Shuguang Chen

University of Houston, 4800 Calhoun Rd, Houston, TX 77004, United States

☐ (+1) 806-459-0724 | ✉ sgchen.cs@gmail.com | 🏠 sgchen.info | 🔗 github.com/nehcgs

EDUCATION

University of Houston
Houston, TX

Ph.D. in Computer Science
Aug. 2018 - Dec. 2022 (Expected)

- Advisor: Prof. Thamar Solorio
- Research Area: Natural Language Processing (NLP)
- Dissertation: Named Entity Recognition on Social Media

Beijing Forestry University
Beijing, China

B.S. in Computer Science
Sep. 2014 - Jul. 2018

- Advisor: Prof. Meng Wei
- Thesis: Music Generation Using Recurrent Neural Networks

RESEARCH INTEREST

My research interest lies in Natural Language Processing (NLP) with a special focus on **neural sequence labeling**. My research aims to overcome linguistic challenges to reduce performance degradation (i.e., a decrease in models' predictive performance) and improve model robustness (i.e., the ability to predict consistently under various conditions).

WORK EXPERIENCE

University of Houston, RiTUAL Lab
Houston, TX

Research Assistant, Advisor: Prof. Thamar Solorio
Aug. 2019 - Present

- Neural sequence labeling on user-generated text
- Data transformation for model robustness
- Knowledge transfer in linguistic code-switching

Microsoft Research, Deep Learning Group
Redmond, WA

Research Intern, Mentor: Dr. Hao Cheng
May. 2022 - Apr. 2022

- Research Topic: Temporal Text Modeling for Event Detection.
- Investigated the impact of temporal information on dialogue data
- Proposed methods to induce temporal information into transformers.

Melax Technologies Inc
Houston, TX

NLP Intern, Mentor: MS Jingqi Wang
May. 2021 - Aug. 2021

- Research Topic: Document Classification and Information Extraction.
- Developed a data annotation platform for NLP tasks, including named entity recognition and question answering.
- Studied named entity recognition and relation extraction with transformers for biomedical data.

SELECTED PROJECT

Data Augmentation for Cross-domain Entity Recognition [Github]
Houston, TX

Advisor: Prof. Thamar Solorio
Feb. 2021 - Sep. 2021

- Research Topic: Data Augmentation for Low-resource Named Entity Recognition.
- Investigated data transformation with GANs to improve model robustness.
- Proposed a new method to transfer the data from low-resource to high-resource domains.

Multimodal Named Entity Recognition on Social Media [Github]
Houston, TX

Advisor: Prof. Thamar Solorio
Sep. 2019 - Sep. 2021

- Research Topic: Multimodal Named Entity Recognition.
- Studied multimodal information representation, extraction, and fusion.
- Proposed methods to reduce performance degradation by fusing visual and textual information.

Named Entity Recognition on Diachronic Twitter Data [Github]
Houston, TX

Advisor: Prof. Thamar Solorio
Jun. 2020 - Apr. 2021

- Research Topic: Named Entity Recognition under Temporal Drift.
- Probed the impact of temporal drift on entity memorization and context generalization.
- Presented a method to efficiently update model parameters with the most informative data.

HONORS & AWARDS

Awards & Scholarships

2022	**Cullen Graduate Student Success Fellowship**, University of Houston	*Houston, TX*
2018	**Outstanding Graduate Awards**, Beijing Forestry University	*Beijing, China*

Academic and Scientific Competitions

2018	**Academic Merit Scholarship**, Beijing Forestry University	*Beijing, China*
2016	**Bronze Metal**, China Collegiate Programming Contest (CCPC)	*Hangzhou, China*
2016	**2nd prize**, The 7th Blue Bridge Cup National Software Competition	*Beijing, China*

PROFESSIONAL SERVICE

Research Activities

- **Program Committee Member** for EMNLP 2020, ACL 2020, MCPR 2021, NAACL 2021, W-NUT 2021.
- **Co-organizer** for the fifth workshop on Computational Approaches to Linguistic Code-Switching **[CALCS]**.
- **Webmaster** for Linguistic Code-switching Evaluation Benchmark **[LinCE]**

Teaching & Mentoring

- **COSC 3320 Algorithms and Data Structures**, Instructor: Prof. Rakesh Verma
- **COSC 3360 Fundamentals of Operating Systems**, Instructor: Prof. Jehan-Francois Paris
- **COSC 6358 Interactive Game Development**, Instructor: Prof. ChangHoon Yun

Licenses & certifications

- **Deep Learning Specialization**, Coursera, Instructor: Andrew Ng **[Certificate]**
- **Build Basic Generative Adversarial Networks (GANs)**, Coursera, Instructor: Sharon Zhou, **[Certificate]**
- **Computer Vision**, Udacity, Instructor: Suzanne Camacho, **[Certificate]**

PUBLICATION

Pre-prints

- **CALCS 2021 Shared Task: Machine Translation for Code-Switched Data**
  Shuguang Chen, Gustavo Aguilar, Anirudh Srinivasan, Mona Diab and Thamar Solorio.
  arXiv:2202.09625

Peer-reviewed Publications

- **Style Transfer as Data Augmentation: A Case Study on Named Entity Recognition**
  Shuguang Chen, Leonardo Neves, Thamar Solorio.
  Under Review

- **A Simple Approach to Jointly Rank Passages and Select Relevant Sentences in the OBQA Context**
  Man Luo, **Shuguang Chen**, Chitta Baral.
  NAACL 2022 SWR

- **Data Augmentation for Cross-Domain Named Entity Recognition**
  Shuguang Chen, Gustavo Aguilar, Leonardo Neves, Thamar Solorio.
  EMNLP 2021

- **Can images help recognize entities? A study of the role of images for Multimodal NER**
  Shuguang Chen, Gustavo Aguilar, Leonardo Neves, Thamar Solorio.
  EMNLP 2021 W-NUT

- **Mitigating Temporal-Drift: A Simple Approach to Keep NER Models Crisp**
  Shuguang Chen, Leonardo Neves, Thamar Solorio.
  NAACL 2021 SocialNLP

- **Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching**
  Thamar Solorio, **Shuguang Chen**, Alan W Black, Mona Diab, Sunayana Sitaram, Victor Soto, Emre Yilmaz.
  NAACL 2021 CALCS