Jie Cao (jie.cao@colorado.edu, 8014485203)   **Resume (December, 2023)**

Postdoc at NSF National AI Institute for Student-AI Teaming | University of Colorado Boulder   Ph.D. at University of Utah

Education

2015 – 2022	**Doctor of Philosophy**
	Computer Science
	University of Utah
	Advisor: Vivek Srikumar
	Dissertation: Inductive Biases for Deep Linguistic Structured Prediction with Independent Factorization [18]
2009 – 2012	**Master of Science**
	Computer Science
	Huazhong University of Science and Technology (HUST)
2005 – 2009	**Bachelor of Engineering**
	Information Security
	Huazhong University of Science and Technolgy (HUST)

Research Experience

NSF-iSAT, University of Colorado Boulder, Boulder   2022 - present

– I am a **post-doctoral** researcher at NSF AI Institute for Student-AI Teaming (iSAT) with **Prof. James Martin and Prof. Martha Palmer**, Boulder NLP Groups, and Institute of Cognitive Science. My research mainly covers conversational modelling [5, 1, 15], speech-aware models [4, 3, 14], education AI on automatic feedbacks to teachers and tutors [16], small group collaborative learning [3, 2].

Utah NLP Group, University of Utah, Salt Lake City   2015 - 2022

– I worked with **Prof. Vivek Srikumar** in Utah NLP Group. I study methods for predicting structured representations of natural language text, e.g., semantic parsing [9], dialogue system [11, 8, 6], self-supervised learning on database workloads [7].

Intern at AWS AI, Amazon Lex, Remote   Summer 2020

– Our paper on schema-guided dialog got accepted to NAACL 2021 [6].

Intern at AWS AI, Amazon Lex, Seattle   Summer 2019

– Cross framework meaning representation parsing, ranked 1st on AMR on CoNLL 2019 shared task [9].

Intern at Tencent, Wechat AI, Palo Alto   Summer 2018

– End-to-End Goal-oriented Dialogue System [11], ranked 2nd in DSTC7 track1

– Rhetorically Controlled Poetry Generation[10] got accetped to ACL 2019.

CGCL Lab, HUST, Wuhan   2008 - 2012

– When I was a master student at HUST, I worked with Prof. Xia Xie and Prof. Hai Jin. My research are around large scale statistical somputing, programming language [12, 13], system virtualization [20, 19]

Teaching & Mentoring

Fall 2023	Lecturer	**Natural Language Processing (CSCI-LING 5832)**, U of Colorado Boulder
2019-2020	Mentor	Bachelor Thesis (Tarun Sunkaraneni), U of Utah
Spring 2018	TA	Structured Prediction (CS 6355), U of Utah
Fall 2016,2018	TA	Machine Learning (CS 6350), U of Utah
Fall 2009	Lecturer	Algorithm and Data Structure, HUST
2006-2008	Leader	Algorithm&Game Team, Unique Studio, HUST

Jie Cao (jie.cao@colorado.edu, 8014485203)   **Resume (December, 2023)**

Postdoc at NSF National AI Institute for Student-AI Teaming | University of Colorado Boulder   Ph.D. at University of Utah

Work Experience

Assitant Researcher at Sohu RDC Lab, Beijing   2014 - 2015

– Hadoop, Spark, Data migration, Data security, Distributed machine learning.

Senior Software Engineer at Zun Club, Beijing   2013 - 2014

– Heterogeneous data intergration, Hotel recommendation system.

Software Engineer at Baidu, Beijing   2012 - 2014

– Baidu Voice Assistant, Query analysis, Dialogue
– Mobile Search Anti-Attack Ecosystem, Speed optimization

Intern at Alibaba, Hangzhou   2010 - 2011

– KV Storage, MySQL, Database Replication, Real-time Computing, Distributed Pub/Sub Data Pipeline.

Honors & Awards

2023	iSAT Trainee Grant
2019	CoNLL Shared Task, MRP, Top 1 on AMR
2019	DSTC7, Top 2 on track 1
2010	VMware Cloud Computing Innovation Cup, Top 50
2009	Google Android Innovative Idea Sharing Award
2007	"Computer World" Magzine Scholarship(50 students per year in China)
2007	Microsoft Imagine Cup
	– Algorithm Challenge, Top 50 World-Wide
	– Visual Gaming Contest, Top 2 in China, 18th World Final
2006	HUST ACM Programming Contest, Top 3

Publications

[1] Jon Z Cai, Brendan King, Margaret Perkoff, Shiran Dudy, **Jie Cao**, Marie Grace, Natalia Wojarnik, Ananya Ganesh, James H Martin, Martha Palmer, et al. "Dependency Dialogue Acts–Annotation Scheme and Case Study". *The 13th International Workshop on Spoken Dialogue Systems Technology* (2023).

[2] **Jie Cao**, Rachel Dickler, Marie Grace, Alessandro Roncone, Leanne Hirshfield, Marilyn Walker, and Martha Palmer. "Designing an AI Partner for Jigsaw Classrooms". *Workshop on Language-Based AI Character Interation with Children* (2023).

[3] **Jie Cao**, Ananya Ganesh, Jon Cai, Rosy Southwell, E Margaret Perkoff, Michael Regan, Katharina Kann, James H Martin, Martha Palmer, and Sidney D'Mello. "A Comparative Analysis of Automatic Speech Recognition Errors in Small Group Classroom Discourse". *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*. 2023, pp. 250–262.

[4] Ananya Ganesh, **Jie Cao**, E. Magerate Perkoff, Rosy Southwell, Martha Palmer, and Katharina Kann. "Mind the Gap between the Application Track and the Real World". *Proceedings of the 61th Annual Meeting of the Association for Computational Linguistics, 2023* (2023).

[5] E. Margaret Perkoff, Abhidip Bhattacharyya, Jon Cai, and **Jie Cao**. "Comparing Neural Question Generation Architectures for Reading Comprehension". *18th Workshop on Innovative Use of NLP for Building Educational Applications, 2023* (2023).

[6] **Jie Cao** and Yi Zhang. "A Comparative Study on Schema-Guided Dialogue State Tracking". *Proceedings of NAACL 2021*. 2021.

[7] Debjyoti Paul*, **Jie Cao***, Feifei Li, and Vivek Srikumar. "Database Workload Characterization with Query Plan Encoders". *Proc. VLDB Endow.* 15.4 (2021), pp. 923–935. issn: 2150-8097. doi: 10.14778/3503585.3503600. url: https://doi.org/10.14778/3503585.3503600.

Jie Cao (jie.cao@colorado.edu, 8014485203)   **Resume (December, 2023)**

Postdoc at NSF National AI Institute for Student-AI Teaming | University of Colorado Boulder   Ph.D. at University of Utah

[8] **Jie Cao**, Michael Tanana, Zac Imel, Eric Poitras, David Atkins, and Vivek Srikumar. "Observing Dialogue in Therapy: Categorizing and Forecasting Behavioral Codes". *Proceedings of ACL 2019*. Florence, Italy, 2019.

[9] **Jie Cao**, Yi Zhang, Adel Youssef, and Vivek Srikumar. "Amazon at MRP 2019: Parsing Meaning Representations with Lexical and Phrasal Anchoring". *Proceedings of the Shared Task on MRP at the CoNLL 2019*. 2019.

[10] Zhiqiang Liu, Zuohui Fu, **Jie Cao**, Gerard de Melo, Yik-Cheung Tam, Cheng Niu, and Jie Zhou. "Rhetorically Controlled Encoder-Decoder for Modern Chinese Poetry Generation". *Proceedings of ACL 2019*. Florence, Italy, 2019.

[11] Shuo Sun*, Yik-Cheung Tam*, **Jie Cao***, Can-Xiang Yan, Zuohui Fu, Cheng Niu, and Jie Zhou. "End-to-end Gated Self-attentive Memory Network for Dialog Response Selection". *Proceedings of AAAI, DSTC7 Workshop*. 2019.

[12] Xijiang Ke, Hai Jin, Xia Xie, and **Jie Cao**. "A distributed SVM method based on the iterative MapReduce". *International Conference on Semantic Computing(ICSC)*. IEEE. 2015.

[13] Xia Xie, **Jie Cao**, Hai Jin, Xijiang Ke, and Wenzhi Cao. "JRBridge: A framework of large-scale statistical computing for R". *Asia-Pacific on Services Computing Conference (APSCC)*. IEEE. 2012.

Preprints & Under-Preparation

[14] **Jie Cao**, Jon Cai, Viet Anh, Rosy Southwell, Jacob Whitehill, Jeff Flanigan, James Martin, and Martha Palmer. "Speech-aware AMR Parsing". 2023.

[15] **Jie Cao**, Jon Cai, E Margaret Perkoff, James Martin, and Martha Palmer. "Dialogue Synthesis on Reinforcing Small Group Discussion with Peer Role Conversational Agents". 2023.

[16] **Jie Cao**, Jennifer Jacobs, and James Martin. "Automatic Feedbacks to Tutors". 2023.

[17] Zhimin Li, Shusen Liu, Xin Yu, Kailkhura Bhavya, **Jie Cao**, Diffenderfer James Daniel, Peer-Timo Bremer, and Valerio Pascucci. ""Understanding Robustness Lottery": A Comparative Visual Analysis of Neural Network Pruning Approaches". *arXiv preprint arXiv:2206.07918* (2022).

Thesis & Patents

[18] **Jie Cao**. "Inductive Biases for Deep Linguistic Structured Prediction with Independent Factorization". University of Utah, 2022.

[19] **Jie Cao**, Xia Xie, and Jin Hai. "A real-time scheduling system on embeded virtualization". Patent ZL201110410689.1 (CN). 2011.

[20] **Jie Cao**. "Cache Performance Evaluation on Xen Virtualization Platform". Huazhong University of Science and Techonology, 2009.

Talks

Slides for selected talks can be found in my slideshare site. https://www.slideshare.net/jiessiecao/

– 11/2023, Invited Talk on Computational Semantics Group Seminar at University of Colorado, Boulder.

– 12/2022, Invited Talk on Database Workload Characterization work at Microsoft's Gray Systems Lab.

– 05/2021, Guest Lecturer @ University of Idaho: CS 501 Seminar: Contemporary Issues, 'Inductive Biases in Deep Linguistic Structured Prediction'

– 06/2021, Conference Presentation on NAACL'2021, 'A Comparative Study on Schema-Guided Dialogue State racking'

– 12/2020, Talk on EMNLP'2020 Watch Party@Amazon Lex, 'Task-oriented Conversational Semantic Parsing'

– 11/2019, Conference Presentation on MRP 2019, 'Amazon at MRP 2019: Parsing Meaning Representations with Lexical and Phrasal Anchoring'

Jie Cao (jie.cao@colorado.edu, 8014485203)   **Resume (December, 2023)**

Postdoc at NSF National AI Institute for Student-AI Teaming | University of Colorado Boulder   Ph.D. at University of Utah

Software

Other code for papers can be found in https://github.com/utahnlp and https://github.com/mlciv

– Therapist-Observer, a family of neural components for various hierarchical dialogue models for motivaitional interviwer, described in our ACL'19 [8].

– LAPA-MRP, our open-sourced system submission to MRP 2019 shared task at CoNLL 2019. Our model ranked 1st in the AMR subtask, 5th in UCCA, 6th in PSD and 7th in DM.

– sgd, which implements a family of neural components for schema-guided dialog used in our NAACL'2021 paper [6], inclduing encoder architectures, description styles, supplementary training and so on.

– dt-proxy, A nodejs-based token proxy to support distributed data copy between secure and insecure hadoop clusters.

– renjin, a JVM-based interpreter for the R language. https://github.com/bedatadriven/renjin

Academic Service

Program Committee Member/Reviewer

– International Conference on Artifical Intelligence in Education (AIED), 2023

– Conference of the Association for the Advancement of Artificial Intelligence (AAAI), 2020–23

– Annual meeting of the Association of Computational Linguistics (ACL), 2020-23

– International Conference on Computational Linguistics (COLING), 2020,22,24

– Conference on Computational Natural Language Learning (CoNLL), 2021-2023

– Conference of the European Chapter of the ACL (EACL), 2021

– Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020-2023

– North American Chapter of the ACL (NAACL), 2021

– Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning (MRP'2019@CoNLL)

– ACL Rolling Review, 2021-23

– Workshop on Innovative Use of NLP for Building Educational Applications (BEA), 2023

– Workshop on NLP for Conversational AI (NLP4ConvAI), 2023

– Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP), 2023

– Workshop on Speech and Language Technology in Education (SLaTE), 2023

– Workshop on Queer in AI, 2023