Guest Editorial for Special Issue KDD'10

This special issue of the ACM *Transactions on Knowledge Discovery from Data* (TKDD) includes extended versions of six papers selected from the works presented at the 16th ACM SIGKDD *Conference on Knowledge Discovery and Data Mining* (KDD) held in Washington, DC from July 25 to 28 in 2010. Only a few papers among the most highly ranked ones were invited to appear in this special issue. They provide a view of the depth and breadth of many of the most interesting areas in data mining research. The invited articles have gone through the standard refereeing process of TKDD to ensure high publication standards. Each one contains important new material in comparison with its conference predecessor.

This special issue is a sample of recent advances and research trends in predictive and descriptive analytics, including fundamental topics such as scalable classification, recently growing topics such as topic modeling, network analysis and multitask learning, and a novel application in mining news articles.

Topic models and latent factor models have emerged as a vital tool for modeling text and dyadic data. Two articles presented in this issue enhance topic models. Iwata et al. in "Sequential Modeling of Topic Dynamics with Multiple Timescales" propose a model for analyzing the evolution over time of themes in document collections. Topic-specific distributions over words are assumed to be generated based on word distributions from multiple timescales. The model is trained by a randomized expectation-maximization (EM) algorithm, in which parameters are sequentially updated using newly obtained data. Huh and Fienberg in "Discriminative Topic Modeling based on Manifold Learning" present a predictive topic modeling method based on learning spaces of reduced dimension, which incorporates both similarities and dissimilarities among the modeled documents through a regularization term. The resulting method outperforms previous state-of-the-art methodologies in a document classification task.

The increasing publication of digital information on the Web has led to many studies of information diffusion and propagation in networks. Gomez-Rodriguez, Leskovec, and Krause in "Inferring Networks of Diffusion and Influence" investigate the way that information flows across the online media space. They develop a method for tracing paths of propagation through networks and for inferring the underlying networks over which contagion occurs. An efficient approximation algorithm for the task scales to large datasets and in practice gives near-optimal performance.

In the past decade, there has been a growing interest in the problem of multitask learning, where classifier accuracy is enhanced by learning multiple classifiers simultaneously. Chen, Liu, and Ye in "Learning Incoherent Sparse and Low-Rank Patterns from Multiple Tasks" propose a linear algebra formulation of multitask learning, and then convert the problem into an efficient convex surrogate task that is solved by a projected gradient scheme.

Yu et al. in "Large Linear Classification When Data Cannot Fit in Memory" address the juncture between two of the most central issues in data mining: classifier learning and massive data. They describe a linear classifier training algorithm that uses a block minimization method to handle data larger than the memory size. Effectiveness is shown through experiments and by proving the guaranteed convergence of the method.

The process of extracting useful knowledge from large datasets has become one of the most pressing problems in today's society. This goal has motivated Shahaf and Guestrin in "Connecting Two (or Less) Dots: Discovering Structure in News Articles" to provide a method for navigating within a collection of news reports. Given two articles, the article shows how to link them together automatically through a coherent

DOI 10.1145/2086737.2086738 http://doi.acm.org/10.1145/2086737.2086738

chain of other articles. For example, the method can recover a chain of events starting with the decline of home prices in the US as reported in January 2007 and ending with the ongoing debate about healthcare reform.

Many thanks go to the authors for contributing articles of the highest quality, to the reviewers for their excellent professional service, and to the editors of TKDD for making this special issue possible.

—CHARLES ELKAN and YEHUDA KOREN