王瑞敏

(+86) 157 0157 4491　　|　　ruiminw128@163.com

北京市海淀区科学院南路 6 号 计算技术研究所 （100190）

博客：http://blog.csdn.net/wrm_nancy　　GitHub：https://github.com/wrm128

教育背景

2014.09 - 2017.06	中国科学院 计算技术研究所	计算机应用技术（**免试推荐**）	硕士	
2010.09 - 2014.06	郑州大学	计算机科学与技术	学士	**Top 1%**

项目/科研经历

2016.01 – 至今　　　整体蛋白质的质谱数据预处理算法研究　　　　　　　　　　　核心设计/开发人员

■ **项目简介**：质谱数据由一系列谱峰信号构成，有用的谱峰信号满足一定的规则成簇出现。蛋白质质谱数据预处理的主要任务是从质谱数据中提取所有有用信号，并尽可能使引入的噪音信号最少。

■ **主要职责**：

- **负责谱峰信号去噪以及特征提取**：使用直方图统计法进行去噪；使用 **EMASS 算法**计算蛋白质的理论同位素分布；从实验谱图中提取同位素峰簇的特征（11 维），用于谱峰分类。
- **负责谱峰分类算法设计**：根据提取到的谱峰的特征信息在线训练 **SVM** 模型并对谱峰信号进行分类处理，报告有用的谱峰信号。也尝试了多元自适应回归样条 **MARS** 和 **logistic** 回归进行离线训练。

■ **研究成果**：在标注数据集上可以达到 95%以上的召回率，使鉴定率提升了 24%，速度比同类预处理软件快 3 倍。

2015.08 - 至今　　　整体蛋白质鉴定与定量研究 （973 项目）　　　　　　　　　核心设计/开发人员

■ **项目简介**：类似文本检索系统，蛋白质搜索引擎以质谱数据为查询进行蛋白质信息的检索。

■ **主要职责**：

- **数据库索引的构建**：熟悉各种索引创建算法，主要包括**倒排索引**（哥德尔编码）、离子索引，**序列标签索引**等。
- **修饰定位算法的实现**：使用优先队列和动态规划算法进行修饰定位。
- **打分评价**：采用 **BM25 算法**进行细打分，使用 **SVM 半监督学习**进行重打分，提供基于 FDR 和基于 E-value 两种过滤策略。
- **引擎的界面开发**：使用 WPF 基于 MVVM 框架和工厂模式进行客户端开发，UI 层使用 XAML 语言，后台逻辑使用 C#。

■ **研究成果**：与国际上知名的完整蛋白质鉴定软件 MS-Align+相比，我们的软件 pTop 与其鉴定精度一致，但速度快 20-100 倍；且目前已支持蛋白质的标记定量。pTop1.2 已在网上(http://pfind.ict.ac.cn/software/pTop/index.html)发布，提供注册下载。

2014.10 - 2014.12　　　基于指纹特征的音频检索系统　　　　　　　　　　　　　核心开发人员

■ **项目简介**：基于音乐指纹特征，实现所谓的"听歌识曲"系统（B/S 架构）。用户输入录音文件，系统从乐库中检索匹配，并返回结果。主要包括预处理、指纹提取、建倒排索引、精确匹配等模块。返回 Top 3 的结果时，准确率可达 **90%**。

■ **主要职责**：负责使用 landmark 算法提取音乐指纹，建立倒排索引表，根据音频特征之间的时序关系进行精确匹配。

2014.02 - 2014.06　　　海量浮点数外部排序算法实现　　　　　　　　　　　　　核心设计/开发人员

■ **项目简介**：实现单机环境下海量浮点数的外部排序算法。

■ **关键部分**：

- 以二进制方式分块读写文件；自定义实现字符串与浮点数的相互转换以取代库函数；根据浮点数在计算机中的存储方式实现**基数排序**；使用**败者树**进行 K 路归并；尝试使用多线程进行加速。
- 单线程版本 2.5 亿浮点数用时 140.3s 完成排序。

专业技能

◆ 熟悉常用的数据结构和算法，熟悉搜索引擎的相关算法。

◆ 熟悉 C++，了解 Java、C#、Python、Matlab、SQL。

◆ 熟悉常用的设计模式，了解数据挖掘、机器学习和信息检索等相关知识。

◆ 英语听说读写能力良好，通过 CET-6， 伯乐在线专栏校稿。

获奖情况

◆ 2014.06　河南省优秀毕业生

◆ 2013.05　河南省第六届 ACM 程序设计大赛 铜奖； 全国大学生英语竞赛 二等奖

◆ 2012.12　第七届全国信息技术水平大赛 省二等奖； 2013 年编程之美挑战赛进入复赛；两次参加百度之星复赛

◆ 一等奖学金（2010-2011,2013-2014），国家励志奖学金（2011-2012,2012-2013）；多次获得郑州大学三好学生、优秀团员等称号

RuiMin Wang

Objective: Software Engineer   **Mobil:** (+86) 157 0157 4491   **Email:** ruiminw128@163.com   **Gender:** Female

Education

2014.09 - 2017.06	Master	Computer Application Technology	Institute of Computing Technology, Chinese Academy of Sciences
2010.09 - 2014.06	Bachelor	Computer Science and Technology	Zhengzhou University **Top 1%**

Project Experience

2016.01 - Now   Preprocessing of Complex Tandem Mass Spectra of Intact Proteins   Designer/Developer

- **Project Description:** Mass spectrometry data contains a series of peak signals and the useful signals occur in clusters based on certain rules. The main task of the preprocessing is to extract useful signals from the MS and MS/MS data and bring in less noise signals.
- **Responsibilities:**
    - **Signal de-noising and feature extraction.** Use histogram-based method for de-noising, use **EMASS** algorithm to compute the theoretical isotopic distribution, analyze the signal clusters and extract the features (11 dimensions) of each signal cluster.
    - **Signal classification algorithm**. Use an online **SVM** to classify the signals and report the useful ones. Multivariate adaptive regression splines (**MARS**) and **logistic regression** were also tried to do signal classification.
- **Results:** On the labeled data set, the recall rate can reach more than **95%**, and the identification rate increased by **24%**. The speed is three times faster than the existing spectra preprocessing software.

2015.08 - Now   Identification and Quantitation of Multi-modified Proteins (NSFC 973 Project)   Designer/Developer

- **Project Description:** Similar to text retrieval, the protein search engine retrievals the input spectra data in the protein database to get the protein sequence information as well as user- defined modifications.
- **Responsibilities:**
    - **Construction of database indexes**. Build tag-based **inverted indexes**, using binary encoding for sequence tags.
    - **Fast algorithm of modifications location**. Use priority queue and dynamic programming to locate modifications.
    - **Scoring and evaluation.** Use **BM25** for scoring and **semi-supervised SVM** for re-rank, use FDR and E-value for filter.
    - **UI of the search engine.** Develop the UI of the engine with **WPF** based on MVVM framework and factory pattern.
- **Results:** Compared with the internationally renowned software MS-Align+, our software pTop could achieve equivalent identification accuracy, while 20 - 100 times faster. pTop 1.2 has been published on the Internet(http://www.topdownproteomics.org/software) for free.

2014.10 - 2014.12   An Audio Retrieval System based on Fingerprint Characteristics   Developer

- **Project Description:** The whole system is a B/S framework. Users upload audio recordings, and then the system searches the database and returns the best matches. There are several core modules including preprocessing, fingerprint extraction, inverted index, exact matching, etc. When the top 3 matches are returned, the accuracy reaches around **90%**.
- **Responsibilities:** Extract fingerprint with landmark algorithm and implement accurate matching based on the temporal relation of the audio information.

2014.02 - 2014.06   External Sort of Massive Floating-point Data   Designer/Developer

- **Project Description:** Sort massive floating-point data under stand-alone environment.
- **Responsibilities:**
    - Read and write data in chunks, implement the conversion functions between string and double type without using STL.
    - Implement **radix sort** based on the storage of floating-point numbers in the computer, use **loser tree** to merge multiple sorted files, try to accelerate the program through **multithreading**.
- **Results:** Finish sorting 250 million floating-point numbers in 140.3s with single thread.

Personal Skills

- Familiar with **C/C++**, know **Python**, Java, C#, Matlab and SQL.
- Familiar with data structures and algorithms, familiar with common **design pattern**.
- Having experience in **search engine** development and machine learning.

Certificates & Awards

- Certified in CET-4, CET-6   ◆ Excellent Graduate of Henan Province (2014.06)
- The 6th Henan Province Programming Contest, **Bronze Medal** (2013.05)
- National English Competition for College Students, **the Second Prize** (2013.05)
- First Prize Scholarship of the university (**top 5%**, 2011, 2014), National Scholarship for Encouragement (**top 3%**, 2012, 2013)