Wei Chu

Personal Information

US Permanent Resident
Nationality: P.R. China
Tel: (86)177-1018-3425

Email: email.chuwei@gmail.com
Homepage: http://chuwei.website/
LinkedIn: https://www.linkedin.com/in/chuwei/

About Me

I am a pragmatic Bayesian, an award-winning researcher, a China's Innovation Talent elected in 2016, and a R&D team leader, with well-balanced academia and industry experience of 15 years. Fascinated by the power of distributed computing for large scale learning tasks in the Internet industry, I am now with Alibaba Group, leading a R&D team of 50+ researchers and engineers to develop distributed machine learning platform for Alibaba Cloud, including distributed deep learning implementation on GPU cluster, online service for predictive models etc. Previously I was a team leader at Microsoft Bing to develop personalized search service. At Yahoo! Labs I worked with colleagues on web-scale user-click stream for content optimization via contextual bandits.

My academic interest is to design and implement statistical learning algorithms, to discover useful patterns in enormous machine-readable data that might otherwise not be found by human inspection. I conducted several research at CCLS, Columbia University, including relational Gaussian processes, SVCR and p-Tucker. I worked with Zoubin Ghahramani and David L. Wild on statistical machine learning as a post-doctoral fellow at the Gatsby Computational Neuroscience Unit, University College London. I received my Ph.D. degree at the National University of Singapore, under the joint guidance of S. Sathiya Keerthi and Chong Jin Ong with a thesis titled "Bayesian approach to support vector machines".

Education

PostDoc, Feb. 2003 – Jan. 2006
Major: Statistical Machine Learning
Gatsby Computational Neuroscience Unit, University College London, UK

Ph.D., Jul. 1999 – Jan. 2003
Major: Machine Learning
National University of Singapore (NUS), Singapore

Master of Engineering, Sept. 1995 – Jan. 1998
Major: Inertial Navigation Technology and Equipment
Harbin Institute of Technology, Harbin, P.R. China

Bachelor of Engineering, Sept. 1991 – Jul. 1995
Major: Automatic Control
Harbin Engineering University, Harbin, P.R. China

Working Experience

Director of Engineering, Nov. 2014 – present
Large Scale Learning, Alibaba Cloud, Alibaba Group, Hangzhou, China

Principal Applied Scientist Lead, Jan. 2014 – Nov. 2014
Senior Applied Researcher, May 2011 – Jan. 2014
Contextual Relevance, Bing, Microsoft, Seattle, USA

Scientist, Jan. 2008 – May 2011
Audience Science, Yahoo! Lab, Sunnyvale, USA

Associate Research Scientist, Jan. 2006 – Jan. 2008
Center for Computational Learning Systems, Columbia University, New York, USA

Journal Article & Book Chapter

1. T. Moon, **W. Chu**, L. Li, Z. Zheng, Y. Chang (2012) Online learning framework for refining recency search results with user click feedback, Transactions on Information Systems 30(4)
2. **W. Chu** and S. S. Keerthi (2007) Support vector ordinal regression, Neural Computation 19(3):792-815
3. **W. Chu**, Z. Ghahramani, A. Podtelezhnikov and D. L. Wild (2006) Bayesian segmental models with multiple sequence alignment profiles for protein secondary structure and contact map prediction, IEEE/ACM Transactions on Computational Biology and Bioinformatics 3(2):98-113
4. **W. Chu**, S. S. Keerthi, C. J. Ong and Z. Ghahramani (2006) Bayesian support vector machines for feature ranking and selection, In I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, editors, Feature Extraction, Foundations and Applications Springer:403-418
5. **W. Chu**, Z. Ghahramani, F. Falciani and D. L. Wild (2005) Biomarker discovery with Gaussian processes in microarray gene expression data, Bioinformatics 2005(21):3385-3393
6. **W. Chu** and Z. Ghahramani (2005) Gaussian processes for ordinal regression, Journal of Machine Learning Research 6(Jul):1019-1041
7. **W. Chu**, C. J. Ong and S. S. Keerthi (2005) An improved conjugate gradient scheme to the solution of least squares SVM, IEEE Transactions on Neural Networks 16(2):498-501
8. **W. Chu**, S. S. Keerthi and C. J. Ong (2004) Bayesian support vector regression using a unified loss function, IEEE Transactions on Neural Networks 15(1):29-44
9. K. Duan, S. S. Keerthi, **W. Chu**, S. K. Shevade and A. N. Poo (2003) Multi-category classification by soft-max combination of binary classifiers, Multiple Classifier Systems (MCS-04) Lecture Notes in Computer Science 2709 Springer:125-134
10. **W. Chu**, S. S. Keerthi and C. J. Ong (2003) Bayesian trigonometric support vector classifier, Neural Computation 15(9):2227-2254

Refereed Conference

11. M. Qiu, et al. (2017) AliMe Chat: a sequence to sequence and rerank based ChatBot engine, Annual Meeting of the Association for Computational Linguistics (ACL-55 Short Paper) in press
12. B. Bi, H. Ma, B. Hsu, **W. Chu**, K. Wang and J. Cho (2015) Learning to recommend related entities to search users, ACM International Conference on Web Search and Data Mining (WSDM-08)
13. J. Yan, **W. Chu**, R. W. White (2014) Cohort modeling for enhanced personalized search, ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-37)
14. H. Wang, X. He, M. Chang, Y. Song, R. W. White, **W. Chu** (2013) Personalized ranking model adaptation for web search, ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-36)
15. R. W. White, **W. Chu**, A. Hassan, X. He, Y. Song, H. Wang (2013) Enhancing personalized search by mining and modeling task behavior, International World Wide Web Conference (WWW-22)
16. H. Wang, Y. Song, M. Chang, X. He, R. W. White, **W. Chu** (2013) Learning to extract cross-session search tasks, International World Wide Web Conference (WWW-22)
17. P. Bennett, R. W. White, **W. Chu**, S. Dumais, P. Bailey, F. Borisyuk and X. Cui (2012) Modeling and measuring the impact of short and long-term behavior on search personalization, ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-35)

18. **W. Chu**, M. Zinkevich, L. Li, A. Thomas, and B. Tseng (2011) Unbiased online active learning in data streams, ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD-17)
19. L. Zhang, J. Yang, **W. Chu**, and B. Tseng (2011) A machine-learned proactive moderation system for auction fraud detection, ACM Conference on Information Retrieval and Knowledge Management (CIKM-20 Short Paper)
20. L. Li, **W. Chu**, J. Langford and X. Wang (2011) Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms, ACM International Conference on Web Search and Data Mining (WSDM-04) 297-306 **Winner of the Best Paper Award**
21. **W. Chu**, L. Li, L. Reyzin, and R. E. Schapire (2011) Contextual bandits with linear payoff functions, International Conference on Artificial Intelligence and Statistics (AISTATS-14)
22. T. Moon, L. Li, **W. Chu**, C. Liao, Z. Zheng and Y. Chang (2010) Online learning for recency search ranking using real-time user feedback, International Conference on Information and Knowledge Management (CIKM-19 Short Paper) 1501-1504
23. L. Li, **W. Chu**, J. Langford and R. E. Schapire (2010) A contextual-bandit approach to personalized news article recommendation, International World Wide Web Conference (WWW-19) 661-670
24. S.-T. Park and **W. Chu** (2009) Pairwise preference regression for cold-start recommendation, ACM Recommender Systems (RecSys-03):21-28
25. **W. Chu** and Z. Ghahramani (2009) Probabilistic models for incomplete multi-dimensional arrays, International Conference on Artificial Intelligence and Statistics (AISTATS-12):89-96
26. **W. Chu** and S.-T. Park (2009) Personalized recommendation on dynamic content using predictive bilinear models, International World Wide Web Conference (WWW-18):692-700
27. **W. Chu**, et al. (2009) A case study of behavior-driven conjoint analysis on Yahoo! Front Page Today Module, ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD-15 Industry Track):1097-1104
28. R. Silva, **W. Chu** and Z. Ghahramani (2007) Hidden common cause relations in relational learning, Neural Information Processing Systems (NIPS-20):1345-1352
29. K. Yu and **W. Chu** (2007) Gaussian process models for link analysis and transfer learning, Neural Information Processing Systems (NIPS-20):1657-1664
30. P. K. Shivaswamy, **W. Chu** and M. Jansche (2007) A support vector approach to censored targets, IEEE International Conference on Data Mining (ICDM-07):655-660
31. V. Sindhwani, **W. Chu** and S. S. Keerthi (2007) Semi-supervised Gaussian process classifiers, International Joint Conferences on Artificial Intelligence (IJCAI-20):1059-1064
32. **W. Chu**, V. Sindhwani, Z. Ghahramani and S. S. Keerthi (2006) Relational learning with Gaussian processes, Neural Information Processing Systems (NIPS-19):289-296
33. K. Yu, **W. Chu**, S. Yu, V. Tresp and Z. Xu (2006) Stochastic relational models for discriminative link prediction, Neural Information Processing Systems (NIPS-19):1553-1560
34. S. K. Shevade and **W. Chu** (2006) Minimum enclosing spheres formulations for support vector ordinal regression, IEEE International Conference on Data Mining (ICDM-06):1054-1058
35. **W. Chu**, Z. Ghahramani, R. Krause and D. L. Wild (2006) Identifying protein complexes in high-throughput protein interaction screens using an infinite latent feature model, Pacific Symposium on Biocomputing (PSB-11):231-242
36. **W. Chu** (2006) Model selection: an empirical study on two kernel classifiers, International Joint Conference on Neural Networks (IJCNN-06):1673-1679
37. S. S. Keerthi and **W. Chu** (2005) A matching pursuit approach to sparse Gaussian process regression, Neural Information Processing Systems (NIPS-18):643-650
38. **W. Chu** and Z. Ghahramani (2005) Preference learning with Gaussian processes, International Conference on Machine Learning (ICML-22):137-144

39. **W. Chu** and S. S. Keerthi (2005) New approaches to support vector ordinal regression, International Conference on Machine Learning (ICML-22):145-152
40. **W. Chu**, Z. Ghahramani and D. L. Wild (2004) A graphical model for protein secondary structure prediction, International Conference on Machine Learning (ICML-21):161-168
41. **W. Chu**, Z. Ghahramani and D. L. Wild (2004) Protein secondary structure prediction using sigmoid belief networks to parameterize segmental semi-Markov models, European Symposium on Artificial Neural Networks (ESANN-05):81-86
42. **W. Chu**, S. S. Keerthi and C. J. Ong (2002) A general formulation for support vector machines, International Conference on Neural Information Processing (ICONIP-09)
43. **W. Chu**, S. S. Keerthi and C. J. Ong (2002) A new Bayesian design method for support vector classification, International Conference on Neural Information Processing (ICONIP-09)
44. S. S. Keerthi, et al. (2002) A machine learning approach for the curation of Biomedical literature - KDD Cup 2002 (Task 1), SIGKDD Explorations Newsletter, 4(2)
45. **W. Chu**, S. S. Keerthi and C. J. Ong (2001) A unified loss function in Bayesian framework for support vector regression, International Conference on Machine Learning (ICML-18):51-58

Refereed Workshop

46. X. Li, C. Guo, **W. Chu**, Y. Wang, J. Shavlik (2014) Deep learning powered in-session contextual ranking using clickthrough data, Workshop on Personalization: Methods and Applications, at Neural Information Processing Systems (NIPS-27)
47. L. Li, **W. Chu**, J. Langford, T. Moon, and X. Wang (2012) An unbiased offline evaluation of contextual bandit algorithms with generalized linear models, Journal of Machine Learning Research - Workshop and Conference Proceedings 26 (JMLR W&CP-26)
48. **W. Chu** and Z. Ghahramani (2005) Extensions of Gaussian processes for ranking: semi-supervised and active learning, Workshop Learning to Rank at (NIPS-18):29-34

Thesis

49. **W. Chu** (2003) Bayesian approach to support vector machines, Doctoral Dissertation, National University of Singapore

US Patents

50. User trustworthiness, US Patent 9519682 B1
51. Determining user preference of items based on user ratings and user features, US Patent 8301624 B2
52. Predicting item-item affinities based on item features by regression, US Patent 8442929 B2
53. Enhanced matching through explore/exploit schemes, US Patent 8244517 B2
54. Dynamic estimation of the popularity of web content, US App. 20100241597 A1
55. Conjoint analysis with bilinear regression models for segmented predictive content ranking, US App. 20100125585 A1
56. Methods and systems relating to ranking functions for multiple domains, US App. 20110087673 A1
57. Contextual-bandit approach to personalized news article recommendation, US App. 20120016642 A1
58. Feature-based method and system for cold-start recommendation of online ads, US App. 20110112981 A1
59. Online active learning in user-generated content streams, US App. 20130111005 A1
60. Personalized recommendations on dynamic content, US App. 20100211568 A1

Honors and Awards

- China's National Innovation Talent, 2016
- Best Paper Award, ACM WSDM, 2011
- Super Star Team Award, Yahoo!, 2008
- Honorable Mention Team, ACM KDD CUP, 2002

Professional Activity

Peer Reviewer: BMC Bioinformatics, IEEE Transactions on Evolutionary Computation, IEEE Transactions on Neural Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Systems, Man, and Cybernetics, Journal of Machine Learning Research, Machine Learning Journal, Neurocomputing, Neural Computation, Operations Research

Program Committee Member: ICML, SIGIR, ECML/PKDD Workshop 2008

Peer Reviewing for Conferences: NIPS, ICML, AISTATS, ECML, ESANN, PSB, WWW

References

Available upon request