

祖松鹏

北京市清华大学 FIT 1-108 ◇ 联系方式: (+86)13810349372

zsp07@mails.tsinghua.edu.cn ◇ <https://github.com/songpeng>

教育

清华大学

2011.09 - 至今

自动化系生物信息学实验室直博生

GPA: 90.84/100

主修课程

- 应用随机过程,统计推断
- 泛函分析, 近世代数
- 贝叶斯统计与蒙特卡洛方法
- 模式识别, 概率图模型基础
- 高维概率及统计

清华大学

2007.09 - 2011.07

生命科学学院

GPA: 86.32/100, 专业排名前1/3

基本技能

掌握主要的机器学习模型

- 熟悉核函数方法, 及支持向量积和高斯过程
- 熟悉决策树, 随机森林, Boosting等方法
- 熟悉回归分析, 如Ridge 回归, LASSO 及广义线性模型等
- 熟悉非监督学习, 如K-means等聚类方法, 主成份分析等
- 熟悉基本的概率图模型, 如贝叶斯网络, 马氏随机场等
- 理解神经网络的基本内容, 如限制玻尔兹曼机模型, 深度学习模型等

掌握主要的统计方法

- 熟悉基本的蒙特卡洛采样策略, 如拒绝采样, Gibbs 采样, 哈密顿采样等
- 熟悉基本的贝叶斯统计, 如贝叶斯回归, 层次贝叶斯模型等
- 熟悉基本的统计策略, 如极大似然估计, 假设检验, EM算法等

掌握基础的凸优化内容

- 理解增广拉格朗日方法
- 熟悉主要的优化算法, 如牛顿法, 拟牛顿法, 遗传算法, 模拟退火等
- 理解ADMM 算法

研究经历

清华大学自动化系博士研究生

2011.07 - 至今

从机器学习角度预测药物与蛋白质的相互作用。

- 利用EM算法估计隐变量状态, 预测药物与蛋白质主要结合位置。该方法的主要贡献在于可以通过表面的数据, 学习出数据背后的模型结构。该方法利用C++实现, 并已发表文章。
- 构建非参数的统计策略, 预测药物与蛋白质的相互作用。我们利用图切割办法, 结合贝叶斯因子, 分析给定数据后, 不同模型的条件概率。
- 构建多任务机器学习模型, 定量预测药物与蛋白质的结合强度。利用层次贝叶斯模型, 将若干线性回归模型组合在一起共同训练, 从而达到有效利用不同任务间数据存在相似性的特点。

哈佛大学统计系访问学者

2014.03 - 2014.09

- 利用一种非参数的贝叶斯统计模型, 检测生物系统中部分变量两两间的依赖性尤其是非线性性。

北京百度公司实习生

2014.01 - 2014.02

- 利用大搜索数据关联地区性的流行性疾病, 如流行性感冒, 艾滋病等。

柳州疾病预防控制中心博士生暑期实习

2013.07 - 2013.08

- 利用病人DNA测序数据分析柳州艾滋病的亚型。
- 协助当地人员构建艾滋病数据分析流程。

论文发表

Zu S., Chen T, Li S. *Global optimization-based inference of chemogenomic features from drug-target interactions*. Bioinformatics, 2015. (已在线出版, SCI: 4.9)

社会工作经历

- 本科生系统生物学导论的助教 2014.09 - 2015.01
- 研究生概率图模型基础的助教 2013.09 - 2014.01
- 本科生带班政治辅导员 2011.08 - 2013.01
- 完成北京国际马拉松全程 2009, 2010
- 2008年北京奥运会志愿者 2008.08

奖项

- 清华大学海外交流奖学金 2014
- 清华大学优秀辅导员 2013
- 清华大学唐仲英奖学金 2008, 2009, 2010

技术与兴趣

常用编程语言	R, Python, Perl, C/C++
常用工具	Emacs, Vim, Latex, Shell
兴趣	国际标准舞, 足球