

Best papers from the Fifth International Conference on Advanced Data Mining and Applications (ADMA 2009)

Jian Pei · Joao Gama · Qiang Yang ·
Ronghuai Huang · Xue Li

Published online: 9 April 2011
© Springer-Verlag London Limited 2011

We are very pleased to present this special issue of the best papers of the Fifth International Conference on Advanced Data Mining and Applications (ADMA 2009). The conference was held in Beijing, China, on August 17–19, 2009. ADMA 2009 was the continuation of the success in the ADMA conference series as one of the leading international forums for data mining researchers, practitioners, developers, and users to exchange edge-cutting ideas, techniques, and experience. The acceptance into ADMA 2009 proceedings was extremely competitive. From the 322 submissions, the program committee selected 39 full papers, 53 short papers, and two invited papers for presentation in the conference and inclusion in the proceedings. The contributed papers cover a wide range of data mining topics and a diverse spectrum of interesting applications.

The five articles included in this special issue are selected from the 39 full papers presented in the conference, and address interesting, important, and challenging data mining problems ranging from data mining applications to systems support.

J. Pei (✉)
Simon Fraser University, Burnaby, Canada
e-mail: jpei@cs.sfu.ca

J. Gama
University of Porto, Porto, Portugal
e-mail: jgama@fe.up.pt

Q. Yang
Hong Kong University of Science and Technology,
Hong Kong, China
e-mail: qyang@cse.ust.hk

R. Huang
Beijing Normal University, Beijing, China
e-mail: huangrh@bnu.edu.cn

X. Li
University of Queensland, Brisbane, Australia
e-mail: xueli@itee.uq.edu.au

The first article authored by Chen, Liu, Furuse et al. titled “*Indexing expensive functions for efficient multi-dimensional similarity search*” tackles the problem of similarity search on high dimensional data, which is very challenging due to the well-known “curse of dimensionality”. The paper develops a new technique to precompute some key values so that evaluating expensive distance functions like L_p norm is made highly efficient.

Social tagging has been popularly used in resource organization and sharing. However, tags are often of low quality. Cui, Liu, He et al. tackled this problem in their article “*Tag-Clus: a random walk-based method for tag clustering*” by developing a random walk-based clustering method to group tags with similar semantics together. Their method exploits the relationship between tags and resources and addresses several challenges in tag clustering.

The third article, written by Tang, Zhou, Li et al. and titled “*Exploring the wild birds migration data for the disease spread study of H5NI: a clustering and association approach*”, presents an interesting data mining application case study. The authors tackled a real biological problem using multiple data mining techniques, including hierarchical clustering and association analysis. They reported some interesting findings that may be meaningful in biology.

Gao, Liu, and Zhong reported another meaningful data mining application case study in their article “*Network immunization and virus propagation in email networks: experimental evaluation and analysis*”. Their study improves a state-of-the-art method and explains from the human dynamics aspect why old virus can survive in nowadays networks.

The fifth article in this special issue is by Feng, Wang, Yu et al. titled “*Extracting common emotions from blogs based on finely grained sentiment clustering*”. The study addresses an interesting and challenging problem—capturing people’s collective sentiments and their distribution on the Web. The authors applied probabilistic latent semantic analysis (PLSA) and developed a clustering analysis approach.

In addition to making good progress in tackling their specific problems, respectively, the five articles in this special issue share two common features. First, they all address some practical problems head-to-head. Second, they all verify their methods using real data sets and provide meaningful insights into the corresponding applications.

We hope that you are entertained and inspired by the articles in this special issue. We also want to take this opportunity to thank again all those who submitted their papers to the ADMA 2009 conference, the members of the program committee for their hard work, the external reviewers for their help in the review process, and the local organizers for making the conference happened in a wonderful way.