

Mini Track: 'Data and Process Mining'

Mini Track Chairs:

Selwyn Piramuthu
University of Florida

H. Michael Chung
California State University, Long Beach

The minitrack covers the broad theory and application issues related to data mining, machine learning, knowledge acquisition, knowledge discovery, information retrieval, data base, and inductive decision-making. Both structured and unstructured data repositories including human expert decisions, environmental/normative datasets, large document collections, and web databases are considered. Theoretical and methodological exploration in the previous years motivates us to further investigate the various and richer data and knowledge representation schemes such as Web, multimedia, and geographic data applied to science as well as management domains.

The minitrack is organized within the Decision Technologies for Management track. Papers submitted to this track were reviewed for their content and appropriateness to this minitrack by researchers from around the world working in the general area of data and/or process mining. In total, seven papers were accepted for presentation at the conference and publication in the proceedings. These papers reflect a variety of issues and perspectives in this emerging space.

In their paper, Landauer and Bellman describe continual contemplation and its use in a wrapping-based system that engenders higher-level models of its own behavior. They show how wrappings and problem posing interpretations of programming languages lead to building systems that have complete low-level models of their own behavior.

Fan and Biagiono present an approach to process and interpret data gathered by sensor networks and visualize these data by utilizing database management technology, geographic information system, and Web development technology. They incorporate the beneficial aspects of several different technologies to build their system including global positioning systems, Voronoi diagrams, dynamic Web programming, and asynchronous collaborative discussion environment.

Nandy and Mahanti review existing approaches and propose a heuristic and an algorithm called Iterative Threshold Search for optimal winner determination in combinatorial auctions. They integrate the beneficial features of IDA and DFBB search algorithms depending on the upper-bound heuristic for determining the revenue maximizing set of winning bids. The proposed method runs like IDA when the heuristic is an (admissible) upper-bound and DFBB when the heuristics is an (inadmissible) lower

bound. They show that their heuristic and algorithm result in improvement in performance over what is available in existing literature.

Gao, Paynter, and Sundaram approach problems associated with spatial decision support systems as well as spatial decision-making in general. They present a generic spatial decision-making process and a domain-independent Flexible Spatial Decision Support System framework and architecture to support this process. They illustrate the proposed system on a location problem using Jade as a complete and fully-integrated system for developing the FDSS prototype.

Chen presents a multi-dimensional fuzzy inductive learning method for knowledge acquisition. Chen integrates fuzzy set theory and conventional multi-dimensional decision tree methods in the paper. The proposed method uses crisp multi-dimensional decision trees generated using traditional multi-decision-tree-induction method as a base and then modifies it by a fuzzification method. Results using the proposed method as applied to four different data sets are promising compared to traditional decision tree approach.

Mittermayer uses text-mining techniques incorporating unstructured textual data to predict intraday stock price trends. The system (NewsCATS) learns categorization rules from archives of press releases and intraday trades and quotes. It then uses these rules to categorize new press releases. He finds that categorization of press releases provides additional information to forecast stock price trends.

Su and Lin present an algorithm for frequent itemset mining for generating association rules that does not degrade in performance when dealing with cases with relatively low support values. They compare their results with Apriori algorithm and find their method to be about an order of magnitude faster.

These papers cover a wide variety of topics that span the general area of data and process mining including data visualization, self-monitoring systems, combinatorial auctions, spatial decision-making, text-mining, association rules, and fuzzy decision trees. These papers also illustrate the trend toward real-time use of data and process mining and their managerial implications using methods and examples from disparate domains.