

## References

- Abney, S. 2002. Bootstrapping. In *Proceedings of ACL-2002*.
- Abney, S. 2004. Understanding the Yarowsky Algorithm. *Computational Linguistics*, 30(3).
- Altun, Y., D. Mcallester, and M. Belkin. 2005. Maximum margin semi-supervised learning for structured variables. In *Proceedings of Neural Information Processing Systems*.
- Ando, R. and T. Zhang. 2004. A Framework for Learning Predictive Structures from Multiple Tasks and Unlabeled Data. Technical report, IBM T.J. Watson Research Labs.
- Ando, R. and T. Zhang. 2005. A High-Performance Semi-Supervised Learning Method for Text Chunking. In *Proceedings of the Association for Computational Linguistics*.
- B. Taskar, C. Guestrin and D. Koller. 2003. Max-margin markov networks. In *NIPS*.
- Bacchiani, M., M. Riley, B. Roark, and R. Sproat. 2006. MAP Adaptation of stochastic grammars. *Computer Speech and Language*, 20(1):41–68.
- Balcan, M. and A. Blum. 2005. A PAC-style Model for Learning from Labeled and Unlabeled Data. In *Proceedings of Computational Learning Theory*.
- Balcan, N., A. Blum, and K. Yang. 2004. Co-Training and Expansion: Towards Bridging Theory and Practice. In *Proceedings of Neural Information Processing Systems*.
- Banko, M. and E. Brill. 2001. Scaling to Very Large Corpora for Natural Language Disambiguation. In *Proc. 39th ACL and 10th EACL*, Toulouse, France, July 9–11.
- Bartlett, P., M. Collins, B. Taskar, and D. McAllester. 2004. Exponentiated gradient algorithms for large-margin structured classification. In *Proceedings of Neural Information Processing Systems*.

- Barzilay, R. and K. R. McKeown. 2001. Extracting paraphrases from a parallel corpus. In *Proc of EACL/ACL-2001*.
- Becker, S. 1995. JPMAX: Learning to recognize moving objects as a model fitting problem. In *Advances in NIPS 7*, pages 933–940.
- Becker, S. and G. E. Hinton. 1992. A self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, 355:161–163.
- Belkin, M., P. Niyogi, and V. Sindhwani. 2004. Manifold Regularization: a Geometric Framework for Learning from Examples. Technical report, University of Michigan.
- Belkin, M., P. Niyogi, and V. Sindhwani. 2005. On Manifold Regularization. In *AISTAT*.
- Bennett, K. P. and A. Demiriz. 1998. Semi-Supervised Support Vector Machines. In *NIPS 12*, pages 368–374.
- Bennett, K. P., A. Demiriz, and R. Maclin. 2002. Exploiting Unlabeled Data in Ensemble Methods. In *Proc. of SIGKDD-2002*, Edmonton, Alberta.
- Blum, A. and S. Chawla. 2001. Learning from labeled and unlabeled data using graph mincuts. In *Proc. 19th International Conference on Machine Learning (ICML-2001)*.
- Blum, A. and T. Mitchell. 1998. Combining Labeled and Unlabeled Data with Co-Training. In *Proceedings of Computational Learning Theory*.
- Bousquet, O., O. Chapelle, and Matthias Hein. 2004. Measure based Regularization. In *Proceedings of Neural Information Proccessing Systems*.
- Brefeld, U., C. Büscher, and T. Scheffer. 2005. Multi-view discriminative sequential learning. In *European Conference on Machine Learning*.
- Brefeld, U. and T. Scheffer. 2004. Co-EM Support Vector Learning. In *Proc of 21st ICML*, Banff, Canada.
- Brefeld, U. and T. Scheffer. 2006. Semi-supervised learning for structured output variables. In *International Conference on Machine Learning*.

- Brill, E. 1997. Unsupervised learning of disambiguation rules for part of speech tagging. In *Natural Language Processing Using Very Large Corpora*. Kluwer Academic Press.
- Callison-Burch, C. 2002. Co-training for statistical machine translation. Master’s thesis, School of Informatics, University of Edinburgh.
- Callison-Burch, C., D. Talbot, and M. Osborne. 2004. Statistical machine translation with word- and sentence-aligned parallel corpora. In *Proc. ACL 2004*.
- Castelli, V. 1994. *The Relative Value of Labelled and unlabelled Samples in Pattern Recognition*. Ph.D. thesis, Stanford University.
- Castelli, V. and T. M. Cover. 1995. On the exponential value of labeled samples. *Pattern Recogn. Lett.*, 16(1):105–111.
- Chapelle, O., J. Weston, and B. Schoelkopf. 2003. Cluster kernels for semi-supervised learning. In *Proceedings of Neural Information Processing Systems*.
- Chapelle, O. and A. Zien. 2005. Semi-Supervised Classification by Low Density Separation. In *AISTAT*.
- Chapelle, O., A. Zien, and B. Scholkopf, editors. 2006. *Semi-Supervised Learning*. MIT Press.
- Chappelle, O., B. Schölkopf, and A. Zien. To appear. Semi-supervised learning. Unpublished book. MIT Press.
- Charniak, E. 1997. Statistical parsing with a context-free grammar and word statistics. In *Proceedings of the 14th National Conference on Artificial Intelligence*, Menlo Park, CA. AAAI Press/MIT Press.
- Chawla, N. V. and G. Karakoulas. 2005. Learning from labeled and unlabeled data: An empirical study across techniques and domains. *Journal of Artificial Intelligence Research*, 23:331–366.
- Collins, M. and Y. Singer. 1999. Unsupervised Models for Named Entity Classification. In *Proceedings of EMNLP*.

- Collins, Michael. 2002. Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms. In *EMNLP*.
- Corduneanu, A. and T. Jaakkola. 2001. Stable Mixing of Complete and Incomplete Informaation. Technical report, CSAIL, MIT.
- Corduneanu, A. and T. Jaakkola. 2002. Continuation Methods for Mixing Hetrogeneous Sources. In *Proceedings of Uncertainty in Artificial Intelligence*.
- Corduneanu, A. and T. Jaakkola. 2003. On Information Regulrization. In *Proceedings of Uncertainty in Artificial Intelligence*.
- Corduneanu, A. and T. Jaakkola. 2004. Distributed information regularization on graphs. In *Proceedings of Neural Information Proccessing Systems*.
- Cucerzan, S. and D. Yarowsky. 2002. Bootstrapping a multilingual part-of-speech tagger in one person-day. In *Proc. 6th Conference on Natural Language Learning 2002 (CoNLL-2002)*.
- Cutting, D., J. Kupiec, J. Pedersen, and P. Sibun. 1992. A practical part-of-speech tagger. In *Proc. of 3rd ANLP Conf.*, Trento, Italy. ACL.
- D. Yarowsky. 1995. Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. In *ACL*.
- Dasgupta, S., D. McAllester, and M. Littman. 2001. PAC Generalization Bounds for Co-Training. In *Proceedings of Neural Information Proccessing Systems*.
- de Sa, V. R. 1994. Learning Classification with Unlabeled Data. In J. D. Cowan, G. Tesuaro, and J. Alspector, editors, *NIPS 8*, San Fransisco, CA. Morgan Kaufmann.
- Dempster, A. P., N. M. Laird, and D.B. Rubin. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 1(39):1–38.

- Elworthy, David. 1994. Does baum-welch re-estimation help taggers? In *Proceedings of 4th ACL Conf on ANLP*, pages 53–58, Stuttgart, October 13–15.
- Goldman, S. and Y. Zhou. 2000. Enhancing supervised learning with unlabeled data. In *Proceedings of ICML-2000*, Stanford University, June 29–July 2.
- Goutte, C., H. Déjean, E. Gaussier, N. Cancedda, and J-M. Renders. 2002. Combining labelled and unlabelled data: a case study on fisher kernels and transductive inference for biological entity recognition. In *Proc. 6th conference on Natural language learning (CoNLL)*, pages 1–7.
- Grandvalet, Y. and Y. Bengio. 2004. Semi-supervised Learning by Entropy Minimization. In *Proceedings of Neural Information Processing Systems*.
- Herbrich, R. 2001. *Learning Kernel Classifiers: Theory and Algorithms*. MIT Press, Cambridge, MA, USA.
- Hindle, D. and M. Rooth. 1993. Structural ambiguity and lexical relations. *Computational Linguistics*, 19(1):103–120.
- Hinton, G. E. and V. Nair. 2005. Inferring motor programs from images of handwritten digits. In *Proc. NIPS-2005*.
- I. Cohen and F. G. Cozman and N. Sebe and M. C. Cirelo and T. S. Huang. 2004. Semi-supervised Learning of Classifiers: Theory, Algorithms for Bayesian Network Classifiers and Application to Human-Computer Interaction. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*.
- I. Tsochantaridis and T. Hofmann and T. Joachims and Y. Altun. 2004. Support Vector Machine Learning for Interdependent and Structured Output Spaces. In *Proceedings of International Conference on Machine Learning*.
- J. Lafferty and X. Zhu and Y. Liu. 2004. Kernel Conditional Random Fields: Representation and Clique Selection. In *Proceedings of International Conference on Machine Learning*.
- Kemp, C., T. Griffiths, S. Stromsten, and J. Tenenbaum. 2003. Semi-supervised learning with trees. In *Proceedings of Neural Information Processing Systems*.

- Krishnapuram, B., D. Williams, Y. Xue, A. Hartemink, L. Carin, and M. Figueiredo. 2005. On semi-supervised classification. In L. K. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems 17*. MIT Press.
- Leskes, B. 2005. The Value of Agreement: A New Boosting Algorithm. In *Proceedings of Computational Learning Theory*.
- Levin, A., P. Viola, and Y. Freund. 2003. Unsupervised Improvement of Visual Detectors using Co-Training. In *International Conference on Computer Vision (ICCV, Nice, France, October*.
- M. Seeger. 2000. Learning with Labeled and Unlabeled Data. Technical report, Edinburgh.
- Maeireizo, B., D. Litman, and R. Hwa. 2004. Co-training for predicting emotions with spoken dialogue data. In *Companion Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL)*.
- McClosky, D., E. Charniak, and M. Johnson. 2006. Effective Self-Training for Parsing. In *Proc. of HLT-NAACL 2006*, New York, NY, June 5–7.
- Merialdo, B. 1994. Tagging english text with a probabilistic model. *Computational Linguistics*, 20(2):155–172.
- Miller, D. J. and H. S. Uyar. 1997. A Mixture of Experts Classifier with Learning Based on Both Labelled and Unlabelled Data. In J. D. Cowan, G. Tesuaro, and J. Alspector, editors, *Proc of NIPS-1997*, San Fransisco, CA. Morgan Kaufmann.
- Mitchell, T. 1999. The role of unlabeled data in supervised learning. In *Proc. of 6th International Symposium on Cognitive Science (invited paper)*, San Sebastian, Spain.
- Müller, C., S. Rapp, and M. Strube. 2002. Applying Co-Training to Reference Resolution. In *Proc. of ACL-2002*, pages 352–359.
- Nigam, K. and R. Ghani. 2000. Analyzing the Effectiveness and Applicability of Co-Training. In *CIKM*.

- Nigam, K., A. McCallum, S. Thrun, and T. Mitchell. 2000. Text Classification from Labeled and Unlabeled Documents using EM. *Machine Learning*, 30(3).
- Niu, Z., D.-H. Ji, and C. L. Tan. 2005. A semi-supervised feature clustering algorithm with application to word sense disambiguation. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT-EMNLP)*.
- Okabe, M., K. Umemura, and S. Yamada. 2005. Query expansion with the minimum user feedback by transductive learning. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT-EMNLP 2005)*.
- Pang, B. and L. Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proc. of ACL-2004*.
- Phillips, W. and E. Riloff. 2002. Exploiting strong syntactic heuristics and co-training to learn semantic lexicons. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*.
- Pierce, D. and C. Cardie. 2001. Limitations of Co-training for Natural Language Learning from Large Datasets. In *Proc. of the 2001 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Ratnaparkhi, A. 1998. Unsupervised Statistical Models for Prepositional Phrase Attachment. In *Proc. of COLING-ACL '98*, pages 1079–1085, Montreal.
- Riloff, E., J. Wiebe, and T. Wilson. 2003. Learning subjective nouns using extraction pattern bootstrapping. In *Proceedings of the 7th Conference on Natural Language Learning (CoNLL-2003)*.
- Rosset, S., H. Zou J. Zhu, and T. Hastie. 2004. A Method for Inferring Label Sampling Mechanisms In Semi-Supervised Learning. In *Proceedings of Neural Information Processing Systems*.
- S. Rüping and T. Scheffer, editor. 2005. *Learning With Multiple Views*.

- Sarkar, A. 2001. Applying co-training methods to statistical parsing. In *Proceedings of the 2nd Meeting of the North American Association for Computational Linguistics: NAACL 2001*, Pittsburgh, PA, June.
- Shahshahani, B. and D. Landgrebe. 1994. The Effect of Unlabeled Samples in Reducing the Small Sample Size Problem and Mitigating the Hughes Phenomenon. *IEEE Transactions on Geoscience and Remote Sensing*.
- Sindhwani, V., P. Niyogi, and M. Belkin. 2005. A Co-Regularization Approach to Semi-supervised Learning with Multiple Views. In *Workshop on Learning with Multiple Views, Proceedings of International Conference on Machine Learning*.
- Steedman, M., R. Hwa, S. Clark, M. Osborne, A. Sarkar, J. Hockenmaier, P. Ruhlen, S. Baker, and J. Crim. 2003a. Example selection for bootstrapping statistical parsers. In *Human Language Technology Conference and the 4th Meeting of the North American Association for Computational Linguistics: HLT-NAACL 2003*, Edmonton, Canada, May.
- Steedman, M., M. Osborne, A. Sarkar, S. Clark, R. Hwa, J. Hockenmaier, P. Ruhlen, S. Baker, and J. Crim. 2003b. Bootstrapping statistical parsers from small datasets. In *11th Conference of the European Association for Computational Linguistics: EACL 2003*, Budapest, Hungary, April.
- Szummer, M. and T. Jaakkola. 2002. Information Regularization with Partially Labelled Data. In *Proceedings of Neural Information Processing Systems*.
- T. Joachims. 1999. Transductive Inference for Text Classification using Support Vector Machines. In *International Conference on Machine Learning*.
- Taskar, B., V. Chatalbashev, D. Koller, and C. Guestrin. 2005. Learning Structured Prediction Models: A Large Margin Approach. In *Proceedings of International Conference on Machine Learning*.
- V. Castelli and T. M. Cover. 1996. The relative value of labeled and unlabeled samples in pattern recognition with an unknown mixing parameter. *IEEE Transactions on Information Theory*.
- V. Vapnik. 2000. *The Nature of Statistical Learning Theory*. Springer.



- Vapnik, V. 1998. *Statistical Learning Theory*. Springer.
- Vincent, P. and Y. Bengio. 2003. Density-sensitive metrics and kernels. In *Snowbird Learning Workshop*.
- X. Zhu and J. Lafferty. 2005. Harmonic mixtures: combining mixture models and graph-based methods for inductive and scalable semi-supervised learning. In *Proc. of ICML-2005*.
- X. Zhu and Z. Ghahramani and J. Lafferty. 2003. Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions. In *Proceedings of International Conference on Machine Learning*.
- Y. Altun, I. Tsochantaridis and T. Hofmann. 2003. Hidden markov support vector machines. In *International Conference on Machine Learning*.
- Yarowsky, D. 1995. Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. In *Proceedings of Association for Computational Linguistics*.
- Zhang, T. and F. J. Oles. 2000. A probability analysis on the value of unlabeled data for classification problems. In *Proc. 17th International Conference on Machine Learning (ICML-2000)*, San Francisco, CA. Morgan Kaufmann.
- Zhou, Y. and S. Goldman. 2004. Democratic co-learning. In *ICTAI 2004*, pages 594–602.
- Zhu, X. 2005. Semi-Supervised Learning Literature Survey. Technical Report 1530, Computer Sciences, University of Wisconsin-Madison. [http://www.cs.wisc.edu/~jerryzhu/pub/ssl\\_survey.pdf](http://www.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf).