Last modificated date: 10/10/2023

Chao Jiang

CONTACT INFORMATION

CODA at 756 W Peachtree St NW
Georgia Institute of Technology
Atlanta, GA 30332

chaojiang@gatech.edu
https://chaojiang06.github.io/
Google Scholar

RESEARCH INTERESTS

Natural Language Processing, Machine Learning, and Social Media

EDUCATION

Georgia Institute of Technology — Atlanta, GA
Ph.D. student in Computer Science — 08/2020 - 08/2024 (expected)
Advisor: Dr. Wei Xu

The Ohio State University (Transfer out) — Columbus, OH
Ph.D. student in Computer Science and Engineering — 08/2018 - 08/2020
Advisor: Dr. Wei Xu

University of Virginia — Charlottesville, VA
Master in Computer Science (GPA: 3.97) — 08/2016 - 05/2018
Advisor: Dr. Kai-Wei Chang

Tianjin University — Tianjin, China
Bachelor of Engineering in Communication Engineering — 09/2011 - 07/2015

SELECTED PUBLICATIONS

[8] arXivEdits: Understanding the Human Revision Process in Scientific Writing
Chao Jiang, Wei Xu and Samuel Stevens
EMNLP 2022, long paper

TL;DR: We provide a complete computational framework for studying text revision in the scientific writing domain, including a human-annotated corpus arXivEdits, an in-depth analysis to unveil "*what common strategies researchers use to improve the writing of their paper*", and a pipeline system for automatic edit extraction and fine-grained intention classification.

[7] Neural semi-Markov CRF for Monolingual Word Alignment
Wuwei Lan*, **Chao Jiang*** and Wei Xu (* indicates equal contribution.)
ACL 2021, long paper

TL;DR: We present a novel neural semi-Markov CRF alignment model, which unifies word and phrase alignments through variable-length spans. Our model achieves SOTA performance in both in-domain and out-of-domain evaluations. We also create a new benchmark with human annotations that cover four different text genres to evaluate this task in more realistic settings.

[6] Neural CRF Model for Sentence Alignment in Text Simplification
Chao Jiang, Mounica Maddela, Wuwei Lan, Yang Zhong and Wei Xu
ACL 2020, long paper (received >100 citations)

TL;DR: We propose a novel neural CRF alignment model for monolingual sentence alignment, which leverages the sequential nature of sentences in parallel documents, and utilizes a neural sentence pair model to capture semantic similarity. It outperforms all prior work by >5 points in F1. Using our CRF aligner, we construct two new text simplification datasets Newsela-Auto and Wiki-Auto, which are much larger and of better quality compared to the existing corpora. A Transformer-based seq2seq model trained on our datasets establishes a new SOTA for text simplification in both automatic and human evaluation.

[5] Learning Word Embeddings for Low-Resource Languages by PU Learning
Chao Jiang, Hsiang-Fu Yu, Cho-Jui Hsieh and Kai-Wei Chang
HLT-NAACL 2018, long paper

TL;DR: We study how to effectively learn a word embedding model on a corpus with only a few million tokens, where the co-occurrence matrix is very sparse. In contrast to existing approaches often only sample a few unobserved word pairs as negative samples, we argue that the zero entries in the co-occurrence matrix also provide valuable information. We then design a Positive-Unlabeled Learning (PU-Learning) approach to factorize the co-occurrence matrix and validate the proposed approaches in four different languages.

OTHER PAPERS

[4] Frustratingly Easy Label Projection for Cross-lingual Transfer
Yang Chen, **Chao Jiang**, Alan Ritter and Wei Xu
Findings of ACL 2023, long paper

[3] Improving Large-scale Paraphrase Acquisition and Generation
Yao Dou, **Chao Jiang** and Wei Xu
EMNLP 2022, long paper

[2] Discourse Level Factors for Sentence Deletion in Text Simplification
Yang Zhong, **Chao Jiang**, Wei Xu and Junyi Jessy Li
AAAI 2020, long paper

PREPRINTS

[1] Multi-task Learning for Universal Sentence Embeddings: A Thorough Evaluation using Transfer and Auxiliary Tasks
Wasi Uddin Ahmad, Xueying Bai, Zhechao Huang, **Chao Jiang**, Nanyun Peng and Kai-Wei Chang
arXiv, long paper

AWARDS

- AAAI-2020 Student Scholarship 2020
- Outstanding Graduate Thesis in Tianjin University (top 5% in university) 2015
- Honorable Mention Prize of Mathematical Contest in Modeling (MCM) 2014
- Second Prize of China Undergraduate Mathematical Contest in Modeling(CUMCM) (top 2% in China) 2013
- Merit Student Scholarship in School of Electronic Information Engineering (top 15% in school) 2013

TECHNICAL TALKS

- **Guest Lecturer** at CSE 5525 Speech and Language Processing 02/2020
- **Poster presentation** at OSU CSE Student Research Poster Exhibition 02/2020
- **Poster presentation** at Mid-Atlantic Student Colloquium on Speech, Language and Learning 05/2017

PROFESSIONAL ACTIVITIES

- **Teaching Assistant**
  - OMSCS 7650 Natural Language Processing
    05/2023 - 12/2023, Georgia Institute of Technology
  - CS 7650 Natural Language Processing
    01/2022 - 05/2022, Head TA, Georgia Institute of Technology
  - CSE 2331 Data Structures and Algorithms
    08/2018 - 12/2018, The Ohio State University
  - CS 4501 Machine Learning
    01/2018 - 05/2018, University of Virginia
- **Reviewer** for ACL 2019 - present; EMNLP 2022 - present; ARR 2023 - present; AAAI 2021; COLING 2020, W-NUT 2021, 2022