

Dan Zhang

Present Address:

1601 Willow Rd
Menlo Park, CA, 94025

Contact:

E-mail: danzhang2008@gmail.com

Current Position Research Scientist at Facebook

Research Interests Advertisement Measurement/Optimization, Text Mining, Machine Learning, Social network analysis, Large scale data analysis, Graph mining, Multimedia Analysis, Web Information Retrieval, Healthcare Data Analysis

Publications

Conference Papers

1. Qifan Wang, Luo Si, Dan Zhang, Learning to Hash with Partial Tags: Exploring Correlation Between Tags and Hashing Bits for Large Scale Image Retrieval, European Conference on Computer Vision (ECCV-14), Zurich, Switzerland, 2014
2. Dan Zhang, Yan Liu, Luo Si, Which Tweets Will Be Headlines? A Hierarchical Bayesian Model for Bridging Social Media and Traditional Media, The 8th KDD Workshop on Social Network Mining and Analysis for Business, Consumer and Social Insights (SNA-KDD), New York, 2014
3. Qifan Wang, Dan Zhang, Luo Si, Weighted Hashing for Fast Large Scale Similarity Search, ACM International Conference on Information and Knowledge Management (CIKM 2013), Burlingame, CA, 2013
4. Qifan Wang, **Dan Zhang**, Luo Si, Semantic Hashing using Tags and Topic Modeling, Annual International ACM SIGIR Conference (SIGIR-13), full paper, Dublin, Ireland, 2013
5. **Dan Zhang**, Jingrui He, Luo Si, Richard Lawrence, MILEAGE: Multiple Instance LEARNING with Global Embedding, In International Conference on Machine Learning (ICML), Atlanta, 2013
6. **Dan Zhang**, Jingrui He, Richard Lawrence, MI2LS: Multi-Instance Learning from Multiple Information Sources , Proceedings of the 19th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining (SIGKDD 13), oral paper, Chicago, IL, 2013
7. Qifan Wang, Luo Si, **Dan Zhang**, A Discriminative Data-dependent Mixture-Model Approach for Multiple Instance Learning in Image Classification, In Proceedings of the 12th European Conference on Computer Vision (ECCV), Firenze, Italy, 2012
8. **Dan Zhang**, Fei Wang, Luo Si, Composite Hashing with Multiple Information Sources, In Proceedings of the 34th Annual International ACM SIGIR Conference (SIGIR-11), full paper, Beijing, China, 2011
9. **Dan Zhang**, Jingdong Wang, Luo Si, Document Clustering with Universum, In Proceedings of the 34th Annual International ACM SIGIR Conference (SIGIR-11), full paper, Beijing, China, 2011
10. **Dan Zhang**, Yan Liu, Richard D. Lawrence, Vijil Chenthamarakshan. Transfer Latent Semantic Learning: Microblog Mining with Less Supervision, Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAI 11), oral and poster presentation, San Francisco, 2011
11. **Dan Zhang**, Yan Liu, Luo Si, Serendipitous Learning: Learning Beyond the Predefined Label Space, Proceedings of the 17th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining (SIGKDD 11), San Diego, CA

12. **Dan Zhang**, Jingrui He, Yan Liu, Luo Si, Richard D. Lawrence, Multi-View Transfer Learning with a Large Margin Approach, Proceedings of the 17th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining (SIGKDD 11), San Diego, CA
13. **Dan Zhang**, Yan Liu, Luo Si, Jian Zhang, Richard D. Lawrence, Multiple Instance Learning on Structred Data, Twenty-Fifth Annual Conference on Neural Information Processing Systems (NIPS), Granada, Spain, 2011.
14. Mohammad Taha Bahadori, Yan Liu, **Dan Zhang**, Learning with Minimum Supervision: A General Framework for Transductive Transfer Learning, IEEE International Conference on Data Mining (ICDM), full paper, Vancouver, Canada, 2011 (**Selected as a best ICDM'11 paper for publication in Knowledge and Information Systems (KAIS).**)
15. **Dan Zhang**, Yan Liu, Richard D. Lawrence, Vijil Chenthamarakshan. ALPOS – Assisted Learning for Partial ObServation: A Machine Learning Approach for Classifying Tweets, Workshop on Optimization Based Methods for Emerging Data Mining Problems, in Conjunction with the IEEE International Conference on Data Mining (ICDM), Sydney, AU, 2010
16. **Dan Zhang**, Fei Wang, Luo Si, Tao Li. M³IC: Maximum Margin Multiple Instance Clustering, the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09), Pasadena, California, USA, 2009
17. **Dan Zhang**, Luo Si, Modeling Search Response Time, in the Proceedings of the 32th Annual International ACM SIGIR Conference (SIGIR-09), postor, Boston, Massachusetts, U.S.A., 2009
18. **Dan Zhang**, Luo Si, Multiple Instance Transfer Learning, Workshop on Optimization Based Methods for Emerging Data Mining Problems, in Conjunction with the IEEE International Conference on Data Mining (ICDM), Miami, Florida, USA, 2009
19. **Dan Zhang**, Jingdong Wang, Fei Wang, Changshui Zhang. Semi-Supervised Classification with Universum, SIAM International Conference on Data Mining (SDM08), Atlanta, Georgia, USA, 2008 .
20. **Dan Zhang**, Fei Wang, Changshui Zhang, Tao Li. Multi-View Local Learning, Twenty-Third AAAI Conference on Artificial Intelligence (AAAI 08), Chicago, Illinois, USA, 2008.
21. **Dan Zhang**, Fei Wang, Zhenwei Shi, Changshui Zhang. Localized Content-Based Image Retrieval Using Multiple Instance Active Learning, the 15th IEEE International Conference on Image Processing (ICIP 08), San Diego, California, U.S.A., 2008
22. **Dan Zhang**, Zhenwei Shi, Yangqiu Song and Changshui Zhang. Localized Content-Based Image Retrieval Using Semi-Supervised Multiple Instance Learning. Asian Conference on Computer Vision (ACCV07), Tokyo, Japan, 2007
23. **Dan Zhang**, Yangqiu Song, Qifeng Qiao, Zhenwei Shi and Changshui Zhang. A Supervised Method to Chart Multiple Manifolds. Asian Conference on Computer Vision (ACCV07) Workshop on Subspace, Tokyo, Japan, 2007.
24. Zhiyao Duan, **Dan Zhang**, Changshui Zhang, Zhenwei Shi, Multi-Pitch Estimation Based On Partial Event and Support Transfer. IEEE International Conference on Multimedia & Expo (ICME07), Beijing, China, 2007
25. Zhenwei Shi, **Dan Zhang**, Changshui Zhang, A Fixed-Point Algorithm For Blind Separation of Temporally Correlated Sources. IEEE International Conference on Multimedia & Expo (ICME07), Beijing, China, 2007

Journal Papers

1. Mohammad Taha Bahadori, Yan Liu, **Dan Zhang**, Learning with Minimum Supervision: A General Framework for Transductive Transfer Learning, Knowledge and Information Systems (KAIS), accepted, 2013
2. **Dan Zhang**, Luo Si, Vernon J. Rego, Sentiment Detection with Auxiliary Data, Journal of Information Retrieval, accepted, 2012
3. **Dan Zhang**, Fei Wang, Luo Si, Tao Li. Maximum Margin Multiple Instance Clustering with its applications to Image and Text Clustering. IEEE Transaction on Neural Networks, vol. 22, no. 5. 739-751, 2011
4. **Dan Zhang**, Fei Wang, Zhenwei Shi, Changshui Zhang. Interactive Localized Content-Based Image Retrieval with Multiple Instance Active Learning. Pattern Recognition, Volume 43, Issue 2, February 2010, Pages 478-484.
5. Yangqiu Song, Changshui Zhang, Jianguo Lee, Fei Wang, Shiming Xiang and **Dan Zhang**. Semi-Supervised Discriminative Method for Tumorous Tissues Segmentation of MR Brain Images. Pattern Analysis and Applications, 2008.
6. Shiliang Sun, Changshui Zhang, **Dan Zhang**. An experimental evaluation of ensemble methods for EEG signal classification. Pattern Recognition Letters, in 2007
7. Zhenwei Shi, **Dan Zhang** and Changshui Zhang, MACBSE: Extracting signals with linear autocorrelations , Neurocomputing , 2007

Patents

- With Yan Liu, Richard D. Lawrence, Vijil Chenthamarakshan, A System and Method for Domain Adaption with Partial Observations , US patent 20130013539, 2011
- With Daniel Jiang, Linji Yang, Jerry Fu, Ads Buying Prediction (not the exact name,. Concrete title is currently confidential), US 14/292,277, May 2014

System Design

- With Dung Trung Hong, Qifan Wang and Luo Si, Query Expansion and Message-passing Algorithms for TREC Microblog Track, **No.5 out of 59 teams in Microblog Track of TREC competition**, 2011

Education

Purdue University, West Lafayette, IN, USA

PHD, Computer Science Department, Aug. 2008 - Dec. 2012

- Advisor: Prof. Luo Si
- PHD Committee members: Prof. Luo Si, Prof. Jiawei Han, Prof. Jennifer Neville, Prof. Vernon Rego, Prof. Walid Aref

Tsinghua University, Beijing, P.R.China

M.S., Department of Automation, Aug. 2005 - July 2008

- Dissertation: "Multiple Instance Learning with Insufficient Supervised Information"
- Advisor: Prof. Changshui Zhang

Honors And Awards

- PhD dissertation nominated for the Outstanding Dissertation Award of Purdue University, 2012
- IBM Ph.D. Fellowship, 2012–2013
- Facebook Ph.D. Fellowship Finalist, 2012
- NIPS Travel Grant, 2011
- ICDM best paper candidate for publication in KAIS, 2011

- AAAI scholarship for Travel, 2011
- SIGIR student travel grant, 2011
- SIGKDD student travel grant, 2011
- IJCAI student Travel Grant, 2009
- Excellent Master's Thesis Award (Top 5%), 2008
- Excellent Research Scholarship for graduate students, 2007
- Excellent Bachelor's Thesis Award, 2005
- "Meritorious Winner" of the (International) Mathematical/ Interdisciplinary Contest In Modeling, 2004
- Awards for the best performance in one specific academic area, 2004
- First-Class Studying Excellence Scholarship (Top 3%), 2003
- Honor of Tri-A Student (Top 2%), 2003
- Rank 2nd in National Mathematical Modeling Contest of College Students in the Province of Jiangsu, 2003
- The first prize of advanced mathematics competition of Jiangsu Province (Top 1%), 2002
- First-Class Studying Excellence Scholarship (Top 3%), 2002
- Awards for the best performance in social activities, 2002

Work Experience Research Scientist in Facebook Research Center, **Dec 2012- Now**

- Apply machine learning and data mining methods to ads delivery and measurement.

Research Intern in IBM TJ Watson Research Center, Yorktown Heights, **May 2012-Aug. 2012** Mentor: Jingrui He, Manager: Richard D. Lawrence
(Led to 1 SIGKDD paper and 1 ICML paper)

- Developed a multi-instance learning method that could tune feature representations adaptively
- Developed a multi-instance multi-view learning method for insider threat detection

Research Intern in IBM TJ Watson Research Center, Yorktown Heights, **May 2010-Aug. 2010** Mentor: Yan Liu, Manager: Richard D. Lawrence
(Led to 1 SIGKDD paper, 1 NIPS paper, 1 AAAI paper, and 1 patent)

- Developed a transfer learning method for Twitter tagging
- Developed a multi-instance method for market analysis
- Developed a multi-view transfer learning method

Research Assistant in Purdue University, West Lafayette, **Aug. 2008- Now**

- Developed a hashing method for text retrieval and healthcare data
- Developed a clustering method that incorporates the irrelevant data
- Proposed and solved a problem that considers the label space differences between training and testing
- Developed a discriminative data-dependent mixture-model approach for multi-instance learning
- Developed a maximum margin multiple instance clustering framework

- Developed a method to model the relationship between the searching time delay and the corresponding query in distributed Information Retrieval
- Developed a multiple instance transfer learning algorithm

Research Assistant in Tsinghua University, **Aug. 2005- July 2008**

- Developed a multiple instance active learning framework
- Developed a semi-supervised universum algorithm
- Developed a multi-view local learning algorithm
- Developed a multiple instance semi-supervised learning method
- Developed a method to chart multiple manifolds
- Co-Developed a method to estimate the pitches in music signal analysis
- Co-Developed a fixed-point algorithm for the Blind Source Separation (BSS)
- Co-Developed a Blind Source Separation method based on MACBSE

Professional Activities

Session Chair of

- ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining, 2014
- ACM Conference on Information and Knowledge Management (CIKM), 2013

Program Committee of

- ACM Multimedia Workshop on Social Behavioral Networked Media Access, 2011
- European Conference on Information Retrieval (ECIR), 2012
- ACM Conference on Information and Knowledge Management (CIKM), 2012
- European Conference on Information Retrieval (ECIR), 2013
- ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining , 2013
- ACM Conference on Information and Knowledge Management (CIKM), 2013
- IEEE International Conference on Big Data, 2013
- International Joint Conference on Artificial Intelligence 2013
- European Conference on Information Retrieval (ECIR), 2014
- ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining , 2014

Reviewer of

- Journal of Information Retrieval, IEEE Transactions on Multimedia, Data Mining and Knowledge Discover, Neurocomputing, Information Sciences, International Journal of Pattern Recognition and Artificial Intelligence, Journal of World Wide Web, Pattern Recognition Letters, Topics in Computer Science, IEEE Transactions on Knowledge and Data Engineering, Journal of System and Software, IEEE Transactions on Neural Networks and Learning Systems

Teaching Experience

Teaching Assistant in Purdue University, West Lafayette, **Aug. 2009- May 2010**

- CS240 - Programming in C, Tasks including teaching a lab, grading and consulting.

Computer Skills

Proficient in C/C++, MATLAB, Perl, Python, PHP, LaTeX etc.
Experience with JAVA, Unix Shell etc.

Certifications

1. Teaching Assistant Certification, Purdue University, West Lafayette, 2008
2. Chinese National Software Designer Certification (Senior Programmer), 2004
3. Chinese National Programmer Certification, 2003

**Immigration
Status**

US Permanent Resident

Homepage

You can find my homepage at:
<http://www.cs.purdue.edu/homes/zhang168/>