天津大学大学生创新创业训练计划

项目申报书（创新训练）

|  |  |
| --- | --- |
| 项目名称： | 基于知识图谱的行业精准营销策略 |
| 所属专业名称： | 数学学院 |
| 项目负责人： | 龙蕲樾 |
| 联系电话： | 13890048528 |
| 指导教师： | 吕良福 |
| 联系电话： | 15522738860 |
| 申报日期： |  |

天津大学教务处制

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **项目名称** | | 基于知识图谱的行业精准营销策略 | | | | | | | |
| **项目所属**  **专业类代码** | |  | | | **申请经费** | | |  | |
| **项目实施时间** | | 起始时间： 年 月 完成时间： 年 月 | | | | | | | |
| **申请人或申请团队** |  | 姓名 | 年级 | 学号 | | 所在院系  /专业 | 联系电话 | | E-mail |
| 负责人 | 龙蕲樾 | 2019 | 3019233059 | | 数学学院(数学与应用数学) | 13890048528 | | [2265530891@qq.com](mailto:2265530891@qq.com) |
| 成员 | 杨雨璋 | 2019 | 3019233075 | | 数学学院(数学类) | 15127431219 | | [yyz1026010894@163.com](mailto:yyz1026010894@163.com) |
| 许佳鸣 | 2019 | 3019201010 | | 数学学院(数学与应用数学) | 13764516962 | | [1156231451@qq.com](mailto:1156231451@qq.com) |
| 朱宏超 | 2019 | 3019244116 | | 智能与计算学部(计算机科学与技术) | 18526161176 | | [zhc\_0409@163.com](mailto:zhc_0409@163.com) |
| 张恩赐 | 2019 | 3019244118 | | 智能与计算学部(计算机科学与技术) | 18676385142 | | [2716548913@qq.com](mailto:2716548913@qq.com) |
| **指 导 教 师** | 第一指导教师 | 姓名 | 吕良福 | | | 专业技术职务 | 副教授 | | |
| 单位 | 天津大学数学学院 | | | 研究方向 | 人工智能算法及应用 | | |
| 联系电话 | 15522738860 | | | E-mail | liangfulv@tju.edu.cn | | |
| 第二指导教师 | 姓名 |  | | | 专业技术职务 |  | | |
| 单位 |  | | | 研究方向 |  | | |
| 联系电话 |  | | | E-mail |  | | |
| **一、项目简介**（200字以内）  近年来，随着互联网的普及和大数据时代的到来,若企业想在市场中不断加大，就必须从营销角度重新审视市场，构建精准营销体系。相较于前人系统中存在的提取大量数据时复杂性较高、用户分析不够准确等问题,本项目从客户精准营销角度，对企业进行数据挖掘与客户行为分析，构建基于知识图谱的用户画像系统，并通过图卷积神经网络优化知识图谱，降低特征提取的复杂度,克服了人工操作标签化的缺陷，做到数据驱动的产品自我推销，从而提高客户忠诚度，实现企业精准营销体系构建。 | | | | | | | | | |
| **二、申请理由**（包括自身具备的知识条件、自己的特长、兴趣、已有的实践创新成果等）  项目组来自不同的学院,对不同的领域有所了解,本项目由数学学院和智能与计算学部联合开发,以数学学院的理论为基础,加之计算技术的支持完成本项目的实现.  1.在课外自学Python并对python中用于数据可视化和数据分析的库有着充分的掌握；  2.我们通过查阅大量的相关资料与文献对于神经网络和用户画像已经有了初步的了解，并再次基础上对于我们研究的课题有了一些创新型的见解。  3.队员已经能熟练的使用PostgreSQL进行数据库处理。  4.项目组成员具备了充足的英文文献与阅读能力。  5.项目组成员热爱计算机技术并具有基础的计算机科学知识，希望能够学以致用。  6.项目组成员具备数学国赛，美赛建模等大赛经验。 | | | | | | | | | |
| **三、项目方案**  具体内容包括：  1、项目研究背景  1.1基于用户画像国内外研究现状  第一句话应该是来一句研究这个问题的意义。用户画像最早由Allan Cooper提出[1]认为构建用户画像的数据需要创建在真实有意义的数据上，能够准确的识别用户的偏好和潜在需求，通过数据模型挖掘用户的现实需求，实现用户的精准营销，降低企业营销成本，提升用户营销的ROI。在国外用户画像技术首先用于个性化检索行业，许多专家研究人员通过各种数字技术方式对客户画像进行了科学研究。客户画像的使用的发展趋势正不断地向上，这放映了其在许多行业中的使用价值，例如客户性别推断，年龄推断等。随着云计算技术的发展趋势越来越快，客户在大数据营销层面，用户画像技术变得更多地被使用；F Abel等基于将Twitter于相关信息文章集成在一起以开发文本挖掘和拓展Twitter内容，从而产生了更精准的客户信息。安东·韦恩斯等人提出了一种全自动的无监督用户肖像构建模式。创建用户画像的方法根据业务场景的不同而发生变化。客户信息的内容不同，主要目的不同，企业的战略定位也不同，匹配的用户肖像实体模型具有不同的方式和维度。  我国的用户画像研究大致分成三个阶段（吴佳琪，2018[2]）：初始阶段、萌芽阶段和发展阶段。初始阶段（2006-2009年）侧重于分析客户的个人行为数据。为此，将服务项目呈现给客户。这段时期刚刚开始应用用户画像技术功能，例如大数据挖掘和客户实体模型。在萌芽期（2010-2014年），此类相关参考文献逐渐变多。特别是谷歌得Google Buzz在2010年被发行，那是在我国刚刚开始普及用户画像。此外，在诸如在线游戏、大数据技术视频和社交网络之类的行业中，用户画像也被慢慢用于新产品开发和设计解决方案中。在发展趋势中（2016年起始），用户画像已经普遍在日常生活中使用，用户画像的科学研究和应用步入到一个新的发展阶段。  根据近年来的发展趋势用户画像的研究已经从早期的计算机信息科学、软件工程和电商等行业发展到自然、人文、医学、经济与管理等各个领域。用户画像的科学研究也开始进入到跨学科的研究。如今，基于互联网大数据，大数据挖掘与用户画像的科学研究紧密联系。张哲（2015）创建了用户画像根据发展的微博数据；丁伟（2016）等通过大数据中数据采集数据加工，数据分析，数据应用等方法对手机用户数据进行清洗挖掘分析，然后沉淀用户标签特征，进而应用到个人征信和用户推荐和精准的广告营销等方面；王智囊（2016）描绘了制药行业的角色，并完成了基于B/S的只能医疗推荐系统软件，病人门诊就医时的等待时间减少了；孙新伟（2017）基于对电子商务企业网购客户数据信息的分析，对客户实现分类。  1.2用户画像在精准营销中的应用研究现状  《大数据时代》的作者Viktor Mayer-Schonberger认为，海量数据的混杂性更能够表达用户信息，相对于小数据量，复杂数据建模对用户的信息会有更好的体现，因此在大数据环境下相关的研究也是目前的大趋所向。现对于挖掘因果关系的为什么，我们更期待相关关系是什么，容忍一定的误差，获取更快更高的业务价值。所以企业未来对数据战略的制定需要有更深的考虑和计划。因而由一系列标签组成的用户画像不仅仅为企业提供了了解用户全貌的机会，同时也反映了很多消费者潜在的业务需求，提供了挖掘消费者的价值，为企业对消费者群体细分的精准营销和业务场景目标客群的投放打下了基础。    2、项目研究目标及主要内容  目标:  通过网络爬虫获取给定网站的用户数据,使用数据库,深度学习,神经网络等技术,分析出用户画像,帮助平台更好的了解用户,掌握平台运行状况,以实现精准营销.项目最后分析实现的思路和方法,给出一个为平台绘制用户画像的一般方法.  主要内容:  2.1数据收集：项目首先需要收集平台的用户数据,用于建立用户画像.数据有两个来源:现有数据和网络收集.网络数据主要依靠网络爬虫技术.现有数据来源行业公司对接.  2.2数据标注：得到适当的用户信息存储至数据库,使用深度学习,自然语言处理等方法标签化用户,获得分类数据.  2.3模型建立：在借鉴现有类似客户画像的基础上，运用图卷积神经网络构建知识图谱（或拟采用决策树算法进行用户分层），并基于此建立客户画像系统.  2.4行为分析：综合所得的用户画像,选取其中具有代表性的部分,通过数学建模和数据可视化技术,制作出简明清晰的营销线路和用户行为分析.  2.5营销方案：基于用户画像的精准营销方案。首先，通过构建用户画像，形成用户画像数据库。构建基于用户画像的精准营销体系平台，最后为企业如何利用用户画像的精准营销体系平台进行精准营销提供具体实施组合策略，希望能够提升企业的营销准确性、服务质量及企业利润。  2.6项目总结：对项目进行总结，并对项目的创新性和不足做出阐述.  3、项目创新特色概述  本项目的创新点在与使用用户画像技术对平台的用户进行分析,相比于传统的行业用户分析,用户画像可以使产品的服务对象更加聚焦，更加的专注。这样说不行啊，得说：对大数据分析能力不够、数据分析不准确等缺点。。。。。对平台进行数据挖掘与客户行为分析，构建基于知识图谱的用户画像系统，并通过图卷积神经网络优化知识图谱，克服人工操作标签化的缺陷，做到数据驱动的产品自我推销，从而提高客户忠诚度，实现企业精准营销体系构建。  4、项目研究技术路线  本项目的主要技术体现在数据采集和建立用户画像两个方面.  数据采集:  本项目计划使用网络爬虫进行数据采集,使用Python语言辅以Scrapy框架,  对指定网站进行数据采集.采集的数据会存入数据库中备用,计划使用PostgreSQL数据管理系统对数据管理,分类.  建立用户画像:  建立用户画像使用机器学习的分类和回归算法（贝叶斯，决策树，逻辑回归，支持向量机等）.采取这些有监督的学习方法，从标注好的训练数据学习到一个预测模型，然后对未标注的数据进行预测。  可能使用的算法和库:  1.libsvm,liblinear库  2.Xgboost库  3.朴素贝叶斯算法 Naive Bayes,k-平均算法 K-Means,线性回归算法 Linear Regression,逻辑回归算法 Logistic Regression.  4.图卷积神经网络  屏幕截图 2021-04-11 203512  图1:项目技术路线图  5、研究进度安排  2021年4月20日-2021年10月1日：广泛搜集、阅读国内外研究文献，学习有关机器学习、数学建模、聚类分析等相关理论知识，为论文奠定理论基础。  2021年10月1日-2022年1月1日：整理搜集资料，设计系统原型。按照预期目标，构建理论模型，使用Python或其他编程语言实现已有算法。  2022年1月1日-2015年2月1日：结合资料文献改进已有算法，运用数据进行实验，优化程序与模型并完成初稿。  2022年2月1日-2022年3月15日：整合导师修改意见及个人资料信息补充修改初稿，进行系统的全面测试，并对可能存在的问题进行修复，完成第二稿。  2022年3月15日-2022年4月30日：汇总并完成结题报告，同时申报论文。  6、项目组成员分工  理论组(杨雨璋,许佳鸣,龙蕲樾):负责精准营销体系的设计和策略制定,学习相关算法和理论基础.  程序组(张恩赐,朱宏超):负责采集数据和算法的代码实现及优化.  广泛阅读学习相关资料：全员  搭建一个简单的产品原型：全员  撰写报告、论文：全员 | | | | | | | | | |
| **四、预期成果（在对应项目后填写明细）**  A、研究报告： 1 篇；(对一个给定平台的用户画像分析报告)  B、论文发表： 1 篇；  C、专利申请： 0 项；  D、产品设计： 0 个；  E、成果实物： 1 个；(研究过程中完成的代码项目)  F、其它预期成果： | | | | | | | | | |
| 1. **经费预算。。。。。。。。这个需要有**  |  |  |  | | --- | --- | --- | | **项目** | **预算（元）** | **计算依据** | | 设备费 |  |  | | 材料费 |  |  | | 测试化验加工费 |  |  | | 燃料动力费 |  |  | | 差旅费 |  |  | | 会议费 |  |  | | 出版/文献/信息传播/知识产权事务费 |  |  | | 其他 |  |  | | 合计： （元） | | | | | | | | | | | | |
| **六、指导教师意见**  签名：  年 月 日 | | | | | | | | | |
| **七、学院意见**  学院评审委员会（小组）负责人签名： （学院盖章）  年 月 日 | | | | | | | | | |
| **八、项目负责人承诺：**  **我保证填报内容的真实性。我已阅知并严格遵守天津大学关于大学生创新创业训练项目管理办法和相关财务管理制度规定。本人将按照申报书的内容负责实施本项目，切实保证学习研究时间，认真开展项目工作，及时报告重大情况变动，按时报送有关材料。**  负责人签名：  年 月 日 | | | | | | | | | |
| **九、学校推荐意见**  学校负责人签名： （学校公章）  年 月 日 | | | | | | | | | |

引用文献:

[1]《The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity》（Sams 出版社，1999）

[2]吴加琪."我国用户画像研究的知识网络与热点领域分析." 现代情报 38.08(2018):130-135+143. doi:CNKI:SUN:XDQB.0.2018-08-018.