Inf Syst Front (2009) 11:345–347
DOI 10.1007/s10796-009-9153-4

Editorial for the special issue of knowledge discovery and management in biomedical information systems

Ying Liu · Lawrence Wing-Chi Chan · Chi-Ren Shyu · Ying Liu

Published online: 18 March 2009

1 Introduction

The rapid growth of research and development using different advanced techniques, e.g. AI and Machine Learning, in biological and medical information systems has drawn worldwide attention on the management of knowledge discovered in, for instance, gene databases and public healthcare portals. However, challenges remain from various fundamental issues, e.g. difficulty in data acquisition and heterogeneity of biological information, to the lack of methodology that supports the fusion, processing and management of abundant biomedical and genomic data and eventually the management and interpretation of knowledge that arises from such discoveries for the purpose of problem solving and decision making in translational medicine, e.g. clinical medical practice. This Special Issue is therefore dedicated to innovative, state-of-the-art research, technology development and applications of knowledge discovery and management in biomedical information systems.

All the papers included in this issue can be basically clustered into two groups. The first group mainly focuses on using data mining in assisting disease prediction, medical diagnosis, and rule discovery in microarray data, etc. "Efficient Mining of Multilevel Gene Association Rules from Microarray and Gene Ontology" by Tseng, Yu and Yang, presents an algorithm that extended the current association rule mining with the assistance of the concept hierarchy in gene ontology to discover multilevel gene association rules from microarray data. In particular, a hierarchy information encoder has been utilized to provide the structured vocabularies which are organized in a rooted directed acyclic graph interpreting the roles of genes and gene products. This encoder serves as an interpreter of domain knowledge and it helps to limit the number of rules generated. Their experimental study shows that the proposed method is able to provide biologists different rule sets which indicate different gene expression patterns existed in cells. The refined version of the proposed algorithm can further interpret the prior biology knowledge into constraints embedded in the rule mining.

"Comparing Data Mining Methods with Logistic Regression in Childhood Obesity Prediction" by Zhang, Tjortjis, Zeng, Qiao, Buchan and Keane, reports a study using non-linear interactions to help improve prediction accuracy of children obesity by comparing the results of logistic regression, a prevailing choice in medical prediction analysis, with some six well established data mining algorithms which are commonly used for either prediction or classification purpose, e.g. decision tree and decision rule, Bayesian network and SVM. The empirical study based on real-world data demonstrates that SVMs can

The first author, Ying Liu, is the corresponding author.

Y. Liu (✉)
Department of Industrial and Systems Engineering,
The Hong Kong Polytechnic University,
Hong Kong, China
e-mail: mfyliu@poly.edu.hk

L. W.-C. Chan
The Hong Kong Polytechnic University,
Hong Kong, China

C.-R. Shyu
University of Missouri-Columbia,
Columbia 65211, USA

Y. Liu
University of Texas at Dallas,
Dallas 75080-3021, USA

produce better sensitivity prediction rate than Bayesian algorithms, while Bayesian algorithms outperform SVMs in terms of overall prediction rate. Therefore, in general, we believe that data mining algorithms can be considered as an alternative choice to logistic regression.

"Image Fusion Enhancement of Deformable Human Structures Using a 2-stage Warping-deformable Strategy: A Content-based Image Retrieval Consideration" by Tang and Ip, describes their efforts in proposing a warping-deformable model in medical image registration and content-based image retrieval. Different from the general non-rigid registration, the proposed approach adopts a 2-stage strategy for image registration, i.e. a global-deformable transformation followed by a warping-transformation. As for the global deformable transformation, it is implemented by least-square fitting of lower order polynomial, while the local warping is developed using the radial basis function. Their experimental study using CT and MR images of abdomen shows that the proposed warping-deformable approach is effective in enhancing the matching rate of images which are basically in deformable nature, such as liver images.

"An Automated Bacterial Colony Counting and Classification System" by Chen and Zhang, tackles the automated bacterial colony counting and classification problem using a data mining approach. Based on the proposed approach, the system first examines the chromatic or color components of the input plate/dish images. Next, counting on a fictitious hierarchical structure concept, i.e. background, container, and colony, the proposed system gradually extracts bacterial colony from the images by removing objects existed in background and container layers. The Watershed algorithm is adopted to detect and separate the possible colony clusters. Finally, a one-class SVM with Radial Basis Function is applied as the classification algorithm. The experimental study shows some promising classification results where the feature combination is based on the pair of solidity and color variance or compactness and color variance.

"Proportional Fault-tolerant Data Mining with Applications to Bioinformatics" by Lee, Peng and Lin, describes their efforts on an interesting topic, i.e. fault-tolerant (FT) pattern mining in biological databases. Their idea is that the number of tolerable faults found in the patterns remains proportional to the length of the patterns. Two algorithms, FT-BottomUp and FT-LevelWise, are introduced where the first one generates all candidate patterns with any number of faults, and the second method divides all patterns into several groups according to the number of tolerable faults. The experimental results based on two reported epitopes of spike proteins reveals that the proposed proportional FT mining approach performs better than a fixed FT mining method, e.g. FT-Apriori.

"Coronary Artery Disease Prediction Method Using Linear and Nonlinear Feature of Heart Rate Variability in Three Recumbent Postures" by Lee, Kim, Noh, Shin, Yun and Ryu, notes down their efforts in developing various feature schemes to represent heart rate variability (HRV) as well as a prediction model to enhance the reliability of medical tests for coronary artery diseases using data mining approach. In their work, both linear features, which are derived from ECG signal processing, time domain analysis, etc, and nonlinear features, which are derived from Poincare plot, approximate entropy, Hurst exponent, etc, are included. The experimental study of disease prediction based on 43 control subjects and 64 patients with coronary artery disease using the proposed feature scheme shows that a good performance in terms of accuracy can be delivered by both SVM and an association rule based algorithm.

The second group of papers reports the current study of deploying advanced information management using RFID, Semantic Web and knowledge integration, sharing and management in biomedical information systems. "Automating Object-Oriented Integration and Visualization of Multidisciplinary Biomedical Data in Radiology Workflow: Compartmental PACS Model" by Wong, Chan and Liu, proposes a novel compartmental Picture Archiving and Communication System (PACS) model for the automatic object oriented integration and visualization of heterogeneous data in the multidisciplinary biomedical studies. Implemented using open source software and structured based on the Digital Imaging and Communication in Medicine (DICOM) model, the proposed system demonstrates a deliberative system design for allocating mission-critical components at the clinical frontline of imaging services and leaving the archiving components at the backend. Currently, the prototype of this system is in operation at the Department of Health Technology and Informatics at the Hong Kong Polytechnic University.

"Interactive Survival Analysis with the OCDM System: From Development to Application" by Klenk, Dippon, Fritz and Heidemann, reports their efforts in providing medical practitioners an interactive system, i.e. Online Clinical Data Mining (OCDM), to deeply analyze medical data without the presence of statisticians and data mining experts. Their field user studies, particularly focusing on the interactive survival analysis in cancer, exhibit a great convenience and support due to the introduction of OCDM. Currently, OCDM is in operation in a few hospitals in Germany, e.g. Robert Bosch Hospital and Stuttgart Cancer Registry.

"Knowledge Management in Biomedical Libraries: A Semantic Web Approach" by Fuentes-Lorenzo, Morato and Gómez, highlights some of their latest work in developing BioSem: a knowledge management system for biomedical libraries using semantic technology as the key enabler for information searching, processing and management. In this work, the comprehensive design principles are explained, including semantic representation, semantic scope, bio-

ontology based semantics creation, navigation and faceted search. The conceptual comparison with other existing systems reveals the advantages of harnessing semantic approach in biomedical library management. More field case studies are expected.

"Design of a RFID-based Healthcare Management System Using an Information System Design Theory" by Ngai, Poon, Suk and Ng, presents their efforts in designing and developing a healthcare management system using Radio Frequency Identification (RFID) technology. The details include the state-of-the-art of RFID application in healthcare information systems, system architecture, theoretical background of information system design principles, and system complexity issue, etc. The extensive user evaluation and field tests of the prototype system have confirmed the practical value in system design and also shed light on several issues that need to be taken care of in the RFID system, e.g. safety management and pharmaceutical inventory operation and management.

The ten articles selected for this Special Issue were chosen after two rounds of reviews. In the first round, each paper was peer reviewed by three reviewers, and in the second round the revised article was reviewed by one former reviewer and one Guest Editor. A third revision was requested for a small number of articles. Due to the language proficiency of authors, some papers had been requested to be proofread by native speakers. Finally, we would like to take this opportunity to thank all the authors for their excellent work contributing to this Special Issue, to the research community and industry. We are very grateful to the reviewers for offering their precious time and efforts and for providing constructive comments. Particularly, we would like to express our sincere thanks to Prof. Ram Ramesh and Prof. H. R. Rao, Editor-in-Chiefs, for giving us the opportunity to make this Special Issue a reality and for their advice and support throughout the whole year. Without all of you, this Special Issue would not be possible.

Ying Liu is presently a Lecturer of the Department of Industrial and Systems Engineering at the Hong Kong Polytechnic University. He obtained his Ph.D. from the Singapore MIT Alliance (SMA) at the National University of Singapore in 2006. His current research interests focus on design informatics, data mining and text mining, intelligent information processing and management, machine learning, and their joint applications in engineering design, manufacturing and medical and healthcare industry for knowledge discovery and management purpose. He is a member with ACM, IEEE\CS, ASME and the Design Society.

Lawrence Wing-Chi Chan is currently Assistant Professor in the Department of Health Technology and Informatics at the Hong Kong Polytechnic University. Dr. Chan received the degrees of Bachelor of Engineering with first class honors and Doctor of Philosophy from the Department of Mechanical Engineering at the University of Hong Kong. He was investigating the combined use of fuzzy logic and neural network in modeling nonlinear stochastic systems in his postgraduate study. After graduation, he worked in Jockey Club Research and Information Center for Landslip Prevention and Land Development as Database Administrator at the University of Hong Kong. From 2002 to 2006, he joined the Hong Kong Polytechnic University as Research Associate and then promoted to Research Fellow in the field of medical imaging. His current research interests include intelligent health knowledge discovery, multidisciplinary clinical decision support and biomedical image and signal processing. He is actively conducting the integrated analysis of the atherosclerotic risk factors and the subsequent risk assessment of diabetic individuals. Dr. Chan has published 10 SCI journal papers, one of which has been non-self cited for 29 times, and two book chapters, presented 21 conference papers or posters, granted or filed three patents, and received two international awards for the innovative mobile computing application in medicine.

Chi-Ren Shyu received his Ph.D. degrees in Electrical and Computer Engineering from Purdue University in 1999. Upon completing one year of post-doctoral training at Purdue, he joined the Computer Science Department at the University of Missouri-Columbia (MU) in October 2000. In addition to holding the Shumaker Endowed Professorship for Informatics, he is currently the Director of the MU Informatics Institute, a Ph.D. program in bioinformatics and health informatics. Dr. Shyu received numerous awards including Outstanding Teaching Awards, MU College of Engineering Research Award, and the National Science Foundation CAREER Award. As the founding director of the Medical and Biological Digital Library Research Lab, Dr. Shyu oversees the work of a crew of students on a variety of research projects in such areas as biomedical informatics, geospatial informatics, data mining, and computer vision. Project sponsors for his research include the National Science Foundation, the National Institute of Justice, the National Geospatial-Intelligence Agency, the National Institute of Health, U.S. Department of Education, and other organizations, both for-profit and non-profit.

Ying Liu received the BS degree in environmental biology from Nanjing University, China, the master's degrees in bioinformatics and computer science from the Georgia Institute of Technology in 2002, and the PhD degree in computer science from the Georgia Institute of Technology in 2005. He is now a tenure-track assistance professor in the Department of Computer Science, Department of Molecular and Cell Biology, University of Texas at Dallas. His research interests include bioinformatics, computational biology, data mining, text mining, and database system. He has publlished more than 40 peer-reviewed research papers in various journals and conferences. His current projects include text mining of medical literature databases, creation of databases for biological applications, computational systems biology, and data mining for better understanding of genomic/proteomic and medical data. He has served as a program cochair/conference cochair and a program committee member of several international conferences/workshops. He is a member of ACM.