

# 匡盟盟

邮箱: kuangmeng@msn.com | 电话: (86) 13270153586

地址: 蒲飞路香港大学黄克竞楼 307

GitHub: <https://github.com/kuangmeng> | 主页: <http://about.meng.uno>

## 教育经历

---

- 2020.12 **研究型硕士**, 计算机科学, 香港大学  
导师: Hing-fung Ting 博士  
毕设: Data-centric Approaches for better Multiple Sequence Alignment  
研究方向: 多序列比对、机器学习、深度学习
- 2018.06 **工学学士**, 计算机科学与技术, 哈尔滨工业大学  
指导教师: 赵铁军教授  
毕设: 面向领域快速移植的高精度汉语分词系统研究  
成绩: 89.6/100

## 工作经历

---

### 研究

- [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.**
- [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 (BCB'20)*.
- [C2] Mengmeng Kuang, Yong Liu and Lufei Gao. "DLPAlign: A Deep Learning based Progressive Alignment Method for Multiple Protein Sequences". *In Proceedings of the Eleventh International Conference on Computational Systems-Biology and Bioinformatics (CSBio 2020)*.

### 实习

- |                   |  |
|-------------------|--|
| 2020.10 至今        | 研究员实习, 思谋科技, 香港<br>致力于通过有条件的 GANs 和自编码器模型为 OCR 任务生成手写文本/图像样本。                    |
| 2018.07 - 2018.08 | 研究助理, 哈工大机器翻译与智能研究室, 哈尔滨<br>负责集成一个医疗信息处理平台, 包括医疗关系提取, 疾病知识图谱, 医疗推荐, 医疗文件翻译和其他功能。 |
| 2017.01 - 2017.04 | 算法工程师实习, 雷掣科技, 哈尔滨<br>参与设计和实现基于 CAMShift 算法和 OpenCV 框架的爬楼车的目标跟踪系统。               |

## 项目

- 2019.10 - 2020.06 | 基于深度学习的蛋白质多序列比对  
为了提高在“低相似度”蛋白质家族上的多序列比对的质量，从以前的十几种多序列比对的渐进方法研究中选择出对结果影响最大的部分，使用 CNN 与双向 LSTM 建立分类决策模型以选择该部分的最佳计算方法，最后发布了一种全新的渐进式多序列比对工具，该工具在 711 个“低相似度”蛋白质家族的测试中，能够得到相对于其他工具，平均 2.8% 的准确度提升。
- 2018.09 - 2019.10 | 基于机器学习和重新调整策略的蛋白质多序列比对  
为了提高在蛋白质家族上的多序列比对的质量，对以前的几种不同策略的多序列比对算法分析，使用随机森林分类器，从中选择针对不同的蛋白质家族最合适的比对策略，得到临时多序列比对结果；为了进一步提高该临时多序列比对结果的质量，从中抽取不同相似度的区间，针对不同的临时多序列比对结果，使用随机森林分类器，选择“可信区间重对齐”或者“不可信区间重对齐”策略；最后得到一种全新的多序列比对流程，该流程在 1356 个蛋白质家族的测试中，能够得到 1.65% 的相对于其他多序列比对算法的准确度提升。
- 2017.11 - 2018.06 | 面向领域快速移植的高精度汉语分词系统  
为了提高中文分词的精确度，使用 1998 年版《人民日报》人工分词结果作为训练数据，用 Java 语言开发了一套基于条件随机场，维特比算法等联合模型的中文分词系统；为了进一步提高该分词系统在特定领域（医学、法律和金融）的适应性，加入启发式规则和特定领域的词典，并使用 N-Gram 对分词结果准确性验证并纠正；最终实现了一套对特定领域文本的平均精度达到 97% 的分词系统。

## 助教

- 2020.01 - 2020.06 | COMP1117 计算机编程  
2019.02 - 2019.07 | COMP7606 深度学习

## 奖学金和证书

### 研究生期间

- 2020.11 | 华为认证 ICT 工程师 - 人工智能  
2018.11 | Certificate of Teaching and Learning in Higher Education  
2018.09 | 全额研究型硕士奖学金

### 本科生期间

- 2018.06 | 企业奖学金  
2016.11 | 国家励志奖学金  
2015.12 | 三好学生  
2015.11 | 国家励志奖学金

## 语言

- 中文 | 母语  
英语 | 流利

## 技能

- 基础 | TensorFlow、编程 (C/C++、Java、Python 等)、办公软件  
中级 | RESTful, Keras, Scikit-Learn  
掌握 | 机器学习与深度学习建模、多序列比对