Mengmeng KUANG

Email: kuangmeng@msn.com | Phone: (+86) 132-7015-3586

GitHub: https://github.com/kuangmeng | Home Page: http://about.meng.uno

EDUCATION

DEC. 2020 **M.Phil** Degree in COMPUTER SCIENCE, **The University of Hong Kong**
Supervisor: Prof. Hing-fung TING
Thesis: Data-centric Approaches for better Multiple Sequence Alignment
Research Interests: Machine Learning, Deep Learning, Bioinformatics

JUN. 2018 **B.Eng** Degree in COMPUTER SCIENCE, **Harbin Institute of Technology**
Thesis Advisor: Prof. Tiejun ZHAO
Thesis: Cross-domain High Precision Chinese Word Segmentation
GPA: 89.6/100

EXPERIENCE

Research

Accepted

[C1] **Mengmeng Kuang** and Hing-fung Ting. "A data-centric pipeline using convolutional neural network to select better multiple sequence alignment method". *In Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics.*

[C2] **Mengmeng Kuang**, Yong Liu and Lufei Gao. "DLPAlign: A Deep Learning based Progressive Alignment Method for Multiple Protein Sequences". *In Proceedings of the 11th International Conference on Computational Systems-Biology and Bioinformatics.*

[C3] **Mengmeng Kuang**, Yinzhe Wu, Lufei Gao, Diego Alonso-Álvarez, David Firmin, Jennifer Keegan, Peter Gatehouse and Guang Yang. "Three-Dimensional Embedded Attentive RNN (3D-EAR) Segmentor for Left Ventricle Delineation from Myocardial Velocity Mapping". *The 11th Conference on Functional Imaging and Modeling of the Heart.*

Submitted

[J1] **Mengmeng Kuang**, Yong Zhang, Tak-wah Lam and Hing-fung Ting. "MLProbs: A Data-centric Approach for better Multiple Sequence Alignment". *IEEE/ACM Transactions on Computational Biology and Bioinformatics.* **The reviewers rated the paper as "Good", and we have submitted the revision and are waiting for the final decision.**

[C4] **Mengmeng Kuang**, Zhenhong Chen, Lie Kang and Qiang Yan. "An efficient two-stage label noise detection method for text classification datasets". *The 2021 Conference on Empirical Methods in Natural Language Processing.*

[C5] **Mengmeng Kuang**, Zhenhong Chen*, Lie Kang, Min Tang, Penghui Hao and Qiang Yan. "MKSM: Multi-task learning based Keywords weighted Siamese Model for semantic retrieval". *The 30th ACM International Conference on Information and Knowledge Management.*

Work

Mar. 2021 - Cur | *Application Researcher,* Tencent (WeChat Group), Guangzhou
Participated in the design and implementation of vector search-related algorithms in WeChat.

Oct. 2020 - Feb. 2021 | *Research Intern,* SmartMore Tech., Hong Kong
Engaged in the data mining, data augmentation, model design and implementation of SMore OCR.

Project

Data-centric Approaches for better Multiple Sequence Alignment | Sept. 2018 - Jun. 2020
To improve the quality of multiple sequence alignment (MSA) construction on protein families, especially the "low similarity" ones, we proposed two mainly high-performance data-centric MSA methods. In the first method, a random forest classifier was used to select the most appropriate strategy for different families to obtain the temporary results and another random forest classifier was used for selecting better column-based realignment strategies. The new pipeline could improve the average accuracy of 1.65% compared with other state-of-the-art MSA tools in the test of 1356 protein families. In the second method, we mentioned a two-stage deep learning-based MSA method by training a decision-making model with CNNs, BiLSTM, Attention to arrange suitable algorithm-centric pipelines for different categories of the protein families. The average accuracy could be improved by 2.8% on 711 "low similarity" protein families through this method.

Cross-domain High Precision Chinese Word Segmentation | Nov. 2017 - Jun. 2018
To improve the accuracy of Chinese Word Segmentation, a system based on conditional random field and Viterbi algorithm was developed with Java training from the artificial word segmentation results of The People's Daily. In order to further improve the adaptability in specific fields (medicine, law and finance), heuristic rules and specific guidelines were added. Finally, this word segmentation system could get an average accuracy of 97% in these specific fields.

Teaching

Jan. 2020 - Jun. 2020 COMP1117 Computer programming
Feb. 2019 - Jun. 2019 COMP7606 Deep learning

Scholarships and Certificates

Postgraduate

Mar. 2021 The Li Ka Shing Prize (Nominated)
Mar. 2021 Outstanding Research Postgraduate Student (Nominated)
Nov. 2020 Huawei Certified ICT Associate – Artificial Intelligence
Nov. 2018 Certificate of Teaching and Learning in Higher Education
Sept. 2018 Postgraduate Scholarship

Undergraduate

Jun. 2018 Enterprise Scholarship
Nov. 2016 National Encouragement scholarship
Dec. 2015 Merit Student
Nov. 2015 National Encouragement scholarship

Languages

Chinese	Mother tongue
English	Fluent