

个人简介

熟练使用 Python 完成数据分析和开发工作。掌握数据挖掘完成过程，包括爬虫获取数据(selenium, beautifulsoup)，清洗整理建立数据库(mongodb, mysql)，数据探索(pandas, matplotlib, ggplot2)，特征选择和建模(sklearn)，以及数据可视化(D3.js)。了解常用机器学习算法的基本原理和参数选择原则(Naive Bayes, SVM, Decision Tree)。

主要技能

Languages: JavaScript, Python, HTML5, CSS3, R, shell

Frameworks/Libraries: jQuery, D3.js, Django, Pyramid, Mako, sklearn, pandas, matplotlib, seaborn, ggplot2, networkx, beautifulsoup

Database: mysql, mongoDB, neo4j.

Other: Git/GitHub, Bitbucket, Jira, Markdown, Letex.

项目经历

Molecular Breeding Knowledge - mbkbase.org

2014 年 10 月

- 建立水稻基因型，表型和种质信息网站，实现数据信息的查询和展示。整体以 pyramid 作为网站框架。
- 实现综合符合查询功能作为数据挖掘工具以及页面的数据可视化和数据的交互式展示。(D3.js, Ajax)
- 优化 MongoDB 数据库查询效率，使某些页面前台响应时间从 1 分钟提高到 3-4 秒。

West Nile Virus Prediction - github.com/kaiwang0112006/mykaggle_westnile

2015 年 5 月

- 用 python 完成 kaggle 竞赛—根据气象和地理信息预测西尼罗病毒在芝加哥未来爆发的可能。
- 用 pandas 完成数据清理，sklearn 完成特征提取并应用 GBDT 迭代决策树建模。

乙酰胆碱酯酶抑制剂的活性预测 - ncbi.nlm.nih.gov/pubmed/22460031

2009 年 5 月

- 利用 R 实现多元线性回归建立乙酰胆碱酯酶抑制剂的活性预测模型。
- 利用 libsvm 实现支持向量机建立乙酰胆碱酯酶抑制剂的活性预测模型。

乙酰胆碱酯酶抑制剂与非抑制剂的分类研究 - ncbi.nlm.nih.gov/pubmed/22263859

2009 年 5 月

- 收集 714 个乙酰胆碱酯酶和 3892 个非抑制剂结构分子建立数据库。
- 用 python 设计和实现了多种特征选择方法，包括基于相关性，基于 F-score 和基于 weka 内置方法的筛选。
- 利用 libsvm 实现支持向量机方法建模，最终模型分类正确率达到 96.5%。

工作经历

中国科学院遗传与发育生物学研究所 - 助理工程师

北京 | 2012 年 7 月—至今

- 基因组与转录组序列组装和分析；
- 数据库和网站开发。

保诺科技（北京）有限公司 - 实习

北京 | 2011 年 7 月—9 月

- 设计完成分子性质计算，结构搜索和反应枚举交互界面；协作支持化合物数据库的更新（MDL® ISIS/Base）

教育经历

Udacity - Data Analyst Nanodegree

2016 年 5 月-2016 年 7 月

MongoDB, Inc - MongoDB for python Developers (Training)

2014 年 11 月-2015 年 2 月

北京化工大学 - 化学工程与技术专业硕士

2009 年 9 月-2012 年 7 月

北京化工大学 - 生物工程专业学士

2005 年 9 月-2009 年 7 月

发表文章

- Dejun Li, Rizhong Zeng, Yan Li, Manman Zhao, Jinquan Chao, Yu Li, **Kai Wang**, Lihuang Zhu, Wei-Min Tian, Chengzhi Liang, Gene expression analysis and SNP/InDel discovery to investigate yield heterosis of two rubber tree F1 hybrids, **Scientific Reports**, 2016. 6:24984.
- **Kai Wang**, Xiaoying Hu, Zhi Wang, Aixia Yan*, Classification of Acetylcholinesterase Inhibitors and Decoys by a Support Vector Machine, **Combinatorial Chemistry & High Throughput Screening**, 2012. 15(6), 492-502.
- Aixia Yan*, **Kai Wang**, Quantitative structure and bioactivity relationship study on human acetylcholinesterase inhibitors, **Bioorganic & Medicinal Chemistry Letters**, 2012, 22(9), 3336-3342.
- Maolin Wang, **Kai Wang**, Aixia Yan*, Changyuan Yu*. Classification of HCV NS5B polymerase inhibitors using Support Vector Machine. **Int.J.Mol.Sci**, 2012, 13(4), 4033-4047.
- Aixia Yan*, Xianglei Nie, **Kai Wang**, Maolin Wang, Classification of Aurora Kinase inhibitors by Self-Organizing Map (SOM) and Support Vector Machine (SVM), **Eur. J. Med. Chem**, 2012(61), 73-83.
- Aixia Yan*, Yang Chong, Liyu Wang, Xiaoying Hu, **Kai Wang**, Prediction of Biological Activity of Aurora-A Kinase Inhibitors by Multilinear Regression Analysis and Support Vector Machine, **Bioorganic & Medicinal Chemistry Letters**, 2011,21(8), 2238-43.
- **Kai Wang**, Aixia Yan*, Introduction of several public databases for chemical activity on the Internet, **Computers and Applied Chemistry**, 2010,21(12):1629-1632.