

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. study 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 • 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 • Supervisor: Prof. Han Sun

Algorithm Researcher Internship at DAMO Academy, Alibaba Group 2019

- Developing emotion-controlled chat robots for e-commerce platform.

🧑‍🔬 PROJECTS & 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 unsupervised 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 is in progress. [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 unsupervised 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 learning and domain adaptation. This work has been accepted in **EMNLP2021 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 crucial for the generalization of neural networks on learning simple algorithm. However, few previous studies focus on understanding the working mechanism 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]