Kexin Wang

✉ kexin.wang.2049@gmail.com ⛓ kwang2049 🐦 @KexinWang2049

👤 Basic Info

- Male   born in May, 1995   living in Darmstadt, Hessen, Germany
- A Ph.D. student at the Ubiquitous Knowledge Processing Lab (UKP-TUDA)
- Final degree of Master of Science in Engineering
- Interested in text representation, information retrieval and low-resource NLP
- Speak English, Chinese and a little German (A2)

🎓 Education & Internship

Department of Computer Science, Technical University of Darmstadt   Sept., 2020 – Now
- Ph.D. student in Computer Science   • Supervisor: Prof. Iryna Gurevych and Dr. Nils Reimers
- Research Field: Natural Language Processing   • Lab: Ubiquitous Knowledge Processing Lab

Institute of Automation, Chinese Academy of Sciences   2017 – 2020
- M.Sc. in Engineering   • GPA: 3.57/4   • Supervisor: Prof. Yu Zhou and Prof. Chengqing Zong
- Research Field: Natural Language Processing   • Lab: National Laboratory of Pattern Recognition

College of Computer Science, Nanjing University of Aeronautics and Astronautics 2013 – 2017
- B.Eng. in Computer Science and Technology   • GPA: 4.1/5   • Supervisor: Prof. Han Sun

Algorithm Researcher Internship at DAMO Academy, Alibaba Group   2019
- Developing emotion-controlled chit-chat robots for e-commerce platform.

👥 & Research Experience

- **GPL: Generative Pseudo Labeling for Unsupervised Domain Adaptation of Dense Retrieval**   2021–
  First author   Previous work shows dense retrievers severely degrades under a domain shift, despite trained on massive labeled data. To face the challenge, we propose a novel unsuperivsed domain adaptation method GPL, which combines query generation with pseudo labeling from a cross-encoder. On six domain-specialized datasets, GPL outperforms an out-of-the-box SOTA dense retriever by up to 8.9 points nDCG@10. Surprisingly, we also find our previously proposed pre-training method TSDAE can further improve GPL by 1.0 point nDCG@10 in average. The work has been accepted by **NAACL 2022**. [Paper][Code]

- **Unsupervised Sentence Embedding Learning for Domain- and Task-Specific Scenarios**   2020–2021
  First author   Learning sentence embeddings often requires large amount of labeled data, which is hardly satisfied in many cases, especially for domain- and task-specific scenarios. We present a new SOTA un-supervised method based on pre-trained Transformers and Sequential Denoising Auto-Encoder (TSDAE) which outperforms previous approaches by up to 6.4 points. We show it also works great in few-shot learn-ing and domain adaptation. This work has been published in **EMNLP 2021 Findings**. The code has been integrated into the popular **Sentence-Transformers** repository. [Paper][Code][Benchmark]

- **Structurally-Comparative Hinge Loss for Dependency-based Neural Text Representation** 2019 – 2020
  First author   Dependency-based graph neural networks (DepGCNs) are proven helpful for modelling text. We find DepGCNs cheat to escape using the structure information under the commonly used cross-entropy loss. We propose a new loss function based on conditional mutual information to solve this problem for DepGCNs with dependency tree from an external parser or induction. New SOTA results on classification tasks are achieved. This work has been published in the journal **ACM TALLIP**. [Paper]

- **Understanding Memory Modules on Learning Simple Algorithms**   2018 – 2019
  First author   Recent work shows memory modules are crutial for the generalization of neural networks on learning simple algorithm. However, few previous studies focus on understanding the working mechansim of them. To face the challenge, we propose a hypothesis-then-verify pipeline supported by visualization and dimension-reduction techniques. This work has been accepted in **IJCAI 2019 workshop XAI**. [Paper]

⚙ RESEARCH-ORIENTED ENGINEERING

- **Sparse-Retrieval: Train and Evaluate All the Sparse-Retrieval Methods in One Stop** 2022–
  Equal contributor: Inference   Recently, besides the success of neural dense retrieval, neural sparse retrieval has drawn more and more attention. One can imagine it as BM25 with learned weights and neural models. And thus exact search with sparse retrieval can be done efficiently with CPUs as BM25, supported by lookup tables. However, inference and evaluation with sparse retrieval are not trivial and there has not been a one-stop soluation for that. And the architectures of different approaches have not been well compared, despite the differences in design are quite small. To face these two challenges, we created the toolkit, *sparse-retrieval*. Specifically for inference, I summarize the common evaluation setup into 6 steps and wrap them into one reusable pipeline (supported by Pyserini/Lucene) for any sparse retrieval model. [Code]

- **UKP-SQuARE: Software for Question Answering Research** 2021–
  Second contributor: Datastores & retrieval   Recent advances in NLP and IR have given rise to a diverse set of question answering tasks that are of different formats (e.g., extractive, abstractive), require different model architectures (e.g., generative, discriminative) and setups (e.g., with or without retrieval). However, there exists no framework where users can easily explore and compare such pipelines and can extend them according to their needs. To address this issue, we present SQuARE, an extensible online QA platform that allows users to conveniently query, analyze and even build themselves a large collection of modern QA pipelines via a user-friendly web interface and integrated behavioural tests. Specifically for datastores in SQuARE that enable retrieval over knowledge sources e.g. Wikipedia, we utilizes (1) Elasticsearch to maintain document collections and support BM25 retrieval and (2) Traefik & Faiss & Docker to support neural retrieval based on vector search in a scalable and efficient way. This work has been accepted in **ACL 2022 System Demonstrations**. [Live demo][Code]

✿ MAIN HONORS & SCHOLARSHIPS

- TU Doctoral Scholarship, TU Darmstadt 2020 – 2023
- Principal's Special Commendation Nomination, Nanjing University of Aeronautics and Astronautics 2016
- Huawei Scholarship, Nanjing University of Aeronautics and Astronautics 2016
- National Scholarship (1/110), Ministry of Education of the People's Republic of China 2015