CrossMark

Healthcare data analytics

**Hui Yang[1] · O. Erhun Kundakcioglu[2] ·
Daniel Zeng[3,4]**

New advances in Information and Communication Technologies (ICT), Operations Research (OR), Artificial Intelligence (AI), and Data Mining (DM) are transforming how healthcare is delivered and managed around the world. Meanwhile, health applications are generating tremendous amounts of data, posing novel technical challenges concerning data processing, knowledge management, analytics, medical decision making, and health policy making, among others. Learning from data also attracts scientists from a large variety of areas such as Mathematics, Statistics, Computer Science, Operations Research, Engineering as well as Physics, Biology, and Medicine. The multidisciplinary approach to healthcare data analytics provides challenging problems and promotes innovative solution methodologies. The articles in this special issue highlight the role that can be played by analytical approaches and powerful solution methodologies for learning from data. These articles also point out challenging problems about data-intensive healthcare applications, which can be tackled by mathematical optimization tools.

✉ Daniel Zeng
zeng@email.arizona.edu

Hui Yang
huy25@engr.psu.edu

O. Erhun Kundakcioglu
erhun.kundakcioglu@ozyegin.edu.tr

[1] The Pennsylvania State University, University Park, PA, USA

[2] Ozyegin University, Istanbul, Turkey

[3] University of Arizona, Tucson, AZ, USA

[4] Chinese Academy of Sciences, Beijing, China

This special issue has three articles on data mining applications in healthcare. The first one, by David Anderson, Bruce Golden, Edward Wasil, and Hao Zhang, proposes a method of diagnosing prostate cancer using magnetic resonance imaging data. A hybrid approach is shown to perform acceptable levels of prediction accuracy and ROC curve. The authors train a logistic regression and a K-nearest neighbor model using MRI data to predict whether a prostate slice has cancer or not. Next, they combine the two methods into an augmented logistic regression model that is shown to outperform both traditional methods. Such information can then be used to determine whether or not a patient should undergo further diagnostic tests, such as a biopsy.

In the second paper by Francesco Folino and Clara Pizzuti, a novel disease prediction mechanism that combines clustering, Markov models, and association analysis is presented. Patient medical records are clustered and a Markov model is generated for each cluster to predict the next disease an individual could incur in the future. Next, association rules are computed for each cluster with certain level of support and confidence. Experimental results show that the combination of models gives better predictive accuracy for new patients.

Kang Zhao, Greta E. Greer, John Yen, Prasenjit Mitra, and Kenneth Portier propose a social network-based classification approach to identify leaders in an online health community for cancer survivors and informal caregivers. The authors generate new neighborhood-based and cluster-based features based on users' features such as contributions, network centralities, and linguistic features. Classification results show that these features can be used for leader identification, which leads to a hybrid approach based on an ensemble classifier.

This issue has one article on data indexing and retrieval. Duy Dinh proposes an algorithm for identifying the most appropriate subdomain of a concept in the context of a document/query. The proposed approach relies on two term sense disambiguation methods for identifying ambiguous terms denoting MeSH concepts: Left-To-Right and Cluster-based. The computational experiments on the OHSUMED corpus show that the proposed approaches based on semantic indexing and retrieval outperform the state-of-the-art approach.

Three papers are from the general area of decision support systems. Christian Wernz, Inga Gehrke, and Daniel R. Ball present the application of real options analysis to a managerial decision-making problem. Despite being commonly used, net present value analysis is shown to lead to a sub-optimal decision, as it does not take into account the value of future options and managerial flexibility. Furthermore, the authors show the capabilities of real options analysis, how it can be applied in practice, data needed to carry out the analysis, and how real options analysis can be integrated into the organizational decision process.

Jennifer Percival, Carolyn McGregor, Nathan Percival, and Andrew James present a prototype architecture for a real-time mobile clinical event data capture application. A pilot study is demonstrated where nursing staff input physiological data streams through an iPod/iTouch based application. This improves the ability of clinical decision support systems and patient monitoring algorithms to detect and adjust for artifacts caused by clinical events. Future research directions are provided to improve the mobile application through increased security, robustness, further

integration into data mining analysis, and future clinical decision support algorithms.

Ivan Gotham, Linh H. Le, Debra L. Sottolano, and Kathryn Schmidt illustrate how a well-established public health informatics framework provides an integrated information system infrastructure. That infrastructure also enhances the efficacy of Public Health Emergency Preparedness actions throughout the phases of the health emergency event life cycle. That is shown to help planning; surveillance; alerting; resource assessment and management; data-driven decision support; and intervention for prevention and control of disease or injury in populations.