Yaqin Yang

CONTACT
INFORMATION

Tel: (781) 654-5811
E-mail: yaqin276@gmail.com
http://cs.brandeis.edu/~yaqin

EXPERTISES

Machine Learning, Natural Language processing, Language Modeling, Information Extraction

EDUCATION

Ph.D., Computer Science (GPA: 3.8/4.0)   Aug. 2009 – Jul. 2014

Brandeis University, Waltham, Massachusetts

B.S. and M.S., Computer Science   Aug. 2002 – Jun. 2009

Beijing University of Posts and Telecommunications, Beijing, China

SKILLS AND
KNOWLEDGE

Python, Java, Pig, Map/Reduce, R, SQL, Linux shell scripting

PROJECTS

Supervised Collection Ranking

Data Scientist, Paypal, *San Jose, CA*   Oct. 2014 – Present

- Modeling and data analysis: developed an end to end model to predict the collectabilities of accounts in negative balance;
- Data pipeline: Clean up and prepare massive data for modeling;
- Feature engineering and selection.

Unsupervised Anti Money Laundering Detection

Data Scientist, Paypal, *San Jose, CA*   Aug. 2014 – Oct. 2014

- Developed a pipeline to extract data set, features to detect anomalies in transactions by using unsupervised machine learning algorithms.

Richer Representations for Machine Translation

Research Assistant, Brandeis University, *Waltham, MA*   Nov. 2009 – Jul. 2014

Project focus is to enhance the performance of Chinese and English machine translation by providing the MT model richer knowledge from both source and target language sides. From the source language side (Chinese), foci are to extract and recover implicit information, such as dropped pronouns and discourse relations.

- Developed a probabilistic approach to detect Empty Categories in Chinese;
- Developed a sentence segmentation method to simplify long Chinese sentences into short independent sentences, using machine learning methods;
- Developed a sentence level discourse segmentation and relation detection method, using machine learning methods.

Scientific Trend Prediction   May. 2013 – Aug. 2013

Intern, Raytheon BBN Technologies, *Cambridge, MA*

Project foci were to automatically predict emerging technologies and influential authors by analyzing a large set of scientific papers and patents.

- Built a database to store information extracted from around 5 million Chinese patents and 5 million English patents submitted/granted over the past twenty years in various fields including Chemistry, Electricity etc;
- Designed and implemented features of scientific terms and significant patent applicants;
- Built regression models to predict the emerging important scientific trends and applicants.

Semantics-Based Machine Translation   Jun. 2012 – Aug. 2012

Visiting Researcher, USC/Information Sciences Institute, *Marina del Rey, CA*

Project foci were to automatically annotate parallel text with Abstract Meaning Representation (AMR), a type of semantic structure, – Natural Language Understanding (NLU); to design automata to map these structures from/to syntactic trees in source/target languages; to generate English language from AMR – Natural Language Generation (NLG), etc.

- Experimented MERT tuning for a new NLG system (AMR to English);
- Automatically extracted rules that substantially improve NLG (morphological rules, and semantics-to-syntax mapping rules);
- Analyzed of re-entrancies in AMR graph structure (breakdown of how many, and causes);
- Designed of an NLU pipeline (English to AMR).

Motorola Advanced Location Services   Jul. 2007 – Oct. 2007

Intern, Global Software Group, Motorola, *Beijing, China*

- Evaluated the performance of Assisted Global Positioning System (AGPS) with high precision and Location Based Services Platform (LSP) with low precision, respectively.;
- Analyzed the efficiency of core functions of AGPS/LSP under UNIX environment, and the datasets collected from several provinces using Motorolas AGPS and LSP;
- Designed test cases according to DoCoMos functional requirements and non-functional requirements including reliability, portability, and exception handling.

PATENTS AND PUBLICATIONS

Yaqin Yang and Fransisco Kurniadi. 2015. Server Device Configurations Based On Machine Learning. United States Patent P3107US1.

Yaqin Yang, Yalin Liu and Nianwen Xue. 2015. Recovering dropped pronouns from Chinese text messages. Proceedigns of ACL. Beijing, China.

Nianwen Xue and **Yaqin Yang**. 2013. Dependency-based empty category detection via phrase structure trees. Proceedings of NAACL HLT. Atlanta, Georgia.

Nianwen Xue, Yuchen Zhang and **Yaqin Yang**. 2013. Distant annotation of Chinese tense and modality. Proceedings of the Workshop on Annotation of Modal Meaning in Natural Language (WAMM). Potsdam, Germany.

Yaqin Yang and Nianwen Xue. 2012. Chinese comma disambiguation for discourse analysis. Proceedings of ACL. Jeju, Republic of Korea.

Elizabeth Baran, **Yaqin Yang** and Nianwen Xue. 2012. Annotating dropped pronouns in Chinese newswire text. Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC). Istanbul, Turkey.

Peter Anick, Pengyu Hong, Nianwen Xue, **Yaqin Yang**. 2011. Coreference Resolution for Electronic Medical Records. Fifth i2b2/VA/Cincinnati Shared-Task and Workshop Challenges in Natural Language Processing for Clinical Data. Washington, DC.

Yaqin Yang, Nianwen Xue and Peter Anick. 2011. A machine learning-based coreference detection system for OntoNotes. Procceedings of CoNLL (Shared Task). Portland, Oregon.

Nianwen Xue and **Yaqin Yang**. 2011. Chinese sentence segmentation as comma classification. Proceedings of ACL. Portland, Oregon.

Yaqin Yang and Nianwen Xue. 2010. Chasing the ghost: recovering empty categories in the Chinese Treebank. Proceedings of COLING. Beijing, China.

HONOR AND AWARDS	**Spot Award** Ebay Inc.	Jan. 2015
	Fellowship Brandeis University	Fall 2009 - Fall 2014
	Outstanding Graduate Student Award Beijing University of Posts and Telecommunications, Beijing, China	Fall 2008
	BUPT Scholarship Beijing University of Posts and Telecommunications, Beijing, China	Fall 2003, 2004, 2005
	BUPT Merit Student Beijing University of Posts and Telecommunications, Beijing, China	Fall 2003

TEACHING EXPERIENCE	**Teaching Assistant**, Internet and Society	Spring 2012
	Teaching Assistant, Human Computer Interaction	Fall 2011
	Teaching Assistant, Information Extraction Systems	Fall 2009 Spring 2010
	Teaching Assistant, Statistical Approaches to Natural Language Processing	Fall 2010
	Teaching Assistant, Algorithms to Computational Linguistics	Spring 2010