Zhengzhong Liu

CONTACT INFORMATION

5507 Gates Hillman Complex
5000 Forbes Avenue
Pittsburgh, PA, 15213

E-mail: hectorzliu@gmail.com
Mobile: 412 980 0646

RESEARCH INTERESTS

I am broadly interested in areas such as Computational Linguistics, Natural Language Processing, Information Retrieval, Information Extraction. I am particularly interested in injecting linguistic and semantic regularities into machine learning models.

EDUCATION

Carnegie Mellon University **Sep, 2014 — 2020 (Expected)**
- PhD Candidate in Language Technologies
- Thesis Topic: Event and Facet Semantics

Carnegie Mellon University **2012 — 2014**
- Master in Language Technologies

The Hong Kong Polytechnic University **2007 — 2012**
- Bachelor of Science in Computing (with First Class Honors)
- Bachelor of Business Administration with a Major in Management

HONORS AND AWARDS

ACL Best System Demonstration Nomination **2019**
- Our system Texar is selected to be one of the top five system demonstration at ACL.

ACL Outstanding Long Paper Award **2016**
- Awarded to the 10 best paper on ACL, less than 1% of the submitted paper are selected.

The Zhi Yuan Scholarship of China Soong Ching Ling Foundation **2007 - 2011**
- For excellent mainland China student pursuing education in Hong Kong

RESEARCH EXPERIENCES

Active Interpretation of Disparate Alternatives (AIDA) **Sep, 2018 — May, 2019**
- Advisors: Prof. Eduard Hovy, Prof. Teruko Mitamura
- The AIDA project aims at overcome the noisy, conflicting, and potentially intentionally deceptive nature of today's data environment via advanced NLP technologies. Our group focuses on interpreting the alternatives of events and participants from text. We have advanced methods to identify the core events and participatns from news articles.

Deep Exploration and Filtering of Text (DEFT) **Sep, 2012 — 2018**
- Advisors: Prof. Eduard Hovy, Prof. Teruko Mitamura
- The DEFT project aims at discovering deeper and implicit information from text using NLP techniques. Our group focuses on understanding event semantic in text. We develop automated systems to detect event mentions, and to detect various types of relations among event mentions, including coreference, hierarchical and procedural.

Entity Linking for Knowledge Based Population **Jun, 2011 — Oct, 2011**
- Advisor: Prof. Qin Lu
- I lead the entity linking task and also contribute to finding the slot fillers of an entity. Besides a system to disambiguate text spans to the appropriate Wikipedia entry, I further contribute to a new clustering algorithm on grouping unknown entities.

Web Person Name Disambiguation **Jun, 2010 — Jun, 2011**
- Advisor: Prof. Qin Lu
- Propose a novel tailored clustering algorithm for the Web Person Name Disambiguation task, which achieved the state-of-the-art performance in the Web Person Name Disambiguation task.
- Best undergraduate project award nomination in Department of Computing, the Hong Kong Polytechnic University in academic year 2010/11.

Working Experiences

Research Scientist at Petuum **May, 2019 — present**
- Lead the research and development of advanced Natural Language Processing systems.
- I lead develop **Forte** and **Stave**, two advanced NLP systems driven by modular design principles.

Research Intern at Google **May, 2015 — Aug, 2015**
- Mentors: Edgar González Pellicer *and* Daniel Gillick
- Worked with the Google NLP team the Verb Phrase Ellipsis (VPE) problem, which is used to help improve a task-based dialogue systems.
- Study the multi-stages of the VPE problem, propose several strategies for joint-training and achieve the state-of-the-art performance on two public datasets. - We clean and contribute one of the datasets and release it to the community.

Google Summer of Code **May, 2012 - Aug, 2012**
- Mentor: Pablo N. Mendes
- Worked on the *DBpedia Spotlight* for DBpedia at Google Summer of Code program 2012
- Developed and delivered a collective entity linking system which considers the coherence of all entities in the document.
- Code base: https://github.com/hunterhector/dbpedia-spotlight F

Open Source Experiences

Forte - A flexible and composable NLP toolkit https://github.com/asyml/forte
- Design a modular system architecture and flexible representations that can be integrated with deep learning systems
- Lead the development team and implement core functions of the system.

Stave - A general purpose, extensible text annotation system https://github.com/asyml/stave
- Design the main functionalities and interfaces, use an ontology system to enable extensible text annotation

Texar- A composable toolkit for Text Generation and NLP models https://texar.io
- Involve in project interface design and implementations of several fundamental modules

OAQA- Open Advancement of Question Answering Systems https://oaqa.github.io/
- Questions Answering systems in High School World History Exam and Alzheimer's disease articles.

DBpedia Spotlight
- Develop a collective disambiguation module in project DBpedia Spotlight using Scala and Hadoop.

Publications

1. Zhiting Hu, Zichao Yang, Haoran Shi, Bowen Tan, Tiancheng Zhao, Junxian He, Xiaodan Liang, Wentao Wang, Xingjiang Yu, Di Wang, Lianhui Qin, Xuezhe Ma, **Zhengzhong Liu**, Devendra Singh, Wangrong Zhu, and Eric P Xing. Texar: A Modularized, Versatile, and Extensible Toolkit for Text Generation. ACL 2019, *best system demonstration nomination*
2. Zhiting Hu, Zichao Yang, Haoran Shi, Bowen Tan, Tiancheng Zhao, Junxian He, Xiaodan Liang, Wentao Wang, Xingjiang Yu, Di Wang, Lianhui Qin, Xuezhe Ma, **Zhengzhong Liu**, Devendra Singh, Wangrong Zhu, and Eric P Xing. 2018. Texar: A Modularized, Versatile, and Extensible Toolbox for Text Generation. In *Proceedings of Workshop for NLP Open Source Software, ACL'18*.
3. **Zhengzhong Liu**, Teruko Mitamura, and Eduard Hovy. 2018. Graph-Based Decoding for Event Sequencing and Coreference Resolution. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING'18)*.
4. **Zhengzhong Liu**, Chenyan Xiong, Teruko Mitamura, and Eduard Hovy. 2018. Automatic Event Salience Identification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP'18)*.
5. Hans Chalupsky, Jun Araki, Eduard Hovy, Andrew Hsi, **Zhengzhong Liu**, Xuezhe Ma, Evangelia Spiliopoulou, and Shuxin Yao. 2018. Multi-lingual Extraction and Integration of Entities,

Relations, Events and Sentiments into ColdStart ++ KBs with the SAFT System. In *Proceedings of the Text Analysis Conference 2017 (TAC'17)*.

6. Teruko Mitamura, **Zhengzhong Liu**, and Eduard Hovy. 2018. Events Detection, Coreference and Sequencing: What's next? Overview of the TAC KBP 2017 Event Track. In *Proceedings of the Text Analysis Conference 2017 (TAC'17)*.
7. Chenyan Xiong, **Zhengzhong Liu**, Jamie Callan, and Tie-Yan Liu. 2018. Towards Better Text Understanding and Retrieval through Kernel Entity Salience Modeling. In *The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'18)*.
8. Chenyan Xiong, **Zhengzhong Liu**, Jamie Callan, and Eduard Hovy. 2017. JointSem: Combining Query Entity Linking and Entity based Document Ranking. In *Proceedings of the 26th ACM International Conference on Information and Knowledge Management (CIKM'17)*.
9. Keyang Xu, **Zhengzhong Liu**, and Jamie Callan. 2017. De-duping URLs with Sequence-to-Sequence Neural Networks. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'17)*.
10. Teruko Mitamura, **Zhengzhong Liu**, and Eduard Hovy. 2017. Overview of TAC-KBP 2016 Event Nugget Track. In TAC 2016.Zhengzhong Liu, Jun Araki, Teruko Mitamura, and Eduard Hovy. 2016. CMU-LTI at KBP 2016 Event Nugget Track. In *Proceedings of the Text Analysis Conference 2016 (TAC'16)*.
11. Zhiting Hu, Xuezhe Ma, **Zhengzhong Liu**, Eduard Hovy, and Eric Xing. 2016. Harnessing Deep Neural Networks with Logic Rules. In *Proceeding of 2016 Annual Meeting of the Association for Computational Linguistics (ACL'16)*.
12. Xuezhe Ma, **Zhengzhong Liu**, and Eduard Hovy. 2016. Unsupervised Ranking Model for Entity Coreference Resolution. In *The 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL'16)*.
13. **Zhengzhong Liu**, Edgar González, and Dan Gillick. 2016. Exploring the steps of Verb Phrase Ellipsis. In *NAACL 2016 Workshop on Coreference Resolution Beyond OntoNotes (CORBON'16)*.
14. **Zhengzhong Liu**, Jun Araki, Teruko Mitamura, and Eduard Hovy. 2016. CMU-LTI at KBP 2016 Event Nugget Track. In *Proceedings of the Text Analysis Conference 2016 (TAC'16)*.
15. **Zhengzhong Liu**, Jun Araki, Dheeru Dua, Teruko Mitamura, and Eduard Hovy. 2015. CMU-LTI at KBP 2015 Event Track. In *Proceedings of the Text Analysis Conference 2015 (TAC'15)*.
16. Teruko Mitamura, **Zhengzhong Liu**, and Eduard Hovy. 2016. Overview of TAC KBP 2015 Event Nugget Track. In *Proceedings of the Text Analysis Conference 2015 (TAC'15)*.
17. **Zhengzhong Liu**, Teruko Mitamura, and Eduard Hovy. 2015. Evaluation Algorithms for Event Nugget Detection: A Pilot Study. In *Proceedings of the 3rd Workshop on EVENTS, NAACL-HLT'15*, pages 53–57.
18. Di Wang, Lenoid Boytsov, Jun Araki, Alkesh Patel, Jeff Gee, **Zhengzhong Liu**, Eric Nyberg, and Teruko Mitamura. 2014. CMU Multiple-choice Question Answering System at NTCIR-11 QA-Lab. In *Proceedings of the 11th NTCIR Conference (NTCIR'14)*.
19. **Zhengzhong Liu**, Jun Araki, Eduard Hovy & Teruko Mitamura. 2014. Supervised Within-Document Event Coreference using Information Propagation. In *Proceedings of The Ninth edition of International Conference on Language Resources and Evaluation (LREC'14)*.
20. Jun Araki, **Zhengzhong Liu**, Eduard Hovy & Teruko Mitamura. 2014. Detecting Subevent Structure for Event Coreference Resolution. In *Proceedings of The Ninth edition of International Conference on Language Resources and Evaluation (LREC'14)*.
21. Xu, J., Lu, Q., & **Zhengzhong Liu**. 2012. Aggregating skip bigrams into key phrase-based vector space model for web person disambiguation. In *Proceedings of KONVENS'12*.

22. Xu, J., Lu, Q., **Zhengzhong Liu**, & Chai, Junyi. 2012. Topic Sequence Kernel. In H. Yuexian, N. Jian-Yun, S. Le, W. Bo, & Z. Peng (Eds.), *Lecture Notes in Computer Science*.
23. Xu, J., Lu, Q., & **Zhengzhong Liu**. 2012. PolyUCOMP: Combining semantic vectors with skip bigrams for semantic textual similarity. In *SemEval'12 Proceedings of the First Joint Conference on Lexical and Computational Semantics*.
24. Xu, J., Lu, Q., & **Zhengzhong Liu**. 2012. Combining Classification with Clustering for Web Person Disambiguation. In *Proceedings of the 21st international conference companion on World Wide Web (WWW'12)*. Lyon.
25. Xu Jian, **Zhengzhong Liu**, Qin Lu, Yu-lan Liu, & Chenchen Wang. 2011. PolyUCOMP in TAC 2011 Entity Linking and Slot-Filling. In *Proceedings of the Fourth Text Analysis Conference (TAC'11)*.
26. **Zhengzhong Liu**, Lu, Qin, & Xu, Jian. 2011. High Performance Clustering for Web Person Name Disambiguation Using Topic Capturing. In *The first International Workshop on Entity Orientation Search, SIGIR'11*.

Professional Activities

Co-organizer, TAC Knowledge Base Population Event Track 2015-2017
Reviewer, ACL, EMNLP, NAACL, CONLL, NLPCC

Tutorials

Modularizing Natural Language Processing — NVDIA GTC 2020, with Zecong Hu
Modularizing Natural Language Processing — AAAI 2020, with Zhiting Hu and Eric P Xing

Teaching

Probablistic Script Induction **2014**
- Talk on *Advanced Seminar in Semantics*, CMU

Coreference Resolution **2015**
- 2015 Guest Lecture for graduate course *Tools for NLP*, CMU

Deep Learning Methods for Computational Semantics **2017**
- 2017 Teaching Assistant and Guest Lecture for Computational Semantics for NLP, CMU

Neural Networks for Discourse Models **2017**
- 2017 Teaching Assistant and Guest Lecture for Neural Networks for NLP, CMU

Students Advising

Adithya Pratapa (CMU MS)	2019.8-Now
Mansi Gupta (CMU MS)	2019.6–2020.3, now Twitter
Wei Wei (CMU MS)	2019.5–2019.8
Zecong Hu (CMU MS)	2019.5–2019.8
Haoran Shi (CMU MS)	2019.5–2019.8

References

Eduard Hovy	Research Professor, LTI, Carnegie Mellon University
Teruko Mitamura	Research Professor, LTI, Carnegie Mellon University
Vincent Ng	Professor, Computer Science, University of Texas Dallas
Edgar Gonzàlez Pellicer	Research Scientist, Google, Inc.
Qin Lu	Professor, Computing, The Hong Kong Polytechnic University