

尊敬的老师，您好：

我打算申请贵校的计算机科学与技术专业全日制非定向就业博士生。我感兴趣的研究领域是自然语言处理，机器学习，文本分类或者是图片处理。

科技改变城市，改善生活。而目前智能还有许多发展的空间，如扫地机器人还有许多改善的空间，如何辨析一道菜中有哪些材料等。我愿意深入研究这片领域，所以向贵校提出了申请。我具备了研究机器学习所需的一些基础知识。在本科阶段，我在东华大学计算机系，综合成绩位于专业第二名，专业成绩排在第一名，曾获得国家奖学金、上海市优秀毕业生称号。数学是我一直最喜欢的学科，其中微积分 93 分，多元微积分 10 分，线性代数 96，数据结构 91。我曾在大一大二参加了全国大学生数学建模竞赛，积极地使用了一些模型，如数据不够线性且少时，用灰度预测或者 bp 神经网络预测；因子多时使用主成分因子分析或者多元线性回归模型；梯度下降求最优解；使用损失函数均方误差评估模型。我最终获得上海市数学建模二等奖并代表学校参加“深圳杯”全国大学生数学建模夏令营。

“纸上得来终觉浅，绝知此事要躬行。”还是缺少实践，于是我在毕业项目设计中，与东华大学校友会合作，我独立完成了校友会在线办公自动系统，使用了 JAVA、JSP、MVA 的 struts2.0 框架，结合 Oracle 数据库。此网页系统包括未读已读信息管理、会议室管理系统和数据库池并发等。毕业项目也幸运的获得了东华大学优秀毕业设计。

本着学习先进新技术的想法，我选择了去爱尔兰最好的大学都柏林圣三一学院深造，攻读网络与分布式系统。我得以了解了分布式系统中的中间件、分布式系统中加密算法、点对点网络与中心型网络的对比等。在课程设计中，实现了 JAVA 与 TCP 点对点聊天室，分析在一个网络中有多少服务器结点可以最好的实现容错性、性能负载均衡等。最值得一提的是，在爱尔兰的大学与花旗银行资助的软件创业大赛，我们组的创业想法是“**Soosokan**”，基于地理位置搜索附近的物品，如加油站、饮用水，并实现了这个项目。我的部分是用 JAVA 实现了 RESTful 增删改查，使用 Lucene API 加索引和搜索并用 ArcGIS API 获取位置信息。我们组面对投资者做了许多展示，最终我们四个中国人在与外国队伍的竞争中获得了第一名和创业投资基金。

我的硕士论文是“A personalised ontology and rule-based approach to managing message overload”。这个项目是与 ADAPT research center 研究的课题，使用当前的语义网研究成果（RDF,OWL），结合规则引擎 JESS，管理推送过来的信息。论文中使用了复杂的规则去处理多变的上下文环境和多变复杂的信息，并用 XPATH、JAVA 模拟多种信息，用 protege 软件生成 ontology 并对信息分类和评优先级，论文最后对规则方法和模型进行了评估。这篇论文中的系统只是模型的第一步，以后可以加入可视化，信息挖掘提取研究等。希望未来可以继续研究。

学校所学的目的还是用到企业里。我毕业后的第一份工作是在 Inflight 做软件开发。Inflight 是一家爱尔兰本土发展很好的航空服务公司，我们团队参与的软件获得了航空公司客户的好评。我曾经独立负责一款模型的实现，使用 twitter 分析用户对航空公司的满意度情感分析。这个系统包含三部分服务，第一部分是基于关键紫和 twitter streaming API 实时获取用户评价并存入 mondoDB，需 24 小时开启，第二部分是 nodejs 后台处理数据并提供给 UI，第三部分是 angularjs 和 d3.js 进行实时数据的可视化。使用了简单的乐观悲观词库提取关键字，进行情感分析，并用 word cloud 进行总体评价。这个项目存在许多可以改善的地方，用关键字库这种简单方法情感分析并不准确，大量数据存入数据库性能欠佳，亟待优化。

后来，经过多轮面试，我进入了跨国集团，世界 500 强 Accenture，并在其总部的创新中心工作。我参与过 Aletheia 项目，对电子文档的可靠度评分，对类别分类。这个项目使用 python,machine learning 的方法做后台，通过作者是否可信、是否有引用、文章内容、和是否有多媒体判断文章的可靠性，对文章进行是否有种族主义和不当内容，我负责的是前端网

页插件。此外，我也参与了基于 AI 的医疗保险改革项目。通过可交互 UI 对医疗申诉文档进行加注修改，我负责了前端网页与介绍网页。

从创业团队，到爱尔兰本土公司，再到跨国公司的工作，我希望我的多元背景对接下来的研究有所帮助。

目前，我在花旗集团担任副经理的工作，作为 Web 团队的元老员工，负责债券网页的设计开发，同时进行前端组件库的开发，特别是组件化开发表单与表格。我目前也负责团队内部的项目的 egg.js 和 koajs 后台，也在新项目中研究了 python 后台和 socketio 通信。

我在业余时间学习机器学习相关的算法，lintcode 编程到了银级并且在万门大学学习了人工智能的课程。

关于机器学习和人工智能，我觉得这片领域包含许多算法，没有四五年，难以了解基本框架。我也知道有许多很聪明的人，并且他们已经付出了许多努力，更坚持，深知自己很有差距，但是不开始就永远错失了，与是向贵校申请研究学习。希望您能考虑我的申请。

## 研究计划

### 一. 文本分析：

**问题：**基于自然语言处理、文本分类和用户偏好，规则 rule-engine 对手机的推送信息过滤处理和选择性推送。包含：1.我们看到的电子信息包含不可靠信息，虚假信息，我们可以根据用户评论，作者，和内容摘要内容用词来判断信息是否可靠，进行推送；2 同时，用户有自己的爱好，把信息进行分类，定优先级后，推送给用户；3.对文本中图片进行解读和分类(如果可以)；4. 用户的爱好随时间改变，根据用户的阅读文章习惯，重新改变用户偏好

**研究意义：**对信息进行文本提取，结合 rule-based 信息处理和机器学习文本处理，提升模型准确率

**预期成果：**一款手机软件，可以获取邮件，短信，手机新闻，可以第一次输入用户的一段时间的偏好，对信息打可靠分和分类，根据时间推送。

### 二. 图片分析

**问题：**根据食物和菜的图片，基于用户特征，分析营养，进行评分和推荐。特别是高血压和孕妇。根据用户喜好，进行优先级打分，推送给用户

**研究意义：**识别度低的图片识别。根据图片，来判断菜的营养。包含：食物经过蒸炸后的图

片识别和根据食物颜色来判断营养元素综合判断。同时，用户的喜好也会变化，根据用户的选择，向用户做出推荐

**预期成果：**识别度低的图片可以用在监控仪上，和菜谱识别上。这个成果是一款手机软件，通过拍菜的图片，帮助高血压用户和其他关注营养的用户，判断是否营养，做出推荐。