Pinar Donmez

Language Technologies Institute
School of Computer Science
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh PA 15213

Office: (412) 251-8270
Fax: (412) 268 6298
pinard@cs.cmu.edu

Education

Ph.D. in Language and Information Technologies, Carnegie Mellon University, 2010 (Expected)
Advisor: Jaime G. Carbonell
Thesis title: *ProActive Learning: Towards Cost-Sensitive Active Learning with Multiple Imperfect Oracles*

M.S. in Language and Information Technologies, Carnegie Mellon University, 2007

B.S. in Computer Science and Engineering, Sabanci University, Turkey, 2004
Summa Cum Laude, with Highest Honors
Minor in Mathematics, Sabanci University, Turkey, 2004

Research Interests

Machine learning, active learning, proactive learning, information retrieval, data mining

Work Experience

Summer Intern, Microsoft Research Text Mining, Search and Navigation Group (TMSN), supervised by Chris Burges and Krysta Svore, Summer 2008

Project: Testing the local optimality of LambdaRank on different IR measures and investigating the relation between matching the training to the testing measure on large-scale ranking datasets.

Selected Research Projects

Proactive Learning
Addressing the omniscent oracle assumptions in active learning. Active learning, thus far, assumes an individual oracle that is infallible (never wrong), indefatigable (always answers) and cost-insensitive. Proactive learning relaxes these assumptions to jointly optimize the oracle and instance selection through a utility maximization framework under budget constraints.

Active Sampling in Rank Learning Applications
The use of direct optimization of rank loss functions for actively sampling the informative examples in the data. A balanced distribution across various relevance levels is aimed. In the presence of complete rank information, exploring ways to reduce the need to user feedback by picking the examples that will lead to faster convergence to the true rank.

Density Sensitive Sampling Methods for Active Learning for Classification
Explored different representations of the data in favor of better separation of the high density regions. Active sampling is adopted in the new representation via a utility-based measure incorporating the value of information (uncertainty) in a local neighborhood.

Ensemble Methods in Active Learning for Classification
Explored dynamic selection among a diverse group of sampling strategies in order to maximize the performance over a full operating range. The relative success of individual active learners based on the estimated future expected error are investigated in order to guide the selection.

Publications

Donmez, P., Lebanon, G., Balasubramanian, K.: Unsupervised Estimation of Classification and Regression Error Rates, submitted to *the Journal of Machine Learning Research*, June 2009

Donmez, P., Carbonell, G.: From Active to Proactive Learning Methods, to appear as Book Chapter *In Recent Advances in Machine Learning, Eds. J. Koronacki, S.T. Wierzchon, Z. Ras and J. Kacprzyk.*, Springer: Studies in Computational Intelligence, 2009

Donmez, P., Svore, K., Burges, C.J.: On the Local Optimality of LambdaRank, to appear *In the Proceedings of 32nd Annual ACM SIGIR Conference on Information Retrieval*, SIGIR '09, Boston, USA 2009

Donmez, P., Carbonell, J.G., Schneider, J.: Efficiently Learning the Accuracy of Labeling Sources for Selective Sampling, to appear *In the Proceedings of 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '09, Paris, France 2009

Donmez, P., Carbonell, J.G.: Active Sampling for Rank Learning via Optimizing the Area Under the ROC Curve, to appear *In the Proceedings of 31st European Conference on Information Retrieval*, ECIR '09, Toulouse, France 2009

Donmez, P., Carbonell, J.G.: Proactive Learning: Cost-Sensitive Active Learning with Multiple Imperfect Oracles, *In the Proceedings of the 17th ACM Conference on Information and Knowledge Management*, CIKM '08, Napa Valley 2008

Donmez, P., Carbonell, J.G.: Optimizing Estimated Loss Reduction for Active Sampling in Rank Learning, *In the Proceedings of the 25th International Conference on Machine Learning*, ICML '08, Helsinki 2008

Donmez, P., Carbonell, J.G.: Paired Sampling in Density-Sensitive Active Learning, *In the Proceedings of The 10th International Symposium on Artificial Intelligence and Mathematics*, ISAIM '08, Florida 2008

Donmez, P., Carbonell, J. G., Bennett, P.N.: Dual Strategy Active Learning, *In Proceedings of the 18th European Conference on Machine Learning*, ECML '07, pp. 116-127, Warsaw, Poland 2007

Donmez, P., Rose C.P., Stegmann, K., Weinberger, A., Fischer, F.: Supporting CSCL with Automatic Corpus Analysis Technology, *In Proceedings of Computer Supported Collaborative Learning*, Taipei, Taiwan 2005

Rose, C., Donmez, P., Gweon, G., Knight, A., Junker, B., Cohen, W., Koedinger, K., Heffernan, N: Automatic and Semi-Automatic Skill Coding with a View Towards Supporting Online Assessment, *Proceedings of AI in Education*, Amsterdam 2005

Oflazer, K., Donmez, P.: SmartReader: An NLP-based Interactive Reading Application for Language Learning, *in Lecture Notes in Computer Science, Proceedings of ISCIS'04*, 2004

Honors and Awards

Yahoo Key Scientific Challenges Award in the area of Machine Learning, April 2009

Best Presentation Award, Student Research Symposium, Langugae Technologies Institute, CMU, 2007

Undergraduate Outstanding Academic Record Scholarship, Sabanci Foundation, Turkey, 2001-2004

Undergraduate Merit Scholarship, Sabanci University, Turkey, 1999-2004

Relevant Course Work

Graduate Level:

Machine Learning	Statistical Foundations of Machine Learning
Intermediate Statistics	Advanced Statistical Learning Seminar
Information Retrieval	Advanced IR Seminar
Software Engineering	Algorithms for NLP

Undergraduate Level:

Artificial Intelligence	Image Processing and Pattern Recognition
Computer Graphics	Finite State Machines & Automata Theory
Theory of Computation	Statistical Modeling
Classical Real Analysis	Number Theory

Programming Skills

C, C++, C#, Java, Perl, Python, Matlab, R, Basic shell, SQL

References

available upon request