Suyu Ge

(+86) 188-1023-8103 • gesy17@mails.tsinghua.edu.cn • Google Scholar • Website

Education

Tsinghua University, Beijing, China 08/2017 – 07/2021 (expected)

B.Eng. in Electronic Engineering, GPA: 3.80/4.00

Selected Publications

- **Suyu Ge**, Chuhan Wu, Fangzhao Wu, Tao Qi, Yongfeng Huang, "*Graph Enhanced Representation Learning for News Recommendation*", in the Web Conference (**WWW**), p. 2863-2869, 2020, arXiv 2003.14292 [PDF]
- **Suyu Ge**, Lu Cheng, Huan Liu, "*Improving Cyberbully Detection with User Interaction*", to appear, in the Web Conference (**WWW**), 2021, arXiv 2011.00449, [PDF]
- **Suyu Ge**, Tao Qi, Chuhan Wu, Fangzhao Wu, Xing Xie, Yongfeng Huang, "*Helpfulness-aware Review Based Neural Recommendation*", in **CCF Transactions** on Pervasive Computing and Interaction, p. 285–295, 2019, [PDF]
- Davy Weissenbacher, **Suyu Ge**, Ari Klein, Karen Oconnor, Graciela Gonzalez-Hernandez, "*Active Neural Networks to Detect Medication Change in Social Media*", in the Intelligent Systems for Molecular Biology (**ISMB**), poster of the Text Mining Session, 2020, [PDF] [Link]
- **Suyu Ge**, Fangzhao Wu, Chuhan Wu, Tao Qi, Yongfeng Huang, Xing Xie, "*FedNER: Medical Named Entity Recognition with Federated Learning*", **Preprint**, 2020, arXiv 2003.09288, [PDF]

For a full publication list, please visit here.

Research Experience

Data Mining Group (DMG), University of Illinois at Urbana-Champaign 07/2020 – Present

Topic: Topic-based Corpus Summarization with Minimal Supervision
Advisor: Prof. Jiawei Han

- Initiated a novel text mining task for topic-based corpus summarization with only seed topics, which facilitated automatic knowledge discovery and organization.
- Mined hierarchical topic structures from large corpora and profiled the corpora by generating sentiment-aware summarization for each topic. *Work in progress*.

Data Mining and Machine Learning lab (DMML), Arizona State University 02/2020 – 10/2020

Topic: Cyberbully Detection
Advisor: Prof. Huan Liu

- Designed a graph attention network based architecture to model the temporal and semantic repetition in cyberbully.
- Outperformed baselines in terms of Recall, F1, and AUC on Instagram. *First-author paper at WWW 2021*.

New Generation Network Group (NGN), Tsinghua University 10/2018 – 05/2020

Topic: Text-based Recommendation and Privacy-preserving Named Entity Recognition (NER)
Advisor: Prof. Yongfeng Huang Mentor: Fangzhao Wu@Microsoft Research Asia (MSRA)

- (News Rec.) Enhanced representation learning by exploring high-order relatedness in the user-news interaction graph. Outperformed baselines by AUC, MRR and nDCG on Microsoft News. *First-author paper at WWW 2020*.
- (Item Rec.) Devised a framework to improve review based recommendation with predicted helpfulness levels of reviews. Achieved lower RMSE scores on four Amazon Reviews datasets with partially labeled helpfulness levels. *First-author paper at CCF Transactions on Pervasive Computing and Interaction.*
- (Privacy-preserving NER) Applied federated learning to protect medical data privacy in the training process of a decomposed NER model. Improved F1 scores over both single- and multi-task counterparts with provided privacy guarantee on three benchmark medical NER datasets. *First-author paper under review*.

Health Language Processing Lab (HLP), University of Pennsylvania 07/2019 – 12/2019

Topic: Medical Non-adherence Discovery
Advisor: Prof. Graciela Gonzalez Hernandez

- Performed semi-automatic analysis of social media with active and transfer learning to detect non-adherence.
- Reduced needed annotation by 50% and discovered 8 non-adherence reasons. *Second-author poster at ISMB 2020.*

Academic Evaluation

Text Mining for Health in Social Media, SMM4H 2019 Shared Task colocated with ACL 2019

System Description: Detecting and Extracting of ADR Mentioning Tweets with Multi-Head Self Attention [PDF]

- Proposed a language model enhanced Self-Att architecture for adverse drug effect classification and extraction.
- Our system ranked 2^{nd} in the adverse effect extraction subtask. *First-author paper accepted by the workshop*.

Contextual Emotion Detection in Text, SemEval 2019 Task 3 colocated with NAACL 2019

System Description: Dialog Emotion Classification using Attentional LSTM-CNN [PDF]

- Proposed an attentional LSTM-CNN model to classify the emotions of short turns of dialogues.
- Our system ranked top 10% among 165 participants. *First-author paper accepted by the workshop*.

Toponym Resolution in Scientific Papers, SemEval 2019 Task 12 colocated with NAACL 2019

System Description: Toponym Detection and Disambiguation on Scientific Papers [PDF]

- Proposed a TagLM-based NER model with various domain features for toponym detection and resolution.
- Our system ranked 2^{nd} in the toponym detection subtask. *Second-author paper accepted by the workshop*.

Awards and Honors

Nanxiang Jiang Scholarship 2019

- 40/3300+ in Tsinghua University

The First Prize of Tsinghua Research Challenge Cup 2020

- 10/200+ in Tsinghua University

Research Excellence Scholarship 2019/2020

- $\sim$20/278 in Electronic Engineering Department

Academic Excellence Scholarship 2019

- $\sim$20/278 in Electronic Engineering Department

Philobiblion Scholarship 2018

- 5/278 in Electronic Engineering Department

Extracurricular

Electronic Engineering Student Association of Science and Technology, Tsinghua University 2018-2020

Committee Member of Software Department

- Designed front-end webpages of 2 coding competitions jointly with other members.

Electronic Engineering Swimming Team, Tsinghua University 2017-2021

Captain

- Participated and organized 6 swimming competitions.

Miscellaneous

English:

- TOEFL: 109 (Reading 29; Listening 28; Speaking 27; Writing 25)
- GRE: 327 (Verbal 157; Quantitative 170; Writing 4.0)

Skills:

- Programming: Python, C/C++, MATLAB, Bash, JavaScript, SQL
- Machine Learning: Keras, Tensorflow, PyTorch, scikit-learn