Dav Zimak

xxxx Washington Street
Apartment x
San Francisco, CA 94115

(xxx) xxx-xxxx
cv@davzimak.com
http://www.davzimak.com/research

Research and Professional Interests

My Ph.D. work focused on Machine Learning and Natural Language Processing. I showed how complex-labeled output problems, such as multi-class classification, ranking problems and structured predictions can be viewed in a common framework to provide both deep theoretic understanding of existing and new learning algorithms and a uniform setting to develop practical applications. While at Illinois, I created state-of-the-art shallow parsing and semantic role labeling systems (both complex NLP tasks). At Yahoo, I have worked on massive classification efforts and core web search relevance. The internet has exposed new ways to interact with data, and in doing so, has exposed interesting and complicated tasks with unknown solutions. I aim to discover how to leverage the structure internet data for new and powerful machine learning based solutions.

Education

Ph.D. in Computer Science — deposited May 2006
University of Illinois at Urbana-Champaign, IL
Dissertation Title: Algorithms and Analysis for Multi-Category Classification Tasks
Dissertation Advisor: Dan Roth

B.S. in Applied Mathematics — February 1998
Columbia University in the City of New York, NY

Research and Work Experience

Research Scientist, *Yahoo! Inc.* — February 2006 - Present
Projects: i. Ranking: Web search relevance. Feature research
ii. Y! Answers: Automatic Question Categorization
iii. Content Analysis: Document Classification
iv. Machine Learning algorithm development
Manager: Cliff Brunk, Ph.D.

Research Assistant, *Computer Science Department, UIUC* — Fall 2000 - Summer 2005
Projects: i. NLP tasks (shallow parsing, semantic role labeling)
ii. Algorithms and theory for complex-labeled output.
Advisor: Professor Dan Roth

Research Assistant, *National Center for Supercomputing Applications* — Fall 1998 - Spring 2000
Project: Clustering of automobile insurance customers.
Mentor: Michael Welge

Research Intern, *Los Alamos National Laboratory, Los Alamos, NM* Summer 1999
Project: Analysis of TCP/IP traffic in high traffic networks.
Mentor: Professor Wu Feng

Research Intern, *IBM Research, Yorktown Heights, NY* Spring 1998 - Summer 1998
Project: Data mining frequent sequential patterns.
Mentor: Murray Campbell, Ph.D.

Programming Experience

- Yahoo! (*perl, Matlab*)
  Web Search Feature Research and Development (`http://search.yahoo.com`).
  Answers Question Categorization (`http://answers.yahoo.com`).
  Document Classification Platform (Used internally).
- Co-created state-of-the-art shallow parsing system (*C++, perl*)
  A shallow parse includes noun and verb phrases. It is a subset of the full parse tree.
  Demo: `http://l2r.cs.uiuc.edu/~cogcomp/shallow_parse_demo.php`
- Co-created state-of-the-art semantic role labeling system (*awk, perl*)
  We label the semantic roles of verbs. For example, *who* ran *where*.
  Demo: `http://l2r.cs.uiuc.edu/~cogcomp/srl-demo.php`
- Constraint Classification as part of the SNoW architecture (*Matlab, C++*)
  A general purpose learning program for multi-class, ranking, and complex-labeled problems.
- Core-set based SVM, an approximate SVM solution (*Matlab, perl*)
  A provably correct approximate maximum margin solution that scales to large datasets.

Ongoing Research Projects

Maximum Margin Core-Sets

Finding the maximum margin hyperplane separating two point sets is an important, but time consuming, tool in state-of-the-art machine learning algorithms (e.g. SVM). We introduce core-sets for this problem to learn an approximate solution, allowing for a scalable extension of the SVM. Additionally, coresets provide a powerful theoretical tool for analyzing learning in the structured output domain, active learning, and learning in the presence of noise.

Web Search and Classification

My work at Yahoo! was focused on the core problem in web search, relevance ranking. There are many opportunities to use machine learning techniques to improve ranking, classification, and user interaction with the the massive amounts of data on the web both within the scope of relevance ranking and beyond. I am interested in utilizing user behavior, structure in information, and ranking in order to find and leverage information to aid the web (and search) experience.

Publications

Journal Articles

- Constraint Classification: A Generalization of Multiclass Classification and Category Ranking
with Sariel Har-Peled and Dan Roth
Journal of Artificial Intelligence Research (JAIR). In Submission.

Refereed Conference and Workshop Articles

- Maximum Margin Coresets for Active and Noise Tolerant Learning
with Sariel Har-Peled and Dan Roth
Proc. of the International Joint Conference on Artificial Intelligence (IJCAI), 2007.
- Learning and Inference over Constrained Output
with Vasin Punyakanok, Dan Roth, and Wen-tau Yih
Proc. of the International Joint Conference on Artificial Intelligence (IJCAI), 2005.
- Learning and Inference over Constrained Output
with Vasin Punyakanok, Dan Roth, and Wen-tau Yih
Learning-05 The Learning Workshop, 2005.
- Learning via Inference over Structurally Constrained Output
with Vasin Punyakanok, Dan Roth, and Wen-tau Yih.
The Conference on Advances in Neural Information Processing Systems (NIPS) Workshop on Learning Structured Output, 2004.
- Semantic Role Labeling via Generalized Inference over Classifiers Shared Task Paper
with Vasin Punyakanok, Dan Roth, Wen-tau Yih and Y. Tu
Proc. of the Annual Conference on Computational Natural Language Learning (CoNLL), 2004
- Semantic Role Labeling via Integer Linear Programming Inference
with Vasin Punyakanok, Dan Roth, Wen-tau Yih
Proc. the International Conference on Computational Linguistics (COLING), 2004.
- Constraint Classification for Multiclass Classification and Ranking
with Sariel Har-Peled and Dan Roth
The Conference on Advances in Neural Information Processing Systems (NIPS), 2002.
- Constraint Classification: a New Approach to Multiclass Classification
with Sariel Har-Peled and Dan Roth
Proc. of the International Conference on Algorithmic Learning Theory (ALT), 2002.
- A Learning Approach to Shallow Parsing
with Marcia Munoz, Vasin Punyakanok and Dan Roth
The Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP-VLC), 1999.

Additional Articles in Preparation/Submission

- Multi-Prototype Margin Perceptron
with Fabio Aiolli and Dan Roth

- Using Multiple Prototypes for Learning Over Multiple Views
  with Fabio Aiolli and Dan Roth

Presentations

- Learning via Inference over Structurally Constrained Output
  IJCAI, Hyderabad, India, January, 2007.
- Learning via Inference over Structurally Constrained Output
  NIPS Workshop on Learning Structured Output, Whistler, Canada, December, 2004.
- Constraint Classification for Multiclass Classification and Ranking
  NIPS, Whistler, Canada, November 2002.
- Constraint Classification: a New Approach to Multiclass Classification
  ALT, Luebeck, Germany, November 2002.
- A Learning Approach to Shallow Parsing
  EMNLP-VLC, University of Maryland, June 1999.

Awards and Honors

- 2005 DCI Postdoctoral Fellowship
- 2005 Beckman Summer Fellowship, University of Illinois
- 2005 IJCAI Travel Scoloarship
- 2002 Conference Travel Grant, University of Illinois

External Referee

- IJCAI 2005, 19th International Joint Conference on Artificial Intelligence
- ACL 2005, 43rd Annual Meeting of the Association for Computational Linguistics
- NIPS 2004, 18th Annual Conference on Neural Information Processing Systems
- ACL 2004, 42nd Annual Meeting of the Association for Computational Linguistics
- JAIR, Journal of Artificial Intelligence Research

References

Available upon request.