

一、专业实践训练整体情况

实践单位名称	十九楼网络股份有限公司	
实践单位地点	杭州市西湖区双龙街 239 号西投创智中心 2 号楼 3 楼	
实践岗位名称	研发	
专业实践训练时间	集中进行	2021 年 10 月 11 日开始 至 2022 年 01 月 20 日结束
		专业实践训练累计 101 天（单位考核前），其中项目研究天数 90 天（单位考核前）
<p>(1) 基本情况（含实践单位简介、实习实践内容等）</p> <p>单位简介：杭州十九楼网络股份有限公司创建于 2006 年 10 月，是杭州日报报业集团和都市快报新媒体战略的重要组成部分。公司以“有城市就有活跃的 19 楼”为创业愿景，始终坚持用户第一，成立以来主营业务发展迅速。主营业务是 19 楼，一个互联网社区网站，提供本地生活消费、情感问题自由交流等活动的平台。</p> <p>实习实践内容：学习机器学习、深度学习等智能推荐相关知识，学习使用决策树、神经网络、deepwalk 等算法去训练模型进行相亲业务的个性化推荐。</p>		
<p>(2) 项目研究概述（含项目名称、项目来源、项目经费、主要研究目标和技术难点等）</p> <p>项目名称：相亲用户智能推荐</p> <p>项目来源：实践单位</p> <p>项目经费：50000 元</p> <p>主要研究目标：根据用户的信息和用户选择合适相亲对象的浏览记录，推断用户最合适的相亲对象</p> <p>技术难点：数据冷启动问题；长尾效应；</p>		

(3) 项目开展情况（含项目研究内容、研究方案及技术路线，研究团队分工、本人承担任务及完成情况，存在问题与改进建议等，不少于 500 字。）

研究内容：根据用户的信息和用户选择合适相亲对象的浏览记录，推断用户最合适的相亲对象

方案及技术路线：

1. 数据集分析：主要由两部分构成，一部分是用户交互记录，总计约 100w 条交互记录，包含登陆用户 A 的 id、被推荐用户 B 的 id、登陆用户是否点击、登陆用户是否选择心动；另外一部分是用户个人信息，总计约 3w 活跃用户，包含用户注册信息：年龄、收入、学历、身高、体重、是否有车有房、星座等信息。

2. 计算用户受欢迎程度 = $(a_1 * \text{被推荐次数} + a_2 * \text{被点击次数} + a_3 * \text{被心动次数}) \div (\text{被推荐次数} + \text{被点击次数} + \text{被心动次数})$ $a_1/a_2/a_3$ 为三种交互行为的权重，满足 $a_1 < a_2 < a_3$ ，目前取值为 $a_1=1, a_2=4, a_3=10$

3. 利用决策树算法计算两位用户的喜爱程度。

4. 使用 deepwalk 算法，根据用户交互记录构建有向带权图，点击次数权重为 1，心动权重为 20。根据前面生成的用户关系图，按照权重占比进行随机游走，得到用户行为序列。根据用户个人资料，比如可以将年龄相同的用户进行连接，得到用户特征序列。将得到用户行为序列和用户特征序列输入到 Word2vec 中生成 Embedding 向量，过滤掉同性别用户后，计算用户向量间的距离，圈出距离最近的一些点进行排序，得到最终的推荐列表。

本人承担以上所描述的任务并且基本完成任务。

问题与改进建议：

1. 数据量还是偏少，可以再积累更多的数据后进行重新训练计算。
2. 受欢迎程度的权重能否学习，而不是通过个人经验估计。
3. 是否有具体目标？比如精确度提升多少？

二、专业实践训练收获

(一) 围绕考核评价指标体系，举例说明以下收获（不少于 800 字）

1. 知识掌握：

(1) 基础以及专业知识掌握：学习掌握了神经网络和机器学习的相关知识，深刻学习掌握了反向传播、梯度下降、损失函数等基本概念。并且根据梯度下降等原理自己去写代码实现了一下；学习并实现了决策树算法。

(2) 行业知识掌握：学习推荐系统相关算法，重点学习了协同过滤和 deepwalk 算法，其中协同过滤侧重于寻找 user 和 item 的联系，deepwalk 更适用于研究 user 和 user 之间的社交网络的关系。

(3) 默会性工程知识掌握：学会了具体业务场景需要具体分析，相亲个性化推荐对比其他推荐的特殊性在于是一个双向选择的过程，并且有较多的客观限制因素，所以不宜采用协同过滤算法，deepwalk 算法更适合。

(4) 跨专业领域知识掌握：从社会学范畴了解调研了现在国内婚姻现状和现代年轻人的婚恋观，从 2013 年起，结婚率持续走低，离婚率持续升高；并且现在随着技术的发展和疫情的影响，相亲的需求再不断增大，婚恋观念有向传统婚恋即门当户对靠拢的迹象。

2. 能力提升：

(1) 环境及岗位适应能力：到新环境和新岗位的适应速度比较快，能较快融入到实践单位的工作流程和企业文化中。能按照计划执行完成指定任务。

(2) 参与工程建设所需的基本技能：学习完成相关算法后能及时实现和应用，例如学会了决策树算法后，进行实现计算后发现准确率并不高，就及时考虑其他算法的可能性。

(3) 技术应用创新及工程实践能力：实现了 deepwalk 算法应用，根据用户交互记录构建有向带权图，点击次数权重为 1，心动权重为 20。根据前面生成的用户关系图，按照权重占比进行随机游走，得到用户行为序列。并且根据用户个人资料，比如可以将年龄相同的用户进行连接，得到用户特征序列。增加了用户特征序列来进一步提升算法准确度。

(4) 团队协作能力：积极和公司的产品经理、项目经理以及网站开发程序员进行沟通，在数据分析收集上一起协作；在技术上也会和技术总监进行各种交流，不断优化算法。

(5) 工程思维养成：提升了问题导向意识，比如为了解决数据冷启动问题，设计了受欢迎程度的参数，按照比例推送给新用户，加速新用户的启动速度。

3. 素质养成：

(1) 品德修养：热爱祖国，爱岗敬业，通过这次专业实践，了解到了国内婚姻现状的严峻性，希望开发的相亲个性化推荐能为改善国内婚姻状况贡献一份力，改善了婚

姻状况以后，也能更好的响应三胎政策，慢慢减轻老龄化的压力。

(2) 科学素养：国内婚姻现状和婚恋观的调研追求真实准确，数据集的收集也是系统获取到的真实的交互记录和用户个人信息，追求科学严谨。

(3) 职业素养：具备良好的职业道德，获取到了大量的用户数据同时不外传，并且只用于课题研究，尽自己最大努力保护用户隐私。

4. 总结：这次专业实践收获颇丰，不论是从知识、能力和素养方面都获得了不小的提升。尤其是工程思维的养成和团队协作的能力等方面只有在具体工程开发过程当中才能有更深刻的领悟。

(二) 取得成效

1. 通过技术应用创新、成果转化、解决企业工程实际问题等取得的经济和社会效益：

(1) 伴随着互联网的飞速发展，传统的婚恋相亲方式已经不能满足生活节奏较快的年轻人了，现在的年轻人对相亲平台的需求逐步增加，相亲平台的个性化推荐是当代年轻人迫切需要的。19楼 app 的相亲板块之前采用的是随机推荐，推荐点击的转化效率低下，成功匹配到人数就更少了。现在使用 DeepWalk 算法实现了个性化推荐，提高了活跃用户的点击转化率。针对新用户推荐的受欢迎程度高的异性，也相对较好地解决了新用户的冷启动问题。提高了 app 会员的充值率，为公司获取了一些经济利益。

(2) 19楼相亲 app 还为会员提供了人工匹配的服务，通过人工比对择偶条件，判断双方合适程度，之后通过电话联系双方进行相亲活动。人工成本很大，而且效率也不高，一个人一天大约只能联系到 20-30 对，其中愿意进行相亲活动的几乎为 0。而现在通过个性化推荐的评分，后台服务人员可以看到用户之间的评分，方便筛选，联系效率变高的同时，双方愿意进行相亲活动的意愿也比原来高。大大降低了人工成本，从侧面也为公司经济利益提升做出了一些贡献。

2. 与学位论文撰写的相关程度（原则上研究生学位论文选题及内容应来源于项目研究内容）：

学位论文题目为相亲平台个性化推荐算法研究与应用，与学位论文撰写的相关程度紧密。学位论文研究的内容为相亲平台个性化推荐算法研究对比与应用，对比了受欢迎程度、决策树、贝叶斯概率、威尔逊取件、DeepWalk 等不同算法的准确率。该专业实践项目使用了受欢迎程度的算法，解决了新用户冷启动问题，使新用户能尽快产生交互。使用 DeepWalk 算法能够提取出活跃用户的偏好和特征，能够得到一个大致满足用户需要的推荐列表，顺利完成开题答辩。后续待解决或者可优化的点：(1) 目前 DeepWalk 算法是比较依赖于用户活跃度的，用户交互记录越少，随机游走到的概率就越低，容易产生长尾效应，后续还需要其他方法去解决该问题。(2) 受欢迎程度以及用户交互记录构建图中的权重可以改用训练获取，而不是通过经验判断来尝试。

3. 在校期间主要研究成果【含产品与样机、专利（含申请）、著作、软件著作权、论

文、标准、获奖、成果转化等】

成果名称	类别[含产品与样机、专利（含申请）、著作、软件著作权、论文、标准、获奖、成果转化等]	发表时间/授权或申请时间等	刊物名称/专利授权或申请号等	本人排名/总人数	学校排名/总参与单位数
------	--	---------------	----------------	----------	-------------

本人承诺


在专业实践训练及考核报告撰写过程中，如实提供材料，严守学术道德、遵循学术规范。

签字: 

2022年6月7日

三、考核评价

<p>校外合作 导师(或现 场导师) 评价</p>	<p>重点对研究生项目研究开展情况、职业素养、行业知识掌握、环境和岗位适应能力、工程实践能力、团队协作能力，以及通过技术应用创新、成果转化、解决工程实际问题等取得的经济和社会效益等方面的评价：</p> <p>赵晓彬同学在本次专业实践中表现优异，学习能力 强，在学习推荐系统相关算法时，有意识地去改造优化 现有算法，为了解决数据冷启动问题，也想出了受 欢迎程度和贝叶斯算法结合做推荐；工作中也积 极和团队中其他人员沟通交流，团队协作能力 较好。</p> <p>校外合作导师（或现场导师）签字：张锡刚 年 6月8日 2022</p>
<p>校内导师 评价</p>	<p>重点对研究生科学素质、基础及专业知识掌握、技术应用创新能力、取得的研究成果、项目研究与学位论文撰写的相关程度等方面的评价：</p> <p>情况属实</p> <p>校内导师签字：吴 年 月 日</p>

实践单位 过程考核 意见	<p>实际实践开始时间: 21年10月11日 实际实践结束时间: 22年1月20日</p> <p>专业实践训练累计天数: 16 其中项目研究天数: 90</p> <p>实践单位过程考核结果: <input checked="" type="checkbox"/>优秀 <input type="checkbox"/>良好 <input type="checkbox"/>合格 <input type="checkbox"/>不合格</p> <p>审核签字并盖公章:  2022年6月8日</p>
最终考核 结果审核 备案	<p>考核总成绩 (由现场答辩考核成绩90%+单位过程考核成绩10%组成):</p> <p>是否重修: <input type="checkbox"/>是 <input type="checkbox"/>否</p> <p>教学管理部 (或相关分院) 审核签字 (公章): _____ 年 月 日</p>

四、相关支撑材料

在校期间主要研究成果【含产品与样机、专利（含申请）、著作、软件著作权、论文、标准、获奖、成果转化等】证明材料原件扫描件，具体提交要求如下：

1. 产品与样机扫描件包含企业证明材料（含产品与样机功能及创新性介绍、社会经济效益、个人贡献说明及相关照片等）。
2. 授权专利扫描件包含专利证书授权页；未授权专利扫描件包含专利受理书扫描件和专利请求书扫描件。
3. 著作扫描件包含封面、封底和版权页。
4. 软件著作权扫描件包含著作权证书和登记申请表。
5. 论文扫描件包含封面、封底、目录和论文全文（含收录证明）。
6. 标准扫描件包含封面、版权页、发布公告、前言和目次。
7. 获奖扫描件包含显示单位和个人排名的获奖证书。
8. 成果转化扫描件包含企业证明材料（含成果技术说明、社会经济效益、个人贡献说明及相关照片等）。