

Chao Ma

VMWare Inc.

Creekside G, 908 Arastradero Rd, Palo Alto, CA 94304

EMAIL nkg114mc@hotmail.com

PHONE (541) 602-4873

WEBSITE <http://nkg114mc.com>

Research Interests

- **Machine Learning on Natural Language Processing:** Multi-task structure prediction; joint entity-linking, entity coreference resolution and named entity recognition.
- **Ads Bidding Price Prediction:** Price bidding techniques for E-commerce Product Listing Ads and Text Ads.
- **AI Ops for Cloud Systems:** Root Cause Analysis, Intelligent Event Management.
- **Computer Board Games:** Reinforcement learning in computer chess and computer gomoku.

Education

Oregon State University, OR, U.S. (09/2011 – 07/2019)

Ph.D. in Computer Science

AI and Machine Learning Group (w/ Prasad Tadepalli and Janardhan Rao Doppa)

Tianjin University, China. (09/2007 – 07/2011)

Bachelor of Engineering in Computer Science

Work Experience

- **VMWare Inc., CA**
 - Member of Technical Staff, Cloud Management Team.** (06/2020 – present)
 - Development of Log Message Root Cause Analysis System - Fault Event Detection/Classification
 - Event Correlation Clustering for Intelligent Event Management
- **Uber Technology, CA**
 - Software Engineer, NeMo Emobility Marketplace Team.** (09/2019 – 6/2020)
 - Bike/Scooter Deployment Hub Targeting - City Hexagon Clustering for Variable Pricing
 - Cross Data-Center Redis Cache for Rebalancing Bounties - Hexagon Cluster Level Demand Forecasting
- **eBay Inc., WA**
 - Intern, Traffic Science Team** (Mentor: Jun Yu) (summer 2017)
 - Implemented a high-accuracy Paid Search Text Ads model from scratch to do the keyword bidding.
- **eBay Inc., WA**
 - Intern, Traffic Science Team** (Mentor: Angelo Restificar) (summer 2016)
 - Feature engineering of item titles for Product Listing Ads with NLP techniques.
 - Implementing an item title Named Entity Recognition system as a byproduct.

Awards and Honors

- **International Students Scholarship of OSU for 2012-2013 Academic Year.** (Aug 2012)
- **Laurel Scholarship of Oregon State University for Winter & Spring Term 2012.** (Dec 2011)
- **IBM Chinese Outstanding Students Scholarship 2010.** (Nov 2010)
- **China National High School Olympiad in Informatics 2005, First Prize.** (Nov 2005)

Projects and Researches

- **Log-Based Root Cause Analysis for VMs in Cloud Systems.** (06/2020 – 03/2022)
 - Developed a root cause analysis (RCA) system to help the cloud customers' troubleshooting from logs. We record log message frequency heatmaps, then detect and output the potential root cause logs based on their Z-scores (compared with the history). The output root cause logs are also clustered according to their semantic similarity.

- **Fault Event Detection in Intelligent Event Management.** (08/2022 – present)
Classify the events to fault/non-fault with the latest NLP methods for the intelligent event management system. Applied VMWare BERT and sequence classification model to featurizing and classify event titles. We achieved an F1 score of 89% and an accuracy score of 91% on various sources of testing data.
- **City Level Hexagon Clustering for Variable Pricing.** (01/2020 – 05/2020)
Employed the agglomerative clustering system for city-level H3 hexagons. Applied production specific cluster distance metrics including cosine distance of demand/shortfall time-series among clusters.
- **Hexagon Trip-Rate Forecasting for Deployment Hub Targeting.** (10/2019 – 01/2020)
Applied GBM model for forecasting daily rush-hour trip counts per H3 hexagon for Seattle bikes. The model was evaluated on 3-month historical data, and reduced wMAPE for 56% compared with the old baseline model.
- **HC-Networks for Search-based Structured Prediction.** (06/2017 – 02/2019)
Developed *HC-Nets* framework to do “Search-based Structured Prediction”. Proposed a generic imitation learning-based formulation with *Heuristic-Network* for guiding search to generate candidate outputs, and *Cost-Network* for selecting the best predict output.
- **Joint Entity-linking, Coreference and NER for Deep Language Understanding.** (03/2014 – 06/2017)
Main developer and maintainer of *OregonStateEntityLinking* system. Applied multi-task structure prediction framework with cross task and cross language (English transliteration) featurizations on joint linking, coreference and NER typing problem in English, Chinese and Spanish, and entered for TAC-KBP Trilingual Entity Discovery and Linking competition in 2014 and 2015.
- **Search Based Approaches in Coreference Resolution for Machine Reading Project.** (06/2012 – 03/2014)
Applied *Prune-and-Score* framework in *OregonStateCoref* system to solve coreference resolution, a classical problem in NLP domain. We formulated coreference resolution into a structure prediction problem and proposed a two-level cascade approach as the solution, and achieved a SOTA accuracy.
- **Tuning Chess Evaluation Function Using Reinforcement Learning.** (02/2012 – 04/2012)
Implemented *RootStrap* and *TreeStrap* algorithms to tune the weight of the chess evaluation function from scratch. Improved the chess program strength from a starter level (ELO2000) to a master level (ELO2400+).

Publications

Conference Papers

Log-Based Root Cause Analysis at Scale

Chao Ma, Darren Brown, Xing Wang, Junyuan Lin, Andrii Malinovskyi, Alexey Zavitaev, Nick Stephen
Conference of *VMWare annual Research and Development Innovation Offsite* (RADIO-2022)

Randomized Greedy Search for Structured Prediction: Amortized Inference and Learning

Chao Ma, Aryan Deshwali, Md Rakibul Islam, Janardhan Rao Doppa, and Dan Roth
Proceeding of International Joint Conference on Artificial Intelligence (IJCAI-2019)

Joint Neural Entity Disambiguation with Output Space Search

Hamed Shahbazi, Xiaoli Fern, Reza Ghaeini, Chao Ma, Rasha Obeidat, Prasad Tadepalli
Proceeding of International Conference on Computational Linguistics (COLING-2018)

Multi-Task Structured Prediction for Entity Analysis: Search-Based Learning Algorithms

Chao Ma, Janardhan Rao Doppa, Prasad Tadepalli, Xiaoli Fern and Hamed Shahbazi
Journal of Machine Learning Research (JMLR), Proceedings Track, Vol 80, 2017 (ACML-2017)

Select-and-Evaluate: A Learning Framework for Large-Scale Knowledge Graph Search

F A Rezaur Rahman Chowdhury, Chao Ma, Md Rakibul Islam, Mohammad Hossein Namaki, Mohammad Omar Faruk, and Janardhan Rao Doppa
Journal of Machine Learning Research (JMLR), Proceedings Track, Vol 80, 2017 (ACML-2017)

Learning Greedy Policies for the Easy-First Framework

Jun Xie, **Chao Ma**, Janardhan Rao Doppa, Prashanth Mannem, Xiaoli Fern, Tom Dietterich, Prasad Tadepalli,
Proceedings of AAAI Conference on Artificial Intelligence (AAAI-2015)

Prune-and-Score: Learning for Greedy Coreference Resolution

Chao Ma, Janardhan Rao Doppa, Xiaoli Fern, Tom Dietterich, and Prasad Tadepalli
Proceedings of International Conference on Empirical Methods in Natural Language Processing (EMNLP-2014)

HC-Search for Multi-Label Prediction: An Empirical Study

Janardhan Rao Doppa, Jun Yu, **Chao Ma**, Alan Fern, and Prasad Tadepalli
Proceedings of AAAI Conference on Artificial Intelligence (AAAI-2014)

Thesis

New Directions in Search-based Structured Prediction: Multi-Task Learning and Integration of Deep Models.
(Ph.D.)

RoboCup Soccer Agents Design Based on HTN Planning. (Bachelor)

Professional Services**Tutorial**

Recent Advances in Structured Prediction at AAAI-18, New Orleans, LA (w/ *Janardhan Rao Doppa and Liping Liu*)

Program Committee/Reviewer:

- Artificial Intelligence: AISTATS (17, 18, 19), AAAI 18, IJCAI (15, 18, 23)
- Machine Learning: NeurIPS (17, 18), ICML 19
- Data Mining: TheWebConf 23

Technical Skills

Programming Language: C, C++, Java, Scala, Python, Golang, MATLAB, Haskell.

Cloud and Large-scale Processing: Hadoop, Apache Spark, Scalding, Hive, Kubernetes, AWS.

Databases: Redis, PostgreSQL, MongoDB, Pestrosql (query engine).

Machine Learning Frameworks: Scikit-learn, PyTorch, Tensorflow, Keras.

Other Tools: Kafka, ElasticSearch, Bash, Git, Open MPI, MPICH3.