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

Ph.D. Student in Computer Science Aug 2018 – Dec 2022 (Expected)
University of Houston, Houston, TX, United States
Research: Natural Language Processing, Advisor: Dr. Thamar Solorio

B.S. in Computer Science and Technology Sept 2014 – July 2018
Beijing Forestry University, Beijing, China
Thesis: Music Generation Using Recurrent Neural Networks

RESEARCH INTEREST

Natural Language Processing, with a special focus on Neural Sequence Labeling, Text Generation, and Machine Translation.

WORK EXPERIENCE

Research Assistant Aug 2019 - Present
University of Houston, RiTUAL Lab, Dr. Thamar Solorio
- Neural sequence labeling on user-generated text
- Machine Translation in linguistic code-switching

NLP Developer Intern May 2021 - Aug 2021
Melax Technologies, Inc, Jingqi Wang
- Developed an annotation platform for named entity recognition and relation extraction task
- Conducted research on document classification and information extraction with biomedical data

RESEARCH EXPERIENCE

Project: Data Augmentation for Cross-domain Entity Recognition [Github] Feb 2021 – Sept 2021
Supervisor: Dr. Thamar Solorio
- Proposed a novel neural architecture to learn the mapping between domains
- Augmented data for low-resource NER by transferring the data from high-resources domains

Project: Multimodal Named Entity Recognition on Social Media [Github] Sept 2019 – Sept 2021
Supervisor: Dr. Thamar Solorio
- Conducted research on multimodal information extraction, fusion and inference
- Worked on analysis of image representations and multimodal fusion techniques

Project: Machine Translation for Code-switched Data Dec 2020 – Aug 2021
Supervisor: Dr. Thamar Solorio
- Created new standard datasets for evaluating machine translation on code-switched data
- Provided baseline systems and analyzed the challenges of code-switched data generation

Project: Named Entity Recognition on Diachronic Twitter Data [Github] June 2020 – Apr 2021
Supervisor: Dr. Thamar Solorio

- Designed a simple method to detect posts that are becoming trends on social media platform
- Presented a strategy to efficiently update model parameters by selecting the most informative data

Project: Reducing Rote Memory Learning of Highly Frequent Entities Sept 2020 – Dec 2020
Supervisor: Dr. Thamar Solorio

- Investigated the performance in entity memorization and contextual generalization of NER models
- Proposed potential solutions to reduce the model's reliance on entity memorization

Project: Handwriting Recognition with Recurrent Neural Networks (RNNs) Mar 2017 – May 2018
Supervisor: Dr. Wei Meng

- Achieved handwriting recognition, study, and generation functionalities with neural networks.
- Trained the RNNs with a mixture dense layer and the source data of different handwriting styles.

HONORS AND AWARDS

Awards & Scholarship

- **School-level Outstanding Graduate Awards**, Beijing Forestry University 2018
- **Academic Merit Scholarship**, School of Information, Beijing Forestry University 2017

Academic and Scientific Competitions

- **Bronze Metal**, Association for Computing Machinery - China Collegiate Programming Contest 2016
- **2nd prize**, The 7th Blue Bridge Cup National Software Competition Heats of Beijing Region 2016

RESEARCH ACTIVITIES AND SERVICE

- Reviewer at EMNLP 2020, ACL 2020, MCPR 2021, NAACL 2021, W-NUT 2021
- Co-organizer of the 5th workshop on Computational Approaches to Linguistic Code-Switching (CALCS)

PUBLICATIONS

- Data Augmentation for Cross-Domain Named Entity Recognition. **Shuguang Chen**, Gustavo Aguilar, Leonardo Neves, Thamar Solorio. Accepted to **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. Accepted to W-NUT at **EMNLP 2021**
- 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. Accepted to CALCS at **NAACL 2021**
- Mitigating Temporal-Drift: A Simple Approach to Keep NER Models Crisp. **Shuguang Chen**, Leonardo Neves, Thamar Solorio. Accepted to SocialNLP at **NAACL 2021**
- A Simple Approach to Jointly Rank Passages and Select Relevant Sentences in the OBQA Context. Man Luo, **Shuguang Chen**, Chitta Baral. **arXiv preprint**
- CALCS 2021 Shared Task: Machine Translation for Code-Switched Data. **Shuguang Chen**, Gustavo Aguilar, Anirudh Srinivasan, Mona Diab and Thamar Solorio. **arXiv preprint**