

Limin Yao

CONTACT INFORMATION	Department of Computer Science University of Massachusetts Amherst, MA 01002 USA	<i>Cell:</i> (413) 835-1423 <i>E-mail:</i> lmyao@cs.umass.edu <i>WWW:</i> www.cs.umass.edu/~lmyao
RESEARCH INTERESTS	Semi-supervised learning, Information extraction (especially relation extraction), and Natural language processing	
EDUCATION	University of Massachusetts , Amherst, MA USA Ph.D. Student, Computer Science, August 2008 (started) <ul style="list-style-type: none">• Advisor: Andrew McCallum Tsinghua University , Beijing, China M.E., Computer Science, July, 2008 <ul style="list-style-type: none">• Advisor: Jie Tang Xi'an Jiaotong University , Xi'an, Shaanxi China B.E., Computer Science, July, 2005	
PROFESSIONAL EXPERIENCE	<i>Research Assistant</i> at University of Massachusetts , Amherst, MA <i>Intern</i> in IR Group, Microsoft Research , Cambridge, UK <i>Intern</i> in NLP Group, Microsoft Research Asia , Beijing, China	August, 2008 - present Jun. 2009 – Aug. 2009 Aug. 2006 – Jan. 2007
	<ul style="list-style-type: none">• Mentors: Bodo Billerbeck and Martin Szummer• Mentors: Mu Li and Changning Huang	
SELECTED PROJECTS	Unsupervised Relation Discovery with Sense Disambiguation Most research methods cluster shallow or syntactic patterns to discover relation types from text, but consider only one possible sense per pattern. In practice this assumption is often violated. For example, “A play B” has multiple senses. “Kate Winslet played Rose” is different from “New York Giants played New England Patriots”. We employ a topic model for pattern sense disambiguation and use hierarchical agglomerative clustering for semantic relations discovery. (To appear in <i>ACL2012</i>)	UMass, Amherst, MA Jul. 2011 – Jan. 2012
	Unsupervised Relation Discovery using Generative Models One relation can be expressed using various particular languages. For example, “X was born in Y, “X, a native of Y, and “X grew up in Y share the same <i>bornIn</i> relation. We present a series of generative models to discover various expressions of relations. We also propose one model to jointly discover relations and the two argument types where they occur. (Published in <i>EMNLP2011</i>)	UMass, Amherst, MA Oct. 2010 – Jul. 2011
	Joint Entity and Relation Extraction without Labeled Data External knowledge base can guide relation extraction without labeled data. We match Freebase facts to a text corpus to “distantly” label the relation instances. Certain relations can only hold	UMass, Amherst, MA Jan. 2010 – Jul. 2010

between certain types of entities, known as “selectional preference”. For example, a *person* can be a director of a *company*, but not of a *location*. We propose a conditional random field model to take this information into account. Our model predicts fine grained entity types and relations jointly. The intuition is that entity and relation prediction can reinforce each other. (Published in *EMNLP2010*)

Constraint-driven Learning for Information Extraction

UMass, Amherst, MA

Nov. 2009 – Jan. 2010

There is plentiful unlabeled data for information extraction tasks, such as segmenting and labeling citations. We incorporate our prior knowledge about the task as constraints on unlabeled data and use a self-training framework to do semi-supervised learning.

Query Alteration

Microsoft Research, Cambridge, U.K.

Summer Intern

Jun. 2009 – Aug. 2009

In information retrieval, the query is crucial for finding the expected results. Users’ queries may not be accurate enough to rank relevant documents highly. We worked on helping users find alternative queries which can retrieve more relevant and accurate documents. We generate candidates for a query and use a classifier to determine whether the alternative query should be used.

Efficient Inference for Topic Models

Univ. of Massachusetts, Amherst MA

Sep. 2008 – Mar. 2009

Topic models are powerful for analyzing large text collections by representing high dimensional data in a low dimensional subspace. Training topic models is computationally expensive. Our idea is to train a topic model on a subset and infer topic distributions for new documents without retraining the model. We empirically evaluate the performance of several methods, Gibbs sampling based methods and a text classification based method. The classification-based method produces results similar to iterative sampling methods, but requires much less computation. (Published in *KDD2009*)

Topic Models for Document Summarization

Tsinghua University, Beijing China

Nov. 2007 – May 2008

Topic models are applied to query oriented document summarization. We score a sentence using the probability that it is generated from a topic model. We rank the sentences and pick the top ones as the summary.

Researcher Profiling

Tsinghua University, Beijing China

Jan. 2007 – Oct. 2007

Researcher profiling aims at extracting all profile information of a researcher from his/her relevant web pages. This work is part of our system (www.arnetminer.org) which aims to build a social network of academic researchers. We proposed a unified approach to researcher profiling, formalized the task as sequence labeling. We utilized Conditional Random Fields and the model can capture the dependencies between tags, offering advantages over other models. (*WI2007, TKDE2010*)

SELECTED
PUBLICATIONS

Limin Yao, Sebastian Riedel, and Andrew McCallum. Unsupervised Relation Discovery with Sense Disambiguation. In *Proc. of ACL*, 2012. (to appear)

Limin Yao, Aria Haghighi, Sebastian Riedel, and Andrew McCallum. Structured Relation Discovery using Generative Models. In *Proc. of EMNLP*, 2011.

Limin Yao, Sebastian Riedel, and Andrew McCallum. Collective Cross-Document Relation Extraction without Labelled Data. In *Proc. of EMNLP*, 2010.

Sebastian Riedel, Limin Yao, and Andrew McCallum. Modeling Relations and Their Mentions without Labeled Text. In *Proc. of ECML/PKDD*, 2010.

Sameer Singh, Limin Yao, Sebastian Riedel, and Andrew McCallum. Constraint-driven Rank based Learning for Information Extraction. In *Proc. of NAACL*, 2010.

Jie Tang, Limin Yao, Duo Zhang, and Jing Zhang. A Combination Approach to Web User Profiling. In *ACM Transactions on Knowledge Discovery from Data*, 2010.

Limin Yao, David Mimno, and Andrew McCallum. Efficient Methods for Topic Model Inference on Streaming Document Collections. In *Proc. of SIGKDD*, 2009

Jie Tang, Limin Yao, and Dewei Chen. Multi-topic based Query-oriented Summarization. In *Proc. of SDM*, 2009.

Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. Extraction and Mining of Academic Social Network, In *Proc. of SIGKDD*, 2008.

Limin Yao, Jie Tang, and Juanzi Li. The Entire Solution Path for Support Vector Machine in Positive and Unlabeled Classification. In *Journal of Tsinghua Science and Technology*, 2008.

Jie Tang, Duo Zhang, and Limin Yao. Social Network Extraction of Academic Researchers. In *Proc. of ICDM*, 2007.

Limin Yao, Jie Tang, and Juanzi Li. A Unified Approach to Researcher Profiling. In *Proc. of ACM/IEEE Web Intelligence*, 2007.

Limin Yao, Mu Li, and Changning Huang. Improving Chinese Chunking with Enriched Statistical and Morphological Knowledge. In *Proc. of NLP-KE*, 2007.

PROGRAMMING

Java, Scala, Perl, C, C++, R, Matlab

HONORS AND AWARDS

KDD Student Travel Award and Google Diversity Award, 2009

First National Prize for excellent performance in study (top 5%)

Excellent Graduate Student Award in Shaanxi Province (top 5%)

First Prize in Mathematical Competition in Shaanxi Province (top 10%)

Enterprise Scholarship sponsored by China National Petroleum Company (top 5%)

First class scholarship of Xi'an Jiaotong University (top 8%)

PERSONAL

Interested in blogging (<http://yln82674826.wordpress.com/>), hiking, traveling, reading, and contra dance