

Victor Wei Yang

CONTACT INFORMATION	330 Phillip Street, N1903, Waterloo, Ontario, Canada, N2L3W9 (1)519-729-1801 w85yang@uwaterloo.ca, victor950118@gmail.com Github, Linkedin, Homepage
INTERESTS	Question Answering, Machine Reading Comprehension, Information Extraction, Information Retrieval, Coreference Resolution, and Chatbot System
EDUCATION	<p>University of Waterloo, Waterloo, ON</p> <p>M.Math., Computer Science Sep 2017 to Apr 2019</p> <ul style="list-style-type: none">• Thesis: End-to-End Neural Information Retrieval• Thesis Supervisor: Jimmy Lin <p>Zhejiang University, Hangzhou, China</p> <p>B.S., Computer Science and Technology Sep 2013 to Jun 2017</p> <ul style="list-style-type: none">• Thesis: Cross Domain Word Embedding• GPA: 3.87 / 4.0 (198.5 credits)• Thesis Supervisor: Lu Wei and Siliang Tang
REFERRED PAPERS	<ol style="list-style-type: none">1. Wei Yang “End-to-End Neural Information Retrieval.” Master thesis.2. Zeynep Akkalyoncu, Wei Yang, Haotian Zhang, and Jimmy Lin “Cross-Domain Modeling of Sentence-Level Evidence for Document Retrieval.”3. Rodrigo Nogueira, Wei Yang, Jimmy Lin, and Kyunghyun Cho “Document Expansion by Query Prediction.”4. Wei Yang*, Yuqing Xie*, Luchen Tan, Kun Xiong, Ming Li, and Jimmy Lin “Data Augmentation for BERT Fine-Tuning in Open-Domain Question Answering”5. Wei Yang*, Haotian Zhang*, and Jimmy Lin “Simple Applications of BERT for Ad Hoc Document Retrieval.” (* equal contribution)6. Peng Shi, Ji Xin, Wei Yang, Jinfeng Rao and Jimmy Lin “Distantly-Supervised Pretraining for Named Entity Recognition.”7. Wei Yang, Kuang Lu, Peilin Yang, Jimmy Lin “Critically Examining the “Neural Hype”: Weak Baselines and the Additivity of Effectiveness Gains from Neural Ranking Models.” <i>SIGIR</i>, 2019.8. Wei Yang*, Yuqing Xie*, Aileen Lin, Xingyu li, Luchen Tan, Kun Xiong, Ming Li, and Jimmy Lin “End-to-End Open-Domain Question Answering with BERTserini.” (* equal contribution) <i>NAACL</i>, 2019.9. Wei Yang, Luchen Tan, Anqi Cui, Han Li, Xi Chen, Chunwei Lu, Muzi Wang, and Jimmy Lin “Detecting Customer Complaint Escalation with Recurrent Neural Networks and Manually-Engineered Features.” <i>NAACL</i>, 2019.10. Linqing Liu, Wei Yang, Jinfeng Rao, Raphael Tang, and Jimmy Lin “Incorporating Compositional and Syntactic Structures Improves Semantic Similarity Modeling.”11. Wei Yang, Junnan Chen, Rui Qiao, Haocheng Qin, Ming Li, and Jimmy Lin “End-to-End Neural Context Reconstruction in Chinese Dialogue.” <i>ACL workshop (NLP for Conversational AI)</i>, 2019.

12. **Wei Yang**, Jinfeng Rao, and Jimmy Lin “Pairwise Loss with Max Sampling: A Neural Meta-Architecture for Answer Selection.”
13. Jinfeng Rao, **Wei Yang**, Yuhao Zhang, Ferhan Ture, and Jimmy Lin “Multi-Perspective Relevance Matching with Hierarchical ConvNets for Social Media Search.” *AAAI*, 2019.
14. **Wei Yang**, Wei Lu, Vincent W. Zheng “A Simple Regularization-based Algorithm for Learning Cross-Domain Word Embeddings.” *EMNLP*, 2888–2894, 2017.
15. Tingting Liang, Liang Chen, Zhining Xie, **Wei Yang**, Jian Wu “CASE: A Platform for Crowdsourcing Based API Search.” *ICSOC*, 482–485, 2015.

RESEARCH EXPERIENCE

Research Assistant Sep 2017 to Apr 2019
Data System Group, University of Waterloo
Supervisor: [Jimmy Lin](#)

- Build an end-to-end neural information retrieval system with Anserini and MatchZoo
- Analyzing negative sampling in the answer selection task
- Explore neural and non-neural baselines for searching short social media posts

Research Assistant Oct 2016 to Feb 2017
Singapore University of Technology and Design, supervised by [Lu Wei](#)

- Work on the cross-domain word embedding project.
- Train a cross-domain embedding to down-stream task given two corpus from different domains. Specifically deal with the problem when the corpus size is small in target domain. [\[Paper\]](#) [\[Code\]](#) (accepted by EMNLP 2017)

Research Assistant Jun 2016 to Sep 2016
The Hong Kong Polytechnic University, supervised by [Ajay Kumar \(IEEE Fellow\)](#)

- Work on the palm identification project.
- Try to apply Convolutional Neural Networks on palm identification task; Implement a real-time and contactless palm identification application through QT5 including both palm localization and classification at real-time video stream.

INDUSTRY EXPERIENCE

Machine Learning Engineer May 2019 to - Now
RSVP.ai, Waterloo, ON, Canada, mentored by [Luchen Tan](#)

- Build open domain question answering system for both Chinese and English.
- Work on the algorithm implementation of proof of concept (POC) projects by integrating the cutting edge NLP algorithms into products, including sentiment analysis, machine reading comprehension, etc.

Research Intern Sep 2018 to Oct 2018
JD.com, Beijing, China

- Build an the complaint detection system on the dialogue data between the JD customer and customer service from collecting and cleaning data through Hive and SparkSQL to proposing deep classification network with Hierarchical LSTM + Attention mechanism and deploy this system through a RESTful API. This product has been deployed online and it saved more than CAD \$50K/month for JD.com according to the report in Oct 2018.

Research Intern May 2018 to Aug 2018
RSVP.ai, Waterloo, ON, Canada

- Work on the contextual conversation system project; Find solution to the English and Chinese coreference resolution task. Design models to detect the coreference

pronoun and zero pronoun, and rank the candidate entities. Run experiments on manually created data with context information.

Software Engineer Intern

Feb 2017 to Jun 2017

Big Data Group, DiDi Research, Hangzhou, China

- Work on self-driving car project; Run experiments for the vehicle's trajectory predictions and pedestrian's intention prediction tasks.

EXCHANGE EXPERIENCE

Exchange Student

Sep 2016 to Dec 2016

GPA: 4.0 / 4.0 in Singapore University of Technology and Design

Courses: Machine Learning (Outstanding prize on this course for full marks on all additional advanced questions), Optimization

AWARDS

- SIGIR Student Travel Grant 2019
- International Student Scholarship, University of Waterloo 2018
- Student Researcher Scholarship, EMNLP 2017 2017
- Graduate with Honors, Zhejiang University & Zhejiang Province 2017
- Champion, Zhejiang University Programming Contest, ACM-ICPC 2016
- Excellent Student Awards, Zhejiang University 2015, 2016, 2017

PRESENTATIONS

- End-to-End Goal-Oriented Question Answering System for Chatbot, Invited Talk, RSVP.ai Mar 2019
- An Anatomy of BERT Model and Its Applications, Invited Talk, RSVP.ai Oct 2018
- Text Matching Models and Their Applications, Invited Talk, RSVP.ai Jul 2018
- Neural Information Retrieval and Text Matching, Group Talk Presentation, Data System Group, University of Waterloo Mar 2018
- Cross-domain Word Embedding, Oral Presentation on EMNLP 2017, Copenhagen, Denmark Sep 2017
- Outstanding Undergraduate Thesis Defence, Zhejiang University, China Jun 2017

TEACHING ASSISTANT EXPERIENCE

CS 137 - Programming Principles (Fall 2017), CS 370 - Numerical Computation (Winter 2018), CS 371 - Computational Mathematic (Spring 2018), CS 651 - Data-Intensive Distributed Computing, CS 651 - Data-Intensive Distributed Computing (Winter 2019)

SKILL

- Programming Language:
 - Over 10000 lines: Python • C • C++ • Java • Shell
 - Over 1000 lines: JavaScript • Scala • R • Matlab
- DL Stack: TensorFlow • Keras • Pytorch
- ML/DS Stack: scikit-learn • pandas
- NLP/IR Stack: Lucene • Elasticsearch • Stanford CoreNLP • Lemur • Matchzoo
- Others: Apache Spark • Hadoop • Hive • SparkSQL • Docker
- Open Source Contribution: Castor • Anserini • NCRFP • Birch • MatchZoo

SERVICE

Sub-reviewer, AAAI 2018, NAACL 2019, SIGIR 2019, CIKM 2019
Reviewer, EMNLP 2019