Positive Unlabeled Learning to Discover Relevant Documents Using Topic Models for Feature Selection

Presenter: Mingzhu Zhu
PhD Candidate- NJIT

Collaborative Knowledge Building in Support of Individual Learning

Presenter: Regina Collins
PhD Candidate- NJIT

DATE: Wednesday October 3, 2012
COFFEE: 2:00-2:30
TIME: 2:30-4:00
LOCATION: GITC Building, Room 1403

Positive Unlabeled Learning to Discover Relevant Documents Using Topic Models for Feature Selection

Mingzhu Zhu

With textual digital information growing exponentially, supervised learning methods have been popularly chosen to help users access personalized information. One problem of the traditional supervised learning methods is that both positive and negative examples are required. However, it is often the case that users provide their information needs with a small set of positive examples but no negative examples. This is a typical scenario where positive and unlabeled learning (PU Learning) is a good alternative to supervised learning. Many PU learning methods exist, but they usually treat a document as a vector, of which the features are keywords in the collections. Such a term-vector based document representation brings high dimensionality problems when the collection is large; or even worse, some noisy features seriously degrade the performance of the learning algorithms. In this research, we propose a framework of using PU learning for relevance feedback using latent topics identified by a topic model for feature dimension reduction. Specifically, we use Latent Dirichlet Allocation (LDA) to reduce the feature dimension of document vectors to a lower dimension of topic vectors. Then the procedure of discovering relevant documents using a PU learning method is conducted in the topic space. Using Mean Average Precision (MAP) and precision at 10 (P@10) measures, experiments on a benchmark dataset indicate that the topic based method has comparable retrieval performances with the term based method in terms of MAP; more importantly, the former outperforms the latter significantly in term of P@10, when the size of positive examples is small. Given that the LDA method has a significantly smaller dimension than the term based method, it is more practical in a relevance feedback setting, where computational efficiency is crucial in providing real time update of search results per the user feedback.

Collaborative Knowledge Building in Support of Individual Learning

Regina Collins

Collaboration has become the norm for knowledge building within organizations and educational communities. Collaborative knowledge building activities are in line with the Web 2.0 culture of participation and sharing that is seen as a prerequisite for success in the global knowledge economy. Existing collaborative knowledge building technologies have provided some level of support for communal learning activities, yet these systems neglect the fact that the ultimate responsibility for learning and contributing to knowledge rests on the individual learner. This research proposes a hybrid learning model that scaffolds communal knowledge building and sensemaking activities while simultaneously supporting individual learner reflection, exploration, and internalization of knowledge.