Xi Victoria Lin

CONTACT
INFORMATION

Ph.D. Student
Department of Computer Science and Engineering
University of Washington
AC486 Paul G. Allen Center
185 Stevens Way
Seattle, WA 98195-2350, USA

Mobile: +1-206-519-8086
EM: xilin@cs.washington.edu

WWW: http://homes.cs.washington.edu/ ˜xilin

RESEARCH
INTERESTS

Natural language processing, machine learning and knowledge representation.

EDUCATION

University of Washington, Seattle, WA
Ph.D. candidate, Computer Science and Engineering, June 2013 - *present*
- Adviser: Professor Luke Zettlemoyer (November 2013 - *present*)
  Professor Benjamin Taskar (June 2013 - November 2013)
- Research Interest: statistical relational learning in natural language processing and its application to web-scale information extraction.

University of Pennsylvania, Philadelphia, PA
Ph.D. graduated as M.Sc., Computer and Information Science, May 2013
- Cumulative GPA: 3.86/4.0
- Adviser: Professor Benjamin Taskar
- Area of Study: Natural Language Processing and Machine Learning

University of Oxford, Kellogg College, Oxford, UK
M.Sc., Computer Science, September 2011
- Thesis Topic: *Fine-grained Named Entity Classification in Machine Reading*
- Adviser: Professor Stephen Pulman
- Area of Study: Computer Science focused on Logic and Semantics

The Hong Kong Polytechnic University, Kowloon, HK
B.S., Electronic and Information Engineering, August 2010
- Cumulative GPA: 4.0/4.0
- Thesis Topic: *Robust Real-time Face Detection*
- Adviser: Professor Kenneth K. M. Lam
- Minor in Computer Science (programming and algorithms)

RESEARCH
PROJECTS

Cross-lingual Distant Supervision for Relation Extraction.
Joint work with Dr. Luke Zettlemoyer and Dr. Sameer Singh
Distant supervision is the paradigm that creates its own training data by heuristically matching the contents of a database to corresponding text. Such paradigm can be naturally extended to a multi-lingual setting since the same fact can be expressed in multiple languages. The widely used common-sense knowledge base, *Freebase*, stores multi-lingual aliases for its entities. A naive approach is hence to use the aliases as anchor texts for mapping relation tuples to text, and train a relation extractor for each language separately. However, for a long tail of Freebase entities, either the foreign translation is missing or they are never mentioned in a foreign language, which makes the distant supervision signal much weaker in many other languages than in English. In this project, we propose bilingual projection for strengthening the distant supervision signal in the secondary language, Chinese. The assumption is that text patterns which are translations to each other should express the same set of relations. On the other hand, patterns that express the same relations are likely to be translations of each other. We are currently investigating the effect of adding

the above assumptions to the distant supervision model. (Spring 2014 - *present*)

Knowledge Extraction for Elementary Physics Question Answering.
Joint work with Dr. Tom Kwiatkowski (Summer 2014)

Multi-label Learning with Posterior Regularization.
Joint work with Dr. Sameer Singh, Luheng He, Dr. Benjamin Taskar and Dr. Luke Zettlemoyer
In many multi-label learning problems, especially as the number of labels grow, it is challenging to gather completely annotated data. In this work, we present a new approach for multi-label learning from incomplete annotations. Specifically, it uses the posterior regularization technique to enforce soft constraints on the output of inductive classifiers, regularizing them to prefer sparse and low-rank (SLR) predictions. By avoiding strict low-rank constraints we enable a learned trade-off leading to better generalization of the classifiers. The resulting joint objective can be optimized efficiently using EM and stochastic gradient descent. Experiments in both image and text domains demonstrate the contributions of each modeling assumption and show that the approach achieves state-of-the-art performance on a number of challenging datasets. (Fall 2013 to Spring 2014)

Aspect-Oriented Event Summarization based on Wiki-Templates.
Supervised by Dr. Benjamin Taskar.
(Winter 2013 to Summer 2013).

Named Entity Linking with Dual Decomposition.
Supervised by Dr. Benjamin Taskar.
(Fall 2012).

What Do They Say About It: Summarizing Quotations in News.
Supervised by Dr. Benjamin Taskar.
(Summer 2012).

Bullet Points: Structured Summarization for News Articles.
Supervised by Dr. Benjamin Taskar.
(Spring 2012 to Summer 2012).

Automatic News Summarization.
Supervised by Dr. Ani Nenkova.
Extractive document summarization deals with the problem of selecting a subset of sentences from the original document to form an informative summary, subjecting to certain length constraints. This problem can be formed as a global optimization, where "information" and "redundancy" are quantified and the objective is to find a subset of text units that maximize "information" while minimize "redundancy". It has been shown that finding the solution for this optimization problem is NP-hard. We studied a series of state-of-the-art methods for this problem and found that greedily optimizing the KL-divergence between the word distributions of the input and the summary performs as well as the optimization-based methods in terms of ROUGE score. We also noticed that in general extractive methods are good at identifying the overview sentences, but they are difficult to be extended to capture the semantic structures of both the input and the summary. For example, the input of multi-document news summarization is often centered around an event or a series of events, and each event contains its own important aspects. For future work we plan to investigate information extraction techniques for summarization as well as authomatically generating novel presentations of the extracted information. (Fall 2011 to Spring 2012).

Fine-grained Named Entity Classification in Machine Reading.
Supervised by Dr. Stephen Pulman.
Psycholinguistic study suggests that our lexical memory organizes named entities and common nouns in a taxonomic hierarchy. The task of fine-grained named entity classification intends to classify the set of named entities in text into entry-level categories. By "entry-level categories" we mean the categories used by humans for understanding the article. For example, besides knowing an entity appeared in text is a *person*, readers are more interested in knowing whether it is a *politician*, a *journalist* or an *actor*. We focus specifically on the *person* domain, where we found that in news reports, a single document is often enough for deciding the right type. We formulate the problem as searching for "is-a" relations between a person and an occupation, and develop a bootstrapping-style classifier which jointly discovers new person/occupation entities as well as new "is-a" relation tuples. The classifier performs competitively on manually annotated sentences. (Spring 2011 to Summer 2011).

PUBLICATIONS	[1] Lin, X.V., Singh, S., He, L., Taskar, B., and Zettlemoyer, L. Multi-label Learning with Posterior Regularization. In *NIPS Modern ML+NLP: Workshop on Modern Machine Learning & Natural Language Processing*, December 08–12, 2014.

[2] Lin, X.V. Fine-grained Named Entity Classification in Machine Reading. *M.Sc. thesis*. Oxford University. 2011.

INTERNSHIPS **Allen Institute for Artificial Inteliigence**, Seattle, WA. (Summer 2014)

TEACHING
EXPERIENCE **University of Pennsylvania**, Philadelphia, PA

Teaching Assistant
- CIS520: Machine Learning Fall 2012
  - Making exam questions; answering Piazza questions; office hours; grading.

HONOURS AND
AWARDS **University of Pennsylvania**
- CIS Department Doctorate Fellowship, 2011–2013.

The Hong Kong Polytechnic University
- Best Academic Performance Award, EIE Department, 2009–2010
- Hong Kong SAR Government Scholarship, 2009-2010
- Hong Kong Polytechnic University Post-entry Scholarship, 2008–2009
- Hong Kong & Kowloon Electrical Appliances Merchants Association Scholarship, 2008–2009
- Apple Inc. WWDC Student Scholarship, 2009

REFERENCES
AVAILABLE TO
CONTACT **Dr. Luke Zettlemoyer** (e-mail: lsz@cs.washington.edu)
- ⋆ *Ph.D. research adviser*
- Assistant Professor , Computer Science and Engineering, University of Washington
- ◇ 185 Stevens Way, Seattle, WA 98195-2350

Dr. Benjamin Taskar (1977-2013)
- ⋆ *former Ph.D. research adviser*
- Boeing Associate Professor , Computer Science and Engineering, University of Washington

Dr. Stephen Pulman (e-mail: sgp@clg.ox.ac.uk; phone: +44-186-561-0800)
- ⋆ *M.Sc. thesis advisor*
- FBA Professor, Computational Linguistics, University of Oxford
- ◇ Wolfson Building, Parks Road, Oxford OX1 3QD

Dr. Kenneth K.M. Lam (e-mail: enkmlam@polyu.edu.hk; phone: +852-2766-6207)
- ⋆ *B.Eng. thesis advisor*
- • Professor, Electronic and Information Engineering, The Hong Kong Polytechnic University
- ⋄ Room DE503c, EIE Department, HKPU, Hung Hom, Kowloon, Hong Kong.