

Toward Optimal Ordering of Prediction Tasks

Yiming Yang

Language Technologies Institute & ML Dept., CMU

<http://www-2.cs.cmu.edu/~yiming/>

Abstract:

In many practical applications, multiple interrelated tasks must be accomplished sequentially through user interaction. The ordering of the tasks may have a significant impact on the overall utility (or performance) of the systems; hence optimal ordering of tasks is desirable. However, manual specification of optimal ordering is often difficult when task dependencies are complex, and exhaustive search for the optimal order is computationally intractable when the number of tasks is large. We present the first attempt at solving the optimal task ordering problem by learning partial order preferences among tasks based on observed system behavior in context, and by reducing the order optimization problem to a well-known Linear Ordering Problem (LOP). For computational tractability of the LOP solution, we further use link analysis (HITS and PageRank) over a partial-order-preference graph as a heuristic approximation. These strategies allow us to find near-optimal solutions with efficient computation, scalable to large applications. We conducted a comparative evaluation of the proposed approach with two practical applications that involve computer-assisted trouble report generation and IT proposal annotation with heterogeneous classification labels (keywords, collaborators, customers, technical categories, etc.), and obtained highly encouraging results in both applications.

Short Bio:

Yiming Yang is a professor in the Language Technologies Institute and the Machine Learning Department within the School of Computer Science at Carnegie Mellon University. She received her Ph.D. in Computer Science from Kyoto University (Japan), and has been a faculty member at Carnegie Mellon University since 1996. Her research has centered on statistical learning methods and their applications to a variety of challenging problems, including automated text categorization, information distillation based on both novelty detection and relevance assessment, joint modeling of heterogeneous sequentially interrelated tasks, user-centric adaptive and collaborative filtering, personalized email filtering and prioritization using social network analysis, and statistical learning from protein/gene expressions in micro-array data and tandem mass spectra, etc.