration of machine learning and data mining will give a wide-scope to interdisciplinary research activities. The main focus of this special issue is to provide the recent developments in the field of machine learning, data mining and its applications, which are contributed by various researchers by developing algorithms, theories, simulations and improved techniques.

Rouhani and Mirsharif presented a medical diagnosis method in gestational diabetes mellitus (GDM) at the initial stage of pregnancy. GDM causes various side effects to both mother and child if it is not treated properly. The authors used artificial neural network and decision tree to reduce the error rate and increase the level of accuracy and prediction.

Jain, Sharma and Agarwal used deep learning technology, called convolutional neural network by including a semantic layer. As a result, they named the resultant model as semantic convolutional neural network. The proposed architecture is evaluated by two corpora, SMS spam dataset and twitter dataset, and the accuracy of the proposed architecture is 98.65% and 94.40% in these datasets in compare to the state-of-the-art results.

Kumar and Sarkar presented a comprehensive performance assessment of state-of-the-art learning algorithms over 14 real world datasets, namely Australian (credit card approval), cancer, credit card (creditable or non-creditable client), dermatology, ecoli, glass, heart (Hungarian), heart (swiss), iris, liver, new thyroid, pima-indians, soyabean (large) and wine. They considered six competent learners,

namely C4.5, RIPPER, Naïve Bayes, C4.5 + RIPPER, C4.5 + Naïve Bayes and Naïve Bayes + RIPPER to investigate their performance in the WEKA 3.7.4.

Virmani, Jain, Parikh, Upadhyaya and Srivastav presented a variant of k-means algorithm to improve clustering, reduce overlapping and cluster formation time. They considered normalization to improve the performance over the traditional k-means algorithm. Kisan, Mishra, Chawda and Nayak estimated the fractal dimension of color images using various color models. Moreover, they proposed a new estimation method for CMY and HSV color spaces and shown the superior performance over MCD method.

Ahsan, Kumari, Singh and Pal analyzed the tweets of U. S. Presidential election 2016. They presented a comprehensive analysis on the sentiments which are encoded in the textual contents. They also explored the relationship between sentiment and information diffusion using random forests. At last, they used support vector and k-neighbors regression models to capture the dynamics of information diffusion.

As guest editors of advances in machine learning and data mining, we hope that the variety of research work covered under this special issue such as machine learning, data mining, neural network, social network analysis, classification, clustering, sentiment analysis and image processing. It has marvelous value for researchers, scientists, developers and practitioners. At the same time, we are also grateful to the authors and co-authors for making their scholarly contributions to this special issue and their patience during crucial revision stages. The technical standard and quality of the published content is based on the strength and expertise of the submitted papers. In response to the call for submissions, advances in machine learning and data mining received 18 submissions. Each submission was peer reviewed by up to three international review committee members (from University of California, CSIRO, University of Technology, Sydney, IITs, NITs, VIT, VSSUT, CVRCE, BVICAM to name a few). After receiving the reviews of the papers, the committee was selected a total of six papers for publication in this special issue of IJKDB. We wish to thank all the committee members for their hard work, dedication and timely submission of the reviews without which it would have been difficult to maintain the special issue schedule. We sincerely thank (Ms) Kayla Bishard, Assistant Development Editor from IGI Global for their cooperation and constant support throughout the publication process.

Our special thanks go to the Editor-in-Chief of the International Journal of Knowledge Discovery in Bioinformatics (IJKDB), Dr. George Perry, University of Texas at San Antonio, USA and Dr. Clyde F. Phelix, University of Texas at San Antonio, USA for all their support, efficiency, and competence rendered to this special issue.

Brojo Kishore Mishra
V. Santhi
Sanjaya Kumar Panda
Ashish Khanna
Vishal Gour
Vishal Jain
Guest Editors
IJKDB