Guokan Shang

Curriculum Vitae

linkedin.com/in/guokan-shang
+33 (0)6 59 94 45 19
guokan.shang@hotmail.com
French, English, Chinese
31 Boulevard Vauban, 78280 Guyancourt, France
outlook.live.com/my-calendar

Work experience & Internships

- Research scientist at LINAGORA Labs, since December 2016
  - tackling research problems related to Spoken Language Understanding / NLP for conversations.
  - working on various research projects / proposals, collaborating with industrial and academic partners, participating in report writing and project review.
  - regularly publishing and presenting scientific articles at international conferences.
  - conducted my Ph.D. entitled "***Spoken Language Understanding for Abstractive Meeting Summarization***" in collaboration with DaSciM, LIX, École Polytechnique, under the supervision of Prof. Michalis Vazirgiannis and Mr. Jean-Pierre Lorré, funded by French CIFRE Ph.D. fellowship.
- Master internship at DaSciM, 6 months
  - *Approximate Reinforcement Learning with Mixture of Experts*

Education

2016–2021	Ph.D.	Computer Science, Data and Artificial Intelligence	École Polytechnique
2015–2016	Master	Machine Learning, Information and Content	Paris-Saclay University
2013–2015	Engineer	Informatics and Information Systems	University of Technology of Troyes
2009–2013	Bachelor	Computer Science and Technology	Harbin Institute of Technology

Research projects & activities

- Member of the SUMM-RE project funded by ANR (2020–2024).
  The objective is to combine expertise in theories of discourse interpretation with recent developments in distant supervision to improve the automatic production of meeting summaries and minutes from spoken data. Additionally, we are building a 100 hour audio-video-text annotated corpus of multi-party, meeting-like interactions in French.
- Member of the LinTO project funded by Bpifrance (2018–2021).
  The goal was to develop a privacy-aware AI-powered open-source conversational meeting assistant for professional use in a corporate environment.
- Member of the OpenPaaS project funded by Bpifrance (2015–2019).
  The aim was to offer an open source alternative to Microsoft 365 & Google Suite, providing all the functionalities for enterprise collaborative work: mail box, instant messaging, video conferencing, etc.

Research papers

1. Shang, G, W Ding, Z Zhang, A Tixier, P Meladianos, M Vazirgiannis, and JP Lorré (2018). Unsupervised Abstractive Meeting Summarization with Multi-Sentence Compression and Budgeted Submodular Maximization. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) - **ACL 2018***.
2. Shang, G, A Tixier, M Vazirgiannis, and JP Lorré (2020). Energy-based Self-attentive Learning of Abstractive Communities for Spoken Language Understanding. In: *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing - **AACL-IJCNLP 2020***.
3. Shang, G, A Tixier, M Vazirgiannis, and JP Lorré (2020). Speaker-change Aware CRF for Dialogue Act Classification. In: *Proceedings of the 28th International Conference on Computational Linguistics - **COLING 2020***.
4. Kamal Eddine, M, G Shang, A Tixier, and M Vazirgiannis (2022). FrugalScore: Learning Cheaper, Lighter and Faster Evaluation Metrics for Automatic Text Generation. In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) - **ACL 2022***.
5. Kamal Eddine, M, G Shang, and M Vazirgiannis (2023). DATScore: Evaluating Translation with Data Augmented Translations. In: ***Findings** of the Association for Computational Linguistics: **EACL 2023***.
6. Rennard, V, G Shang, J Hunter, and M Vazirgiannis (2023). Abstractive meeting summarization: A survey. *arXiv preprint arXiv:2208.04163 - accepted for publication in **TACL Journal***.

Technical skills: Python, Keras, Scikit-learn, spaCy, Transformers, Apache Spark, NLTK, Gensim, Networkx, Git, SQL, HTML, Linux, Celery, Redis, FastAPI, etc.