

Syllabus for LING 575
Unsupervised learning: a case study on unsupervised POS tagging
Winter 2008

Class time & location: T 3:30-5:50pm, MGH 284

Instructor: Fei Xia
Office: Padelford A-210G
Office Hours: Fr: 10:30-11:30am
Phone: (206) 543-9764
Email: fxia at u.washington.edu
(include "Ling575" in the subject line).

Website:
<http://courses.washington.edu/ling575x>

Course Description:

The existing work on unsupervised POS tagging can be divided into three categories: the first starts with the forward-backward algorithm and improves performance by using a “filtered” lexicon; the second clusters words and tags the words with the cluster labels; the third takes advantage of bilingual data by projecting POS information from one language to the other. In this course, we will discuss each approach and build systems aiming at improving the state of the art.

Required Textbooks:

None. Reading materials are online.

Some background reading can be found in

“Foundations of statistical natural language processing” (**M&S**) by Manning and Schütze and/or

Speech and Language Processing (**J&M**) by Jurafsky and Martin (the Table of Content is available at

<http://www.cs.colorado.edu/~martin/slp2.html#top>)

Prerequisites:

- CS 326 (Data Structures) or equivalent
- Stat 391 (Prob. and Stats for CS) or equivalent
- Programming in Perl, C, C++, Java, or Python
- LING 570

- LING 572

Grading:

- **Assignments (45-60%):** there will be four assignments plus a final report. All the assignments are due at 2:30pm on Tuesday.
- **Presentation (15-25%):** each student will present one or two papers and give a final presentation.
- **Reading (5-10%):** it will be checked but not graded.
- **Class participation (10-20%)**

Course policies:

- **Course website:** All course information on this web page is tentative and can change at any time. Confirm crucial dates or information with me in person during class.
- **Attendance:** Students are expected to attend all classes: announcements about assigned readings and assignments will be made available at the start of each class. Such announcements may not be made on this web page, so don't rely on information here instead of attending class.
- **Late assignment submission:** There will be a **1% penalty for every hour** after the deadline. For instance, suppose the assignment is due at 2:30pm and you turn in the assignment at 2:30am the next morning, your grade would be $x * 0.88$, where x is the grade you would get if you have turned in before the 1pm deadline. **No assignments will be accepted four (4) days after the due date.**
- **Reading: You need to read papers before class starts.** Write your questions on a piece of paper and bring it to class.
- **"Incomplete":** According to UW policy, "incomplete grades may only be awarded if you are doing satisfactory work up until **the last two weeks** of the quarter." Therefore, it is crucial for you to hand in your homework on time. An "incomplete" grade is given only under extremely unusual circumstances (e.g., health issues, family emergency).
- **Emails: Use the prefix "ling575: " on all the subject line of your messages.** If you do not include the prefix, then the mail might go unanswered. If you don't receive a reply from me within 48 hours, please send me a reminder.

Tentative schedule (subject to changes):

Week	Date	Topic	Reading	Hw	Due
1	1/8	Introduction Recap on HMM : by Will	M&S 9.3	Hw1	
2	1/15	EM approach (I): by Fei <ul style="list-style-type: none"> • Forward-background algorithm 	M&S 9.3.3		
3	1/22	EM approach (II) by Fei <ul style="list-style-type: none"> • Will EM help? 	(Merialdo, 1994) (Elworthy, 1994)	Hw2	Hw1
4	1/29	EM approach (III) : by Alex <ul style="list-style-type: none"> • Using lexicon (w/ or w/o filtering) • Constraint EM 	(Banko and Moore, 2004) (Wang and Schuumans, 2005)		
5	2/5	Clustering approach: by Fei	(Schutze, 1995) (Clark, 2000)		
6	2/12	Prototype learning : by Ryan	(Haghighi and Klein, 2006)		
7	2/19	Cross-lingual approach (I) : by Ankit	(Yarowsky and Ngai, 2001) (Cucerzan and Yarowsky, 2002)		Hw2
8	2/26	Bayesian approach	(Blei et al.,	Hw3	

		(I): by Kristina Toutanova from Microsoft	2003 (Toutanova and Johnson, 2007)		
9	3/4	Cross-lingual approach (II) by Fei	(Hana et al., 2004) (Feldman et al., 2006)		
10	3/11	Bayesian approach (II) by Fei	(Johnson, 2007) (Goldwater and Griffiths, 2007)		Hw3 (3/19)

Reading material: (*: papers discussed in class)

- EM approach:
 - [\(Baum et al., 1970\)](#) L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statistics*, 41(1):164–171, 1970.
 - [\(Cutting et al., 1992\)](#) D. Cutting, J. Kupiec, J. Pedersen, and P. Sibun. A practical part-of-speech tagger. In *Proc. of the 3rd on Applied Natural Language Processing (ANLP 1992)*, 1992.
 - [\(Elworthy, 1994\)*](#) David Elworthy. Does Baum-Welch Re-estimation :Help Taggers? In *Proc. of ANLP*, 1994.
 - [\(Merialdo, 1994\)*](#) B. Merialdo. Tagging English text with a probabilistic model. *Computational Linguistics*, 20(2), 1994.
 - [\(Banko and Moore, 2004\)*](#) M. Banko and R. C. Moore. Part of Speech Tagging in Context. In *Proc. of the 20th International Conference on Computational Linguistics (Coling 2004)*, pages 556–561, Geneva, Switzerland, 2004.
 - [\(Smith and Eisner, 2005\)](#) N. A. Smith and J. Eisner. Contrastive Estimation: Training Log-Linear Models on Unlabeled Data. In *Proc of the 43th Annual Conference of the Association for Computational Linguistics (ACL-2005)*, Ann Arbor, Michigan, 2005.
 - [\(Wang and Schuurmans, 2005\)*](#) Q. I. Wang and D. Schuurmans. Improved Estimation for Unsupervised Part-of-Speech Tagging. In *Proc. of 2005 IEEE*

International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE 2005), 2005.

- Clustering approach:
 - ([Schutze, 1995](#))* H. Schutze. Distributional part-of-speech tagging. In Proc. of the EACL, pages 141–148, 1995.
 - ([Clark, 2000](#))* A. Clark. Inducing syntactic categories by context distributional clustering. In Proc. of the 4th Conference on Computational Natural Language Learning (CoNLL-2000), pages 91–94, 2000.
 - ([Clark, 2003](#)) A. Clark. Combining distributional and morphological information for part of speech induction. In Proc. of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2003), 2003.
 - ([Biemann, 2006](#)) C. Biemann. Unsupervised part-of-speech tagging employing efficient graph clustering. In Proceedings of the COLING/ACL 2006 Student Research Workshop, pages 7–12, Sydney, Australia, July 2006. Association for Computational Linguistics.
 - ([Dasgupta and Ng, 2007](#)) S. Dasgupta and V. Ng. Unsupervised part-of-speech acquisition for resource-scarce languages. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pages 218–227, 2007.
- Bayesian approach:
 - ([Blei et al., 2003](#))* Latent Dirichlet Allocation. Journal of Machine Learning Research (3), 2003.
 - ([Toutanova and Johnson, 2007](#))* A Bayesian LDA-based model for semi-supervised part-of-speech tagging, NIPS 2007.
 - ([Johnson, 2007](#))* Why doesn't EM find good HMM POS-taggers? EMNLP 2007.
 - ([Goldwater and Griffiths, 2007](#))* A fully Bayesian approach to unsupervised POS tagging. ACL 2007.
- Cross-lingual approach:
 - ([Yarowsky and Ngai, 2001](#))* D. Yarowsky and G. Ngai. Inducing multilingual pos taggers and np bracketers via robust projection across aligned corpora. In Proc. of the 2001 Meeting of the North American chapter of the Association for Computational Linguistics (NAACL-2001), pages 200–207, 2001.
 - ([Cucerzan and Yarowsky, 2002](#))* S. Cucerzan and D. Yarowsky. Bootstrapping a multilingual part-of-speech tagger in one person-day. In Proc. of the 6th Conference on Computational Natural Language Learning (CoNLL-2002), 2002.

- [\(Hana et al., 2004\)*](#) J. Hana, A. Feldman, and C. Brew. A resource-light approach to Russian morphology: Tagging Russian using Czech resources. In Proc. of EMNLP 2004, Barcelona, Spain, 2004.
- [\(Xi and Hwa, 2005\)](#) C. Xi and R. Hwa. A backoff model for bootstrapping resources for non-English languages. In Proc. of HLT-EMNLP, pages 851–858, Vancouver, British Columbia, Canada, 2005.
- [\(Feldman et al., 2006\)*](#) A. Feldman, J. Hana, and C. Brew. Experiments in morphological annotation transfer. In A. Gelbukh, editor, Proceedings of Computational Linguistics and Intelligent Text Processing (CICLing), Lecture Notes in Computer Science. Springer-Verlag, 2006.
- Prototype learning:
 - [\(Haghighi and Klein, 2006\)*](#) A. Haghighi and D. Klein. Prototype-driven learning for sequence models. In Proceedings of the Human Language Technology Conference of the NAACL (HLT/NAACL 2006), pages 320–327, New York City, USA, June 2006. Association for Computational Linguistics.