

# 陈星强

厦门大学理学硕士

居住地：北京·朝阳区望京南嘉润花园  
电 话：15001334125 (手机&微信)  
E-mail：[joy6677@qq.com](mailto:joy6677@qq.com)  
籍 贯：新疆五家渠市工业园



个人目前维护的GitHub 人脸聚类项目：

[https://github.com/chensexingqiang/facenet\\_face\\_clustering](https://github.com/chensexingqiang/facenet_face_clustering)

## 教育背景

2014.09 ~ 2017.06	厦门大学	生物物理与凝聚态物理	硕士研究生
2010.09 ~ 2014.06	厦门大学	理论物理学专业	本科

相关课程学习：

2014-09 ~ 2015-06	人工智能	厦门大学 智能系专业课程	史晓东教授
2015-09 ~ 2016-06	数据挖掘	厦门大学 计算机系专业课程	林琛教授
2015-09 ~ 2016-06	模式识别	厦门大学 信息科学学院专业课程	杨晨晖教授
2015-09 ~ 2016-06	机器学习	厦门大学 信息科学学院专业课程	王程教授

## 专业技能

1. 熟练使用 C, C++, Python, MATLAB 编程;
2. 有使用C/C++有蛋白组学大数据(10T 数据)处理的项目经验。
3. 数学建模，数学语言转换为计算机算法的能力突出，本科二年级获得“高社杯”数学建模比赛国家一等奖，本科三年级获得MCM 北美大学生数学建模一等奖。
4. 熟悉各类图像检测算法（SSD, YOLO2, faster-rcnn 等）的原理，图像分类算法（AlexNet, VGGNet, GoogLeNet, ResNet等）的原理；视频预测，基于RNN+CNN的结构进行短时雷达图降雨预报建模。
5. 熟悉掌握机器学习的核心算法，SVM，决策树，随机森林，熟练使用XGBOOST和LightGBM等框架。
6. 良好的数理背景，在专业上从事动力学结合量子力学计算模型的构建，对量子力学熟悉，具有物理、化学、生物学及计算医学的综合研究背景。

## 语言能力

英语水平：CET-4/CET-6；

良好的外语阅读能力，能熟练的阅读英语专业文章、说明书、专利等；

能使用英语撰写专业学术论文、专利申请/科研项目申请文书等。

## 项目经历

### 2018.03~至今 北京心知科技有限公司 机器学习工程师

工作内容：围绕心知科技基于精准全面的气象数据，落地新零售需求预测、基于地理位置和气象条件的广告精准投放、以及短时降雨预报的算法开发和模型应用交付工作。

主要职责：算法研发，工程实现，算法产品交付，算法团队搭建。

主要难点：机器学习项目全周期管理与实现。

项目内容：

- 1.基于前期工作经验，应用XGBOOST模型为超市发提供销量预测的解决方案。
2. 广告点击率预测和广告投放策略，使用 DeepFM 模型和 LR+GBDT 模型优化，也考虑用强化学习实现DSP平台的RTB竞价模式下投放策略优化的解决方案。
3. 雷达图短临降雨预报，使用RNN（GRU）结合CNN的网络结构，在MXNET框架下实现无监督的雷达短临降雨预报系统。

平台及技术：MXNET, XGBOOST, Tensorflow .

### 2017.06~2018.03 便利蜂虫极科技(北京)有限公司 算法工程师

项目一：千店千面，基于机器学习的门店销量预测和订货计划

项目描述：50 家线下门店销量预测模型，实现千店千面的精准需求预测。使用GBDT模型进行商品中分类和热餐销量预测。后期改进模型，基于LSTM算法和DeepFM算法框架进行销量的时间序列预测，改进模型和提高准确度。

主要职责：算法研发，工程实现。

主要难点：构建合适的特征工程，提高代码运行速度，算法嵌入spark hive数据平台每日自动执行。

解决方案：特征工程的构建通过不断组合和逐一增删的策略验证其有效性和对应的特殊情形分析，如节假日等。鉴于模型需要为众多门店的诸多SKU提供预测，前期使用XGBOOST在服务器上运行，后期改写为基于Spark分布式的执行流程，提高模型训练速度。任务需求配合 Jenkins job调度系统自动执行完成。

平台及技术：python pandas 模块的高效利用，用python写爬虫获取网络数据，Python Keras, Hadoop MySQL 数据库交互。

## 项目二：门店人顾客脸识别与绑定

项目描述：进店结账人脸检测与绑定；门店购物结账时商品自动检测。

主要职责：算法研发，工程实现。

主要难点：人脸识别算法的准确度已经够高，就是判断人脸绑定比较困难。商品检测中的热餐检测是难点。

解决方案：人脸识别通过对Facenet 算法框架中提取网络抽提的人脸embedding 进行聚类 and 多次逻辑判断，从而绑定顾客的APP ID 和 顾客面貌。门店购物高峰期排队严重，需要快速收银解决方案。通过初期实地调研和采集实时收银的数据，从中获取购物场景进行数据标注，SSD等模型训练和检测，从而验证实时检测收银系统的可靠性。热餐在盒子中，不好判断盒子中的菜品，为此考虑进行图像的语义分割，后期考虑到计算成本，并未实施。

平台及技术：Tensorflow caffe facenet, SSD, YOLO, faster-RCNN 等模型。

## 2016.07~2016.09 北京乐普医疗有限公司 研发工程师 实习

项目名称：液体活检试剂盒开发 ctDNA 数据分析 负责技术研究和甲基化基因芯片数据分析

项目描述：液体活检需要通过特定的生物标记物去筛查潜在患者。发现并快速判断找到准确的生物标记物是格外重要的。通过借助已有的基因芯片数据，应用机器学习的算法，对其进行数据建模，能够分析得出比较有效的潜在标记物，为实验给出更准确的指导建议。

主要职责：负责数据收集，文献重现，国外相关专利解读，美国FDA证书内容分析。建立癌症患者甲基化DNA数据分析库，收集4万多病人的数据（2T数据），对数据进行预处理。

主要难点：数据量巨大，需要并行多核处理，应用SVM等机器学习算法对其进行分析。

平台及技术：Python 数据分析与建库, Linux, MPI

## 2016.10~2017.07 厦之医生物科技有限公司 公司创始人

搭建智能医疗影像数据分析平台并申请专利

项目描述：通过整合各类组学数据，包含基因组学，蛋白组学，表观组学，医疗影像组学的的数据，结合机器学习和深度学习的算法，搭建基于云平台的DeepOmics系统，为医院，医生提供智能化的服务和高效准确的医疗专家资源支持。

主要职责：团队运营负责和算法开发。

主要难点：数据收集 算法集成 组织运营。

解决方案：通过各种渠道主动和医院建立联系，并以合作的方式获取一定数据，从而开展算法开发。学习运营管理公司，带团队。

平台及技术：Python/Matlab C++; tensorflow/caffe; Linux; GTX 1080 服务器，FPGA 快速算法研发。

## **2016.01~2017.04 厦门大学 韩家淮院士试验室 项目负责人**

项目名称：蛋白质网络动力学分析 质谱数据分析课题

项目描述：通过质谱分析获取的蛋白数据对细胞反应体系进行建模和分析

主要职责：算法研究

主要难点：复杂的蛋白相互作用网络，需要通过数据挖掘和信息理论去解决网络中隐藏的信息。

解决方案：通过图结构模型，进行蛋白互相作用的分析，利用互信息等手段提高相互作用的准确关联。

平台及技术：用MATLAB 实现（自洽场）SCF算法 SDG（随机梯度下降） Huffman树

## **2013.03~2016.03 厦门大学 物理系 生物物理计算医学实验室**

项目名称：蛋白质动力学模拟

项目描述：组氨酸激酶的自磷酸化是微生物的重要反应之一。通过研究该反应的机制，能够对微生物引起 的疾病进行很好的药物设计，从而帮助患者抵御该类微生物。

主要职责：蛋白质动力学模拟算法研究

主要难点：生物体系的大规模模拟，需要大量生物和物理，化学，计算机知识和能力的积累。反应体系涉及到化学反应和金属离子，需要通过量子化学计算去完成。

解决方案：主要的知识都是通过阅读大量文献，结合蛋白质动力学计算模拟软件 amber 开发说明书实现，其中为了学习量子化学计算，专程自费到清华大学参与15th国际量子化学计算大会，并主动认识领域专家高加力院士，后期到高老师所在吉林大学理化所交流学习三个月，从而掌握到动力学结合量子化学模拟的能力。

平台及技术：Linux 系统，Amber， Gauss09， NAMD 。

## 科研成果

与人合作发表 SCI论文 4 篇，其中3篇SCI

- 1.Conformational Dynamics of Response Regulator RegX3 from Mycobacterium tuberculosis. IF 3.48
2. Suppressing effect of  $\text{Ca}^{2+}$  blips on puff amplitudes by inhibiting channels to prevent recovery. IF 3.5
3. Conformational dynamics of the essential sensor histidine kinase WalK. IF 2.5
4. Chiral discrimination of amino acid enantiomers based on different interactions with  $\text{Cu}^{2+}$ . IF 5.27

发明专利1项：《一种基于深度神经网络的肺癌识别系统》专利号：2017107522379。

## 所获奖项

---

2010/12 获得厦门大学物理机电学院“优秀团员”称号；  
2012/09 获得“高社杯”全国大学生数学建模国家“一等奖”；  
2013/04 获得北美大学生数学建模一等奖；  
2013/07 参加北京大学生物理定量研究中心的直博资格认证，获得“优秀营员”称号；  
2013/09 参与国际太阳能全能十项竞赛，耗时数月，负责完成编辑400页的英文版工程规划书(Manual)；比赛最终获得国际第六名，其中节水模块等获得数项第一。  
2014/06 邀请参加由中国科学院物理所举办的生物物理暑期学校，并获得“优秀学员”称号；  
2015/06 与生命科学院韩爱东老师课题组合作发表文章Conformational Dynamics of Response Regulator RegX3 from Mycobacterium tuberculosis  
2016/04 获厦门大学校庆奖学金。  
2016/06 福建省自然辩证法研究生演讲比赛“三等奖”。

## 自我评价

学习能力强，实践动手能力强。  
良好的算法、数据结构及编程基础。

熟悉深度学习、人工智能、数学建模、统计、物理、量子化学计算等专业知识。

熟练的阅读中英文文献及撰写中英文论文、科技项目申请书等，能够独立复现论文中的各种算法，工程能力过硬。

团队协作能力和组织能力。