Toronto, ON, Canada
+1 (519)-729-1801
victoryang147@gmail.com

Wei (Victor) Yang

Machine Learning Engineer

Google Scholar
LinkedIn
Homepage

Machine Learning Engineer with 4 years of experience applying ML algorithms to industrial applications (e.g. e-commerce, finance, chatbot). **Areas of expertise:** 1). recommendation, ranking and personalization, information retrieval 2). NLP applications such as question answering, semantic parsing, 3) data driven machine learning, transfer learning. **Academic Impact:** publications in top ML/IR/NLP conferences such as WWW/ACL/EMNLP/SIGIR/AAAI/NAACL, receiving 1500+ citations according to Google Scholar.

Experience

Machine Learning Engineer, Wish — San Francisco (Remote), Oct 2021 - Now
Research Engineer, Borealis.ai — Toronto, Dec 2019 - Sep 2021
Machine Learning Engineer , RSVP.ai — Waterloo, May 2018 - Aug 2019
Researcher Intern, DiDi Research — Hangzhou, Feb 2017 - Jun 2017
Research Assistant, Data System Group, University of Waterloo — Waterloo, Sep 2017 - Apr 2019
Research Assistant, Singapore University of Technology and Design — Singapore, Oct 2016 - Feb 2017
Research Assistant, The Hong Kong Polytechnic University — Hong Kong, Jun 2016 - Sep 2016

Education

M.Math, Computer Science, University of Waterloo — **Sep 2017 — Dec 2018**
- Thesis: End-to-End Neural Information Retrieval **[Paper]** **[Slide]**
- Thesis Supervisor: Jimmy Lin

B.Sci, Computer Science, Zhejiang University, GPA: 3.9/4.0 — **Sep 2013 — Jun 2017**
- Thesis: Cross Domain Word Embedding (advised by Lu Wei and Siliang Tang)
- Minor: Advanced Honors Class of Engineering Education (ACEE), Chu Kochen Honors College, Zhejiang University

Academic Service

- Sub-reviewer for conferences/journals: ACL 2020, NAACL 2019, SIGIR 2019, CIKM 2019, AAAI 2018
- Reviewer for conferences/journals: EMNLP(2022, 2021, 2020, 2019), ACL 2021, NAACL 2021, COLING 2020, AAAI(2023, 2022, 2021, 2020), IJCAI (2022, 2021), AACL (2022, 2020), TOIS 2021, ACL Rolling Review (2021-06, 2021-09, 2022-01, 2022-04, 2022-07, 2022-09)

Publications & Patent

1. Semantic Parsing, Text-to-SQL
   - "System and method for transferable natural language interface" U.S. Patent Application No. 17/508,914.
   - "Hierarchical Neural Data Synthesis for Semantic Parsing" arXiv e-prints (2021) (first author)
   - "A Globally Normalized Neural Model for Semantic Parsing" *ACL workshop*, 2021.
   - "TURING: an Accurate and Interpretable Multi-Hypothesis Cross-Domain Natural Language Database Interface" *ACL*, 2021.
   - "Optimizing Deeper Transformers on Small Datasets" *ACL*, 2021. **[Paper]**
2. Information Retrieval, Document Ranking, Semantic Textual Similarity
   - "Approximate Nearest Neighbor Search and Lightweight Dense Vector Reranking in Multi-Stage Retrieval Architectures" *ICTIR*, 2020. (first author) **[Paper]**
   - "Multi-stage document ranking with BERT" **[Paper]** **[Code]**
   - "Capreolus: A Toolkit for End-to-End Neural Ad Hoc Retrieval." *WSDM demo*, 2020. **[Paper]** **[Code]**
   - "Applying BERT to Document Retrieval with Birch." *EMNLP demo*, 2019. **[Paper]** **[Code]**
   - "Bridging the Gap of Relevance Matching and Semantic Matching with Hierarchical Co-Attention Network." *EMNLP*, 2019. **[Code]**
   - "Incorporating Compositional and Syntactic Structures Improves Semantic Similarity Modeling." *EMNLP*, 2019. **[Paper]** **[Code]**
   - "Cross-Domain Modeling of Sentence-Level Evidence for Document Retrieval." *EMNLP*, 2019. **[Paper]** **[Code]**
   - "Document Expansion by Query Prediction." **[Paper]** **[Code]**
   - "Simple Applications of BERT for Ad Hoc Document Retrieval." (first author) **[Paper]** **[Code]**

- "Critically Examining the "Neural Hype": Weak Baselines and the Additivity of Effectiveness Gains from Neural Ranking Models." *SIGIR*, 2019. (first author) (**Best Paper Honorable Mention**) **[Paper]** **[Code]**
- "Multi-Perspective Relevance Matching with Hierarchical ConvNets for Social Media Search." *AAAI*, 2019. **[Paper]** **[Code]**

3. NLP Applications, Industrial Demos

- "End-to-End Neural Context Reconstruction in Chinese Dialogue." *ACL workshop* (NLP for Conversational AI), 2019. (first author) **[Paper]**
- "Detecting Customer Complaint Escalation with Recurrent Neural Networks and Manually-Engineered Features." *NAACL industry track*, 2019. (first author) **[Paper]**
- "Distant Supervision for Multi-Stage Fine-Tuning in Retrieval-Based Question Answering" *WWW*, 2020. (first author)
- "End-to-End Open-Domain Question Answering with BERTserini." (first author) *NAACL*, 2019. **[Paper]** **[Code]**
- "A Simple Regularization-based Algorithm for Learning Cross-Domain Word Embeddings." *EMNLP*, 2017. (first author) **[Paper]** **[Code]**

Projects

Ranking & Personalization Wish **Oct 2021 — Now**

- Lead and launch the ranking model project for the shoppable video (Wish Clip) and significantly improved the user engagement and business metrics (+172% for add to cart and +109% for add to cart users in video feed). (100+ QPS, 500k+ DAU)
- Own the ranking service and improve the ranking system's performance, stability and latency. Propose the stability improvement plan of the core ranking service and reduce the latency by 50ms/request. (2k+ QPS, 3M+ DAU).
- Design the system for the first version of video ranker and make plans for two consecutive quarters and drive the discussion of the system design plan with teammates. Design the video feed metrics for the A/B test with the data science team.
- Design and implement the pretrain-finetune baseline of ranking models and increase the recommendation GMV by 5.86%.
- Increase the diversity of user impressions for the homepage feed based on category CTR; increase long-term user retension.
- Fix a cache issue in feed service and increase core business metrics such as company wide GMV by 1.76%.

Credit Capacity Estimation, Borealis AI **Jun 2021 — Sep 2021**

- Analyze the effect of missing values in features; find better processing methods for missing values
- Design training strategies for the capacity estimation model to make the predictions on OOD data more conservative

Cross-domain Text-to-SQL Borealis AI **Dec 2019 — Jun 2021**

- Research tasks: 1). build various model architectures for semantic parsing including RoBERTa + RAT encoder, grammar-guided LSTM/transformer decoder, meta learning and neural data augmentation; 2). lead a small research group to investigate the contributions of data from different sources in a proof-of-concept (POC) project.
- Engineering tasks: 1). build and maintain a safe-guard module for out-of-domain (OOD) and unanswerable question detection in a question answering system; 2). build an end-to-end inference system for natural language database interface and deploy it into the market research domain inside RBC; 3). write template-matching grammar in the answer retrieval module of a FAQ system.

Open Domain Question Answering, RSVP.AI **Jan 2019 — Nov 2019**

- Build a open-domain question answering system for intelligent chatbot
- Propose a data augmentation method for open-domain question answering via distant supervision

Efficient and Effective Document Retrieval, University of Waterloo **Jan 2018 — Dec 2018**

- Investigate and summarize 130 papers in top IR conferences in the 20 years to justify the neural hype arugment in document retrieval.
- Propose an effective cross-domain sentence-level training method for document retrieval
- Propose an algorithm to improve the retrieval effectiveness of a search engine is to expand documents by query prediction
- Propose a multi-stage algorithm for document retrieval (monoBERT-duoBERT)

Awards & Honors

- Rank 1st in the Spider Cross-domain Semntic Parsing leaderboard 2020
- Rank 1st in the MS-MARCO Passage Retrieval leaderboard 2019
- Student Researcher Scholarship, EMNLP/SIGIR/NAACL 2017/2019/2019
- Best Paper Honourable Mention, SIGIR (2 out of 108) 2019
- Graduate with Honors, Zhejiang University & Zhejiang Province 2017
- Champion, Zhejiang University Programming Contest, ACM-ICPC 2016

Technical Skills

Tools Airflow • K8S • Apache Spark • Solr • Pytorch & Tensorflow • Grafana • Prometheus • Superset
Open Source Project Castor • Anserini • NCRFpp • Birch • MatchZoo • BERT-Fine_tune • Capreolus• Bertserini