Special issue devoted to papers presented at the second INFORMS workshop on artificial intelligence and data mining, Seattle, November 3, 2007

Wei Jiang · Anurag Agarwal

The Second INFORMS Workshop on artificial intelligence and data mining (WAID-07) was held in Seattle on November 3, 2007. The workshop was a pre-conference workshop to the National INFORMS meeting. Twenty papers were presented at the workshop, divided in two tracks—AI and Data Mining. A special issue of *Information Technology and Management* was devoted to papers from WAID-07. Twelve papers were submitted for possible publication in the special issue, of which six have been accepted for publication. Three papers were published in Volume 10 issue 1, 2009 of *Information Technology and Management*. Three more papers are being published in this issue.

The first paper, titled "A Finite-Sample Simulation Study of Cross Validation in Tree-Based Models" by Seoung Bum Kim, Xiaoming Huo and Kwok-Leung Tsui, explores the behavior of cross validation in tree-based classifiers. The performance of cross-validated tree-based classifier and Bayes classifier are empirically tested. The authors find that the differences between the testing and training errors from a cross-validated tree classifier and the Bayes classifier follow a simple linear regression model. Based on their experience, the authors recommend using cross-validated tree-based classifiers when sample size is relatively small.

The second paper, titled "Antecedents of Open Source Software Defects: A Data Mining Approach to Model Formulation, Validation and Testing" by Uzma Raja and Marietta J. Tretter, uses data mining and text mining techniques to develop, test and validate a model for the antecedents of Open-Source Software (OSS) defects. The authors look at over 5,000 active and mature OSS projects to validate their model. Results indicate that variables such as project type, end-user activity, process quality, team size and project popularity have a significant impact on the defect density of operational OSS projects.

The third paper, titled "Identifying Fall-Related Injuries: Text Mining the Electronic Medical Record" by Monica Chiarini Tremblay et al. uses data mining and text mining techniques to investigate if unstructured text-based information included in the electronic medical records can validate and enhance those records in the administrative data that should have been coded as fall related injuries. The authors identify many challenges involved in any study involving text mining.

W. Jiang
Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
e-mail: jiangwei08@gmail.com

A. Agarwal (✉)
Department of Information Systems and Decision Sciences, College of Business, University of South Florida, Sarasota, Sarasota, FL, USA
e-mail: agarwala@sar.usf.edu