



王海清

各种研究助理经历，帮教授、博士实现算法。

软件系统工程 · 亚琛工业大学

男 | 26岁 | 硕士 | 应届毕业生 | 深圳

18682067832 | whqwill@126.com

■ 实习经历

弗劳恩霍夫协会-智能分析和信息系统研究所

研究助理（硕士论文）

2015.08-2016.09

用 spark 实现基于词义分配的一词多义词向量算法，并进行参数调优。算法是基于 skipgram with negative sampling，根据词义向量评分动态调整句子中每个词的词义，再进行分布式梯度下降训练词义向量。

算法特色：每次调整词义后根据词义的更新频率调整低频词义，不同 executor 上的词向量合并后做 normalization。

项目特色：JVM 分布式框架上的一词多义向量算法的尝试。

难点：参数调优，空间优化以及分布式的参数更新，参数包括 Spark 自带的一些系统参数(executor 的数量，driver memory 的大小)以及算法参数(词典大小，多义词的词义数量)。

遇到的麻烦：机器间的数据传输太慢，最后只是用单机32核做一个任务的训练。

附加工作：词义向量可视化(Python)，t-SNE 非线性降维以及相关系数计算(Spark)。

项目地址：<https://github.com/whqwill/SenseAssignment>

亚琛工业大学计算机六所模式识别组

学生助理

2014.10-2015.08

项目1：对一个鸟类数据库做图像 transformation 归一化。用 Python 实现自动选取鸟类特征点(人工标注)选取算法，然后根据选取特征部位的空间相对位置归一化所有图片(旋转，翻折，缩放)，然后训练用 Caffe 载入训练好的 AlexNet 进行参数微调训练鸟类数据库。项目地址：

<https://github.com/whqwill/CaffeTask1>

项目2：借助 Caffe 实现图像分层分类模型，训练模型为 GoogleNet，训练数据库为 cifar100。先做一次粗粒度物体分类(100类)，根据混淆矩阵做聚类(10个大类)，聚类后重新修改训练集，并重新训练网络(100类变10类)。在每一个大类里面再做一次细粒度的物体分类。难点：参数调优，大部分时间是在尝试不同参数和等待训练。遇到的麻烦：计算集群中每台机器有至少两个 GPU，Caffe 有个 bug 会经常占用两个 GPU 但实际只用一个。解决办法：在每个提交的任务前加一个脚本检测空闲的一个 GPU 并置顶它只用这个。附加工

作：调用 Caffe 图像数据库生成工具，聚类后，修改标签，生成新数据集。 项目地址：

<https://github.com/whqwill/CaffeTask2>

亚琛工业大学计算机8所-计算机视觉组

学生助理

2014.03-2014.09

物体检测算法以及3D 场景构建算法的代码测试工作。网上找寻相关算法的 C++ 源代码，修改，编译，测试。项目背景是德国 KIT 发起的研究汽车自动驾驶项目，数据集是来自于测试汽车的双摄像头拍摄的交通行驶视频。先完成物体检测(汽车，行人，楼房)，再根据两个摄像头检测的相同物体，绘制整个3D 深度图。

浙江大学普适计算实验室

研究助理（本科论文）

2012.07-2012.09

用 NMF (非负矩阵分解) 方法分析交通数据。数据是一个月的杭州市所有 路口摄像头记录的车辆牌照和时间。预处理后得到每一辆车的平均每天每个时间段的大致行驶距离，并用向量表示。非负矩阵分解后得到 特征向量, 然后用特征向量进行分析。（MATLAB 实现）

项目意义：非负矩阵分解这个方法的最大的优势就在于它分解后的两个矩阵的值都是非负的，也就是说它的值是“可读的”，有实际价值的。所以分解出来的行驶特征也是“可读的”。

附加工作：用 C++ 进行数据预处理，读文本文件，做统计和累加。

■ 教育经历

亚琛工业大学

硕士 · 软件系统工程

2017年毕业(预计)

中山大学

本科 · 信息与计算科学

2013年毕业

■ 项目经验

生物信息作业提交客户端

客户端实现

2012.03-2012.07

用 JAVAWEB 为中山大学广东省计算科学重点实验室做了一个生物信息的网页客户端的原型。用户可以通过网页登录，并保存自己的作业提交信息(作业的参数信息和结果信息)，数据库由 mybatis 实现。选择提交后系统根据作业优先级排序，并访问计算集群的空闲资源，然后先后提交用户作业(JAVA 封装的 SSH 实现)，作业计算完毕后系统会保存下来每个作业的结果信息或是错误信息以便用户可以随时登录查看。

■ 作品展示

<http://t.cn/Rfa04Vy>

简历来自：拉勾网 - 最专业的互联网招聘网站 - www.lagou.com

硕士论文，本科论文，成绩单，获奖证明，项目内容。

■ 自我描述

本科参加过 ACM 比赛拿过省一等奖，数学建模比赛拿过全国二等奖，年级排名前15%。大四暑期去浙江大学的普适计算实验室实习做研究助理，帮教授实现非负矩阵分解算法做交通车辆出行数据分析，并根据其实习过程完成了毕业论文。硕士阶段主要课程是机器学习相关和软件工程相关，有不少算法实现和参数调优(机器学习，深度学习)相关的研究助理工作。对基本机器学习，深度学习，计算机视觉算法都有一定了解和实战。研究生尾声在 Fraunhofer IAIS 做研究助理实习，主要课题用工业大数据框架实现一词多义的词向量表示，并完成硕士论文。

本科主要课程：主要课程:数学分析, 实变函数, 数值分析, 偏微分方程, 几何与代数, 操作系统, 算法与数据结构。 硕士主要课程：数据挖掘, 机器学习, 计算机视觉, 语音识别, 模式识别与神经网络, UML 原理, 数据库实现, 面向对象软件构建。

■ 期望工作

算法工程师， 全职， 深圳， 25k-50k

■ 技能评价

看论文

精通

推公式

精通

■ 编程语言及工具

JAVA

：接口，多态，数据结构，基本工具库，JAWWEB。 Scala：函数编程，词向量算法，用 SBT 添加算法库
Python：图像库，聚类算法，脚本，几何变换算法。 C：位运算，线段树，指针，链表，Hash 函数，搜索算
法。 C++：Caffe 代码修改，图像数据库修改，物体检测算法编译，链接 OpenCV 库 MATLAB：非负矩阵分
解，图像处理 Spark(Scala)：相关系数计算，t-SNE 降维 Caffe：卷积网络定义 Linux：文件系统操作