

# 叶俊杰

1991年, 男, 现居上海

电话: 15867835270

邮箱: junjieyemain@outlook.com

## 教育经历

### 英国爱丁堡大学

理学硕士 2014年9月 - 2015年11月  
信息学院人工智能专业  
专业绩点: 65 (一等荣誉学位为70)

### 浙江财经大学

本科双学位 2010年9月 - 2014年6月  
信息学院计算机科学与技术专业  
专业绩点: 88/100  
财政学院税务专业  
专业绩点: 89/100

## 个人主页

LinkedIn: Junjie Ye

## 技术技能

### 认证证书

- 中级软件工程师职称
- 斯坦福大学机器学习课程证书

### 技术领域

- 机器学习 / 深度学习
- 贝叶斯推断 / 概率图模型
- 强化学习 / 机器人决策
- 自然语言处理
- 语音识别
- 知识图谱
- 大数据处理
- 网络爬虫

### 技术项目

- 自动化测试项目 (工作)
- HMM-GMM语音识别系统 (硕士)
- Kaggle识别驾车司机 (硕士)
- 强化学习机器人寻路算法 (硕士)
- 机器人博弈论算法 (硕士)
- 信息论压缩解压算法实现 (硕士)
- 图书馆图书管理系统 (本科)
- 地方税务发票系统 (本科)

### 技术框架

- Tensorflow / Theano / Keras
- Hadoop / Spark / Hbase
- Docker
- Scrapy
- Flask
- MongoDB
- Neo4j / ArangoDB
- Redis
- Weka

### 编程语言

- Python
- Java
- Shell
- Matlab
- C / C++
- SQL
- JavaScript
- HTML / CSS
- Markdown
- LaTeX

## 工作经历

### 竹间智能科技（上海）有限公司

研发工程师

中国上海

2016年4月 - 至今

- Bot Memory Team:  
使用多模态 (规则引擎 + 机器学习/深度学习) 架构设计实现多个机器人记忆模块, 包括个人基本信息, 情感状态, 生活习惯等方面。机器人记忆模块在自然语言交互界面中自动识别解析并存储相关信息, 使得机器人能够更加深入的理解用户, 理解对话语境, 从而使机器人回复更加贴近用户, 更加智能。  
线上系统使用Java搭建, 多种模型通过RESTful API集成, 使系统具有良好的可扩展性。深度学习模块使用TensorFlow系统。模块的输入通过word2vec处理为特征向量, 输入深度学习模型。深度学习模型使用了包括RNN, LSTM, GPU等在内的多种神经网络模型。多个模块的离线准确率超过90%, 在线准确率超过80%。其中情感状态记忆模块正在申请专利。
- Knowledge Graph Team:  
搭建大规模知识图谱, 包括通用知识图谱 (百度百科, 维基百科等百科知识) 和领域知识图谱 (包括音乐, 影视, 体育, 电商等)。使用Scrapy框架搭建网络爬虫, 爬取网页数据存储在Hbase / MongoDB中。使用Map-Reduce进行数据清洗和数据解析, 使用Spark进行数据挖掘。处理后的数据以图谱形式存入Neo4j / ArangoDB 图数据库中。系统的爬虫端用Python实现, 解析端用Java实现。目前知识图谱已有超过一千万实体, 几百万关系。

### 渣打银行

信息与网络安全部门实习生

中国宁波

2013年10月 - 2014年1月

- 协助企业银行部评估企业信用, 跟踪客户信用评估报告。
- 协助维护银行金融系统。

### 宁波市信息产业局

网络与通信管理处实习生

中国宁波

2013年1月 - 2013年2月

- 学习了解IT产业的相关技术政策, 技术体制和技术标准等。

### 施乐公司

技术部和调度中心实习生

中国宁波

2012年7月 - 2012年8月

- 技术部学习计算机维护和故障处理; 调度中心学习产品和人员管理。

### 中国电信

网络运行维护事业部实习生

中国宁波

2011年7月 - 2011年8月

- 学习局域网组建, 网络参数配置, ISP与局域网连接配置等。

## 研究经历

### 递归神经网络声学模型

硕士毕业论文项目

2015年6月 - 2015年8月

本项目使用递归神经网络RNN构建自动语音识别系统中的声学模型。项目目标是使用RNN声学模型获得比DNN声学模型更优秀的语音识别表现, 并且探索进一步提高语音识别系统表现的其他方法。本项目使用Theano构建RNN模型, 使用Kaldi构建语音识别系统的其他部分。本项目成功证明了项目假设, 即使用RNN构建声学模型, 能够使语音识别系统获得更优的总体表现。项目也探索了影响RNN表现的其他因素。

### 车辆保险用户预测模型

数据挖掘项目

2015年3月 - 2015年5月

本项目基于真实数据预测用户是否会为车辆购买保险, 使用Python和Weka为工具。抽样解决了数据不平衡问题; 使用信息增益率进行特征选择和特征提取; 训练了多个模型 (决策树 / 朴素贝叶斯 / 逻辑回归 / SVM)。

### 商业银行信用评估系统的设计与实现

本科毕业论文项目

2014年2月 - 2014年5月

本项目是受渣打银行实习经验启发, 旨在实现商业银行信用评估系统的流程自动化。一般商业银行的信用评估系统大都需要人工逐个输入和检查数据, 常有易出错, 效率低下, 界面复杂难用等问题。项目采用Java EE技术构建金融系统, 使用业界通用的框架工具 (Hibernate + Struts + Spring) 进行开发。项目系统具备一般信用评估系统的基本功能, 并加入了自动上传报表文件, 自动信用评估 (初级) 等新功能。