Danqing Wang

Computer Science - University of California, Santa Barbara - CA
✉ dqwang122@gmail.com   ●   🏠 dqwang122.github.io

Research Interest: Interested in various generation tasks, including natural language and biological sequence. Devoted to helping the machine generate new content that benefits human life. Mainly focus on (but not limited to):
- Natural Language: Human Alignment, Personalization, Controllable Generation, etc.
- Biological Sequence: Peptide Design, Antibody Discovery, etc.

Github: `https://github.com/dqwang122`

Education

PhD in Computer Science   **UC, Santa Barbara**
Advisor: Prof. Lei Li   *2022.9 – Current*

Master in Computer Science   **Fudan Univerisity**
Advisor: Prof. Xipeng Qiu and Prof. Xuanjing Huang   *2018.9 – 2021.1*
Ranking: Top 5%

Bachelor in Computer Science and Technology   **Fudan Univerisity**
Ranking: Top 10%   *2014.9 – 2018.6*

Experience

Student Visitor   **CMU (Language Technologies Institute)**
Advisor: Lei Li   *2023.9 – Current*

Research Scientist Intern   **Meta AI (FAIR)**
Advisor: Yuandong Tian   *2023.6 – 2023.9*

Research Scientist   **ByteDance Research (AI Lab)**
Advisor: Jiaze Chen, Hao Zhou, and Lei Li   *2020.4 – 2022.8*

Main Publication

Learning Personalized Story Evaluation %
Danqing Wang, Kevin Yang, Hanlin Zhu, Xiaomeng Yang, Andrew Cohen, Lei Li, Yuandong Tian   ***Submit to ICLR***
- Investigate personalization in open-ended text generation. Propose two personalized story evaluation datasets.
- Develop a personalized story evaluation model PerSE to give detailed review-specific feedback on the story plot. It achieves a higher correlation with humans than GPT-4 by 15.8% on the Kendall correlation.

Learning from Mistakes via Interactive Study Assistant for Large Language Models %
Danqing Wang, Lei Li   ***EMNLP 2023***
- Propose a general framework SALAM to assist LLM to learn from mistakes. It introduces a study assistant to provide global analysis and guidelines for mistakes.
- Develop the model-specific study assistant via the imitation learning of successful experiences, making feedback better aligned with specific LLM behaviors.

Instructscore: Towards Explainable Text Generation Evaluation with Automatic Feedback %
*Wenda Xu, **Danqing Wang**, Liangming Pan, Zhenqiao Song, Markus Freitag, William Yang Wang, Lei Li*   ***EMNLP 2023***
- Introduce an interpretable metric INSTRUCTSCORE for text generation tasks. It provides comprehensive evaluation by diagnostic reports, culminating in a well-justified final score.
- It significantly surpasses strong baselines in four generation tasks and achieves the best performance in unseen keyword-to-dialogue generation.

ALGO: Synthesizing Algorithmic Programs with Generated Oracle Verifiers %
*Kexun Zhang, **Danqing Wang**, Jingtao Xia, William Yang Wang, Lei Li*   ***NeurIPS 2023***
- Introduce a novel framework for algorithm synthesis (ALGO) verified by LLM-generated oracles.
- Equipped with ALGO, we achieve an $8\times$ better one-submission pass rate over the Codex model and a $2.6\times$ better one-submission pass rate over CodeT.

On Pre-training Language Model for Antibody %
Danqing Wang, Fei Ye, Hao Zhou   ***ICLR 2023***

- The first comprehensive antibody benchmark: AnTibody Understanding Evaluation (ATUE).
- Explore the representation capability of pre-trained language models on antibody problems with different specificity, especially the influence of the introduction of evolutionary information (EATLM).

○ **Accelerating Antimicrobial Peptide Discovery with Latent Structure** ⸽

Danqing Wang, Zeyu Wen, Fei Ye, Lei Li, Hao Zhou　　　　　　　　　　　　　　　　　　　*KDD 2023*

- Sample peptides from the latent secondary structure space to control the peptide properties.
- The generated peptides have a high AMP probability (93.62%) and 2/21 show high activity in wet laboratory experiments.

○ **Contrastive Aligned Joint Learning for Multilingual Summarization** ⸽

Danqing Wang, Jiaze Chen, Hao Zhou, Xipeng Qiu, Lei Li　　　　　　　　　　　　　　*ACL 2021 Finding*

- A large-scale multilingual summarization corpus MLGSum with 1.1 million articles and summaries in 12 languages.
- Propose two tasks, contrastive sentence ranking and sentence-aligned substitution, for multilingual summarization.

○ **Heterogeneous Graph Neural Networks for Extractive Document Summarization** ⸽

Danqing Wang*, Pengfei Liu*, Yining Zheng, Xipeng Qiu and Xuanjing Huang　　　　　　　　*ACL 2020*

- Introduce word nodes to model the cross-sentence relationship for extractive summarization.
- Easily adapt the graph model from single to multiple document summarization.

○ **Extractive Summarization as Text Matching** ⸽

Ming Zhong*, Pengfei Liu*, Yiran Chen, Danqing Wang, Xipeng Qiu and Xuanjing Huang　　　　　*ACL 2020*

- Formulate extractive summarization as a semantic text matching problem and select sentences in the summary level.
- Achieve superior performance on six benchmark datasets, including state-of-the-art extractive results on CNN/DailyMail.

○ **Searching for Effective Neural Extractive Summarization: What Works and What's Next** ⸽

Ming Zhong*, Pengfei Liu*, Danqing Wang, Xipeng Qiu, Xuanjing Huang　　　　　　　　　　*ACL 2019*

- Models with an autoregressive decoder are prone to achieving better performance against non-auto-regressive ones.
- LSTM is more likely to suffer from the architecture overfitting problem while Transformer is more robust.

Academic Services

○ Program Committee of ACL (2020,2021), EMNLP (2020,2021), AAAI (2020-), ICML(2023)
○ TA for CS190I Deep Learning (Winter 2023) and CS165B Machine Learning (Spring 2023)
○ Local Organization Chair of Socal NLP Symposium 2022

Honor

○ **Sept 2022** Academic Excellence Fellowship
○ **May 2021** Shanghai Outstanding Graduates (5% of graduates)
○ **Nov. 2020** Venustech Scholarship (1% of Fudan students)
○ **Sept. 2019** Scholarship for Outstanding Students (First Prize)
○ **Dec. 2017** Fudan's Undergraduate Research Opportunities Program