**项目介绍、岗位及对学生的要求**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 企业团队负责人姓名 | 陈婉斯 | 邮箱 | chenwansi@corp.netease.com | |
| 联系人电话 | 0571-89852957 | 招生人数 | | 根据报名情况确定 |
| 团队简介 | 伏羲实验室自然语言处理组是一个致力于将最新的自然语言处理技术转化到实际产品中的研发团队。我们目前孵化了网易众包任务平台(zb.163.com)、风言舆情平台和小西机器人平台（暂定名称）三个产品。网易众包任务平台目前已经为多个实验室内部以及其他部门的项目提供数据标注服务。风言舆情平台为网易公司的大量游戏产品，包括《倩女幽魂》手游、《逆水寒》、《阴阳师》等等提供了舆情数据分析及其他服务。小西机器人平台目前已经完成了第一期开发工作，正在对接及已经对接的游戏产品包括《倩女幽魂》手游，《忘川风华录》，《倩女隐世录》等等。其中《倩女幽魂》手游的智能养育系统是MMORPG中首个能够进行开放式对话的NPC系统。目前我们共有12名全职员工，若干名实习生待入职。  我们在发力产品的同时，也在研究上进行了一定的探索。我们和清华大学建立了学术合作关系，并且在aaai等会议上投稿6篇，提交了7篇专利申请。在产品逐渐成熟之后，我们会进一步加大研究的投入。 | | | |
| 项目介绍 | 1. 风言舆情平台-抓取数据去噪   风言舆情平台需要对从外部信息源（微博、贴吧、taptap、NGA等）抓取的数据来进行分析。目前我们通过渠道来判断抓取数据是否是某个产品相关的评论，比如说我们认为“逆水寒”贴吧的抓取数据一定是关于《逆水寒》的评论。这显然会引入大量的噪声，因为用户可能会在“逆水寒”贴吧里聊任何事情，这降低了我们分析结果的准确性。因此，我们希望能够开发一套算法来准确地识别那些不与我们关心的产品相关的评论，从而在统计时去除这些评论。   1. 网易众包任务平台-通过主动学习减少数据标注需求量   实现一个可用的机器学习系统需要大量的数据标注，这是软件成本的重要组成部分。为了减少数据标注的数量，研究者提出了主动学习的方法。我们希望能够在众包任务平台的某些任务上实践这一思想。   1. 风言舆情平台-评论摘要生成   目前风言舆情平台使用关键字来对评论进行摘要。这种方式不足以准确地呈现玩家对于游戏的大致看法。我们希望能够对玩家的评论生成简明的摘要，从而更准确清晰地掌握整体舆情。   1. 小西机器人平台-离线用户画像抽取   我们希望能够通过分析用户和机器人对话的日志，获取用户的一部分特征信息，从而让我们的机器人能够针对性地进行对话，表现得更“熟络”。 | | | |
| 实习岗位情况 | 1. 从事与自然语言处理相关产品的研发工作 2. 开发并维护相关平台及工具，提高研发效率 3. 改善现有的自然语言处理相关算法，适用于各类业务 4. 根据业务需求设计并开发自然语言处理相关算法，并撰写论文 | | | |
| 对学生的要求 | 1. 具有机器学习的基础知识，了解常见机器学习算法 2. 具有深度学习的基础知识 3. 具有自然语言处理(NLP)的基础知识和DL4NLP的相关知识 4. 能够熟练使用主流深度学习框架，如pytorch, tensorflow, keras等 5. 计算机基础知识扎实，熟悉算法数据结构，具有良好的编程能力 6. 有在linux上进行开发的经验，能使用Spark等系统进行海量数据处理 7. 了解、能使用各种数据库软件，如MariaDB，MongoDB，Redis等 | | | |