

## Victor Wei Yang

---

CONTACT INFORMATION	872 Creekside Dr, Waterloo, Ontario, Canada, N2V2S7 w85yang@uwaterloo.ca, victor950118@gmail.com Github, LinkedIn	(1)519-729-1801
INTERESTS	Question Answering, Machine Reading Comprehension, Information Extraction, Information Retrieval, Coreference Resolution, and Chatbot System	
EDUCATION	<b>University of Waterloo</b> , Waterloo, ON  M.Math., Computer Science Sep 2017 to Apr 2019 <ul style="list-style-type: none"><li>Score: 90.0 / 100 on 3 courses: Machine Learning, Data Cleaning, Data-Intensive Distributed Computing</li><li>Thesis Supervisor: Jimmy Lin</li></ul> <b>Zhejiang University</b> , Hangzhou, China  B.S., Computer Science and Technology Sep 2013 to Jun 2017 <ul style="list-style-type: none"><li>GPA: 3.84 / 4.0 ( 198.5 credits ) with major GPA: 3.87 / 4.0</li><li>Thesis Supervisor: Lu Wei and Siliang Tang</li></ul>	
REFERRED PAPERS	<ol style="list-style-type: none"><li><b>Wei Yang</b> “End-to-End Neural Information Retrieval.” Master thesis.</li><li>Zeynep Akkalyoncu, <b>Wei Yang</b>, Haotian Zhang, and Jimmy Lin “Cross-Domain Modeling of Sentence-Level Evidence for Document Retrieval.”</li><li>Rodrigo Nogueira, <b>Wei Yang</b>, Jimmy Lin, and Kyunghyun Cho “Document Expansion by Query Prediction.”</li><li><b>Wei Yang*</b>, Yuqing Xie*, Luchen Tan, Kun Xiong, Ming Li, and Jimmy Lin “Data Augmentation for BERT Fine-Tuning in Open-Domain Question Answering”</li><li><b>Wei Yang*</b>, Haotian Zhang*, and Jimmy Lin “Simple Applications of BERT for Ad Hoc Document Retrieval.” (* equal contribution)</li><li>Peng Shi, Ji Xin, <b>Wei Yang</b>, Jinfeng Rao and Jimmy Lin “Distantly-Supervised Pretraining for Named Entity Recognition.”</li><li><b>Wei Yang</b>, 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.</li><li><b>Wei Yang*</b>, 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.</li><li><b>Wei Yang</b>, 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.</li><li>Linqing Liu, <b>Wei Yang</b>, Jinfeng Rao, Raphael Tang, and Jimmy Lin “Incorporating Compositional and Syntactic Structures Improves Semantic Similarity Modeling.”</li><li><b>Wei Yang</b>, 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.</li></ol>	

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
- Design a multi-perspective relevance matching model with hierarchical ConvNets for social media search. [[Paper](#)] (accepted by AAAI 2019)

**Research Assistant** Oct 2016 to Feb 2017

StatNLP group, Singapore University of Technology and Design

Supervisor: [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

Department of Computing, The Hong Kong Polytechnic University

Supervisor: [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 (which includes palm localization and classification in the camera image).

INDUSTRY  
EXPERIENCE

**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.

	<b>Software Engineer Intern</b> Big Data Group, DiDi Research, Hangzhou, China <ul style="list-style-type: none"> <li>• Work on self-driving car project</li> <li>• Run experiments for the vehicle's trajectory predictions and pedestrian's intention prediction tasks.</li> </ul>	Feb 2017 to Jun 2017
EXCHANGE EXPERIENCE	<b>Exchange Student</b> 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	Sep 2016 to Dec 2016
AWARDS	<ul style="list-style-type: none"> <li>• SIGIR Student Travel Grant</li> <li>• International Student Scholarship, University of Waterloo</li> <li>• Student Researcher Scholarship, EMNLP 2017</li> <li>• Graduate with Honors, Zhejiang University &amp; Zhejiang Province</li> <li>• Champion, Zhejiang University Programming Contest, ACM-ICPC</li> <li>• Honorable Mention, MCM-ICM</li> <li>• Excellent Student Awards, Zhejiang University</li> </ul>	2019 2018 2017 2017 2016 2015 2015, 2016, 2017
PRESENTATIONS	<ul style="list-style-type: none"> <li>• End-to-End Goal-Oriented Question Answering System for Chatbot, Invited Talk, RSVP.ai</li> <li>• An Anatomy of BERT Model and Its Applications, Invited Talk, RSVP.ai</li> <li>• Text Matching Models and Their Applications, Invited Talk, RSVP.ai</li> <li>• Neural Information Retrieval and Text Matching, Group Talk Presentation, Data System Group, University of Waterloo</li> <li>• Corss-domain Word Embedding, Oral Presentation on EMNLP 2017, Copenhagen, Denmark</li> <li>• Outstanding Undergraduate Thesis Defence, Zhejiang University, China</li> </ul>	Mar 2019 Oct 2018x Jul 2018 Mar 2018 Sep 2017 Jun 2017
TEACHING EXPERIENCE	Teaching Assistant: CS 137 - Programming Principles Teaching Assistant: CS 370 - Numerical Computation Teaching Assistant: CS 371 - Computational Mathematics Teaching Assistant: CS 651 - Data-Intensive Distributed Computing Teaching Assistant: CS 651 - Data-Intensive Distributed Computing	Fall 2017 Winter 2018 Spring 2018 Fall 2018 Winter 2019
SKILL	<ul style="list-style-type: none"> <li>• Programming Language:             <ul style="list-style-type: none"> <li>• Over 10000 lines: Python • C • C++ • Java • Shell</li> <li>• Over 1000 lines: JavaScript • Scala • R • Matlab</li> </ul> </li> <li>• DL Stack: TensorFlow • Keras • Pytorch</li> <li>• ML/DS Stack: scikit-learn • pandas</li> <li>• NLP/IR Stack: Lucene • Elasticsearch • Stanford CoreNLP • Lemur • Matchzoo</li> <li>• Others: Apache Spark • Hadoop • Hive • SparkSQL • Docker</li> <li>• Open Source Contribution: Castor • Anserini • NCRFpp</li> </ul>	
SERVICE	Sub-reviewer, AAAI 2018 Sub-reviewer, NAACL 2019 Sub-reviewer, SIGIR 2019	Oct 2017 Jan 2019 Mar 2019