Guest editorial

Guest editorial for DKE special issue on "Biomedical Data Mining"

Data mining aims at identifying valid, novel, potentially useful, and ultimately understandable patterns in data. The data mining processes used to extract and verify patterns in data form the core for knowledge discovery and decision support. Data mining bridges several areas, including databases, statistics, machine learning and human–computer interaction.

Modern biology and medicine are data-intensive sciences that generate large volumes of complex and heterogeneous data. Undoubtedly they represent one of the most important application domains for data mining. Biomedical informatics is a discipline related to bioinformatics and has its roots in medical informatics. It studies the use of information technology and advanced research computing in the practice of biomedical sciences and medicine. Biomedical data mining may be considered as one of the most important sub-areas in biomedical informatics, the goal of which is the study of the discovery of important patterns or regularities in the huge volumes of biomedical data.

IEEE CBMS is an annual event organized by the IEEE Technical Committee on Computational Medicine (TCCM) of the IEEE Computer Society, and IEEE CBMS 2008 was the 21st conference in the series. The Call for Papers at IEEE CBMS 2008 generated 246 submissions, out of which 86 manuscripts were accepted as full papers. Out of these seven papers with the focus on knowledge discovery and data mining were invited to submit extended and revised papers to this special issue. All papers have undergone the blind peer review, resulting for some of them in further substantial revision and improvement.

The first three papers represent the field of medical imaging. Lee et al. in "CBIR of spine X-ray images on inter-vertebral disc space and shape profiles using feature ranking and voting consensus" consider a novel technique for content-based retrieval of spine X-ray images that makes use of both inter-vertebral disc space and vertebral shape profiles. The technique assists in retrieving vertebra pairs that exhibit a specified disc space narrowing and inter-vertebral disc shape. Ribeiro et al. in "Supporting content-based image retrieval and computer-aided diagnosis systems with association rule-based techniques" advocate the use of association rule mining for Content-Based Image Retrieval (CBIR) and Computer-Aided Diagnosis (CAD) systems. Two novel techniques are introduced and studied in the paper; Feature selection through Association Rules (FAR) aimed to improve CBIR systems, and IDEA (Image Diagnosis Enhancement through Association rules) aimed for use in CAD imaging systems. Then, Aydogan et al. in "2D texture based classification, segmentation and 3D orientation estimation of tissues using DT-CWT feature extraction methods" develop and validate four different 2D dual-tree complex DT-CWT wavelet-based methods for feature extraction from texture. Applications to segmenting and classifying tissues are considered; also an interesting application to estimate 3D orientation of tissues is suggested and empirically studied.

The next paper by Soda et al. "On the use of classification reliability for improving performance of the one-per-class decomposition method" addresses the problem of multiclass classification in biomedicine. The authors proposed a new approach that is based on the decomposition method that instead of learning a single multiclass classification learns an ensemble of binary classifiers. The novelty of their approach is in taking into account the quality of the sample under test and the use of classification reliability estimators in combining binary classifiers outputs. The authors experimentally demonstrated that their approach compares favorably with other existing approaches with a set of medical datasets.

Kalyakin et al. in "Extraction of the mismatch negativity elicited by sound duration experiments: A comparison of three procedures" employ advances from the field of biomedical signal processing. The paper compares three techniques for extracting the brain event related potentials (ERPs) – brain responses to stimuli recorded using EEG; conventional difference wave (DW) with average standard sweep, optimal digital filtering (ODF), and independent component analysis (ICA). The responses are studied to learn how the synchronization of brain electrical responses is associated with cognition such as how the brain detects changes in the auditory world.

Next, Soini et al. in "Predicting forensic admission among the mentally ill in a multinational setting: A Bayesian modelling approach" explore protective and risk factors for a forensic admission among the mentally ill. The main contributions of this study include the use of multinational setting, a Bayesian modeling technique, generalized country-independent factors, a merger model with high predictive performance, various measures of predictive performance including relative predictive value, and independent training and test sets.

Last, Gardner and Li Xiong in "An integrated framework for de-identifying unstructured medical data" present a framework and a prototype system for de-identifying health information that may include both structured and unstructured data. The empirical studies in extracting identifying attributes from unstructured data consider a simple Bayesian classifier, a Bayesian classifier with data sampling, and a conditional random field based classifier.

In summary, this special issue presents advances in biomedical data analysis in 7 selected and extended papers from the 21st IEEE International Symposium on Computer-Based Medical Systems, IEEE CBMS 2008. We hope that these papers could be interesting for both academics and practitioners working in the fields of both data mining and biomedicine.

Finally, we would like to thank the authors for their high-quality contributions to this special issue and the referees for their generous and prompt help and valuable remarks and suggestions.

Guest Editors

Mykola Pechenizkiy
Eindhoven University of Technology,
Dept. of Computer Science – Information Systems Group,
P.O. Box 513, 5600 MB Eindhoven,
The Netherlands
Tel.: +31 40 247 4977
E-mail address: m.pechenizkiy@tue.nl

Alexey Tsymbal
Siemens AG,
Corporate Technology,
CT SE 5 SCR2,
Günther-Scharowsky-Str. 1,
91058 Erlangen,
Germany
E-mail address: alexey.tsymbal@siemens.com

Available online 4 July 2009