APPENDIX A

Nomenclature Reference

"What's the use of their having names," the Gnat said, "if they won't answer to them?"

"No use to them," said Alice; "but it's useful to the people who name them, I suppose. If not, why do things have names at all?"

— *Lewis Carroll*

$\mathcal{L}$	labeled data set
$\mathcal{U}$	unlabeled data set
x, y	input data instance and corresponding label
$\mathcal{H}$	hypothesis space (i.e., model class)
$\mathcal{V}$	version space (i.e., subset of $\mathcal{H}$ consistent with $\mathcal{L}$)
h	hypothesis
θ	parameter(s) of a particular hypothesis or model
$\phi_A(\cdot)$	utility measure A
x_A^*	best query instance according to $\phi_A(\cdot)$
$\mathrm{DIS}(\cdot)$	region of disagreement
ξ	disagreement coefficient
$H(\cdot)$	entropy
$KL(\cdot\Vert\cdot)$	Kullback-Liebler (KL) divergence
$I(\cdot;\cdot)$	mutual information / information gain
∇x	Fisher score
F	Fisher information matrix

Bibliography

N. Abe and H. Mamitsuka. Query learning strategies using boosting and bagging. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 1–9. Morgan Kaufmann, 1998. Cited on page(s) 29

K.S. Alexander. Rates of growth and sample moduli for weighted empirical processes indexed by sets. *Probability Theory and Related Fields*, 75(3):379–423, 1987. DOI: 10.1007/BF00318708 Cited on page(s) 59

S. Andrews, I. Tsochantaridis, and T. Hofmann. Support vector machines for multiple-instance learning. In *Advances in Neural Information Processing Systems (NIPS)*, volume 15, pages 561–568. MIT Press, 2003. Cited on page(s) 71

D. Andrzejewski, X. Zhu, M. Craven, and B. Recht. A framework for incorporating general domain knowledge into latent dirichlet allocation using first-order logic. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1171–1177. AAAI Press, 2011. DOI: 10.5591/978-1-57735-516-8/IJCAI11-200 Cited on page(s) 68

D. Angluin. Queries and concept learning. *Machine Learning*, 2:319–342, 1988. DOI: 10.1023/A:1022821128753 Cited on page(s) 6, 32

D. Angluin. Queries revisited. In *Proceedings of the International Conference on Algorithmic Learning Theory*, pages 12–31. Springer-Verlag, 2001. DOI: 10.1007/3-540-45583-3_3 Cited on page(s) 6

S. Arora, E. Nyberg, and C.P. Rosé. Estimating annotation cost for active learning in a multi-annotator environment. In *Proceedings of the NAACL HLT Workshop on Active Learning for Natural Language Processing*, pages 18–26. ACL, 2009. DOI: 10.3115/1564131.1564136 Cited on page(s) 67

L. Atlas, D. Cohn, R. Ladner, M. El-Sharkawi, R. Marks II, M. Aggoune, and D. Park. Training connectionist networks with queries and selective sampling. In *Advances in Neural Information Processing Systems (NIPS)*, volume 3, pages 566–573. Morgan Kaufmann, 1990. Cited on page(s) 7

J. Attenberg and F. Provost. Why label when you can search? alternatives to active learning for applying human resources to build classification models under extreme class imbalance. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 423–432. ACM, 2010a. DOI: 10.1145/1835804.1835859 Cited on page(s) 72

J. Attenberg and F. Provost. Inactive learning? difficulties employing active learning in practice. *SIGKDD Explorations*, 12(2):36–41, 2010b. DOI: 10.1145/1964897.1964906 Cited on page(s) 73

J. Attenberg, P. Melville, and F. Provost. A unified approach to active dual supervision for labeling features and examples. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, pages 40–55. Springer, 2010. DOI: 10.1007/978-3-642-15880-3_9 Cited on page(s) 69

M.F. Balcan, A. Beygelzimer, and J. Langford. Agnostic active learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 65–72. ACM, 2006. DOI: 10.1145/1143844.1143853 Cited on page(s) 61, 62

M.F. Balcan, A. Broder, and T. Zhang. Margin based active learning. In *Proceedings of the Conference on Learning Theory (COLT)*, pages 35–50. Springer, 2007. Cited on page(s) 62

M.F. Balcan, S. Hanneke, and J. Wortman. The true sample complexity of active learning. In *Proceedings of the Conference on Learning Theory (COLT)*, pages 45–56. Springer, 2008. Cited on page(s) 61

J. Baldridge and M. Osborne. Active learning and the total cost of annotation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9–16. ACL, 2004. Cited on page(s) 65, 76

J. Baldridge and A. Palmer. How well does active learning *actually* work? Time-based evaluation of cost-reduction strategies for language documentation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 296–305. ACL, 2009. DOI: 10.3115/1699510.1699549 Cited on page(s) 65

A.L. Berger, V.J. Della Pietra, and S.A. Della Pietra. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71, 1996. Cited on page(s) 39

A. Beygelzimer, S. Dasgupta, and J. Langford. Importance-weighted active learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 49–56. Omnipress, 2009. DOI: 10.1145/1553374.1553381 Cited on page(s) 61, 62

C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006. DOI: 10.1117/1.2819119 Cited on page(s) xi

D.M. Blei, A.Y. Ng, and M. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003. Cited on page(s) 52

M. Bloodgood and V. Shanker. A method for stopping active learning based on stabilizing predictions and the need for user-adjustable stopping. In *Proceedings of the Conference on Natural Language*

Learning (CoNLL), pages 39–47. ACL, 2009. DOI: 10.3115/1596374.1596384 Cited on page(s) 77

A. Blum and T. Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the Conference on Learning Theory (COLT)*, pages 92–100. Morgan Kaufmann, 1998. DOI: 10.1145/279943.279962 Cited on page(s) 53

L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996. DOI: 10.1023/A:1018094028462 Cited on page(s) 28, 30, 38

K. Brinker. Incorporating diversity in active learning with support vector machines. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 59–66. AAAI Press, 2003. Cited on page(s) 44

R. Burbidge, J.J. Rowland, and R.D. King. Active learning for regression based on query by committee. In *Proceedings of Intelligent Data Engineering and Automated Learning (IDEAL)*, pages 209–218. Springer, 2007. DOI: 10.1007/978-3-540-77226-2_22 Cited on page(s) 31

A. Carlson, J. Betteridge, R. Wang, E.R. Hruschka Jr, and T. Mitchell. Coupled semi-supervised learning for information extraction. In *Proceedings of the International Conference on Web Search and Data Mining (WSDM)*, pages 101–110. ACM, 2010. DOI: 10.1145/1718487.1718501 Cited on page(s) 74

K. Chaloner and I. Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3): 237–304, 1995. DOI: 10.1214/ss/1177009939 Cited on page(s) 41, 42

D. Cohn. Neural network exploration using optimal experiment design. In *Advances in Neural Information Processing Systems (NIPS)*, volume 6, pages 679–686. Morgan Kaufmann, 1994. DOI: 10.1016/0893-6080(95)00137-9 Cited on page(s) 43

D. Cohn, L. Atlas, and R. Ladner. Improving generalization with active learning. *Machine Learning*, 15(2):201–221, 1994. DOI: 10.1007/BF00993277 Cited on page(s) 7, 8, 24, 58, 62

D. Cohn, Z. Ghahramani, and M.I. Jordan. Active learning with statistical models. *Journal of Artificial Intelligence Research*, 4:129–145, 1996. Cited on page(s) 6, 43

T. H. Corman, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. MIT Press, 1992. Cited on page(s) 17

T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley, 2006. Cited on page(s) 41

A. Culotta, T. Kristjansson, A. McCallum, and P. Viola. Corrective feedback and persistent learning for information extraction. *Artificial Intelligence*, 170:1101–1122, 2006. DOI: 10.1016/j.artint.2006.08.001 Cited on page(s) 65

A. Culotta and A. McCallum. Reducing labeling effort for stuctured prediction tasks. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 746–751. AAAI Press, 2005. Cited on page(s) 17, 65

I. Dagan and S. Engelson. Committee-based sampling for training probabilistic classifiers. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 150–157. Morgan Kaufmann, 1995. Cited on page(s) 8, 28

A. Das and D. Kempe. Algorithms for subset selection in linear regression. In *Symposium on Theory of Computing (STOC)*, pages 45–54. ACM, 2008. DOI: 10.1145/1374376.1374384 Cited on page(s) 45

S. Dasgupta. Two faces of active learning. *Theoretical Computer Science*, 412(19):1767–1781, 2010. DOI: 10.1016/j.tcs.2010.12.054 Cited on page(s) xiii

S. Dasgupta and D.J. Hsu. Hierarchical sampling for active learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 208–215. ACM, 2008. DOI: 10.1145/1390156.1390183 Cited on page(s) 50, 52, 61

S. Dasgupta, D. Hsu, and C. Monteleoni. A general agnostic active learning algorithm. In *Advances in Neural Information Processing Systems (NIPS)*, volume 20, pages 353–360. MIT Press, 2008. Cited on page(s) 61, 62

S. Dasgupta, A.T. Kalai, and C. Monteleoni. Analysis of perceptron-based active learning. *Journal of Machine Learning Research*, 10:281–299, 2009. DOI: 10.1007/11503415_17 Cited on page(s) 62

V.R. de Sa. Learning classification with unlabeled data. In *Advances in Neural Information Processing Systems (NIPS)*, volume 6, pages 112–119. MIT Press, 1994. Cited on page(s) 53

T. Dietterich, R. Lathrop, and T. Lozano-Perez. Solving the multiple-instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89:31–71, 1997. DOI: 10.1016/S0004-3702(96)00034-3 Cited on page(s) 71

P. Donmez and J. Carbonell. Proactive learning: Cost-sensitive active learning with multiple imperfect oracles. In *Proceedings of the Conference on Information and Knowledge Management (CIKM)*, pages 613–622. ACM, 2008. DOI: 10.1145/1458082.1458165 Cited on page(s) 68

P. Donmez, J. Carbonell, and J. Schneider. Efficiently learning the accuracy of labeling sources for selective sampling. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 259–268. ACM, 2009. DOI: 10.1145/1557019.1557053 Cited on page(s) 74

P. Donmez, J. Carbonell, and J. Schneider. A probabilistic framework to learn from multiple annotators with time-varying accuracy. In *Proceedings of the SIAM Conference on Data Mining (SDM)*, 2010. Cited on page(s) 74

G. Druck, G. Mann, and A. McCallum. Learning from labeled features using generalized expectation criteria. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 595–602. ACM, 2008. DOI: 10.1145/1390334.1390436 Cited on page(s) 69

G. Druck, B. Settles, and A. McCallum. Active learning by labeling features. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 81–90. ACL, 2009. DOI: 10.3115/1699510.1699522 Cited on page(s) 69

R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley-Interscience, 2001. Cited on page(s) xi

V. Federov. *Theory of Optimal Experiments*. Academic Press, 1972. Cited on page(s) 18, 41

P. Felt, E. Ringger, K. Seppi, K. Heal, R. Haertel, and D. Lonsdale. First results in a study evaluating pre-annotation and correction propagation for machine-assisted syriac morphological analysis. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, pages 878–885. ELRA, 2012. Cited on page(s) 66

P. Flaherty, M. Jordan, and A. Arkin. Robust design of biological experiments. In *Advances in Neural Information Processing Systems (NIPS)*, volume 18, pages 363–370. MIT Press, 2006. Cited on page(s) 42

Y. Freund and R.E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997. DOI: 10.1006/jcss.1997.1504 Cited on page(s) 28

Y. Freund and R.E. Schapire. Large margin classification using the perceptron algorithm. *Machine learning*, 37(3):277–296, 1999. DOI: 10.1023/A:1007662407062 Cited on page(s) 24

Y. Freund, H.S. Seung, E. Shamir, and N. Tishby. Selective samping using the query by committee algorithm. *Machine Learning*, 28:133–168, 1997. DOI: 10.1023/A:1007330508534 Cited on page(s) 28, 61, 62

E. Friedman. Active learning for smooth problems. In *Proceedings of the Conference on Learning Theory (COLT)*, pages 3–2, 2009. Cited on page(s) 61

A. Fujii, T. Tokunaga, K. Inui, and H. Tanaka. Selective sampling for example-based word sense disambiguation. *Computational Linguistics*, 24(4):573–597, 1998. Cited on page(s) 8, 49

K. Ganchev, J. Graça, J. Gillenwater, and B. Taskar. Posterior regularization for structured latent variable models. *Journal of Machine Learning Research*, 11:2001–2049, 2010. Cited on page(s) 68

S. Geman, E. Bienenstock, and R. Doursat. Neural networks and the bias/variance dilemma. *Neural Computation*, 4:1–58, 1992. DOI: 10.1162/neco.1992.4.1.1 Cited on page(s) 40

Y. Grandvalet and Y. Bengio. Semi-supervised learning by entropy minimization. In *Advances in Neural Information Processing Systems (NIPS)*, volume 17, pages 529–536. MIT Press, 2005. Cited on page(s) 53

C. Guestrin, A. Krause, and A.P. Singh. Near-optimal sensor placements in gaussian processes. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 265–272. ACM, 2005. DOI: 10.1145/1102351.1102385 Cited on page(s) 45

Y. Guo and R. Greiner. Optimistic active learning using mutual information. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*, pages 823–829. AAAI Press, 2007. Cited on page(s) 39

Y. Guo and D. Schuurmans. Discriminative batch mode active learning. In *Advances in Neural Information Processing Systems (NIPS)*, number 20, pages 593–600. MIT Press, Cambridge, MA, 2008. Cited on page(s) 45

R. Haertel, K. Seppi, E. Ringger, and J. Carroll. Return on investment for active learning. In *Proceedings of the NIPS Workshop on Cost-Sensitive Learning*, 2008. Cited on page(s) 67, 68

S. Hanneke. A bound on the label complexity of agnostic active learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 353–360. ACM, 2007. DOI: 10.1145/1273496.1273541 Cited on page(s) 59, 61, 62

S. Hanneke. *Theoretical Foundations of Active Learning*. PhD thesis, Carnegie Mellon University, 2009. Cited on page(s) 59, 62

A. Hauptmann, W. Lin, R. Yan, J. Yang, and M.Y. Chen. Extreme video retrieval: joint maximization of human and computer performance. In *Proceedings of the ACM Workshop on Multimedia Image Retrieval*, pages 385–394. ACM, 2006. DOI: 10.1145/1180639.1180721 Cited on page(s) 9

D. Haussler. Learning conjunctive concepts in structural domains. *Machine Learning*, 4(1):7–40, 1994. DOI: 10.1023/A:1022698821832 Cited on page(s) 26

S.C.H. Hoi, R. Jin, and M.R. Lyu. Large-scale text categorization by batch mode active learning. In *Proceedings of the International Conference on the World Wide Web*, pages 633–642. ACM, 2006a. DOI: 10.1145/1135777.1135870 Cited on page(s) 9, 43, 44

S.C.H. Hoi, R. Jin, J. Zhu, and M.R. Lyu. Batch mode active learning and its application to medical image classification. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 417–424. ACM, 2006b. DOI: 10.1145/1143844.1143897 Cited on page(s) 45

R. Hwa. On minimizing training corpus for parser acquisition. In *Proceedings of the Conference on Natural Language Learning (CoNLL)*, pages 1–6. ACL, 2001. DOI: 10.3115/1117822.1117829 Cited on page(s) 76

R. Hwa. Sample selection for statistical parsing. *Computational Linguistics*, 30(3):73–77, 2004. DOI: 10.1162/0891201041850894 Cited on page(s) 17

P. Jain, S. Vijayanarasimhan, and K. Grauman. Hashing hyperplane queries to near points with applications to large-scale active learning. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R.S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems (NIPS)*, volume 23, pages 928–936. 2010. Cited on page(s) 24

J. Kang, K. Ryu, and H.C. Kwon. Using cluster-based sampling to select initial training set for active learning in text classification. In *Proceedings of the Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD)*, pages 384–388. Springer, 2004. DOI: 10.1007/978-3-540-24775-3_46 Cited on page(s) 49

A. Kapoor, E. Horvitz, and S. Basu. Selective supervision: Guiding supervised learning with decision-theoretic active learning,. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*, pages 877–882. AAAI Press, 2007. Cited on page(s) 66

R.D. King, K.E. Whelan, F.M. Jones, P.G. Reiser, C.H. Bryant, S.H. Muggleton, D.B. Kell, and S.G. Oliver. Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature*, 427(6971):247–52, 2004. DOI: 10.1038/nature02236 Cited on page(s) 7, 32, 66

R.D. King, J. Rowland, S.G. Oliver, M. Young, W. Aubrey, E. Byrne, M. Liakata, M. Markham, P. Pir, L.N. Soldatova, A. Sparkes, K.E. Whelan, and A. Clare. The automation of science. *Science*, 324(5923):85–89, 2009. DOI: 10.1126/science.1165620 Cited on page(s) 7

D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009. Cited on page(s) 12, 17

C. Körner and S. Wrobel. Multi-class ensemble-based active learning. In *Proceedings of the European Conference on Machine Learning (ECML)*, pages 687–694. Springer, 2006. DOI: 10.1007/11871842_68 Cited on page(s) 63

A. Krause. *Optimizing Sensing: Theory and Applications*. PhD thesis, Carnegie Mellon University, 2008. Cited on page(s) 45

V. Krishnamurthy. Algorithms for optimal scheduling and management of hidden markov model sensors. *IEEE Transactions on Signal Processing*, 50(6):1382–1397, 2002. DOI: 10.1109/TSP.2002.1003062 Cited on page(s) 8

S. Kullback and R.A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22: 79–86, 1951. DOI: 10.1214/aoms/1177729694 Cited on page(s) 29

G. Kunapuli, R. Maclin, and J. Shavlik. Advice refinement for knowledge-based support vector machines. In *Advances in Neural Information Processing Systems (NIPS)*, volume 24, pages 1728–1736. 2011. Cited on page(s) 68

J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 282–289. Morgan Kaufmann, 2001. DOI: 10.1038/nprot.2006.61 Cited on page(s) 39

K. Lang. Newsweeder: Learning to filter netnews. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 331–339. Morgan Kaufmann, 1995. Cited on page(s) 49

K. Lang and E. Baum. Query learning can work poorly when a human oracle is used. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, pages 335–340. IEEE Press, 1992. Cited on page(s) 6, 7

D. Lewis and J. Catlett. Heterogeneous uncertainty sampling for supervised learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 148–156. Morgan Kaufmann, 1994. Cited on page(s) 11, 76

D. Lewis and W. Gale. A sequential algorithm for training text classifiers. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3–12. ACM/Springer, 1994. DOI: 10.1145/62437.62470 Cited on page(s) 9

P. Liang, M.I. Jordan, and D. Klein. Learning from measurements in exponential families. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 641–648. Omnipress, 2009. DOI: 10.1145/1553374.1553457 Cited on page(s) 69

R. Liere and P. Tadepalli. Active learning with committees for text categorization. In *Proceedings of the Conference on Artificial Intelligence (AAAI)*, pages 591—597. AAAI Press, 1997. Cited on page(s) 28

Y. Liu. Active learning with support vector machine applied to gene expression data for cancer classification. *Journal of Chemical Information and Computer Sciences*, 44:1936–1941, 2004. DOI: 10.1021/ci049810a Cited on page(s) 9

Z. Lu and J. Bongard. Exploiting multiple classifier types with active learning. In *Proceedings of the Conference on Genetic and Evolutionary Computation (GECCO)*, pages 1905–1906. ACM, 2009. DOI: 10.1145/1569901.1570228 Cited on page(s) 77

D. MacKay. Information-based objective functions for active data selection. *Neural Computation*, 4 (4):590–604, 1992. DOI: 10.1162/neco.1992.4.4.590 Cited on page(s) 43, 45

G. Mann and A. McCallum. Efficient computation of entropy gradient for semi-supervised conditional random fields. In *Proceedings of the North American Association for Computational Linguistics (NAACL)*, pages 109–112. ACL, 2007. Cited on page(s) 17

G. Mann and A. McCallum. Generalized expectation criteria for semi-supervised learning of conditional random fields. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 870–878. ACL, 2008. Cited on page(s) 69

G.S. Mann and A. McCallum. Generalized expectation criteria for semi-supervised learning with weakly labeled data. *Journal of Machine Learning Research*, 11:955–984, 2010. Cited on page(s) 68

C. D. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999. Cited on page(s) 31, 52

O. Maron and T. Lozano-Perez. A framework for multiple-instance learning. In *Advances in Neural Information Processing Systems (NIPS)*, volume 10, pages 570–576. MIT Press, 1998. Cited on page(s) 71

A. McCallum and K. Nigam. Employing EM in pool-based active learning for text classification. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 359–367. Morgan Kaufmann, 1998. Cited on page(s) 9, 28, 49, 54

P. Melville and R. Mooney. Diverse ensembles for active learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 584–591. Morgan Kaufmann, 2004. DOI: 10.1145/1015330.1015385 Cited on page(s) 29

P. Melville, S.M. Yang, M. Saar-Tsechansky, and R. Mooney. Active learning for probability estimation using Jensen-Shannon divergence. In *Proceedings of the European Conference on Machine Learning (ECML)*, pages 268–279. Springer, 2005. DOI: 10.1007/11564096_28 Cited on page(s) 31

M. Mintz, S. Bills, R. Snow, and D. Jurafsky. Distant supervision for relation extraction without labeled data. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 1003–1011. ACL, 2009. Cited on page(s) 74

T. Mitchell. Generalization as search. *Artificial Intelligence*, 18:203–226, 1982. DOI: 10.1016/0004-3702(82)90040-6 Cited on page(s) 8, 21

T. Mitchell. *Machine Learning*. McGraw-Hill, 1997. Cited on page(s) xi, 57

R. Moskovitch, N. Nissim, D. Stopel, C. Feher, R. Englert, and Y. Elovici. Improving the detection of unknown computer worms activity using active learning. In *Proceedings of the German Conference on AI*, pages 489–493. Springer, 2007. DOI: 10.1007/978-3-540-74565-5_47 Cited on page(s) 9

I. Muslea, S. Minton, and C.A. Knoblock. Selective sampling with redundant views. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 621–626. AAAI Press, 2000. Cited on page(s) 29

I. Muslea, S. Minton, and C.A. Knoblock. Active + semi-supervised learning = robust multi-view learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 435–442. Morgan Kaufmann, 2002. Cited on page(s) 54

G.L. Nemhauser, L.A. Wolsey, and M.L. Fisher. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming*, 14(1):265–294, 1978. DOI: 10.1007/BF01588971 Cited on page(s) 45

G. Ngai and D. Yarowsky. Rule writing or annotation: Cost-efficient resource usage for base noun phrase chunking. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 117–125. ACL, 2000. DOI: 10.3115/1075218.1075234 Cited on page(s) 31

H.T. Nguyen and A. Smeulders. Active learning using pre-clustering. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 79–86. ACM, 2004. DOI: 10.1145/1015330.1015349 Cited on page(s) 49

F. Olsson and K. Tomanek. An intrinsic stopping criterion for committee-based active learning. In *Proceedings of the Conference on Computational Natural Language Learning (CoNLL)*, pages 138–146. ACL, 2009. DOI: 10.3115/1596374.1596398 Cited on page(s) 77

G. Paaß and J. Kindermann. Bayesian query construction for neural network models. In *Advances in Neural Information Processing Systems (NIPS)*, volume 7, pages 443–450. MIT Press, 1995. Cited on page(s) 44

B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up: Sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 79–86. ACL, 2002. DOI: 10.3115/1118693.1118704 Cited on page(s) 69

G.J. Qi, X.S. Hua, Y. Rui, J. Tang, and H.J. Zhang. Two-dimensional active learning for image classification. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE Press, 2008. DOI: 10.1109/CVPR.2008.4587383 Cited on page(s) 76

H. Raghavan, O. Madani, and R. Jones. Active learning with feedback on both features and instances. *Journal of Machine Learning Research*, 7:1655–1686, 2006. Cited on page(s) 68

R. Rahmani and S.A. Goldman. MISSL: Multiple-instance semi-supervised learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 705–712. ACM, 2006. DOI: 10.1145/1143844.1143933 Cited on page(s) 71

S. Ray and M. Craven. Supervised versus multiple instance learning: An empirical comparison. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 697–704. ACM, 2005. DOI: 10.1145/1102351.1102439 Cited on page(s) 71

R. Reichart, K. Tomanek, U. Hahn, and A Rappoport. Multi-task active learning for linguistic annotations. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 861–869. ACL, 2008. Cited on page(s) 76

E. Ringger, M. Carmen, R. Haertel, K. Seppi, D. Lonsdale, P. McClanahan, J. Carroll, and N. Ellison. Assessing the costs of machine-assisted corpus annotation through a user study. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, pages 3318–3324. ELRA, 2008. Cited on page(s) 67

D. Roth and K. Small. Active learning for pipeline models. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 683–688. AAAI Press, 2008. Cited on page(s) 76

N. Roy and A. McCallum. Toward optimal active learning through sampling estimation of error reduction. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 441–448. Morgan Kaufmann, 2001. Cited on page(s) 38, 39, 49

S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, second edition, 2003. Cited on page(s) xi

A.I. Schein and L.H. Ungar. Active learning for logistic regression: An evaluation. *Machine Learning*, 68(3):235–265, 2007. DOI: 10.1007/s10994-007-5019-5 Cited on page(s) 43, 44, 63

M.J. Schervish. *Theory of Statistics*. Springer, 1995. DOI: 10.1007/978-1-4612-4250-5 Cited on page(s) 42

G. Schohn and D. Cohn. Less is more: Active learning with support vector machines. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 839–846. Morgan Kaufmann, 2000. Cited on page(s) 24

H. Schütze, E. Velipasaoglu, and J.O. Pedersen. Performance thresholding in practical text classification. In *Proceedings of the Conference on Information and Knowledge Management (CIKM)*, pages 662–671. ACM, 2006. DOI: 10.1145/1183614.1183709 Cited on page(s) 20

R. Schwartz and Y.-L. Chow. The N-best algorithm: an efficient and exact procedure for finding the N most likely sentence hypotheses. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 81–83. IEEE Press, 1990. DOI: 10.1109/ICASSP.1990.115542 Cited on page(s) 17

B. Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009. Cited on page(s) xi, xiii

B. Settles. Closing the loop: Fast, interactive semi-supervised annotation with queries on features and instances. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1467–1478. ACL, 2011. Cited on page(s) 69, 70

B. Settles and M. Craven. An analysis of active learning strategies for sequence labeling tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1069–1078. ACL, 2008. DOI: 10.3115/1613715.1613855 Cited on page(s) 9, 17, 43, 44, 49, 63

B. Settles, M. Craven, and L. Friedland. Active learning with real annotation costs. In *Proceedings of the NIPS Workshop on Cost-Sensitive Learning*, 2008a. Cited on page(s) 4, 67, 68

B. Settles, M. Craven, and S. Ray. Multiple-instance active learning. In *Advances in Neural Information Processing Systems (NIPS)*, volume 20, pages 1289–1296. MIT Press, 2008b. Cited on page(s) 71, 72

H.S. Seung, M. Opper, and H. Sompolinsky. Query by committee. In *Proceedings of the ACM Workshop on Computational Learning Theory*, pages 287–294. ACM, 1992. DOI: 10.1145/130385.130417 Cited on page(s) 28, 62

C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423,623–656, 1948. Cited on page(s) 14

V.S. Sheng, F. Provost, and P.G. Ipeirotis. Get another label? improving data quality and data mining using multiple, noisy labelers. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 614–622. ACM, 2008. DOI: 10.1145/1401890.1401965 Cited on page(s) 74

V. Sindhwani, P. Melville, and R.D. Lawrence. Uncertainty sampling and transductive experimental design for active dual supervision. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 953–960. Omnipress, 2009. DOI: 10.1145/1553374.1553496 Cited on page(s) 69

K. Small, B.C. Wallace, C.E. Brodley, and T.A. Trikalinos. The constrained weight space SVM: Learning with ranked features. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 865–872. Omnipress, 2011. Cited on page(s) 68, 69

B.C. Smith, B. Settles, W.C. Hallows, M.W. Craven, and J.M. Denu. SIRT3 substrate specificity determined by peptide arrays and machine learning. *ACS Chemical Biology*, 6(2):146–157, 2011. DOI: 10.1021/cb100218d Cited on page(s) 4

R. Snow, B. O'Connor, D. Jurafsky, and A. Ng. Cheap and fast—but is it good? In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 254–263. ACM, 2008. DOI: 10.3115/1613715.1613751 Cited on page(s) 74

M. Sugiyama and N. Rubens. Active learning with model selection in linear regression. In *Proceedings of the SIAM International Conference on Data Mining*, pages 518–529. SIAM, 2008. Cited on page(s) 77

C.A. Thompson, M.E. Califf, and R.J. Mooney. Active learning for natural language parsing and information extraction. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 406–414. Morgan Kaufmann, 1999. Cited on page(s) 9

K. Tomanek and U. Hahn. Semi-supervised active learning for sequence labeling. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 1039–1047. ACL, 2009. Cited on page(s) 54

K. Tomanek and U. Hahn. A comparison of models for cost-sensitive active learning. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, volume Posters, pages 1247–1255. ACL, 2010. Cited on page(s) 68

K. Tomanek and K. Morik. Inspecting sample reusability for active learning. In *Active Learning and Experimental Design*, volume 15 of *JMLR Workshop and Conference Proceedings*, pages 169–181. Microtome Publishing, 2011. Cited on page(s) 76

K. Tomanek and F. Olsson. A web survey on the use of active learning to support annotation of text data. In *Proceedings of the NAACL HLT Workshop on Active Learning for Natural Language Processing*, pages 45–48. ACL, 2009. DOI: 10.3115/1564131.1564140 Cited on page(s) 64

K. Tomanek, J. Wermter, and U. Hahn. An approach to text corpus construction which cuts annotation costs and maintains reusability of annotated data. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 486–495. ACL, 2007. Cited on page(s) 76

K. Tomanek, F. Laws, U. Hahn, and H. Schütze. On proper unit selection in active learning: Co-selection effects for named entity recognition. In *Proceedings of the NAACL HLT Workshop on Active Learning for Natural Language Processing*, pages 9–17. ACL, 2009. DOI: 10.3115/1564131.1564135 Cited on page(s) 20

S. Tong and E. Chang. Support vector machine active learning for image retrieval. In *Proceedings of the ACM International Conference on Multimedia*, pages 107–118. ACM, 2001. DOI: 10.1145/500141.500159 Cited on page(s) 9

S. Tong and D. Koller. Support vector machine active learning with applications to text classification. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 999–1006. Morgan Kaufmann, 2000. DOI: 10.1162/153244302760185243 Cited on page(s) 9, 23, 24

G. Towell and J. Shavlik. Knowledge-based artificial neural networks. *Artificial Intelligence*, 70: 119–165, 1994. DOI: 10.1016/0004-3702(94)90105-8 Cited on page(s) 68

G. Tür, D. Hakkani-Tür, and R.E. Schapire. Combining active and semi-supervised learning for spoken language understanding. *Speech Communication*, 45(2):171–186, 2005. DOI: 10.1016/j.specom.2004.08.002 Cited on page(s) 9, 54

A. Tversky and D. Kahneman. Extension versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4):293—315, 1983. DOI: 10.1037/0033-295X.90.4.293 Cited on page(s) 69

L.G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984. DOI: 10.1145/1968.1972 Cited on page(s) 2, 57

V. Vapnik. *Statistical Learning Theory*. Wiley, 1998. Cited on page(s) 23, 57, 58

V.N. Vapnik and A. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and Its Applications*, 16:264–280, 1971. DOI: 10.1137/1116025 Cited on page(s) 58

S. Vijayanarasimhan and K. Grauman. What's it going to cost you? Predicting effort vs. informativeness for multi-label image annotations. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2262–2269. IEEE Press, 2009a. Cited on page(s) 67, 68, 72

S. Vijayanarasimhan and K. Grauman. Multi-level active prediction of useful image annotations for recognition. In *Advances in Neural Information Processing Systems (NIPS)*, volume 21, pages 1705–1712. MIT Press, 2009b. Cited on page(s) 72

A. Vlachos. A stopping criterion for active learning. *Computer Speech and Language*, 22(3):295–312, 2008. DOI: 10.1016/j.csl.2007.12.001 Cited on page(s) 77

B.C. Wallace, K. Small, C.E. Brodley, and T.A. Trikalinos. Active learning for biomedical citation screening. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 173–182. ACM, 2010a. DOI: 10.1145/1835804.1835829 Cited on page(s) 20, 72

B.C. Wallace, K. Small, C.E. Brodley, and T.A. Trikalinos. Who should label what? Instance allocation in multiple expert active learning. In *Proceedings of the SIAM Conference on Data Mining (SDM)*, pages 176–187, 2011. Cited on page(s) 74

Byron C. Wallace, Kevin Small, Carla E. Brodley, Joseph Lau, and Thomas A. Trikalinos. Modeling annotation time to reduce workload in comparative effectiveness reviews. In *Proceedings of the ACM International Health Informatics Symposium (IHI)*, pages 28–35. ACM, 2010b. DOI: 10.1145/1882992.1882999 Cited on page(s) 67, 68

L. Wang. Sufficient conditions for agnostic active learnable. In *Advances in Neural Information Processing Systems (NIPS)*, volume 22, pages 1999–2007. 2009. Cited on page(s) 61

J.H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58:236–244, 1963. DOI: 10.2307/2282967 Cited on page(s) 52

M. Warmuth, K. Glocer, and G. Rätsch. Boosting algorithms for maximizing the soft margin. In *Advances in Neural Information Processing Systems (NIPS)*, volume 20, pages 1585–1592. MIT Press, 2008. Cited on page(s) 24

Z. Xu, R. Akella, and Y. Zhang. Incorporating diversity and density in active learning for relevance feedback. In *Proceedings of the European Conference on IR Research (ECIR)*, pages 246–257. Springer-Verlag, 2007. DOI: 10.1007/978-3-540-71496-5_24 Cited on page(s) 44, 49

R. Yan, J. Yang, and A. Hauptmann. Automatically labeling video data using multi-class active learning. In *Proceedings of the International Conference on Computer Vision*, pages 516–523. IEEE Press, 2003. DOI: 10.1109/ICCV.2003.1238391 Cited on page(s) 9

D. Yarowsky. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 189–196. ACL, 1995. Cited on page(s) 53

H. Yu. SVM selective sampling for ranking with application to data retrieval. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 354–363. ACM, 2005. DOI: 10.1145/1081870.1081911 Cited on page(s) 8

K. Yu, J. Bi, and V. Tresp. Active learning via transductive experimental design. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 1081–1087. ACM, 2006. DOI: 10.1145/1143844.1143980 Cited on page(s) 54

C. Zhang and T. Chen. An active learning framework for content based information retrieval. *IEEE Transactions on Multimedia*, 4(2):260–268, 2002. DOI: 10.1109/TMM.2002.1017738 Cited on page(s) 9

T. Zhang and F.J. Oles. A probability analysis on the value of unlabeled data for classification problems. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 1191–1198. Morgan Kaufmann, 2000. Cited on page(s) 43

Y. Zhang. Multi-task active learning with output constraints. In *Proceedings of the Conference on Artificial Intelligence (AAAI)*, pages 667–672. AAAI Press, 2010. Cited on page(s) 75

Z.H. Zhou, K.J. Chen, and Y. Jiang. Exploiting unlabeled data in content-based image retrieval. In *Proceedings of the European Conference on Machine Learning (ECML)*, pages 425–435. Springer, 2004. Cited on page(s) 54

X. Zhu and A. Goldberg. *Introduction to Semi-Supervised Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool, 2009. Cited on page(s) 4, 53

X. Zhu, J. Lafferty, and Z. Ghahramani. Combining active learning and semi-supervised learning using Gaussian fields and harmonic functions. In *Proceedings of the ICML Workshop on the Continuum from Labeled to Unlabeled Data*, pages 58–65, 2003. Cited on page(s) 39, 54

Author's Biography

BURR SETTLES

Burr Settles is a Postdoctoral Research Scientist in the Machine Learning Department at Carnegie Mellon University. He received a PhD in computer sciences from the University of Wisconsin-Madison in 2008, with additional studies in linguistics and biology. His research focuses on machine learning technology that interacts with humans as part of knowledge acquisition and training, with applications in natural language processing, biological discovery, and social computing.

Burr is the author of more than 20 research papers, including a popular literature survey on active learning (which was the genesis for this book). He has organized workshops at the International Conference on Machine Learning (ICML) and the North American Chapter of the Association for Computational Linguistics (NAACL) on active learning and related topics in cost-sensitive machine learning. In his spare time, he runs a website community for musicians (FAWM.ORG), prefers sandals to shoes, and plays guitar in the Pittsburgh pop band Delicious Pastries.

Index