Wei Chu

Personal Information

U.S. Permanent Resident
Phone: 206-471-4732
Email: email.chuwei@gmail.com
Homepage: https://weichu.github.io/

About Me

I am a seasoned R&D team leader and machine learning researcher, with ∼20 years of well-balanced academic and industry experience. I have published 50+ papers at top-tier conferences and journals and have received over 10,000 citations according to Google Scholar.

I am an expert in statistical machine learning and have completed a 3-year postdoc training in probabilistic kernel machines at the Gatsby Unit, UCL. I also did research at Columbia University and Yahoo! Labs. I gained experience on recommender systems at Yahoo! Labs and Microsoft Bing, to model user behavior from web-scale clickstream data and design an unbiased offline evaluator for content optimization.

I have extended my knowledge scope to congitive intelligence through deep learning techniques. At Alibaba Group, I led a team to develop a distributed large-scale learning platform, and deliver the platform product to Alibaba Cloud. Now I lead a team at Ant Group to develop multi-modal learning solutions for computer vision, natural language understanding and knowledge graphs.

I am looking for opportunities to design and deliver learning algorithms that transform large-scale machine-readable data into human-comprehensible knowledge that not only has a major impact on human life, but also makes machine intelligence more equitable and trustworthy.

Working Experience

Senior Director of Engineering, Jul. 2018 – present
Director of Engineering, Jul. 2017 – Jul. 2018
Cognitive Computing, Ant Group, Alibaba Group, Bellevue, USA
Leading an R&D team of 150+ researchers and engineers to develop multi-modal learning solutions for computer vision, natural language understanding and knowledge graph.

Director of Engineering, Nov. 2014 – Jul. 2017
Large Scale Learning, Alibaba Cloud, Alibaba Group, Hangzhou, China
An R&D team leader to develop a distributed machine learning platform, and deliver the platform product PAI 2.0 to Alibaba Cloud, including the implementation of hundreds of distributed learning algorithms on clusters and online services of predictive models.

Principal Applied Scientist Lead, Jan. 2014 – Nov. 2014
Senior Applied Researcher, May 2011 – Jan. 2014
Contextual Relevance, Bing, Microsoft, Seattle, USA
A team leader at Microsoft Bing to deliver personalized search service.

Scientist, Jan. 2008 – May 2011
Audience Science, Yahoo! Lab, Sunnyvale, USA
Working with colleagues on web-scale user click streams for content optimization.

Associate Research Scientist, Jan. 2006 – Jan. 2008
Center for Computational Learning Systems, Columbia University, New York, USA
Conducting independent research on pragmatic Bayesian techniques.

Education

Post Doctoral, Feb. 2003 – Jan. 2006
Gatsby Computational Neuroscience Unit, University College London (UCL), UK
Advisor: Prof. Zoubin Ghahramani

Ph.D., Jul. 1999 – Jan. 2003
National University of Singapore (NUS), Singapore
Advisor: Prof. Sathiya Keerthi and Prof. Chong Jin Ong

Master of Engineering, Sept. 1995 – Jan. 1998
Harbin Institute of Technology, Harbin, China

Bachelor of Engineering, Sept. 1991 – Jul. 1995
Harbin Engineering University, Harbin, China

Publications

1. W. Hong, J. Lao, W. Ren, J. Wang, J. Chen, **W. Chu** (2022) Training object detectors from scratch: An empirical study in the era of vision transformer, in Proc. of CVPR 2022
2. H. Wang, T.-W. Chang, T. Liu, J. Huang, Z. Chen, C. Yu, R. Li, **W. Chu** (2022) ESCM2: Entire space counterfactual multi-task model for post-click conversion rate estimation, in Proc. of SIGIR 2022
3. K. Ji, J. Liu, W. Hong, L. Zhong, J. Wang, J. Chen, **W. Chu** (2022) CRET: Cross-modal retrieval transformer for efficient text-video retrieval, in Proc. of SIGIR 2022
4. M. Li, X. Lin, X. Chen, J. Chang, Q. Zhang, F. Wang, T. Wang, Z. Liu, **W. Chu**, D. Zhao and R. Yan (2022) Keywords and instances: A hierarchical contrastive learning framework unifying hybrid granularities for text generation, in Proc. of ACL 2022
5. F. Yu, K. Huang, M. Wang, Y. Cheng, **W. Chu**, and C. Li (2022) Width & depth pruning for vision transformers, in Proc. of AAAI 2022
6. H. Huang, Y. Wang, Z. Chen, Y. Zhang, Y. Li, Z. Tang, **W. Chu**, J. Chen, W. Lin, and K.-K. Ma (2022) CMUA-Watermark: A cross-model universal adversarial watermark for combating deepfakes, in Proc. of AAAI 2022
7. L. Chao, J. He, T. Wang and **W. Chu** (2021) PairRE: Knowledge graph embeddings via paired relation vectors, ACL 2021: 4360-4369
8. F. Xu, M. Wang, W. Zhang, Y. Cheng and **W. Chu** (2021) Discrimination-aware mechanism for fine-grained representation learning, CVPR 2021
9. W. Hong, P. Guo, W. Zhang, J. Chen and **W. Chu** (2021) LPSNet: A lightweight solution for fast panoptic segmentation, CVPR 2021
10. W. Hong, K. Ji, J. Liu, J. Wang, J. Chen and **W. Chu** (2021) GilBERT: Generative vision-language pre-training for image-text retrieval, SIGIR 2021: 1379-1388
11. C. Jiang, K. Huang, S. He, X. Yang, W. Zhang, X. Zhang, Y. Cheng, L. Yang, Q. Wang, F. Xu, T. Pan and **W. Chu** (2021) Learning segment similarity and alignment in large-scale content based video retrieval, ACM MM 2021
12. K. Chen, W. Xu, X. Cheng, X. Zou, Y. Zhang, L. Song, T. Wang, Y. Qi and **W. Chu** (2020) Question directed graph attention network for numerical reasoning over text, EMNLP 2020:6759-6768
13. L. Chao, J. Chen and **W. Chu** (2020) Variational connectionist temporal classification, ECCV 2020:460-476
14. X. Chen, W. Xu, K. Chen, T. Wang, S. Jiang, F. Wang, **W. Chu** and Y. Qi (2020) SpellGCN: Incorporating phonological and visual similarities into language models for Chinese Spelling Check, ACL 2020:871-881

15. X. Lin, W. Jian, J. He, T. Wang, and **W. Chu** (2020) Generating informative conversational response using recurrent knowledge-interaction and knowledge-copy, ACL 2020:41-52
16. F. Xu, W. Zhang, Y. Cheng and **W. Chu** (2020) Metric learning with equidistant and equidistributed triplet-based loss for product image search, WWW 2020:57-65
17. S. Wang, B. Zhu, C. Li, M. Wu, J. Zhang, **W. Chu**, and Y. Qi (2020) Riemannian proximal policy optimization, Computer and Information Science 13(3)
18. W. Zhang, Y. Cheng, X. Guo, Q. Guo, J. Wang, Q. Wang, C. Jiang, M. Wang, F. Xu and **W. Chu** (2020) Automatic car damage assessment system: reading and understanding videos as professional insurance inspectors, AAAI 2020:13646-13647 Demonstration Track
19. W. Huang, X. Cheng, K. Chen, T. Wang, **W. Chu** (2020) Towards fast and accurate neural Chinese word segmentation with multi-criteria learning, COLING 2020:2062-2072
20. C. Li, X. Yan, X. Deng, Y. Qi, **W. Chu**, L. Song, J. Qiao, J. He and J. Xiong (2019) Latent dirichlet allocation for Internet price war, AAAI 2019:639-646
21. X. Cheng, W. Xu, T. Wang, **W. Chu**, W. Huang, K. Chen and J. Hu (2019) Variational semi-supervised aspect-term sentiment analysis via transformer, CoNLL 2019:961-969
22. W. Huang, X. Cheng, T. Wang and **W. Chu** (2019) BERT-based multi-head selection for joint entity-relation extraction, NLPCC (2) 2019:713-723
23. W. Sui, Q. Zhang, J. Yang and **W. Chu** (2018) A novel integrated framework for learning both text detection and recognition, ICPR 2018:2233-2238
24. T. Yin, X. Deng, Y. Qi, **W. Chu**, J. Pan, X. Yan and J. Xiong (2018) Personalized behavior prediction with encoder-to-decoder structure, NAS 2018:1-10
25. J. Yu, M. Qiu, J. Jiang, J. Huang, S. Song, **W. Chu** and H. Chen (2018) Modelling domain relationships for transfer learning on retroeval-based question answering systems in E-commerce, ACM International Conference on Web Search and Data Mining (WSDM-11):682-690
26. M. Qiu, P. Zhao, K. Zhang, X. Shi, X. Wang, J. Huang and **W. Chu** (2017) A short-term rainfall prediction model using multi-task convolutional neural networks, IEEE International Conference on Data Mining (ICDM)
27. F. Li et al. (2017) AliMe Assist: an intelligent assistant for creating an innovative E-commerce experience, ACM International Conference on Information and Knowledge Management (CIKM) *Winner of the Best Demo Award*
28. M. Qiu, F.-L. Li, S. Wang, X. Gao, Y. Chen, W. Zhao, H. Chen, J. Huang and **W. Chu** (2017) AliMe Chat: A Sequence to Sequence and Rerank based Chatbot Engine, Annual Meeting of the Association for Computational Linguistics (ACL-55 Short Paper)
29. J. Yang, Y. Chen, S. Wang, L. Li, C. Meng, M. Qiu, **W. Chu** (2017) Practical lessons of distributed deep learning, Workshop on Principled Approaches to Deep Learning, at ICML
30. 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):139-148
31. 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)
32. 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)
33. 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)
34. 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)

35. 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):1353-1364
36. T. Moon, **W. Chu**, L. Li, Z. Zheng, Y. Chang (2012) An online learning framework for refining recency search results with user click feedback, Transactions on Information Systems 30(4)
37. 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)
38. 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)
39. **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)
40. 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)
41. 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*
42. **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)
43. 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
44. 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
45. S.-T. Park and **W. Chu** (2009) Pairwise preference regression for cold-start recommendation, ACM Recommender Systems (RecSys-03):21-28
46. **W. Chu** and Z. Ghahramani (2009) Probabilistic models for incomplete multi-dimensional arrays, International Conference on Artificial Intelligence and Statistics (AISTATS-12):89-96
47. **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
48. **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
49. R. Silva, **W. Chu** and Z. Ghahramani (2007) Hidden common cause relations in relational learning, Neural Information Processing Systems (NIPS-20):1345-1352
50. K. Yu and **W. Chu** (2007) Gaussian process models for link analysis and transfer learning, Neural Information Processing Systems (NIPS-20):1657-1664
51. 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
52. **W. Chu** and S. S. Keerthi (2007) Support vector ordinal regression, Neural Computation 19(3):792-815
53. V. Sindhwani, **W. Chu** and S. S. Keerthi (2007) Semi-supervised Gaussian process classifiers, International Joint Conferences on Artificial Intelligence (IJCAI-20):1059-1064
54. **W. Chu**, V. Sindhwani, Z. Ghahramani and S. S. Keerthi (2006) Relational learning with Gaussian processes, Neural Information Processing Systems (NIPS-19):289-296

55. 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
56. 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
57. **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
58. **W. Chu** (2006) Model selection: an empirical study on two kernel classifiers, International Joint Conference on Neural Networks (IJCNN-06):1673-1679
59. **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
60. **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
61. **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
62. **W. Chu** and Z. Ghahramani (2005) Gaussian processes for ordinal regression, Journal of Machine Learning Research 6(Jul):1019-1041
63. **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
64. S. S. Keerthi and **W. Chu** (2005) A matching pursuit approach to sparse Gaussian process regression, Neural Information Processing Systems (NIPS-18):643-650
65. **W. Chu** and Z. Ghahramani (2005) Preference learning with Gaussian processes, International Conference on Machine Learning (ICML-22):137-144
66. **W. Chu** and S. S. Keerthi (2005) New approaches to support vector ordinal regression, International Conference on Machine Learning (ICML-22):145-152
67. **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
68. **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
69. **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
70. **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
71. **W. Chu** (2003) Bayesian approach to support vector machines, Doctoral Dissertation, National University of Singapore
72. 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
73. **W. Chu**, S. S. Keerthi and C. J. Ong (2003) Bayesian trigonometric support vector classifier, Neural Computation 15(9):2227-2254
74. **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)
75. **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)

76. 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) *Honorable Mention*
77. **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

US Patents

78. User trustworthiness, US Patent 9519682 B1
79. Determining user preference of items based on user ratings and user features, US Patent 8301624 B2
80. Predicting item-item affinities based on item features by regression, US Patent 8442929 B2
81. Enhanced matching through explore/exploit schemes, US Patent 8244517 B2
82. Character recognition method and device, US Patent 10872274 B2
83. Segmentation-based damage detection, US Patent 10783643 B1
84. Methods and systems relating to ranking functions for multiple domains, US Patent 10019518 B2
85. Personalized recommendations on dynamic content, US Patent 9600581 B2
86. Segmentation-based damage detection, US Patent 11004204 B2
87. Character recognition method and device, US Patent 10872274 B2
88. Online active learning in user-generated content streams, US Patent 99673218 B2
89. Methods and apparatuses for building data identification models, US App. 20180365522 A1
90. Text information clustering method and text information clustering system, US App. 20180365218 A1
91. Multi-sampling model training method and device, US App. 20180365525 A1
92. Question recommendation method and device, US App. 20180330226 A1
93. Feature data processing method and device, US App. 20180341801 A1
94. Text information clustering method and text information clustering system, US App. 20180365218 A1
95. Multi-sampling model training method and device, US App. 20180365525 A1
96. Method and system for training model by using training data, US App. 20180365521 A1
97. Question recommendation method and device, US App. 20180330226 A1
98. Feature data processing method and device, US App. 20180341801 A1

Honors and Awards

- Best Demo Award, ACM CIKM, 2017
- Best Paper Award, ACM WSDM, 2011
- Super Star Team Award, Yahoo!, 2008
- Honorable Mention Team, ACM KDD CUP, 2002

Professional Services

Reviewer for the following journals:

- 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

Reviewer for the following conferences: ICML, SIGIR, NIPS, AISTATS, ECML, ESANN, PSB, WWW

Ongoing Projects

- Logical graph neural networks for symbolic reasoning
- Event logical graph construction and inference
- Insurance claim automation systems by structural extraction from images
- Object recognition and segmentation for satellite imagery
- Video content understanding by multi-modality learning
- Latent confounder discovery in casual inference

References

Available upon request

Date: 2022/06/18