分类号_____　　密级　_____
U D C _____　　编号　10741

硕 士 学 位 论 文

(专业学位)

论文题目　移动终端用户网络行为分析及应用

研　究　生　姓　名：杜金娥

指导教师姓名、职称：庞智强　教授

学 科、专 业 名 称：统计学　应用统计硕士

研　究　方　向：市场研究

提　交　日　期：2022 年 5 月 30 日

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：杜金娥 签字日期：2021年5月30日

导师签名：[illegible] 签字日期：2021年5月30日

关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定，同意（选择“同意”/“不同意”）以下事项：

1.学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；

2.学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入 CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名：杜金娥 签字日期：2021年5月30日

导师签名：[illegible] 签字日期：2021年5月30日

Mobile Terminal Users Network Behavior Analysis and Application

Candidate : Du Jine

Supervisor: Pang Zhiqiang

摘要

随着移动互联网的飞速发展，移动终端设备已经完全融入到我们的日常生活、工作学习、娱乐以及社交活动中，随之而来的是用户难以从海量的行为信息中获取有效资源。如何挖掘移动终端用户的网络行为并且进行精准的个性化推荐服务已成为运营商亟待需要解决的问题，然而传统的用户行为分析从单一场景出发，未考虑用户的行为习惯可能由于时间因素或者外部环境的变化发生改变，导致个性化推荐的信息不符合用户实际的行为习惯。

因此，本文从上网时段、访问内容两方面来分析移动终端用户的网络行为特征，并且利用 BP 神经网络优化协同过滤推荐算法中基于平均分的预测评分公式，对于运营商而言，可以利用预测结果为用户制定个性化网络服务的推荐列表，有助于提升用户对网络应用的黏度和满意度。为了实现移动终端用户网络行为特征的挖掘，本文从用户的上网时段、访问内容两方面进行分析。首先，选用欧氏距离度量用户上网时间的相似性，通过验证发现用户上网时间在最大上网时段处具有相似性规律，基于此利用层次聚类挖掘用户上网时间的 4 种行为模式；其次，采用 k−means 聚类算法挖掘用户访问内容的 7 种行为模式；最后关联用户上网时段的 4 种行为模式和访问内容的 7 种行为模式，得到用户在不同时间段的上网行为特征。

为提高移动终端用户的上网体验，使其实现更加精准的个性化网络服务。本文利用 TF−IDF 算法计算用户对网络服务类别的喜好程度，针对传统推荐算法的预测评分准确性较低的问题，本文将 BP 神经网络与协同过滤推荐算法相结合来优化预测评分，并将其与改进的协同过滤推荐算法、基于情感分析的协同过滤推荐算法进行对比。研究结果表明：通过 BP 神经网络改进协同过滤推荐算法，其均方根误差和平均绝对误差均小于改进相似度的协同过滤推荐算法、基于情感分析的协同过滤推荐算法，因此本文所使用的个性化推荐算法能够更有效地提高评分预测的准确性且得到了更好的效果。最后提出个性化推荐算法在广告精准投放、网络信息精准推送与电子商务营销三个方面的应用价值。

关键词：移动终端用户 网络行为 BP 神经网络 聚类算法 个性化服务

Abstract

With the rapid development of the mobile Internet, mobile terminal devices have been fully integrated into our daily life, work, study, entertainment and social activities, and it is difficult for users to obtain effective resources from massive behavioral information. How to mine the network behavior of mobile terminal users and provide accurate personalized recommendation services has become an urgent problem for operators to solve. However, the traditional users behavior analysis starts from a single scenario and does not consider the user's behavioral habits due to time factors or external environment. changes, resulting in personalized recommendations that do not conform to the actual behavior of users.

Therefore, this thesis analyzes the network behavior characteristics of mobile terminal users from the aspects of Internet access time and access content, and uses the back propagation neural network to optimize the average score-based prediction scoring formula in the collaborative filtering recommendation algorithm. For operators, the prediction results can be used. Developing a recommendation list of personalized network services for users can help improve users' stickiness and satisfaction with network applications. In order to realize the mining of mobile terminal users' network behavior characteristics, this thesis analyzes the users' online time period and access content. First, the Euclidean distance is

used to measure the similarity of users' online time. Through verification, it is found that the users' online time has a similarity law at the maximum online period. Based on this, hierarchical clustering is used to mine four behavioral patterns of users' online time. Second, k-means clustering algorithm mines 7 behavior patterns of users accessing content; finally, correlates the 4 behavior patterns during the users' surfing period with the 7 behavior patterns of accessing content, and obtains the surfing behavior characteristics of users in different time periods.

In order to improve the Internet experience of mobile terminal users and enable them to achieve more accurate personalized network services. In this thesis, the termfrequency–inversedocumentfrequency algorithm is used to calculate the users' preference for the network service category. In view of the low accuracy of the prediction score of the traditional recommendation algorithm, this thesis combines the back propagation neural network with the collaborative filtering recommendation algorithm to optimize the prediction score, and use it to optimize the prediction score.Compared with improved collaborative filtering recommendation algorithm and the collaborative filtering recommendation algorithm based on sentiment analysis. The research results show that the root mean square error and mean absolute error of the collaborative filtering recommendation algorithm improved by the back propagation neural network are smaller than those of the collaborative filtering

recommendation algorithm based on the improved similarity and the collaborative filtering recommendation algorithm based on sentiment analysis. Therefore, the personalized recommendation algorithm used in this thesis is the recommendation algorithm can improve the accuracy of rating prediction more effectively and get better results. Finally, the application value of personalized recommendation algorithm in three aspects: advertising precision placement, network information precision push and e-commerce marketing is proposed.

Keywords: Mobile terminal users;Network behavior;Back propagation neural network;Clustering algorithm;Personalized service

目 录

1 绪论

1.1 研究背景与意义

互联网技术的出现使我们步入了信息化时代，网络平台提供的信息也日益丰富，无论是移动终端用户行为分析，还是 Web 服务器用户行为分析，都是通过得到用户的网络行为规律来优化网络服务。互联网技术的出现有利有弊，我们在享受获取信息便捷的同时，无不被“信息过载”[1]所重压，用户的网络行为研究必然对于各大运营商及用户有着重要的意义。

1.1.1 研究背景

由于互联网与移动通信有机融合诞生的新兴市场，政府加大对互联网的支持力度，运营商通过降低上网资费来巩固老客户并吸引新的用户，移动网络的覆盖率也越来越高，从而使用户的上网门槛越来越低。根据中国互联网络信息中心（CNNIC）发布的报告[2]可知，截止 2021 年 6 月，我国手机用户规模达 10.07 亿，其中使用手机上网的用户比例达 99.6% [61]，从图 1.1 可看出，手机上网的使用率显著高于其他设备的使用率。又根据 APP Annie[3]发布的数据显示，2021 年全球移动终端使用量达到 3.8 万小时，创历史新高。

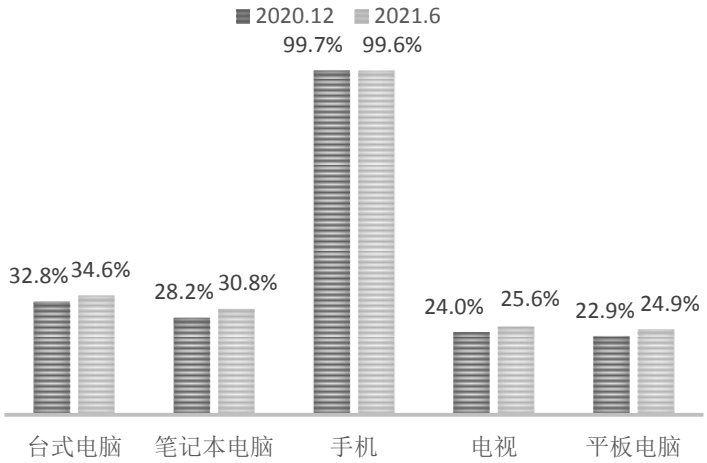

图 1.1 互联网接入设备使用情况

[1] 信息过载：是指社会信息超过个人或系统所能接受、处理和有效利用的范围，并导致故障的状况。
[2] 中国互联网信息中心报告：第 47 次《中国互联网网络发展状况统计报告》。
[3] APP Annie 报告：移动数据分析公司 APP Annie 发布的《2022 年移动状态报告》。

移动网络技术的快速发展使手机终端步入智能化时代，网络资源也变得越来越丰富，用户获取网络信息便捷甚至实现了信息共享，进而使移动终端用户规模不断扩大。我们的生活、娱乐社交、工作等与移动智能化设备紧密融合，例如网络购物、在线银行、远程教育、即时聊天、电子商务、线上医疗咨询等各种网络服务，增加了人与人之间的沟通交流以及节省了大量时间和金钱。与此同时，人们使用移动终端设备时生成了海量的用户网络行为“痕迹”，通过利用数据挖掘技术采集并分析这些行为数据，获得用户使用网络时的偏好，从而有助于网络服务提供商发现隐藏的商业价值，更有助于推动移动互联网技术的发展。

目前，针对移动终端上各种类型的 APP 越来越多元化，运营商更希望根据用户的需求提供服务，同样用户也希望得到别样的网络体验。从服务提供商的角度出发，设计者通过量化分析用户的行为习惯、时间动态及访问频率等，调整系统功能和系统内容，为用户提供更有针对性的网络服务，从而提高用户的信任度和粘性，增加营业收入；从用户的角度出发，同样希望在网络应用中快速获取自己喜欢的内容或者服务，例如，服务商将整个网络应用的全站式搜索框设置在最显眼的位置，这样更方便用户发现可能喜欢的新事物，也使用户享受到更便捷的上网体验和有价值的娱乐服务。从服务提供商和用户的需求出发，分析个体或者群体潜在的网络行为偏好，已经成为个性化网络发展的一种必然趋势[27]，移动终端用户网络行为分析已经成为一个重要的热点研究课题，移动互联网的服务仍有待提升。

网络行为的定义是伴随现代网络技术出现的，它是发生在互联网上的一种虚拟行为，用户为了实现某种目的，借助互联网平台进行有意识的活动。用户的个体网络行为虽然在短时间内表现不出明显的行为规律，但是在长时间内可以表现出其规律性[52]。用户行为分析就是研究和分析用户在网络应用活动中所表现出来的规律，借此预测用户的兴趣偏好。目前，运营商及服务提供商更关注的是用户留在网络平台的“痕迹”，通过分析这类型的数据发现他们的行为偏好，有助于发现潜在商机，在电子商务营销中最重要的应用之一便是“行为定向”[17]。

虽然移动互联网给人们的生活带来了很多便利，但随之而来的信息大爆炸，海量信息不断涌入移动终端，进而导致了“信息过载”问题的出现，针对“信息过载”用户获取有效信息困难的困扰，学者们提出了多种多样的解决方法，比如

刘建国等通过历史的选择和相似性挖掘出每个用户潜在感兴趣的信息，继而进行差异化服务[39]；个性化推荐系统的应运而生，胡明珠根据关键词筛选出与用户相关的信息，并结合个性化推荐系统有效的为用户提供更精准的个性化网络服务[29]。如今，个性化推荐技术已然成为互联网公司获取用户、创造收益的关键技术，国内外各大公司纷纷推出个性化推荐服务，例如国外的 Twitter、Facebook、You Tube 等，国内的淘宝、京东、WeChat 等，将推荐系统与这些大规模平台相互融合，可以让用户的操作变得更加方便，企业也可以通过对推荐系统的分析和优化来完善自己的推荐系统，进一步提高推荐效果的精确度，从而获得更多新用户及收益。

由于移动终端设备的网络服务各式各样，导致各种各样的信息蜂拥而至，并且在这些推送中用户往往对很多内容都不感兴趣，过多呈现无用或者不相关的网络信息可能会浪费用户时间，甚至降低用户体验，为了满足移动终端用户的不同需求，从而为其提供更为优质的服务，如何对用户的网络行为偏好进行准确和全面挖掘？为了解决这一问题，本文针对移动终端用户历史网络行为数据进行画像分析，并结合个性化推荐系统，能够最大限度匹配不同用户的需求，从而帮助运营商提升服务和开展精准营销策略。

1.1.2 研究意义

随着人们对电商、新闻、游戏、生活服务等网络应用的依赖程度逐步加深，以及在互联网上随时随地获取各种网络信息。用户通过访问网络应用给服务提供商留下访问足迹，而这些足迹中包含大量宝贵信息，它们可以帮我们回答很多问题，比如用户在什么时间段使用网络，对何种类型的网络服务比较感兴趣，通过对这些网络行为数据进行有效合理的分类，从而获得用户在移动智能终端设备的网络行为习惯，最终有利于更多企业为用户制定差异化的网络需求、个性化的新型产品和精准的广告推荐等方面做出贡献，并为其带来更多的增值服务。准确来说，对移动终端用户网络行为的研究意义主要体现在以下几个方面：

（1）就运营企业或服务提供商而言，挖掘移动终端用户的网络行为习惯有助于其进行网络营销或者网络应用设计。一方面，企业针对挖掘的用户行为习惯为用户提供个性化的网络需求，不仅满足了用户的需求，也提升用户对网络应用

的满意度及粘性，最终实现双赢，这也是一种提高市场竞争能力行之有效的方法。另一方面，分析用户的网络行为习惯有助于企业精准投放广告，为目标客户提供精确的市场推广，还可以挖掘出潜在的目标消费人群进行产品的搭配营销和服务。

（2）就移动终端用户而言，移动互联网给我们的生活带来了很多便利的同时也伴随着信息大爆炸，相关互联网企业通过网络行为分析优化用户的网络需求，用户可以更顺畅、更方便地使用互联网，甚至是体验个性化的网络服务。

（3）就关联企业而言，关联其他不同类型的网络行为数据，发现多角度的网络行为模式之间的相关性。例如，关联移动终端用户的上网时间与访问内容的行为偏好，有助于企业发现用户在不同时间的潜在网络需求；还可与旅游交通企业拥有的关于地理位置的网络行为数据相关联，可以更加全面准确地描述用户在移动终端的上网规律，分析用户的行为偏好，对网络应用的内容安排与广告投放具有重要意义。

综上可知，移动终端用户网络行为分析的意义重大，结合数据挖掘的深度探究，利用移动终端用户的网络行为轨迹，精准定位和挖掘用户的行为偏好，从而满足用户的不同需求以及给用户提供更优质的网络服务，这对研究移动终端的应用体验以及提升用户的满意度有着不言而喻的意义。

1.2 国内外研究现状

随着网络服务的升级和无线网络的不断发展，用户的网络行为分析也由最初的科研服务转向了商用服务。当前，针对复杂多样的网络行为，国内外许多学者进行了大量的研究与分析。国外相较于国内研究网络用户行为较早，20 世纪 90 年代就有针对网络用户行为的研究，主要是通过对各类用户的访问内容和用户行为的影响因素的分析[11]，以达到优化网络应用资源配置，提高网络系统性能的目的，更加注重实际应用。国内在 21 世纪才开始进行网络用户行为的研究，主要是从理论上探讨或从 Web 服务角度来分析用户的行为特征，基于此为用户定制个性化网络服务。

1.2.1 用户网络行为分析

移动通信与互联网的相结合而产生的新兴市场被称作移动互联网，伴随着移动互联网技术的蓬勃发展，移动终端设备已然深深的融入到我们的日常生活、工作学习、娱乐休闲以及社交活动等领域中，用户在访问移动终端设备的同时会留下大量的网络行为数据，许多学者通过分析用户的网络行为数据来挖掘用户的行为习惯，进而为用户制定个性化服务以及提升用户对网络应用的粘度更具有现实意义。目前，在互联网发展的同时，移动互联网的发展也呈现出爆发式增长，导致移动终端用户规模保持平稳增长，而且用户在使用移动终端设备的同时也产生了海量的网络行为数据，而这些海量信息中隐藏着巨大的商业价值，同时也引发了学术界和业界对移动终端用户网络行为分析的关注。

基于用户社交圈和日常生活的条件下，MontjoyeYaD 等采用用户的手机通话记录，分析用户的通话时间和活动范围，以此预测用户的性格特点[10]。王飞飞等为探究移动社交网络用户信息发布行为统计特征，以“微信”为研究对象，从发布时间间隔的角度对用户的信息发布行为规律进行研究，发现大多数用户的发布行为会在较短时间内密集发生[53]。刘丽娟通过分析天涯论坛的用户行为，发现论坛帖子热度不仅与帖子内容有关，同时也与用户的活跃程度有关，并设计出基于论坛的用户行为分析系统[40]；杨彬从统计学的角度出发，通过对 QQ 空间数据挖掘来研究和分析手机用户群，探究不同品牌的手机用户量、用户作息时间变化和用户情感状态，从这三方面发现大数据对市场营销、社会学和心理学等方面提供新的依据，也可以为政府部门的决策提供支持[59]。

基于用户地理位置的条件下，越来越多的运营商或者企业通过用户的地理位置预测用户的行为习惯，例如，刘树栋等提出了计算用户实际地理位置偏好相似度的方法，有效提高了网络服务推荐的准确性和可靠性[41]。符绕利用社交网络随时随地记录自己的位置及分享身边的事，通过用户的签到时间和次数，提出一种基于用户签到信息的潜在好友推荐系统，利用该系统向用户推荐潜在好友[25]。

1.2.2 用户网络行为预测

CNNIC 将网络用户定义为 6 周岁及以上的中国公民平均每周使用互联网至

少一小时。由于文化背景、生活背景、年龄和性别等因素差异，网络用户往往会表现出不同的行为偏好，其主要受个体用户差异、网络用户态度、网络服务需求等因素的影响，用户实时的行为习惯可以通过用户网上的浏览行为得到很好地反映，我们可以根据用户上网的历史数据去推测用户未来的行为倾向，然后更深层次的了解用户的行为偏好，有助于即时了解用户的网络服务需求而高效快速的制定个性化推荐。用户的浏览行为与用户的行为习惯密切相关，例如，关注时事政治的用户会经常去浏览一些新闻类网络应用，而偏爱游戏类的用户则会经常访问一些游戏类网络应用或游戏直播。目前针对用户网络行为的预测主要包括以下几个方面：

（1）对于用户行为习惯衰减的研究

在现实生活中，用户的行为可能随时发生变化，许多学者认为最近的行为对用户的喜好影响最大，而较长时间之前的行为对当前的喜好影响较小。Windmer 和 Ku-bat 针对兴趣动态偏好时，他们认为用户只对最近的访问行为感兴趣，而且是最能够准确反映出用户的兴趣，并利用最近一段时间的行为数据进行建模，其余时间的数据均无需考虑[2]。Crabtree 和 Soltysiak 设计的模型在不需要用户提供关键字的情况下，通过设立不同的时间因子来挖掘用户兴趣，随时间的推移，可以跟踪生成兴趣主题，并利用相似性度量可轻松识别用户兴趣[6]。Cheng 和 Qiu 等利用服从指数分布的遗忘函数，根据原始数据计算得到权重来反映用户的兴趣程度，提出了利用维基百科分类图生成用户层次兴趣模型[1]。

（2）对于用户长短期行为习惯的研究

由于受互联网上各种信息的影响，在不同的时段，用户的行为习惯会发生变化，可分为长期兴趣和短期兴趣，黄令贺等选用百度百科 6 个月近 50 万条的数据进行研究，对用户的突出兴趣和稳定兴趣进行聚类，并深入分析几种主要的用户类型，探索用户兴趣的动态变化[30]。但当用户兴趣发生偏移时，怎样将用户最近的行为偏好和过去的行为偏好结合起来，形成一个新的行为习惯，针对这种情况，宋丽哲等提出了一种基于概念相关性的用户兴趣漂移法，采用混合模型将用户兴趣分为长期模型和短期模型，能够比较准确的预测用户兴趣，继而给用户推荐感兴趣的信息[51]。Li 和 Yang 等从历史数据中了解用户兴趣变化，提出一种新的日志排名机制，分别为短期和长期的用户偏好设计独立模型，以适应用户偏

好和退化变化，从而创建个性化的 Web 视图[7]。

（3）对于用户行为预测模型的研究

线性预测模型的原理简单且易于操作，其具有易于进行数据拟合的优点，主要适用于时间序列的线性数据，一般地，普遍用到的模型包括自回归模型（AR）、自回归平均模型（ARMA）、二元分类模型、马尔科夫模型和泊松模型等。张玉成等根据用户使用网络的行为，通过分析用户行为特征、最优状态分类和用户行为迁移情况，构建基于加权马尔科夫链的用户行为预测模型，从而预测用户未来的网络行为[62]；党小超等利用用户行为状态特征值确定异常、偏正常、较正常和不正常等行为状态，以规范化的各阶自相关系数为权重，提出模糊状态的马尔科夫链预测模型[24]；胡璨利用二元分类模型预测用户是否对特定的行为主题感兴趣[31]。而对于非线性数据而言，神经网络的诞生正好将一组毫无关联的原始数据转化为一定的规律性，李旭阳等人以某商城真实的月销售量为实验数据集，选用 LSTM 神经网络提取用户的行为动态数据和利用随机森林对用户的购买行为进行预测，提出适合动态和静态的用户行为预测模型，研究证明该模型具有更好的稳定性和更高的准确率[42]。

1.2.3 个性化推荐

近年来，互联网的出现使我们步入了信息化时代，随之而来的传统网络模式被数字时代的网络服务逐渐取代，网络上的信息超载问题也随之出现，因此，如何通过用户的行为习惯为用户制定精准的个性化推荐，提高用户的满意度一直是广大学者研究的热点。1995 年 3 月，Marko Balabanovic 等人首次提出个性化推荐系统[38]，推荐系统的出现有效缓解了信息过载问题。

目前，国内外普遍用到的推荐算法有协同过滤推荐、基于内容的推荐、基于关联规则的推荐。其中，基于内容的推荐是以项目的内容信息为基础而产生推荐，主要用于对文本数据的处理；基于关联规则的推荐是基于历史行为统计不同规则之间的联系，其思想与以项目为基础的协同过滤推荐“啤酒和尿布”的关联分析相似，常用于电子商务平台的购物车分析；相较于前两种算法，协同过滤推荐算法在各个领域的普及程度最高且应用程度最广[26]，算法的实现简单有效且推荐效果精准，能够满足用户的个性化需求，而本文针对移动终端用户的网络行为分

析并进行个性化推荐，故围绕协同过滤推荐算法进行研究。

然而，传统的协同过滤推荐算法产生的推荐效果的精确度不高，其直接影响用户对网络服务的满意度，这也是国内外学者一直追求更精准的推荐结果的原因。Liangmin 等提出基于信任和情感的协同过滤推荐算法，首先采用基于显式和隐式满意度缓解数据稀疏，其次利用客观信任和主观信任建立用户之间的信任关系获取近邻用户，最后进行评分预测，该方法不仅能够提高数据稀疏情况下的推荐精度，而且针对冷启动用户其推荐精度更高[8]。He 等提出贝叶斯对偶神经网络模型补齐用户—项目评分矩阵，通过引入用户和物品的偏置项来提高评分预测的准确性[3]，Lee 利用用户评分信息熵改进现有的相似度量法，以反映用户对物品的整体评分行为，改善了协同过滤推荐算法因数据稀疏而影响预测结果[9]。王岩等将 Pearson 相关系数与用户评分差异度、用户评分倾向相结合，从而提高推荐结果的精确度[54]。

1.2.4 文献评述

根据已有的研究可知，国内外研究者从不同场景、不同角度以及利用不同方法分析用户的网络行为规律，并设计了合理的网络用户行为预测模型，基于此为用户制定个性化的网络服务，那么用户网络行为分析与个性化推荐效果的重要性显而易见。一方面，对于用户行为分析，学者们只从个人用户角度来研究用户的社交圈及生活情况，或者从各类型的网络应用了解用户的行为特点，能够为产品开发和运营提供更加有力的支持，对网络运营商和内容提供商是非常必要的，然而得到的用户行为特征与实际情况有偏差，由于移动终端用户在网络中的行为具有较高的复杂性和较大的不确定性，现有的用户行为分析一般只针对某一个特定场景而忽略了多角度的网络行为特征，这可能直接忽略了用户的个性化而只关注用户的共同性；另一方面，在现实生活中，由于受时间或者外部环境的变化等因素的影响，用户的网络行为可能会发生变化，而且许多学者根据不同场景的用户行为数据构建了合理的网络行为预测模型，通过预测用户的行为习惯并进行个性化推荐，但其个性化推荐效果的精确度不太理想。

因此，通过对国内外研究现状的综述与分析，可将本文的研究思路总结为：首先将时间因素引入网络用户行为分析的体系中，从移动终端用户访问 APP 的

时间和内容出发，基于不同情景下深入挖掘移动终端用户的网络行为模式；然后针对两个场景的分析结果，对其进行关联分析，发现移动终端用户行为模式之间的关联；最后针对移动终端用户的历史行为习惯，为了得到更精准的推荐结果，本文立足于 TF-IDF 思想和 BP 神经网络理论，引入协同过滤推荐系统，搭建基于 BP 神经网络的移动终端用户网络行为评分预测模型，挖掘出移动终端用户更精准的个性化信息，进而实现移动终端用户的个性化推荐。

1.3 研究内容及可能创新点

1.3.1 研究内容

本文针对移动终端用户的上网时间和访问内容挖掘用户的网络行为特征与提高个性化推荐的精度研究，主要研究内容包括两部分：一是移动终端用户网络行为分析，二是移动终端网络服务个性化推荐。

（1）移动终端用户网络行为分析

用户浏览的网页蕴含了与用户相关的信息，包括浏览网页的时间戳和网络地址等，在现实生活中，网络运营商通过用户的浏览记录对网络行为进行分类，以此分析移动终端用户的行为特点。

首先，通过阅读大量的国内外文献可知，学术界对用户的上网时间行为通过设置阈值进行分析，换而言之，在某一上网时间之内，如果用户访问网络的时长超过了规定的阈值，则被界定为在此期间的用户偏好上网，该处理方法大部分取决于人的主观意识，又因用户的生活习惯不同，继而访问网络的时间规律也有所不同，故本文结合用户上网时间的相似性规律，采用小时作为划分标准，将用户的上网时段划分为 24 个时段，并且通过层次聚类法分析得到用户上网时段的 4 种喜好。

其次，在研究用户访问内容的喜好时，为了便于分析，本文结合移动互联网对 APP 的分类方式，通过提取每个用户上网产生的 URL 的关键字，将提取出的关键字进行分类，即对用户访问的网络应用分类，分别为社交通信、网络视频、网络购物、网络音乐、游戏、摄影、新闻资讯等共 18 个类别，再利用 k-means 聚类算法，以用户累计的访问时长为度量标准，对每个用户访问 18 个类别的网

络应用访问时长所占的百分比进行分类，将用户划分为 7 个不同的簇，并分析各个用户类型的特征，即用户访问内容的喜好。

最后，通过利用数据挖掘技术得到移动终端用户上网时间的行为习惯与访问内容的行为习惯，再利用关联法分析这两种行为模式的关联性，得到用户的网络行为习惯，继而为运营商或广告商提供在某网络应用的某时间段进行重点营销的战略支撑，有助于降低企业的投入成本和提高营销效率。

（2）移动终端网络服务个性化推荐

在互联网信息化时代，移动终端用户所接受的信息大多是普遍性和无差别的，为了使移动终端用户享受到制定的个性化网络服务，用户的网络行为预测显得格外重要，通过预测获得用户更精准的网络行为偏好，基于此运用个性化推荐系统为用户制定个性化的网络需求。

首先，为了提高个性化推荐的准确度，根据移动终端用户的网络行为模式，经由 TF−IDF 算法计算用户访问 APP 的累计使用时长来反映用户对该网络应用的喜爱程度，在此基础上构建用户—类别评分矩阵，其次，按照 4:1 的比例划分训练集和测试集，并利用余弦相似度为目标用户筛选近邻用户集，最后利用 BP 神经网络构建评分预测模型，基于预测结果，结合协同过滤推荐算法评估个性化推荐效果。

1.3.2 可能创新点

本文主要围绕移动终端用户的网络行为展开研究，研究内容包括网络行为分析和网络服务个性化推荐，与国内外现有的研究工作相比，本文可能的创新点有：一方面，本文将时间特征引入移动终端用户网络行为分析的体系之中，从移动终端用户访问 APP 的时间戳和内容出发，基于不同情景深入挖掘移动终端用户的网络行为偏好模式；另一方面，以往研究网络用户的行为往往集中于某一个或某一类网络应用的行为数据，本文通过分析移动终端用户访问 APP 的行为偏好模式为用户定制差异化服务。

1.4 论文组织结构

为了提高移动终端个性化网络信息的推荐精度，本文结合协同过滤推荐算

法，利用 BP 神经网络搭建移动终端用户关于网络行为类别的评分预测模型，全文共分为五章，相关章节内容安排如下：

第 1 章是绪论。主要阐述了本文的研究背景及意义，针对本文的研究意义分别阐述了移动终端用户网络分析、网络行为预测以及个性化推荐的相关研究现状，最后阐述了本文的主要研究内容及论文组织结构。

第 2 章是相关理论基础。主要介绍了网络用户、网络行为特点及网络行为分类相关的基础定义，然后对本文中所使用的研究方法做了详细介绍，包括层次聚类和 k-means 聚类算法，最后介绍了 BP 神经网络和协同过滤推荐算法。

第 3 章是移动终端用户网络行为分析。首先分析移动终端用户上网时间的行为习惯，根据用户生活习惯的不同，本章首先将用户上网时段以每小时为间隔划分为 24 个时间段，然后按照用户上网时间的近相似性规律对用户进行初始分组，最后采用层次聚类法将上网时段相似的归为一类，进而得到 4 种用户上网时段的喜好模式。其次分析移动终端用户访问内容的行为习惯，结合移动互联网报告中对 App 的分类方式，通过提取每个用户上网产生的 URL 的关键词，将提取出的关键词进行分类，将用户的访问内容进行分类，共划分为 18 个类别，再利用 k−means 聚类算法，以访问时长之和为度量标准，统计每一位用户在访问 18 种不同类型的网络应用时上所花费时长的比例，将用户划分为 7 个不同类型，并分析各个类型的用户特征，即用户访问内容的行为模式特征；最后，采用关联法分析这两种行为之间存在的相关性，得到用户的网络行为规律。

第 4 章是移动终端网络服务的个性化推荐。为给用户提供更精准的个性化网络信息，根据上一章分析得到的移动终端用户的网络行为综合模式，本章利用 TF−IDF 算法得到用户对各个网络应用类别的权重，并利用 BP 神经网络优化协同过滤推荐算法的评分预测值，从而实现更精准的个性化推荐。

第 5 章是总结与展望。总结本文所作的主要工作，并指出当前研究的不足以及未来需要进一步做的工作。

2 用户网络行为分析理论与方法

2.1 网络用户界定

近几年，互联网和移动通信技术的跨越式发展，使得电子及智能设备在各个方面影响着我们的生活和工作方式，使用互联网的用户规模正在不断壮大，伴随着网络用户的行为越来越复杂，用户的需求也越来越多样化，那么什么是网络用户？顾名思义，网络用户既是网络信息的创造者也是使用者，但是从不同的角度出发，网络用户的定义并未统一，有学者表示“网络用户是指在各项实践活动中利用互联网获取和交流信息的个人”[55]，也有学者提出“网络用户就是在一定条件下，一段时间内正在利用网络获得信息的个人和团体”[32]，还有学者认为“网络用户是指在科研、教学、生产、管理、生活以及其他活动中需要和利用网络信息的个人和群体”[43]。网络用户的内涵是从以上三个视角来界定的，而在本文的研究中，可以从用户与网络服务的关系、用户与网络的行为模式两个层面界定其内涵。

（1）网络服务是网络用户的提供者

1994 年我国正式加入国际互联网（WTO），数字网络技术的快速发展和广泛应用，不仅我国网络用户的数量呈稳定式增长，网络基础设施也逐步实现全面覆盖，还产生了许多互联网服务业，互联网在为网络用户获取便利的同时，用户的网络需求也在不断提高。就网络运营商而言，其具备网络用户和网络信息提供者的双重身份，因为时间和精力的有限性，用户倾向于将更多的时间花费在有价值的信息上，而对垃圾信息一扫而过，故运营商需要不断的提升网络信息质量，只有不断地找到新需求，通过优化信息加强用户忠诚度甚至吸引新用户；就用户而言，通过接受网络服务，获取网络信息并加以推广和利用，继续扩大用户群，进一步促进了互联网的发展。

（2）网络用户的行为方式决定用户分类

网络用户通常在自我意识、网络使用态度、网络信息偏好上表现出一定的特点，他们是由互联网行为组成的虚拟群体，不管用户出于何种目的访问网络，无论是娱乐休闲还是教育科研，他们都是为了获取网络信息，为自己的学习、工作或生活提供服务。尽管网络用户是特殊用户群体且都有自己特定的行为特征，但

更多的是具有趋同性，互联网实验室根据用户的网络行为，归纳总结当今网络用户类别，即：纯信息网民、纯沟通网民、基础网民、泛娱乐网民、典型娱乐网民、信息娱乐网民、泛娱乐网民的特征、网络工作网民、次全能网民、全能网民[33]共十种，网络用户的分类能更好的挖掘用户行为偏好模式。

由以上分析可知，网络用户就是通过耗费精力和时间获取网络上有价值的内容与创造个性化服务的个体，从个体的行为特征推测具有相似行为特征的一类用户群体的网络行为习惯。

2.2 用户网络行为分析

网络行为的研究是一门综合性学科，它与数据挖掘、计算机网络、心理学等所有与网络行为相关的学科密切相关，主要目的是研究用户的网络行为规律，即运用多学科研究用户行为的特点、用户构成及其行为表现出来的规律。

2.2.1 用户网络行为界定

互联网的出现使网络行为分析应运而生，而其网络行为具有虚拟性的特点，是用户为了实现某种目的，借助互联网平台进行有意识的活动，因此，网络行为可以定义为：网络信息时代的到来，使得互联网为人们创建了新的活动场所，可以通过终端设备主动或被动获得信息的行为即为网络行为，例如在线购物、QQ聊天、即时通信、网络游戏、在线影音等都属于网络行为。那么，用户网络行为分析是通过统计与分析用户访问互联网时产生的路径轨迹，挖掘用户在网络应用过程中反映的行为规律，借此规律控制并预测用户的未来行为偏好，企业或网络服务商可以提供更精准的网络信息及产品。由此可见，将网络行为分析定义为收集用户访问网络应用的行为数据，利用数据挖掘技术统计并分析相关数据，挖掘出用户在网络平台上表现出来的规律，借以控制并预测网络用户的未来行为，有助于运营商发现用户潜在的网络需求，从而提升用户对网络应用的粘性及信任度。

依据研究目的与行为主体差异，网络行为包括个体用户网络行为与群体用户网络行为。个体用户网络行为[23]是个体用户在互联网上的行为，由于每个人都有自己独特的个性，导致个体在网络平台上的行为具有多样性，结合用户的特性

建立行为偏好模型能够更好的体现用户网络行为的变化趋势，进而为用户提供差异化的服务需求。群体用户是两个或多个相互影响、交互依赖的个体集合，其群体用户的网络行为是将个体用户在网络活动中相似且互动的网络行为叠加为一个整体，确切的说是用来研究多个个体组成的某类网络群体的网络行为习惯。

2.2.2 用户网络行为特点

相对于现实生活而言，网络是一个虚拟的领域，我们可以通过互联网来满足某种需要而与计算机或移动终端进行的交互式访问行为，如在线聊天、网络购物、娱乐社交、上传和下载等，故网络行为具备其特殊性，但是用户网络行为具有异于其他现实行为的特点，主要表现为以下几个方面[23]：

（1）主动性强。由于网络的开放性、自由性和平等性，继而使得网络行为完全不受用户的性别、年龄、文化背景、工作学习及生活环境的限制，可以完全自由地按照自己的意愿选择网络服务。

（2）性质复杂多样。由于互联网提供的是一个虚拟场所，该场所具备复杂且虚拟性，使得用户在不同于物理空间的场所中发生网络行为，用户之间也存在差异性，因此，网络行为的性质具备复杂且多样性。

（3）判断标准不同。互联网连通为越境数据流创建了稳固载体，用户在虚拟网络空间中发生的行为不受区域性的限制，不同的网络区域就有不同的判断标准。

（4）高度隐匿性。一方面，用户在互联网平台上的主体身份大多是隐匿的，这也导致发生的网络行为存在匿名性，换而言之，任何用户利用终端设备获得网络信息的过程不需要用户的真实身份。另一方面，网络行为能主动隐匿行为特征，由于网络信息以数字化形式出现在互联网上，用户在获得信息的过程中可以不留痕迹的改变其内容或者形式。

2.2.3 用户网络行为分类

由于用户的网络行为发生在虚拟空间中，目前对这一行为还没有统一的标准，对网络行为的分类形式同样也未统一，则其行为分类具有模糊性与多样性，那么根据不同的研究角度和需要解决的具体问题进行网络行为的分类，主要分为

以下几类：

（1）根据对象数目分类

由于研究目的与行为主体的差异，将用户网络行为细分为个体网络行为与群体网络行为，即用户的数目不同，所表现出的网络行为模式也大不一样。个体用户由于受其生活背景、经济地位或教育背景等影响，网络行为迥异，所体现出的性格、习惯、心理变化、价值观及兴趣爱好千差万别，而由多个个体网络行为的共性组合起来反映了群体行为特征。

（2）根据应用角度分类

互联网平台所提供的网络服务种类繁多，因而互联网实验室从用户的角度出发，对当前国内用户的网络行为进行总结分析。目前，从应用角度划分的用户网络行为包括基础网络服务和扩展网络服务，其中，交流沟通类和信息获取类即为基础网络服务，而网络娱乐休闲类和电子商务服务类即为扩展网络服务。进一步细分用户网络行为，交流沟通类即为网上聊天，包括微信、QQ、电子邮件、网络电话等；信息获取类服务是由于网络的共享性和开放性而产生的，包括搜索新闻、网络杂志、网上办公等；网络娱乐休闲类是依托互联网而进行的日常生活项目，例如网络游戏、摄影、在线音乐、网络视频等；而电子商务服务类是利用网络的便捷性获取有形和无形的网络服务，包括网络购物、网上支付、旅游服务、网络教育、医疗服务等。在当前的一段时间内，无论用户的网络行为是现有的网络服务还是未来新生的网络服务，都可以按照它们的特征归类到这四种类型中，进而研究网络用户的行为偏好模式，为用户制定差异化服务。

（3）根据网络类型分类

网络是以相互交流信息资源为目的，可以将其划分为移动网络用户行为和互联网用户行为。由于终端的特性、网络宽带、网络质量、网络资费和终端应用的差异，使得手机和互联网用户表现出了各自的特性。移动网络行为是用户使用电信网而诞生的行为，主要表现为短信、语音、电话等基础应用，而互联网行为是用户使用互联网而产生的行为，主要表现为文件传输、即时通信等应用。

从以往学者们研究的文献中发现，研究者们多从应用角度出发分析用户的网络行为习惯，而且当前个性化的网络需求已成为一种趋势，故本文以移动终端用户为基础，从网络行为分类、行为预测以及个性化网络服务三个方面出发，为移

动智能终端个性化服务提供支撑。

2.2.4 用户网络行为表示

由于用户的网络信息需求千差万别，使得用户行为的表示方式也是多种多样。Humberto、Marques 等利用统计图的方式表示用户行为特征[4]，白友东用向量来表示用户的网络行为特征[20]，从以往学者们研究的文献中发现，大多利用向量来表示网络行为，很显然向量的方式更易于表达，故本文利用向量来描述用户的行为特征，n 个属性的用户行为可以表示为：[属性 1，属性 2，…，属性 n]，其中 n 个属性分别为此行为中 n 个特征，以用户上网时段喜好的分析为例，用户的时间分布向量可表示为：[IMEI，访问网络应用时长 1，访问网络应用时长 2，…，访问网络应用时长 n]，则本文选取的特征指标如下：

（1）访问网络应用的类型

为了定义用户访问网络应用的类别，本文结合移动互联网对 App 的分类方式，通过提取每个用户上网产生的 URL 关键词并进行分类，如网络视频类、游戏类、系统工具类、新闻类等网络应用，通过分析用户访问各种类型的网络应用的时长，可以在一定程度上发现移动终端用户日常访问网络时的行为习惯。

（2）访问网络应用的时长

从用户开始访问某网络应用到退出该网络应用时所持续的时间，即被称作访问网络应用的时长，其停留时间的长短反映了用户对何种网络信息更感兴趣，何种网络应用更具有吸引力，使用数据挖掘技术分析用户访问网络应用的时长，从而挖掘出用户访问网络时所表现出来的行为偏好，为网络运营商进行精准广告投放和网络应用设计提供策略支持。

（3）用户兴趣度

由于网络行为属于用户对互联网的隐性反馈行为，无法明确反映用户对某类网络服务的喜好程度，故通过用户访问网络应用的时长量化喜好程度，继而引入用户兴趣度。用户兴趣度是用户针对网络服务量化出的权重，它代表用户对某类网络应用的喜好程度，一般情况下，不同用户对其网络服务的喜好程度有所不同，故利用[0, 1]之间的数值量化用户兴趣度，0 表示用户对此类网络服务不感兴趣，1 表示用户非常喜欢此类网络服务，本文选择用户访问某类型网络应用的使用时

长和网络应用标签来估计用户兴趣度，准确估计用户兴趣度既能有效的反映用户的行为趋向，也能决定运营商的个性化服务质量。

2.3 用户网络行为分析方法

2.3.1 聚类分析

聚类分析是最常用的数据挖掘手段，是一种无监督多变量分类的统计算法。它不需要事先确定分类规则，通常按照相似度将数据集划分为不同“簇”，“簇”是指相似对象的集合，经过聚类后发现数据集中存在的规律，即同簇间的对象最相似，异簇间的对象较相异。通常情况下，聚类算法包括基于划分的聚类、层次聚类和基于密度的聚类。

（1）基于划分的聚类。基于划分的聚类有别于其他聚类算法，该算法需要事先确定研究对象的聚类簇和初始聚类中心，通过反复迭代确定最佳聚类簇及初始聚类中心，最终划分为互不相交的簇，每个对象只属于一个簇，每个簇至少包含一个数据对象。其中，最常用到的算法是 k-means 聚类算法，该算法原理简单且易于实现，其缺点是不同的初始聚类中心可能会产生不同的聚类结果。

（2）层次聚类。层次聚类是基于簇间的相似度在不同层次上分析数据，从而形成树形的聚类结构，它不需要事先设定聚类数目，Tan 等认为层次聚类能够产生较高质量的聚类[14]。

（3）基于密度的聚类。通常前两种聚类算法只能适用于凸形的簇，针对其他形状的簇，则可使用基于密度的聚类，它是利用数据集之间的密度决定聚类簇，即将密度较大的区域相连接，DBSCAN 是其中最典型的算法。

目前，用户的网络行为分析方面普遍使用基于划分和基于层次的聚类算法，故本文选用层次聚类和 k-means 聚类算法对移动终端用户的网络行为进行挖掘，得到用户普遍的网络行为特征。

2.3.2BP 神经网络

BP(back propagation)神经网络由 Rumelhart 和 McClelland 等在 1986 年提出，是一种按照误差反向传播算法的前馈型监督学习的神经网络[48]，是目前应用最

广泛的神经网络。

输入层、隐含层和输出层构成的多层感知器——BP 神经网络，其中隐含层可以有一个或多个，但一般情况下采用三层网络结构。该算法的主要思想是以网络误差平方为目标函数，通过利用最快梯度下降法的学习规则进行学习和训练，以此确定输入输出之间的多维函数映射关系，以期使网络的实际输出值越来越逼近期望值。BP 神经网络的应用需要对网络进行训练，使其具备联想和预测的能力，分为正向传递和反向传递两个过程，其训练过程如下：

（1）网络初始化。确定输入层、隐含层以及输出层的神经元个数，输入层与隐含层之间的权重矩阵为$w_{ik}=(\vec{w}_1,\vec{w}_2,\cdots,\vec{w}_m)$，隐含层与输出层之间的权重为$v_{ho}=(\vec{v}_1,\vec{v}_2,\cdots,\vec{v}_m)$，输入层向量为$u=(\vec{u}_1,\vec{u}_2,\cdots,\vec{u}_m)$ $u=(\vec{u}_1,\overrightarrow{u_2},\cdots,\overrightarrow{u_m})$，实际输出层向量为$y=(\overrightarrow{y_1},\overrightarrow{y_2},\cdots,\overrightarrow{y_m})$，期望输出向量为$d=(\overrightarrow{d_1},\overrightarrow{d_2},\cdots\overrightarrow{d_m})$。为了控制更新的幅度，设置学习速率$\eta$。隐含层与输入层的每个神经元都分为输入和输出，且每个神经元都需要一个激活函数，而本文的兴趣度数值范围在[0, 1]之间，故激活函数一般取 sigmoid 函数，其形式为$f(x)=\frac{1}{1+e^{-x}}$。

（2）计算隐含层的输出。隐含层神经元k的输入h_{ik}通过权重w与输入层的输入u进行加权求和，其公式如（2-1）所示：

$$h_{ik}=\sum_{i=1}^{m}w_{ik}u \tag{2-1}$$

在得到神经元k的输入值之后，其神经元k的输出通过激活函数激活所得，因此神经元k的输出h_{ok}的计算如公式（2-2）所示：

$$h_{ok}=f_k(h_{ik})=\frac{1}{1+e^{-h_{ik}}} \tag{2-2}$$

（3）计算输出层的输出。在回归预测中，BP 神经网络通常设置一个神经元的输出层，再利用与（2）相同的方法计算输出层神经元的输入y_i和输出y_o，其公式如（2-3）和（2-4）所示：

$$y_i=\sum_{t=1}^{p}v_{ho}h_{ok} \tag{2-3}$$

$$y_o=f_y(y_i)=\frac{1}{1+e^{-y_i}} \tag{2-4}$$

（4）计算误差e。根据 BP 神经网络模型得到的预测值为y_o，而该输入值对

应的输出值为d，那么 BP 神经网络训练目标函数的精度如公式（2-5）计算得到：

$$e=\frac{1}{2}(y_o-d)^2 \tag{2-5}$$

（5）权值更新。步骤（1）至步骤（4）为神经网络的正向传递阶段，而权值的更新则为神经网络的反向传递阶段。其反向传递的过程为：利用梯度下降法修正输出层与隐含层、隐含层与输入层之间的权重，其公式如（2-6）和（2-7）所示：

$$w_{ik}=w_{ik}-\eta\frac{\partial e}{\partial w_{ik}} \tag{2-6}$$

$$v_{ho}=v_{ho}-\eta\frac{\partial e}{\partial v_{ho}} \tag{2-7}$$

其中，$\frac{\partial e}{\partial w_{ik}}$和$\frac{\partial e}{\partial v_{ho}}$的计算如式（2-8）和（2-9）所示：

$$\frac{\partial e}{\partial w_{ik}}=\frac{\partial e}{\partial y_o}\frac{\partial y_o}{\partial y_i}\frac{\partial y_i}{\partial h_{ok}}\frac{\partial h_{ok}}{\partial h_{ik}}\frac{\partial h_{ik}}{\partial w_{ik}} \tag{2-8}$$

$$\frac{\partial e}{\partial v_{ho}}=\frac{\partial e}{\partial y_o}\frac{\partial y_o}{\partial y_i}\frac{\partial y_i}{\partial v_{ho}} \tag{2-9}$$

（6）判断误差e是否达到预先设定的精度要求。通过步骤（5）得到的权重参数重复循环正向传递阶段得到神经网络的预测值，得到误差e小于预先设定的精度值，若达到要求，则训练结束，否则返回步骤（2）再重复学习直到误差e达到要求，即结束学习。

2.3.3 协同过滤推荐算法

信息化时代的到来，信息技术不断发展，用户很难从海量的信息中获取需求信息，个性化推荐系统的出现解决了信息过载的困扰，帮助用户快速的找到有效信息，而对于收集用户的个性化信息，协同过滤推荐算法具有更显著的优越性，它是当前推荐系统中被广泛应用且技术最成熟的推荐算法。

协同过滤推荐算法的核心思想是从用户的历史行为中挖掘网络行为习惯，并基于此进行个性化推荐。一般地，最常用的协同过滤推荐算法是基于用户的协同过滤推荐（UBCF）和基于项目的协同过滤推荐（IBCF），其中 UBCF 利用一定

的规则寻找与目标用户对网络信息或物品的喜好具有相似偏好的“邻居”，基于“邻居”的行为喜好程度评估目标用户对某一未知行为的喜好度；而 IBCF 基于用户对网络信息或项目的偏好，通过项目与项目之间的相似性，为目标用户提供个性化推荐服务。这两种算法的区别在于：第一，两者的分析对象不同，前者是寻找相似用户，后者是寻找相似项目；第二，推荐效果不同，前者根据相邻用户生成推荐，后者根据相邻项目生成推荐，但前者的推荐效果更符合实际。

3 移动终端用户网络行为分析

随着互联网的兴起，我们步入了一个信息化的时代，传统的网络模式已经逐渐被数字化的网络服务所代替，且为用户提供的网络信息愈来愈多，这使得我们在获取有效信息时变得更加困难。因此，新时代下的网络模式转变使得用户对网络服务的质量不断提升，运营商挖掘用户的网络行为规律应由盲目性发展走向科学化，以及传统的网络服务模式从被动向主动的个性化服务转变，那么，个性化推荐服务就需要挖掘出与实际相符的用户网络行为模式，故本章从用户访问网络的时间和网络内容两个角度分析用户的网络行为习惯，从而为用户制定个性化服务以及运营商调整服务需求策略提供了双重便利，更有助于缓解信息过载与获取有效信息困难的问题。

3.1 数据介绍

3.1.1 数据描述

本文所使用数据来源于 Y 公司，该司数据处理系统完备，能够收集到用户访问移动终端的数据，相关字段包括设备 IMEI（IMEI）、开始访问时间（start_time）、结束访问时间（end_time）、统一资源定位符（URL），此外，为了保护用户的隐私，设备 IMEI 用数字编号 1，2，…代表，移动终端用户上网行为数据如表 3.1 所示。

表 3.1 移动终端用户网络行为相关属性

属性名称	具体描述
IMEI	识别移动设备的唯一标识
start_time	用户开始访问网络应用的时间
end_time	用户结束访问网络应用的时间
URL	统一资源定位符

3.1.2 数据预处理

1、上网时段

由于移动终端用户的生活习惯和工作性质不同，用户的上网时间也因此发生变化，通过分析用户上网时间的行为习惯，可以发现用户在什么时段喜爱访问何种网络，有助于运营商或者网络服务提供商发现潜在的目标用户，以此制定个性化的网络服务来提高市场营销的精确度。另外，在科研领域中，大多数研究都是通过设定阈值的方式挖掘用户上网时间的行为模式，换而言之，在某一上网时间内，如果用户访问网络的时长超过了规定的阈值，则被界定为在此期间的用户偏好上网，反之亦然。虽然这种方法操作简单，但这种设定阈值的方式往往大部分取决于个人的主观意识，缺乏基础理论的支撑，无法确保分析结果的精准度，又因用户的生活习惯不同，继而访问网络的时间规律也有所不同，因此，本文分析移动终端用户的上网时间特征时，将上网时间划分为 24 时段的向量形式，该方法相对于阈值法得到的结果较合理。

在本章中引入了基于小时的时间划分方法，即把每天的 24 个小时划分为 24 组，以每小时为间隔记作一个分组，则时间段为 0:00～1:00，1:00～2:00，…，23:00～24:00，并将各个时段的名称标记为$T_0,T_1,\cdots,T_{23}$，用户上网时段的划分如表 3.2 所示。

表 3.2 划分移动终端用户的上网时段

时段	时段名称	时段	时段名称	时段	时段名称
0:00-1:00	T_0	8:00-9:00	T_8	16:00-17:00	T_{16}
1:00-2:00	T_1	9:00-10:00	T_9	17:00-18:00	T_{17}
2:00-3:00	T_2	10:00-11:00	T_{10}	18:00-19:00	T_{18}
3:00-4:00	T_3	11:00-12:00	T_{11}	19:00-20:00	T_{19}
4:00-5:00	T_4	12:00-13:00	T_{12}	20:00-21:00	T_{20}
5:00-6:00	T_5	13:00-14:00	T_{13}	21:00-22:00	T_{21}
6:00-7:00	T_6	14:00-15:00	T_{14}	22:00-23:00	T_{22}
7:00-8:00	T_7	15:00-16:00	T_{15}	23:00-24:00	T_{23}

将移动终端用户上网时间划分为 24 个时段，以此统计用户 28 天内在不同时段所花费的时间总量，根据计算加工得到一个 24 维的用户上网时段的向量矩阵，具体的计算方式为：

首先根据移动终端用户上网时间的行为数据，计算在 24 个时段上用户访问不同网络应用的时长，其次统计持续一段时间内的累计上网时长，最终转化为百分比的形式，基于此类行为数据挖掘用户上网时间的行为习惯，得到的时间分布向量构成 24 维的用户上网时间的向量矩阵，用公式(3−1)表示。

$$U=[V_{T_0},V_{T_1}\cdots,V_{T_m},\cdots V_{T_{23}}] \tag{3-1}$$

其中，V_{T_m}表示用户在T_m时段的平均上网时长百分比。由于数据比较多，只列举了部分用户的数据处理结果，如表 3.3 所示。

表 3.3 移动终端用户在各个时段的上网时长比例

用户编号	T_0	…	T_{18}	T_{19}	T_{20}	T_{21}	T_{22}	T_{23}
1	0	…	0.09098	0.12447	0.11815	0.12149	0.05307	0.03285
2	0.01288	…	0.07232	0.07515	0.05666	0.04433	0.04958	0.02708
3	0	…	0.08872	0.11954	0.13884	0.11122	0.14427	0.03103
4	0.07683	…	0.03266	0.06540	0.09006	0.08580	0.07773	0.09852
5	0.01139	…	0.05814	0.07114	0.13323	0.14122	0.10676	0.03283
6	0	…	0.08075	0.01469	0.05349	0.08121	0.08666	0.07336
7	0.01191	…	0.05255	0.07436	0.09232	0.09617	0.10846	0.04861
8	0	…	0.08842	0.12138	0.11233	0.10781	0.13710	0.07851
9	0	…	0.03425	0.07723	0.10012	0.02770	0.07872	0.06369
10	0	…	0.06771	0.07390	0.07817	0.08243	0.07198	0.03969

2、访问内容

研究移动终端用户访问内容的行为特征，我们需要利用用户的上网信息 URL 来推测用户的行为习惯，首先我们利用关键词将用户的访问内容划分为不同类别，其次利用用户的访问时长来对用户的行为进行聚类，最后根据聚类结果具体分析用户的行为模式。

由于用户的访问记录具有多样性，且 URL 比较复杂，为了便于处理数据我

们需要截取 URL 信息中的关键词，进而将提取出的关键词进行统计与分类，例如某用户访问记录中的 URL 为：https://mail. qq. com/、https://news. qq. com/，我们从中提取出的第一个链接的关键词为“mail. qq. com”、第二个链接的关键词为“news. qq. com”，通过搜索查找资料发现：第一个关键词代表 QQ 邮箱，第二个关键词代表腾讯新闻。为了便于了解用户对不同类别的应用软件是否感兴趣，我们需要对关键词进行分类，换而言之，我们不必知道用户是否使用爱奇艺或其他网络应用观看视频，只需要知道他们是否喜欢观看视频。通过提取 URL 中的关键词，结合移动互联网报告中对应用软件现有的分类方式，将用户访问的网络应用划分为 18 个类别，得到的网络应用分类如表 3.4 所示。

表 3.4 网络应用类型分类

编号	应用类别	应用软件代表
S_1	娱乐休闲	快手、抖音、小红书、火山视频、爱奇艺、腾讯、优酷、韩剧 TV
S_2	新闻资讯	搜狐新闻、腾讯新闻、今日头条、新浪新闻、网易新闻
S_3	社交通信	微信、QQ、新浪微博、百度贴吧、绿洲、QQ 空间
S_4	网络购物	淘宝、京东、蘑菇街、拼多多、闲鱼、得物、考拉海购、天猫
S_5	摄影摄像	秒拍、美拍、水印相机、美颜相机、美图秀秀
S_6	游戏	腾讯游戏、网易游戏
S_7	金融理财	支付宝、各类银行 APP
S_8	交通出行	高德地图、百度地图、携程网、去哪儿旅行、滴滴打车
S_9	居家生活	58 同城、前程无忧、boss 直聘、赶集网、墨迹天气、天气预报
S_{10}	浏览搜索	搜狗浏览器、百度搜索、谷歌搜索、知乎
S_{11}	效率办公	电子邮箱、印象笔记、WPS Office
S_{12}	图书阅读	QQ 阅读、书旗小说、七猫小说、番茄小说、喜马拉雅
S_{13}	教育学习	扇贝、网易有道词典、百度翻译、百度网盘、作业帮、粉笔职教
S_{14}	快餐服务	美团、饿了么、大众点评
S_{15}	医疗健康	平安健康、京东健康、薄荷健康、小豆苗、蜗牛睡眠
S_{16}	体育运动	Keep、直播吧、虎扑、腾讯体育
S_{17}	汽车服务	瓜子二手车、易车网、汽车之家
S_{18}	系统工具	手机管家、主题、时钟日历、应用商店

根据用户访问移动终端的数据，将其划分为 18 个不同的网络应用类型，并处理用户使用相同类型网络应用的访问时长，我们将用户的上网行为偏好定义为长度为 18 的向量，根据计算加工便得到一个 18 维的用户上网行为的向量矩阵，如式（3-2）所示：

$$S=[S_1,S_2,\cdots S_m,\cdots S_{18}] \tag{3-2}$$

其中，S_m表示用户访问某类型网络应用的平均访问时长，由于数据比较多，只列举了数据处理的部分结果，如表 3.5 所示。

表 3.5 移动终端用户访问网络应用的时长

用户编号	S_1	S_2	S_3	S_4	S_5	S_6	…	S_{18}
1	0.32849	0.09853	0.29590	0.00764	0	0.14671	…	0
2	0.09519	0.25720	0.34580	0.03832	0.00245	0.03878	…	0.00468
3	0.31127	0.09632	0.28124	0.01477	0.00665	0.14375	…	0.00186
4	0.33579	0.11723	0.24543	0.03491	0.00473	0.13871	…	0.00520
5	0.15887	0.21486	0.19503	0.01642	0.00883	0.13808	…	0.00740
6	0.30095	0.09028	0.26429	0.03783	0.00570	0	…	0.00935
7	0.03865	0.04431	0.10062	0.02965	0.00275	0.01416	…	0.00058
8	0.10586	0.07992	0.15819	0.00719	0.00079	0	…	0.00431
9	0.09580	0.26484	0.35657	0.03701	0	0.03749	…	0.00199
10	0.00765	0.10892	0.37064	0.02565	0.00683	0.38339	…	0.00045

3.2 用户上网时间近相似性

3.2.1 近相似性定义

用户使用网络终端时所留下的时间信息，与用户群体的日常生活作息特征密切相关，根据用户的网络行为数据分布特征挖掘用户上网时间的行为模式规律，从而发现用户之间的相似性。

用户上网时间的近相似性代表两个用户使用网络终端时最大上网时长对应的上网时段相同，其上网时间分布向量越相似[18]，而最大上网时段表示移动终

端用户的上网时间向量的最大值所对应的访问网络的时段，为了衡量用户上网时间的相似性，本文的用户上网时间是用向量表示的，故选用欧氏距离测度用户间的相似程度。由于 Yan Hao 等人是基于城域网用户行为数据发现在最大上网时段附近存在相似性的规律，该类型数据属于非线性数据，是持续一段时间内发生的网络行为，具有很强的规律性[18]，而本文是基于移动终端用户的网络行为数据，与上述文献中宽带数据的特征比较相似，因此，基于移动终端上网时间行为数据的特殊性和复杂性，充分利用上网时间的近相似性规律，挖掘移动终端用户上网时间的网络行为习惯。

3.2.2 近相似性规律验证

为了验证用户上网时间的近相似性规律在移动终端用户行为数据中的可行性，且由于所研究的移动终端用户较多，采用随机抽样技术，能够确保总体中每一个样本取样的可能性都一样，而且还保证了所抽取的样本具有代表性，故随机抽取了 10%的样本，利用欧式距离测度用户之间的相似性，而所涉及到的一些基础理论知识如下[18]：

（1）用户i与用户j的相似度：利用用户i与用户j上网时间分布向量的欧式距离表示用户的相似度，其欧式距离愈短，两用户的上网时间愈接近。

（2）用户i与分组j的相似度：计算用户与分组间的相似度，那么只需要计算用户i与分组j内所有用户的欧氏距离，基于此得到用户与分组的相似度，如果用户存在于分组内，只需要计算他们之间的欧式距离的平均值。

（3）分组与分组之间的距离：由于移动终端用户的上网时间特征是按照时间段划分的，那么计算分组与分组的距离时，只需要计算与分组相对应的时段的距离。此外，为了便于计算，本文采用坐标反向计算分组间隔，例如，时段T_2与T_{23}的分组对应的分组距离是−3 而不是 21。

（4）用户i与分组j的距离：用户i所对应的分组与分组j的间隔距离代表了用户i与分组j的距离，用D表示。

（5）用户与用户之间的距离为两用户所属分组间的距离，用d表示。

利用随机抽取的样本数据，计算用户与用户、用户与分组、分组与分组之间的相似度，其结果如图 3.1 所示。

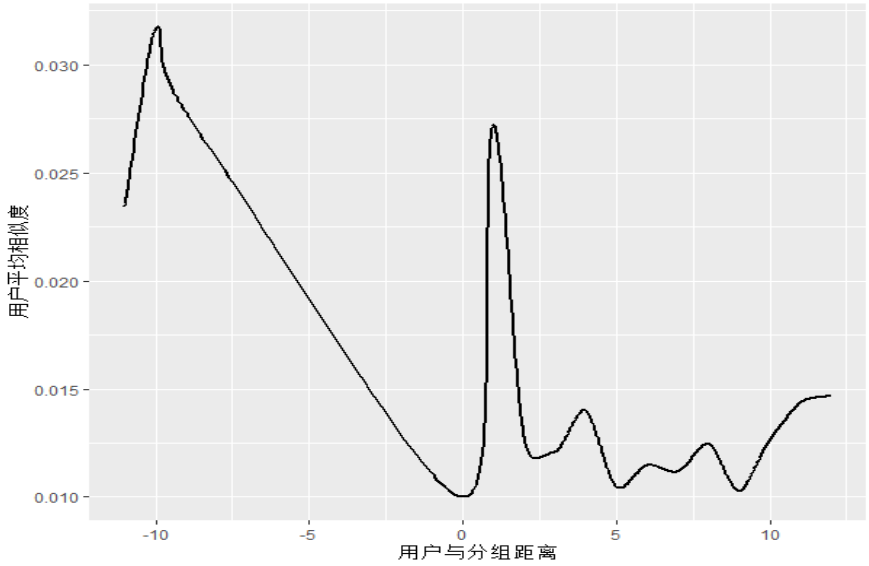

图 3.1 用户平均相似度与分组距离(D)的分布

通过计算不同间隔间的用户平均相似度，其结果由图 3. 1 可知，当两用户之间的间隔距离最小时，两用户的欧式距离最小，也即两用户的上网时间分布向量最相似，因此证明了移动终端用户上网时间分布向量存在近相似性规律。

通过计算用户与其欧式距离最小用户之间的距离分布情况，计算结果如表 3.6 所示。

表 3.6 用户与欧式距离最小的用户分布情况

d	$d=0$	$d>0$
占抽样用户数比例	83.3%	16.7%

由表 3.6 可知，通过计算用户与用户之间的欧式距离，当用户间的上网时间分布向量的欧氏距离最小时，有 83.3%的移动终端用户能在同一分组内发现与其最相似的用户，从而说明上网时间的近相似性规律在移动终端用户中是广泛存在的。

3.3 用户上网时间行为分析

近年来，用户的网络行为画像分析越来越受关注，用户的上网时间也属于上网行为方式，且其是一种持续性的网络行为，分析移动终端用户的上网时间的行为偏好规律，可以帮助我们了解用户在不同时段使用移动终端设备的上网情况，

就运营企业而言，能够更有针对性的制定广告投放的营销策略，提高广告的曝光率，降低运营成本；就用户而言，能够为用户提供可能更有趣的网络服务，增加用户对网络应用的粘性，或者更高效地获得有用的信息，实现用户与网络应用运营商的共赢。基于用户和企业两个角度出发，分析用户上网时间的行为模式的重要性不言而喻，目前，针对移动终端用户上网时间的行为习惯分析，在科研领域通常采用的阈值法具备主观性，得到的用户上网时间偏好存在偏差，故本小节利用划分时段的方式分析用户上网时间的行为模式规律。

3.3.1 层次聚类算法

基于时间特征为分析用户的网络行为提供了新的机会，随着网络行为分析的发展，时间划分显得越来越重要，它有助于降低数据处理的复杂性，并且对获取准确的行为信息具有重要意义。T.Yamakami 将一天划分为 8 个时间段，用来研究网络用户访问模式的稳定性[15]；Halvey 在研究移动互联网的用户行为，将上网时间划分为三个主要跨度，即白天、晚上及周末[5]；而校园网的运行一般都会有一个时间规律，贺雯静研究了校园网络运行时间段的规律，将一天划分为 19 个时段，即 0～6 点为一个时段，而其余时段以一小时为间隔来划分，研究校园网的使用高峰期更有助于校园网络的管理[34]。为了研究移动终端用户上网时段的喜好情况，一般采用聚类的方式将具有相似上网时段的用户划分到一个用户群，挖掘群体用户上网时间的喜好特征，可以考虑根据用户之间欧式距离的大小对用户进行合并，由于层次聚类的原理与该思路相似，故本文选取层次聚类法分析用户的上网时间特征。

由于研究对象在较多的情况下，层次聚类每次合并两个用户时都要更新用户间的相似度，使得聚类算法的时间复杂度较高，而章节 3. 2. 2 验证了移动终端用户上网时间的相似性规律，基于此规律，将移动终端用户的上网时间划分为 16 个初始分组，其组内的用户都普遍相似，而组间用户相异，基于平均连接规则在分组内部进行层次聚类。

其中，分组时间分布向量是对用户进行初始分类之后，再计算分组中用户时间分布向量的算数平均值，如式（3-3）所示：

$$C=\frac{1}{N_C}\sum_{T\in C}T=[\overline{M_{T_0}},\overline{M_{T_1}},\cdots,\overline{M_{T_m}},\cdots\overline{M_{T_{23}}}] \tag{3-3}$$

而且C代表分组的时间分布向量，N_C表示分组内的用户数，$\overline{M}_{T_m}$表示分组用户在时段T_m的用户平均上网时间。

3.3.2 确定聚类簇

根据用户使用移动智能终端访问网络应用时不同访问时间所对应的累计时长，同时，利用向量形式表示用户的上网时间，即上网时间分布向量，然后将用户上网时间分布向量按照最大上网时段划分到相应的组内，由于用户实际的上网时间在 0:00～7:00 不存在最大上网时段，故本文初始分组为 16 组，每组用户所占的比例如表 3.7 所示。

表 3.7 每个时段的分组及用户占比

时段名称	用户分组	用户比例	时段名称	用户分组	用户比例
T_8	G_8	8.30%	T_{16}	G_{16}	6.99%
T_9	G_9	6.33%	T_{17}	G_{17}	2.84%
T_{10}	G_{10}	6.77%	T_{18}	G_{18}	1.30%
T_{11}	G_{11}	5.9%	T_{19}	G_{19}	5.9%
T_{12}	G_{12}	6.55%	T_{20}	G_{20}	6.11%
T_{13}	G_{13}	6.33%	T_{21}	G_{21}	9.83%
T_{14}	G_{14}	6.33%	T_{22}	G_{22}	13.18%
T_{15}	G_{15}	6.33%	T_{23}	G_{23}	5.02%

根据层次聚类算法计算簇间距离的方法不同，层次聚类可分为单连接聚合聚类、全连接聚合聚类以及平均聚合聚类。由于考虑到基于平均连接规则的层次聚类具有较好的鲁棒性，所以我们采用平均连接法来度量簇间距离。随着聚类数目的增多，每个分类中的用户越来越少，簇间的最小距离也越来越小，当簇位于高纬空间与低维空间的交界点时，簇间最小距离下降的速率最快，则可认为进一步增大聚类数目，聚类效果也得不到加强，其中该分界点为“拐点”，该“拐点”处的聚类效果最好。为了寻找这一“拐点”，我们用α来表示最小簇间距离，用

k 表示聚类数目，那么 α 与 k 之间的变化关系如图 3.2 所示。

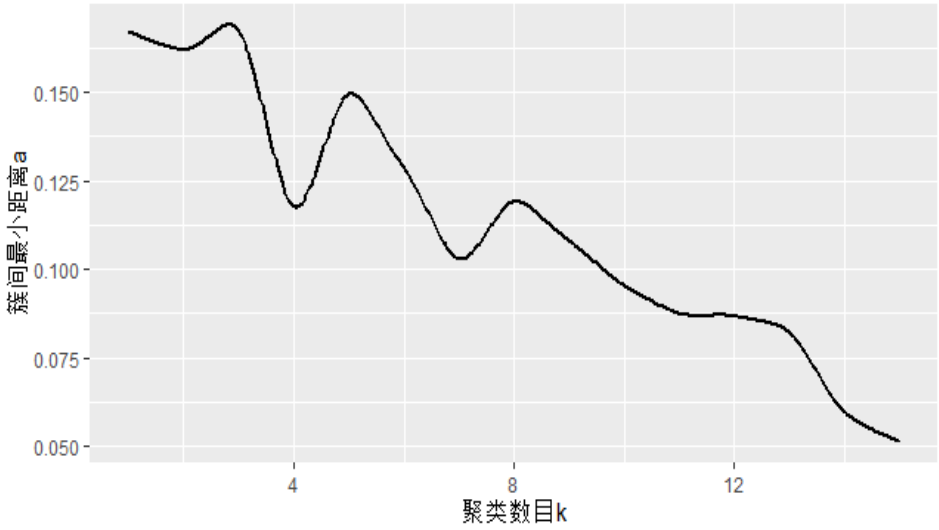

图 3.2 簇间最小距离与聚类数目的评估曲线图

为了便于确定簇间最小距离下降速率最快的“拐点”，我们计算图 3. 2 中各个点之间的斜率，然后按照簇间最小距离的递减速率选取最优聚类簇，以此确定最佳聚类数，得到的斜率如表 3.8 所示。

表 3.8 簇间最小距离的斜率变化表

起点	终点	斜率	起点	终点	斜率
1	2	-0.0050	8	9	-0.0111
2	3	0.0052	9	10	-0.0126
3	4	-0.0496	10	11	-0.0080
4	5	0.0318	11	12	-0.0007
5	6	-0.0209	12	13	-0.0046
6	7	-0.0256	13	14	-0.0224
7	8	0.0162	14	15	-0.0085

由表 3.8 可知，簇间最小距离有 4 个比较明显的拐点，对应的聚类数目分别为 3、4、5、7，其下降速率由高到低依次为 4、7、5、3，因此本文选取的最佳聚类簇为 4 时，此时的聚类效果最好。

3.3.3 聚类结果分析

通过移动终端用户上网时间分布向量在最大上网时段附近存在的近相似性规律，按照最大上网时段对用户进行初始分组，进而得到分组时间分布向量，利用基于平均连接规则的层次聚类算法将相似的分组进行了聚类合并，并根据拐点法得到的最佳聚类簇为 4，那么得到的原始分组结果为：$\{G_{12}, G_{13}\}$，$\{G_8, G_9, G_{10}, G_{11}\}$，$\{G_{17}, G_{18}, G_{20}, G_{21}, G_{22}, G_{23}\}$，$\{G_{14}, G_{15}, G_{16}\}$，其层次聚类过程如图 3.3 所示。

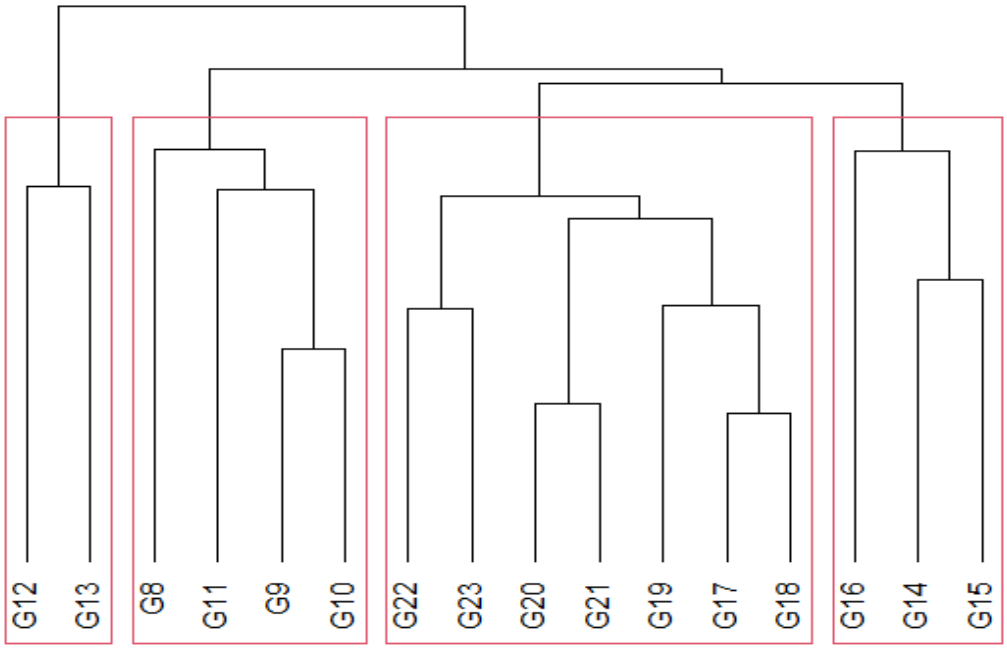

图 3.3 层次聚类过程

利用层次聚类算法得到用户的原始分组结果，由图 3. 3 可以得到具体结果，为了进一步分析聚类结果与原始分组结果，我们将原始分组（虚线）和聚类结果（实线）的分布趋势放在同一图中进行比较，其中，原始分组中的用户在一天的时间跨度上具有相似的偏好，而聚类结果组呈现出每个群体对原始分组的平均偏好，我们发现同一图中的曲线具有比较明显的波峰和波谷，但是原始分组与聚类结果组对时间的划分有所差别，并且每个聚类结果用户组只表示用户对时间段的偏好，基于以上分析，我们利用聚类结果组的平均偏好来分析用户上网时间的行为模式规律。用户的聚类结果组的时间分布向量与原始分组的时间分布向量之间的关系表示如图 3.4 所示。

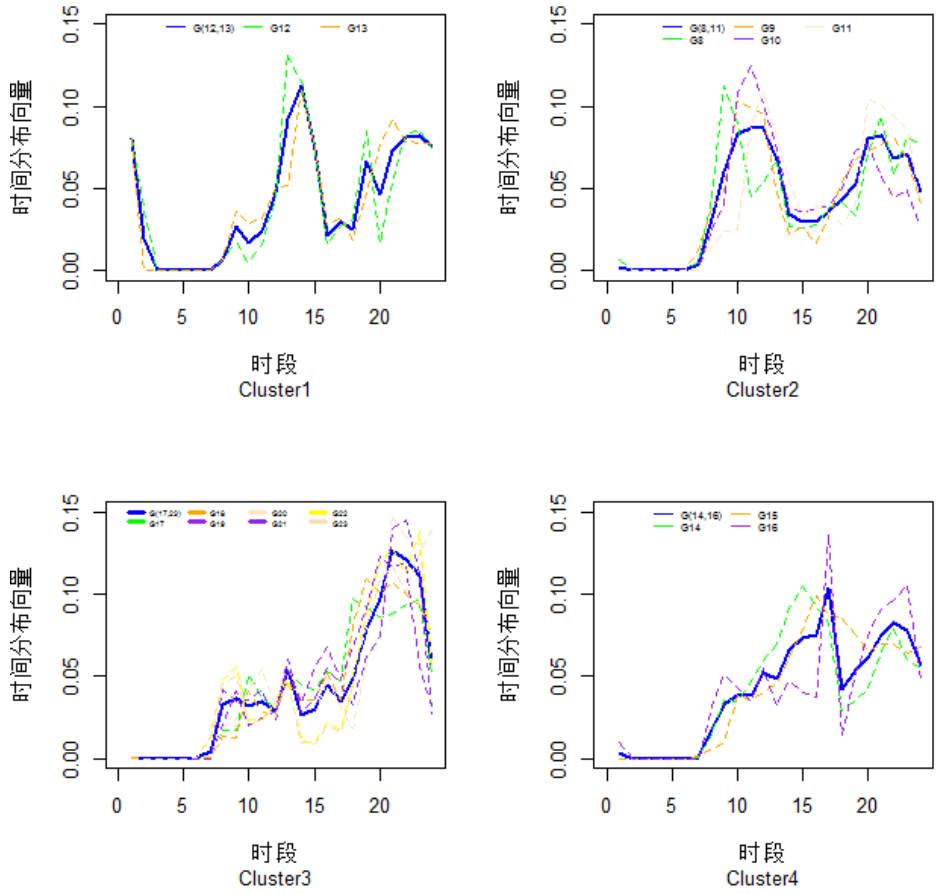

图 3.4 移动终端用户上网时间分布图

由图 3.4 可知，其聚类结果与原始分组的时间分割有一些不同，我们利用平均偏好将用户上网时间划分为 4 种行为模式，其具体分析如下：

类型 1：这一群体的用户最喜欢在 12 至 13 点时段上网，人数占比约为 12.88%，此段上网时间位于午休时段，属于午休上网用户群。

类型 2：这一群体的用户最喜欢在 9 至 11 点时段上网，而在午夜和下午时段的上网偏好变弱，用户喜好的上网时间主要位于上午时段，人数占比约为 27.29%，属于上午上网用户群。

类型 3：这一群体的用户最喜欢在 18 至 22 点时段上网，而且用户在的上网时间整体呈现出上升的趋势，此段上网时间主要位于晚间时段，人数占比约为 40.18%，属于晚间休闲上网用户群。

类型 4：这一群体的用户最喜欢在 17 点时段上网，人数占比约为 19.65%，故可称之为傍晚上网用户群。而在 18 点时段用户的上网偏好处于波谷期，且在晚间时段用户的上网时间偏好有所回升。

综上所述，移动终端用户的上网时间在最大上网时段附近存在近相似性规律，根据此规律将用户的上网时间划分为 4 种偏好模式。就总体而言，移动终端用户不同的上网时间行为模式存在显著的差异，其中，第一簇和第二簇的用户最喜欢在白天访问移动终端的网络应用，其占比达到 40.17%，而第三簇和第四簇的用户最喜欢在傍晚至晚间时访问移动终端的网络应用，其占比达到 59.83%，且在四个簇中的占比最大，则说明大多数用户喜好的上网时间在晚上。根据各个类型的用户上网时间喜好特征，运营商可通过实际需求挖掘潜在用户并定制差异化服务，同时，需要更加重视喜好晚上上网的用户，可根据他们的不同需求提供多种网络服务选择。

3.4 用户访问内容行为分析

随着移动互联网的不断发展，各种新型的网络应用在移动终端平台上层出不穷，各种各样的网络应用所产生的网络信息使用户应接不暇，而且不同类型的网络应用已经渗透到用户日常生活的各个领域，为用户带来了极大的便利。用户在访问网络应用的过程中会产生大量的浏览记录，而且这些记录中隐藏着用户在上网过程中的行为习惯，对刻画用户画像有很大的帮助，根据用户的访问记录，挖掘不同类别的用户行为特征。Welke、Andone I 等通过分析用户访问网络应用的行为特征，主要将 500 个网络应用软件的用户行为作为研究对象，其结果表明，用户对应用程序的使用习惯存在显著差异[12]。Zhao、Ramos 等以用户访问应用软件留下的网络足迹为研究对象，通过分析用户的网络行为得到 382 种用户类型，如夜晚通信者、晚间学习者、爱车人士等，最后根据分析结果为不同类型的用户定制差异化服务[19]。

3.4.1k-means 聚类算法

移动终端用户访问内容的分析是指用户通过网络应用获取网络信息的偏好分析，是运营商识别用户兴趣的主要手段，通过分析用户访问内容的偏好，一方面，有助于运营企业根据用户的喜好制定差异化服务提供依据，从而提高用户对该网络应用的粘度和满意度；另一方面，根据用户兴趣偏好特征，以便协助广告商制定合理的广告推送策略。由于移动终端用户的网络行为复杂且多样，则不同

用户访问网络内容时所表现出的行为偏好规律也存在差异性，那么利用机器学习算法提取用户访问内容时的行为偏好规律，有助于我们研究移动终端用户上网时间与访问内容之间的相关性，进而依据用户兴趣向用户推送相关网络服务。目前，分析移动终端用户访问内容的偏好一般采用 k-means 聚类算法，该算法得到的结果不仅结构特性优良，而且显著表明用户上网的行为特征，最重要的是能灵活划分网络行为类型，故在本节中，我们采用 k-means 聚类算法分析用户访问内容的喜好。

在对每个用户的网络行为数据进行统计和分析后，我们需要利用相关的机器学习算法分析用户对网络应用的粘度和喜好情况。聚类分析将相似的对象归入同一簇，针对用户的网络行为分析，本文首先在 k-means 算法的基础上，运用肘部法（elbow-method）和轮廓系数（Silhouette-Coefficient）选取合适聚类簇，其次利用 URL 的关键词信息将每个用户在 18 种网络行为类别所占的百分比进行分类，将用户划分为不同类型，最后对每种用户类型进行分析研究。

3.4.2 确定聚类簇

在确定 k-means 聚类算法的最佳聚类数目 k 时，最普遍使用的方法是 elbow-method，在移动终端用户访问内容的行为数据中，设置的 k 值越大，划分的网络应用类型簇越多，则会导致簇内的用户逐渐减少，样本距离簇的重心会更近，平均畸变程度随之减小，畸变值随着 k 的增大而减小，其下降幅度最大的位置即为肘部 k，而 k 值是最佳的聚类数，但是一般情况下肘部法的拐点并不是总能清晰的找到拐点位置。但是轮廓系数也可以用于确定最佳聚类数目 k，该方法结合了聚类的凝聚度和分离度，用于评价聚类效果的好坏，系数越大，表示聚类效果越好。

如图 3.5 所示，当 k 从 2 到 9 时，平均畸变程度变化最大，当 $k>6$ 时，畸变程度趋于稳定，但其 elbow-method 的拐点并不明显，故而本文结合肘部法则和轮廓系数进行 k 值的选取，而且拐点对应 k 值的范围在 2～9 之间，通过计算此范围内 k 值所对应的轮廓系数，从而进行最终聚类簇的选取，k 值与对应得到的轮廓系数如表 3.9 所示。

表 3.9 k 值所对应的轮廓系数

k值	Silhouette-Coefficient	k值	Silhouette-Coefficient
2	0.3736	6	0.4848
3	0.3544	7	0.4939
4	0.3938	8	0.4762
5	0.4600	9	0.4684

由表 3.9 可知，当$k=7$时，轮廓系数达到最大，即为 0.4939，此时移动终端用户访问网络应用的时长能够得到最好的聚类效果，因此，当本文选取聚类簇为 7 时，利用 k−means 聚类算法分析移动终端用户访问内容的行为模式。

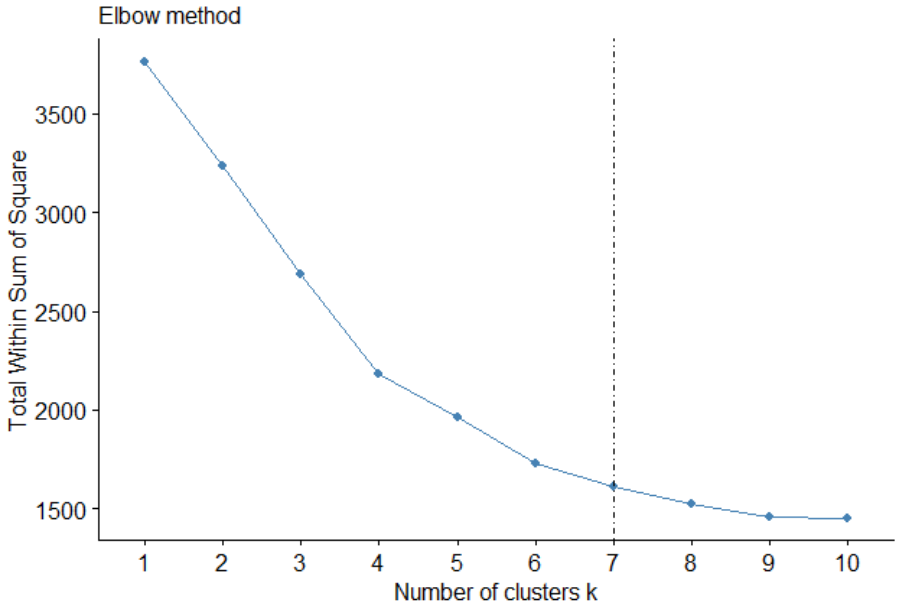

图 3.5 elbow−method 选取聚类数 k

3.4.3 聚类结果分析

根据上述分析，利用 k−means 聚类算法对移动终端用户的访问内容向量进行聚类，通过轮廓系数与肘部法确定最优聚类簇，得到 7 个用户类型分组，分群的结果如表 3.10 所示，根据聚类结果返回原始数据，其中每行表示分类后的 7 种用户类型，每列的数值表示每个网络应用类型在每个用户类型中所占的百分比，分析每个用户对应的不同类别特征。

表 3.10 各个簇中各类型网络应用的访问时长占比（%）

	类型 1	类型 2	类型 3	类型 4	类型 5	类型 6	类型 7
娱乐休闲	14.366	13.537	29.558	10.939	13.089	5.654	8.087
新闻资讯	8.133	7.467	10.066	26.124	8.956	1.685	28.981
社交通信	26.206	30.077	29.197	33.139	20.507	28.991	11.413
网络购物	2.817	4.175	2.312	3.593	28.501	0.759	3.208
摄影摄像	0.332	0.396	0.495	0.190	0.527	0.048	0.334
游戏	7.396	32.749	12.728	4.812	7.633	0.983	1.441
金融理财	1.039	1.137	0.930	1.185	3.381	1.932	0.605
交通出行	0.382	0.465	0.497	0.382	2.812	0.052	0.214
居家生活	0.064	0.005	0.051	0.031	0.106	0.010	0.592
浏览搜索	1.824	2.071	1.198	2.109	2.782	3.878	3.501
效率办公	0.930	0.891	0.635	0.564	1.129	5.522	0.462
图书阅读	25.956	1.379	0.948	11.870	1.091	0.083	4.595
教育学习	4.264	1.096	0.463	0.775	0.460	46.298	28.413
快餐服务	2.177	3.589	3.034	1.751	4.567	3.353	3.250
医疗健康	0.007	0.002	0.021	0.001	0.007	0.002	0
体育运动	3.789	0.795	7.438	2.090	4.092	0.498	4.887
汽车服务	0.044	0.019	0.005	0.089	0	0	0
系统工具	0.274	0.151	0.424	0.355	0.363	0.253	0.016
聚类个数	31	59	151	66	54	56	41

总体上来看，各个群体对某种兴趣的偏好不一，社交通信、娱乐休闲两类兴趣倾向在 7 个簇中的占比达到 71.43%，这在一定程度上反映了大部分用户的日常生活状况。此外，相对小众的兴趣，群体特征表现的很明显，例如，喜欢社交阅读类的用户对图书阅读类网络应用的使用时长会明显偏高于其他不常用的用户。就聚类簇的用户数而言，用户最多的是类型 3，而用户最少的是类型 1，由此可以看出，最多的用户群是最少用户群的 5 倍，其余的用户类型包含的用户相差不多，因此，类型 3 代表移动终端用户的普遍行为偏好。

由表 3.10 可知，对于这 7 种用户类型中每一类用户类型的特征，具体分析

如下：

类型 1：社交阅读型用户（Social network and Reading users），占总人数的 6.77%。该类型的用户主要访问的是社交通信和图书阅读类网络应用，所占百分比分别约为 26.21%和 25.96%，另外在娱乐休闲、新闻资讯、游戏类网络应用上的访问时长相对于其他类较多。

类型 2：游戏社交型用户（Games and Social network users），占总人数的 12.88%。该类型的用户特别爱好游戏类网络应用，所占百分比约为 32.75%，同时社交通信类所占百分比约为 30.77%，另外对娱乐休闲类的访问量也比较多。

类型 3：娱乐社交型用户（Entertainment and Social network users），用户数量最多，占总人数的 32.97%，属于一种主要的上网行为模式。该类型用户主要访问娱乐休闲和社交通信类网络应用，所占百分比分别为 29.56%、29.20%，其次是游戏、体育运动以及餐饮服务类网络应用。

类型 4：社交新闻型用户（Social network and News users），占总人数的 14.41%。该类型用户主要访问的网络应用是社交通信和新闻阅读类网络应用，所占百分比分别为 33.14%、26.12%，其次是图书阅读和娱乐休闲类网络应用，而在游戏、网络购物、浏览搜索类网络应用上花费的时间很少。

类型 5：网络购物型用户(Shopping users)，占总人数的 11.79%。该类型用户主要访问的网络应用是网络购物类，所占百分比约为 28.50%，其次是社交通信、娱乐休闲类网络应用，很少有用户访问新闻资讯、游戏、快餐服务以及体育运动类的网络应用。

类型 6：学习型用户（Education users），占总人数的 12.23%。这类用户主要访问关于教育类的网络应用，所占百分比约为 46.3%，其次喜好访问社交通信类网络应用。

类型 7：新闻关注型用户（News and Education users），占总人数的 8.96%。该类型用户主要访问新闻资讯和教育学习类网络应用，所占百分比分别为 28.98%、28.41%，其次是社交通信和娱乐休闲类网络应用。

综合上述分析，我们可以挖掘出移动终端用户访问网络应用的一些特点为：

（1）社交通信类网络应用访问量大。分析表明社交通信类是移动终端用户最为普遍使用的网络应用，而且访问量相当大，用户通过使用社交通信类网络应

用积极参与网络信息的传播，也是用户信息交换的最主要途径。

（2）主体兴趣倾向相似。根据用户访问网络应用的时长可知，用户最喜欢使用的网络应用是社交通信和娱乐休闲类，而且用户占比是最大的。

（3）主体兴趣倾向明显。通过 k-means 聚类算法将移动终端用户划分为 7 种类型，每种类型的用户访问网络应用的时长均要明显高于其他类型的用户，且其占比都在 25%以上，有些甚至在 45%以上，每类用户都有别于其他类别用户的网络兴趣倾向。

（4）上网兴趣倾向相对比较固定。就聚类结果而言，约 71.43%的用户群体使用 4 种以内的网络应用，其中约 60%的用户群体只使用 2-3 种网络应用。

3.5 用户网络行为关联分析

就移动终端用户的网络行为而言，用户的上网时间代表用户在不同时段访问网络的规律，而用户访问的网络内容则代表用户从移动终端平台获取网络信息的偏好。通过分析用户上网时间的行为习惯，我们将上网行为识别为 4 种行为模式；通过分析用户访问内容的喜好分析，我们将用户的网络行为识别为 7 种主要行为模式。本文利用关联分析法将两方面挖掘出的用户行为模式相融合，分析移动终端用户在不同时段对应访问网络应用的行为模式偏好，通过关联共得到 28 种用户行为模式组合，如图 3.6 所示。

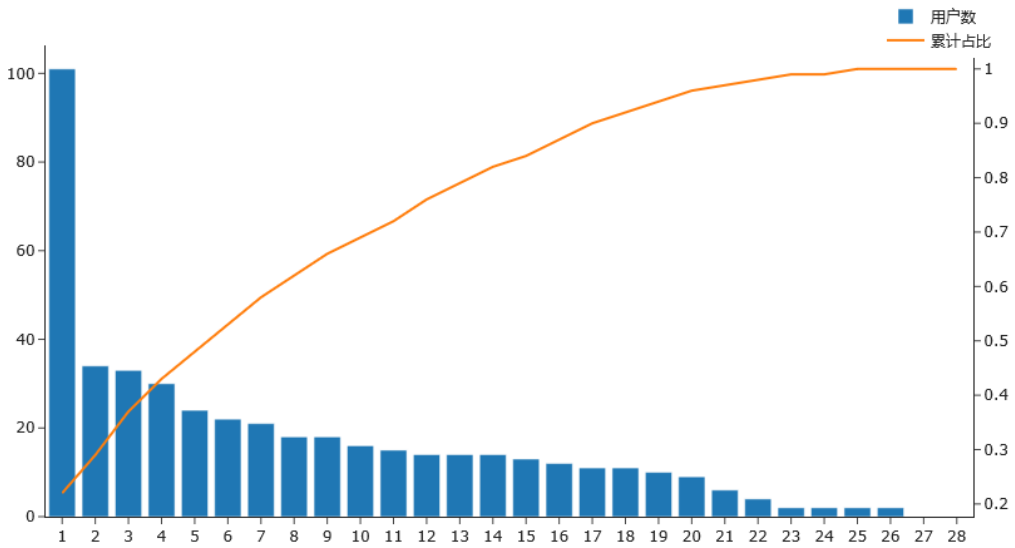

图 3.6 用户网络行为组合模式

由图 3.6 可知，移动终端用户的综合网络行为模式比较多，但是用户普遍的

网络综合行为相对较少，其总体表现为用户逐渐减少，因此，我们只需要分析前10 个用户的综合网络行为模式，而且这些行为模式是移动终端用户最常出现的行为偏好模式，其占比达到 72.49%，那么各个组合网络行为的用户占比如表 3.11 所示。

表 3.11 组合行为模式排名前 10 的特征

组合排序	时段喜好模式	内容喜好模式	用户数	占比（%）
1	3	3	101	22.05
2	2	6	34	7.42
3	4	4	33	7.21
4	3	2	30	6.55
5	3	5	24	5.24
6	4	4	22	4.80
7	2	7	21	4.59
8	2	2	18	3.93
8	2	3	18	3.93
9	4	7	16	3.49
10	2	1	15	3.28

我们对表 3.11 的 10 种网络行为组合模式进行逐一分析:

（1）晚间娱乐社交型用户，用户比例为 22.05%，就上网时间而言，喜欢在晚上上网的用户群，即晚间休闲用户群，这类用户最喜欢在访问娱乐休闲和社交通信类的网络应用。

（2）上午学习型用户，用户比例为 7.42%，就上时间而言，喜欢在上午上网的用户群，这类用户偏重于使用教育学习和社交通信类的网络应用。

（3）傍晚社交新闻型用户，用户比例为 7.21%，就上网时间而言，喜欢在傍晚上网的用户群，这类用户最喜欢使用社交通信、新闻阅读以及娱乐休闲类的网络应用。

（4）晚间游戏社交型用户，用户比例为 6.55%，就上网时间而言，喜欢在晚上上网的用户群，这类用户最喜欢使用游戏和社交通信类的网络应用。

（5）晚间网络购物型用户，用户比例为 5.24%，喜欢在晚间上网的用户群，这类用户偏重于使用网络购物、社交通信、娱乐休闲类的网络应用。

（6）傍晚社交新闻型用户，用户比例为 4.80%，喜欢傍晚上网的用户群，这类用户偏重于使用社交通信、新闻阅读以及娱乐休闲类的网络应用。

（7）上午新闻关注型用户，用户比例为 4.59%，喜欢上午上网的用户群，这类用户最喜欢使用新闻资讯和社交通信类的网络应用。

（8）上午社交娱乐型与新闻型用户，用户比例为 3.93%，喜欢上午上网的用户群，这类用户最喜欢使用游戏、社交通信、娱乐休闲、体育运动以及快餐服务类的网络应用。

（9）傍晚新闻关注型用户，用户比例为 3.49%，喜欢傍晚上网的用户群，这类用户最喜欢使用新闻资讯和教育学习类的网络应用。

（10）上午社交阅读型用户，用户比例为 3.28%，喜欢上午上网的用户群，这类用户最喜欢使用社交通信和图书阅读类的网络应用。

综合上述分析，较多的移动终端用户喜好在晚上访问娱乐休闲类和社交通信类网络应用，这两类网络应用是集社交、娱乐、交友等多功能糅合的 APP，增加了用户与网络应用之间的黏度，由此可见，用户偏爱使用多功能的网络应用，同时，需要更加重视喜好晚上使用这两类网络应用的用户，根据他们的需求提供这两类网络应用的服务选择。

3.6 本章小结

本章基于移动终端的数据分析了用户的网络行为特征，主要内容分为三部分，其一，分析用户上网时段的行为特征：使用 24 维向量表示用户的上网时间，基于用户上网时间的相似性规律，将其划分为 16 个初始用户组，并通过层次聚类分析，得到了用户上网时段的 4 种行为模式规律：午休上网用户群、上午上网用户群、晚间休闲用户群以及傍晚上网用户群；其二，分析用户访问内容的行为特征，结合肘部法则和轮廓系数选取聚类簇，利用 k-means 聚类算法将用户化分为社交阅读型用户、游戏社交型用户、新闻关注型用户等 7 种行为模式特征，各个用户类型的偏好各不相同，各个网络应用类别在每个用户类型中所占的百分比也不相同。最后关联上述两场景的网络行为模式，全面描述用户的网络行为模式。

4 移动终端网络服务个性化推荐

移动互联网和大数据技术的迅猛发展，不仅使网络应用类别十分丰富，网络信息也十分繁杂，导致信息过载现象不断加剧，以至于用户难以在短时间内精准地筛选出符合自己要求的信息，为了解决用户无法快速获取有效信息的难题，本章利用 TF-IDF 算法计算用户访问 APP 的兴趣程度，结合协同过滤推荐算法，筛选出用户感兴趣的网络信息，针对推荐结果的准确性，提出利用 BP 神经网络优化个性化推荐算法的准确度。

4.1 个性化推荐系统设计

4.1.1 流程概述

在第三章主要分析了移动终端用户的上网时间和访问内容的行为模式规律，根据数据挖掘技术获得的网络综合行为模式为用户制定个性化网络信息推荐列表。为了提升用户的个性化信息推荐效果，本章利用 BP 神经网络优化协同过滤推荐算法中的评分预测值，针对移动终端用户进行网络服务个性化推荐，并验证该算法推荐效果的精准度，本章通过真实的移动终端用户网络行为数据进行了实验，主要流程如 4.1 所示。

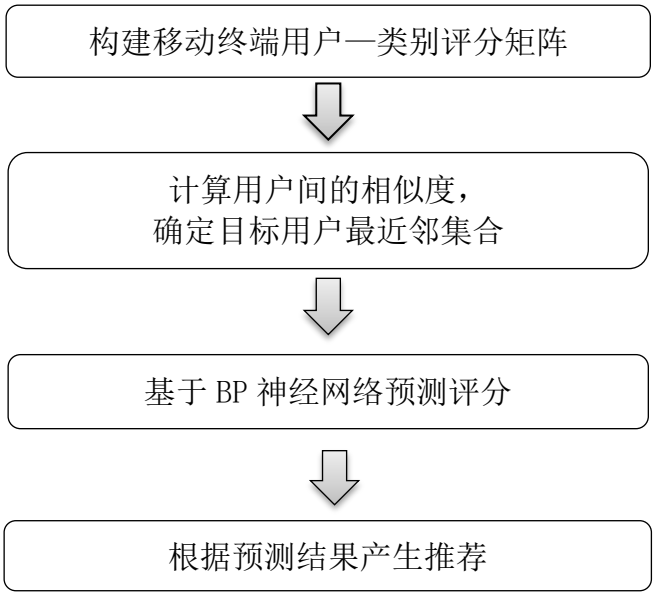

图 4.1 协同过滤推荐算法流程图

4.1.2 构建用户—类别评分矩阵

协同过滤推荐算法利用用户的历史行为为目标用户制定个性化网络信息推荐，它主要依赖用户—类别评分矩阵来预测评分，并产生推荐列表。用户使用移动终端浏览网页时的行为反映了用户的兴趣偏好，那么了解用户的行为偏好分类及兴趣程度是个性化推荐的基础，用户的兴趣度即为用户行为偏好的权重，代表用户对某一类网络应用的喜好程度，目前用户兴趣度的获得具有两种方式，其中，一种方法是用户自行标注，这种方法一般能得到较为准确的行为偏好，但在实际应用中，很少有人愿意标注网络应用页面的偏好度[35]；另一种方法是利用用户访问某类网络应用的时长来估计兴趣度，这种方式不需要用户的直接参与，同时，用户的网络行为就是用户的兴趣偏好的表现[44]。因此，本小节基于用户访问某类 APP 的时长来估计用户的兴趣度，采用 0 到 1 的实数来表示，其中，0 代表用户对该类网络应用完全不感兴趣，1 代表用户对该类网络应用非常感兴趣。

通常而言，如果用户访问了某 APP，就可以认为其对该类 APP 感兴趣，累计使用时长越久表示越感兴趣，通过计算用户一天内使用该 APP 的时长占使用所有 APP 的总时长来表示兴趣度，但是会产生误差，因为不同类型的网络应用决定了用户的访问时长。据统计，系统工具类 APP 的使用时长要远远低于娱乐休闲类 APP，如果按照这种方式估算用户兴趣度，那么很显然用户更倾向于喜欢娱乐休闲类 APP，故无法有效地表示用户对 APP 的喜好度。

为了解决这一问题，黄良发提出借鉴 TF-IDF 的思想来估计用户对某个类别的兴趣度 I [36]。TF-IDF 是统计分析法，常用来计算某文档的字词在该文档的重要程度，其中词频（TF）是某一词在该文档中出现的频率，而 IDF 用于衡量某一特定词的重要程度[13]。提出该兴趣度计算公式的文献与本文的数据特征类似，都是用户访问移动终端的数据，故本文借助 TF-IDF 的思想来估计用户对某个 APP 类别的兴趣度，首先定义用户的兴趣指数 II，可反映用户在特定兴趣类别中的兴趣程度，其数值越大表示兴趣程度越大，利用公式（4-1）表示。

$$II_{ij} = \frac{S_{ij}}{sum(S_i)} / \frac{T_j}{sum(T)} \tag{4-1}$$

其中，II_{ij}表示用户 i 对 j 类兴趣偏好的兴趣指数，S_{ij}表示用户 i 访问 j 类兴趣偏好的特征值，$sum(S_i)$表示用户 i 访问所有兴趣偏好的特征值之和，T_j表示所

有用户访问 j 类兴趣偏好的特征值之和，$sum(T)$表示所有用户访问所有网络行为类别的特征值之和，其特征值为对应用户访问 APP 的使用时长。

最后，利用公式（4−2）得到用户使用 APP 的兴趣度，其中，$sum(II_i)$表示用户 i 的兴趣指数之和，其定义式如下：

$$I_{ij} = II_{ij} / sum(II_i) \tag{4-2}$$

下面本文利用第三章的移动终端用户网络综合行为，从实验数据中随机抽取 9 名用户的网络行为分类特征值，通过利用公式（4−1）和（4−2）来计算用户的兴趣度，得到用户访问不同网络应用的兴趣度权重如表 4.1 所示。

表 4.1 用户兴趣度权重

类别	8	19	26	34	58	75	110	128	133
娱乐休闲	0.050	0.074	0.061	0.024	0.071	0.100	0.095	0.020	0.017
新闻资讯	0.032	0.024	0.146	0.080	0.020	0.091	0.047	0.100	0.002
社交通信	0.038	0.051	0.116	0.015	0.048	0.081	0.043	0.070	0.086
网络购物	0.014	0.014	0.101	0.037	0.091	0.281	0.029	0.015	0.010
摄影摄像	0.028	0.064	0	0.066	0.067	0	0.045	0.149	0
游戏	0	0.021	0.043	0.008	0.077	0.038	0.345	0	0
金融理财	0.029	0.045	0.079	0.019	0.018	0.064	0.036	0.128	0.120
交通出行	0.031	0.065	0.058	0.025	0.061	0.026	0.060	0.140	0
居家生活	0	0	0	0.39736	0	0	0	0	0
浏览搜索	0.098	0.024	0.044	0.050	0.046	0.037	0.131	0	0.109
效率办公	0.223	0.017	0.022	0.008	0.009	0.015	0.032	0.001	0.256
图书阅读	0	0.108	0.243	0.030	0	0.218	0	0.012	0
教育学习	0.216	0	0	0.081	0	0	0	0	0.260
快餐服务	0.100	0.066	0.030	0.057	0.016	0.024	0.109	0.006	0.110
医疗健康	0	0	0	0	0.380	0	0	0	0
体育运动	0.046	0	0	0.102	0.039	0	0	0.240	0
汽车服务	0	0.356	0	0	0	0	0	0	0
系统工具	0.094	0.071	0.057	0	0.057	0.024	0.028	0.119	0.030

由表 4.1 可知，虽然用户在社交通信类的使用时长多于系统工具类，但是通过借助 TF-IDF 的思想来计算用户兴趣度，发现用户在社交通信类上的兴趣偏好要大于系统工具类。最根本的原因在于，社交通信类在整个兴趣群中的使用时长都很大，因此，计算结果也反映了其在社交通信类上不是最大的，故而验证了兴趣度计算的合理性。

每个用户访问 APP 的行为代表用户的行为习惯，且兴趣度表示用户对网络行为的评分，那么所有用户对移动终端发生的网络行为的兴趣度构成一个矩阵，即该矩阵为用户—类别评分矩阵$R_{m\times n}$，表示有n个用户对m个网络行为类别的兴趣程度，如表 4.2 所示。其中r_{ij}表示用户u_i对网络行为类别$item_j$的喜好程度，由[0, 1]之间的数值构成，其数值的大小表示用户对 APP 的喜好程度，1 表示特别感兴趣，0 表示不感兴趣。

表 4.2 用户—类别评分矩阵$R_{m\times n}$

	$item_1$	$item_2$	...	$item_m$
u_1	r_{11}	r_{12}	...	r_{1m}
u_2	r_{21}	r_{22}	...	r_{2m}
...	...	...	...	...
u_n	r_{n1}	r_{n2}	...	r_{nm}

4.1.3 筛选近邻用户

个性化推荐是通过分析大量用户的历史行为资料而产生有价值的资讯及服务，那么，利用第三章得到的移动终端用户网络综合行为，以用户间的相似度为标准，随后根据目标用户与“邻居”用户之间的相似性为目标用户建立近邻用户集 N，继而利用近邻用户对项目的评价给目标用户提供推荐。筛选近邻用户时需要以构建的用户—类别评分矩阵$R_{m\times n}$为基础，计算目标用户与其余用户的评分信息之间的相似度，将相似度按照递减次序排列，选取前K个用户作为目标用户的“邻居”。

通常而言，为了实现精准化的个性化推荐效果，在协同过滤推荐算法中，使用相似度来进行“邻居”用户的筛选是一个非常关键的部分，这与推荐结果的精

确度有直接关系，故需深入挖掘用户对网络服务类别评分向量之间的相似性，而相似性反映了用户之间的差异程度，差异程度越小相似度越大，反之亦然。计算用户之间的相似性的方法主要有以下四种，其中，公式中的参数代表的含义为：$\overline{R_u}$、$\overline{R_v}$分别代表用户u和v对所有已评分网络服务的平均值，$sim(u,v)$代表用户u与v的相似度，R_{vi}代表近邻用户v对网络服务类别i的评分，$N(u)$、$N(v)$分别代表用户u、v交互网络服务类别的集合：

（1）杰卡德系数（Jaccard）：在计算相似度时，不考虑具体的评分值，只考虑用户是否产生行为记录，用于比较两个用户之间的差异性，其计算公式如（4-3）所示。

$$sim(u,v)=\frac{\left|N(u)\cap N(v)\right|}{\sqrt{\left|N(u)\right|\cup\left|N(v)\right|}} \tag{4-3}$$

（2）皮尔森系数（Person Correlation）：该系数主要用来计算变量间的相关程度，其计算公式如式（4-4）所示。

$$sim(u,v)=\frac{\sum(R_{ui}-\overline{R_u})(R_{vi}-\overline{R_v})}{\sqrt{\sum(R_{ui}-\overline{R_u})^2}\sqrt{\sum(R_{vi}-\overline{R_v})^2}} \tag{4-4}$$

（3）余弦相似度（Cosine）：两个向量的相似性是由两个向量之间的角度来决定的，角度愈小相似性愈大，常被用来计算两个向量的相关性。其计算公式如（4-5）所示。

$$sim(u,v)=\cos(\vec{u},\vec{v})=\frac{\sum R_{ui}\times R_{vi}}{\sqrt{\sum R_{ui}^{\ 2}}\sqrt{\sum R_{vi}^{\ 2}}} \tag{4-5}$$

（4）修正的余弦相似度（Adjusted Cosine）：由于不同用户对项目或者物品的评分存在主观性和差异性，而余弦相似度并没有关注该性质，因此提出利用平均值修正余弦相似度，其计算公式如（4-6）所示。

$$sim(u,v)=\frac{\sum(R_{ui}-\overline{R_u})(R_{vi}-\overline{R_v})}{\sqrt{\sum(R_{ui}-\overline{R_u})^2}\sqrt{\sum(R_{vi}-\overline{R_v})^2}} \tag{4-6}$$

由于章节 4.1.2 计算得到的用户访问不同类型 APP 的兴趣度，其计算方式具有合理客观性，与用户对项目的传统评分不同，不需要考虑不同用户的行为差异性和主观性，因此本文选择余弦相似度来计算用户之间的相似性，针对移动终端用户网络行为数据能够表现出良好的性能。

4.1.4 基于 BP 神经网络预测评分

个性化推荐系统中最影响推荐效果的是评分预测环节，当为目标用户u选取出近邻用户集N之后，利用近邻用户集N的历史评分信息评估目标用户对某一网络服务类别的偏好程度。推荐系统的项目评分预测方法是 Resnick 在 1994 年基于平均分提出的，该方法主要以用户间的相似度和近邻用户对项目的平均评分为基础而产生的预测结果，根据预测的结果，由高到低依次排列，进而产生一个推荐列表，迄今为止，在个性化推荐算法中，这种平均分的预测方法仍然是最普遍使用的一种，其公式如（4−7）所示：

$$P_{ui}=\overline{R_u}+\frac{\sum sim(u,v)\times(R_{vi}-\overline{R_v})}{\sum|sim(u,v)|} \tag{4-7}$$

由于评分预测结果直接关系到推荐效果的优劣，而传统的评分预测法仅从用户之间的相似性出发，忽视了不同网络服务标签属性对用户的影响，又因为移动终端用户网络行为数据是非线性的，传统的线性预测模型不适用于处理此类数据，而神经网络的出现能够很好的解决非线性问题，为了使个性化推荐服务的结果更精准，本文将协同过滤推荐算法中基于平均分的评分预测公式替换为 BP 神经网络。

将 BP 神经网络与协同过滤推荐算法相融合，那么需要构建 BP 神经网络预测模型，即利用梯度下降法迭代更新连接神经元的权重参数，主要思想为：首先将近邻用户对网络服务类别的评分作为输入层，其神经元个数为f，将目标用户对网络服务类别的实际值作为输出层，由于本文是回归预测类问题，故可将输出层的神经元设置为 1 个，但是选取隐含层的神经元个数p是没有规定的理论经验，一般地按照实际情况而定，随后进行训练网络，最后根据训练完的神经网络预测目标用户对某一网络服务类别的评分。章节 2.3.2 详细地讨论了 BP 神经网络的训练过程，结合协同过滤推荐算法完成对目标用户的个性化推荐服务。

4.2 个性化推荐效果评估

4.2.1 实验准备

在第三章中为了降低数据的稀疏性，利用聚类和关联算法处理了移动终端用

户网络行为数据，分析得到用户在不同时段访问不同类型 APP 的网络行为分类。虽然传统的预测技术已经成熟，但是用户行为数据比较复杂，为了提高评分预测的准确度，利用 BP 神经网络改进协同过滤推荐算法，本节将第三章分析得到移动终端用户网络综合行为数据按照 4:1 的比例划分训练集和测试集，进而检验算法的优劣。

4.2.2 评价标准

在为移动终端用户进行个性化推荐服务时，需要估算用户对某网络应用的显性或隐性评分，在本文中利用 TF−IDF 算法估算的评分比较客观合理，属于用户对网络应用的隐性评分，随后利用个性化推荐算法预测用户对某网络应用的未来评分，并根据估算的未来评分判断是否推荐该网络信息。在个性化推荐领域，通常采用特定的评估方法来验证推荐系统的优劣，准确度是评价推荐系统的一个重要指标，一般情况下，选用均方根误差（RMSE）和平均绝对误差（MAE）来判断推荐系统的可行性，当 RMSE、MAE 值越小时，其说明 BP 神经网络训练的实际输出值与期望值之间的误差越小，继而证明基于 BP 神经网络的个性化推荐系统的评分预测越接近用户的实际评分值，推荐效果更佳。

假设n为测试集的评分记录数，y_i为推荐算法对测试集第i条评分记录的预测值，r_i为测试集第i条评分记录的实际评分值，则 RMSE 的计算公式（4−8）：

$$RMSE = \sqrt{\frac{\sum_{i}^{n}(y_i - r_i)^2}{n}} \tag{4-8}$$

而平均绝对误差 MAE 的计算公式（4−9）：

$$MAE = \frac{\sum_{i}^{n}|y_i - r_i|}{n} \tag{4-9}$$

4.2.3 参数选取

为了得到基于 BP 神经网络的协同过滤推荐算法的最优推荐结果，应该进行控制变量实验，根据评分预测误差大小来确定输入层和隐含层的神经元个数，用平均绝对误差 MAE 来衡量优劣，实验结果如下所示：

（1）输入层神经元个数 f

BP 神经网络评分预测模型的对应输入为行为类别对应的属性特征向量，因此输入层神经元个数 f 与算法选择的行为类别个数一致，本文采用移动终端用户网络行为作为构建 BP 神经网络评分预测模型时考虑的行为类别特征，由章节 3.1.2 将移动终端用户网络行为划分为 18 个类别，因此输入层与输出层的神经元个数共计 18，而本文主要处理回归预测问题，故输出层神经元个数设置为 1，则输入层神经元个数设置为 17。

（2）隐含层神经元个数 p

隐含层的神经元个数 p 的选取是没有规定的理论经验，因此需要通过实验测试来选择一个合适的较优经验值，本文按照 4:1 的比例划分训练集和测试集，并在给定近邻用户为 50 的情况下进行测试，观察在不同神经元个数下的测试性能，从而选择较优的神经元参数值，实验结果如图 4.2 所示。

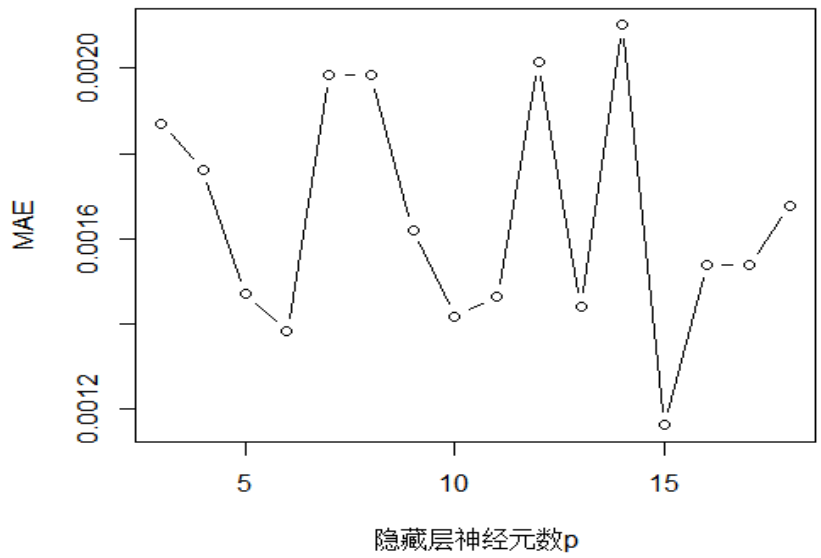

图 4.2 隐含层神经元 p 对应的 MAE

由图 4.2 可知，在给定近邻用户 $K=50$ 的情况下，当隐含层的神经元个数为 15 时，基于 BP 神经网络改进的评分预测结果的误差最小，即隐含层神经元的个数设置为 15。

4.2.4 结果分析

本文针对传统协同过滤推荐算法中评分预测准确性较低的问题，为了有效提

升评分预测效果，挖掘出移动终端用户更精准的个性化差异信息，提出基于 BP 神经网络改进的协同过滤推荐算法，通过控制变量实验，神经网络输入层神经元个数设置为 17 个，设定一个隐含层且隐含层神经元个数设置为 15 个，输出层神经元个数设置为 1 个。为了验证本文基于 BP 神经网络的协同过滤推荐算法（BPCF）优化评分预测的真实效果，设置对照实验组，将齐晶[50]等提出在杰卡德系数中加入惩罚因子改进协同过滤推荐算法（Jaccard−CF）、雷鸣[47]等提出的基于情感分析的协同过滤推荐算法（SACF）作为对照组，当近邻用户K从 5 增至 50 个时，得到的实验结果如图 4.3 所示。

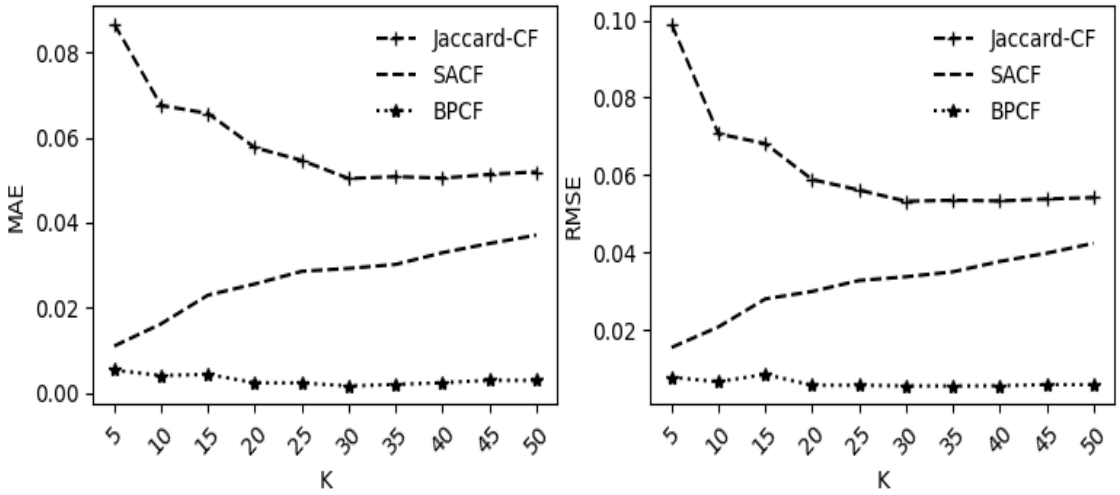

图 4.3 不同算法的 MAE、RMSE 值对比

观察图 4.3 可得，与改进相似度的协同过滤推荐算法、基于情感分析的协同过滤推荐算法相比，基于 BP 神经网络改进评分预测的协同过滤推荐算法的评分预测误差更低，具有更高地评分预测准确性和稳定性，有效提高了其他改进算法的评分预测准确性。因此，对于提高移动终端用户个性化网络服务的推荐效果，本文采用 BP 神经网络进行评分预测可以改善协同过滤推荐算法推荐效果的精确度，且更接近实际情况。

4.3 个性化推荐系统应用

在 APP 的实际应用中，搜索引擎能够满足有明确目标的用户主动查找网络服务，其输出更贴近用户需求，而个性化推荐引擎则可以在用户不确定的情况下，以用户的历史行为资料为基础，帮助用户找到有价值的资讯及服务，更有助于提升用户对网络应用的满意度及体验质量。个性化推荐引擎以应用的形式存在于各

类网络应用中，通过分析用户的行为偏好并智能推荐与众不同的个性化服务，由于用户的需求具有不确定性以及信息的爆炸式增长使得个性化推荐系统的应用经久不衰，根据第三章分析的移动终端用户的网络行为偏好和第四章移动终端网络服务的精准定位，本节将从广告投放、网络信息推送及电子商务营销三个方面实现移动终端个性化推荐系统的应用价值。

4.3.1 实现广告精准投放

随着数字媒介在人们日常生活和工作中的作用日益突出，利用数字媒介进行广告投放已成为一种潮流，然而，随着互联网广告规模的迅速增长和网络广告平台的不断探索，广告的表现形式也日趋多样化，同质化的内容充斥，使得网络广告的效益日益下降，在日益激烈的同质化竞争中，如何让广告获得受众的认同和消费者的购买意愿，除了要有创新的内容外，还必须与互联网相结合，而利用个性化推荐系统精准推送广告，无疑是解决广告投放的新途径。

1、优化时段投放

根据挖掘得到的移动终端用户的上网时间喜好，得到用户在不同时段访问不同类型的网络应用的行为特征，利用 BP 神经网络与协同过滤推荐系统的融合算法定位这一群体对不同网络服务的喜好程度，那么设置广告在某网络应用的投放时间前，寻找目标用户的有效活跃时段，屏蔽其他不相干的上网时段，就可以精准地向对应时段的目标用户投放广告，因此，就企业而言，可以减少广告成本的投入，提高广告的曝光率。

2、优化内容投放

全覆盖式广告推广营销早已与网络平台营销不相匹配，根据网络广告进行精准营销是大势所趋，需要对目标群体进行定位，从而极大地提升了广告的转化率。因此，运营商要想具备较强的商业竞争能力，必须将互联网技术与个性化推荐系统的相结合，通过互联网平台分析移动终端用户的行为特点，并利用个性化推荐系统实时预测用户未来的网络行为，从而基于预测结果将相应的广告内容定向曝光给目标群体，不仅满足用户的需求，也提升用户对网络应用的满意度，最终实现双赢。

4.3.2 网络信息精准推送

在传统传媒时代，大众主要通过电视、报纸、杂志等渠道获取网络信息及资讯，其获得方式相对落后且用户无法自主挑选网络需求，移动互联网技术的兴起，网络媒体可以随时收集并快速传递给用户，使得用户获取信息来源的渠道增多，实现了网络与用户的双向沟通，尽管互联网时代为我们提供了很多方便，但同时也带来了信息超载和难以获取有效的信息。因此，为给移动终端用户带来个性化推荐服务和网络体验的双提升，基于 BP 神经网络的协同过滤推荐技术（个性化推荐引擎）应被广泛推广。个性化推荐引擎技术首先会采集移动终端用户的网络行为信息，结合数据挖掘技术进行智能分析用户的行为偏好，并将其应用到用户的社交和浏览记录中。在使用时根据用户的喜好，对个性化推荐算法进行优化，从而达到个性化的推荐效果，更符合用户的个性化需求，最终提升用户的活跃度与粘性，加快运营商发展网络业务并提升数据运营竞争力。

4.3.3 加速电子商务营销

网络平台上的商品，包括实体商品之外的虚拟物品，例如在线课程、影音会员、游戏装备、网络小说等都属于虚拟物品的范畴，在电子商务领域，亚马逊作为“推荐系统之王”，它的推荐结果是基于对用户的历史行为进行挖掘所分析出的，其推荐结果就能为电子商务平台提供源源不断地商业价值，这也对个性化推荐系统的推广起到了很大的作用。

1、产品策略

对于电商平台的实物商品，首先收集用户的历史行为（包括行为类别、访问时间、访问时长），结合用户的行为估算用户对商品的兴趣度（即评分），在离线阶段，个性化推荐系统会根据用户的反应，通过推荐算法，获得所需的产品清单，甚至提供适当的折扣，这种营销策略不仅能够留存老用户还能发现新用户。

2、增值服务

电商平台中的增值服务属于特色服务，提供超越传统的个性化服务，并确保基本服务，即虚拟商品的个性化推荐。通过个性化推荐系统挖掘出用户对各类型增值服务的兴趣点或者需求，提供相对应的产品曝光给用户，例如为学习型用户

提供在线课程类服务、游戏类用户提供游戏装备、图书阅读类用户提供网络小说等等，另外，个性化推荐还能为更多冷门且高质量的增值内容提供曝光机会，有利于用户的转化与留存双提升。

4.4 本章小结

本章针对传统协同过滤推荐算法中评分预测准确性较低的问题，利用 BP 神经网络优化评分预测的协同过滤推荐算法。首先将移动终端用户历史网络行为数据划分为训练集和测试集，运用 TF-IDF 思想估计移动终端用户访问网络的兴趣度，并在此基础上构建用户—类别评分矩阵；其次，使用余弦相似度为目标用户建立近邻用户群，然后，利用 BP 神经网络算法对传统的协同过滤推荐算法中的评分预测进行优化，且通过控制变量实验确定输入层和隐含层较优的神经元个数；最后，与传统的协同过滤推荐算法、改进相似度的协同过滤推荐算法进行比较，其实验结果表明本文提出的改进算法能够有效提高传统算法对评分预测的准确度，并从广告投放、网络信息推送及电子商务营销三个方面提出个性化推荐系统的应用价值。

5 总结与展望

5.1 总结

移动互联网、大数据、云计算、人工智能等新一代信息技术的快速发展为移动智能终端的应用提供了新动力，不仅给我们的生活带来了巨大改变也让我们淹没于海量信息之中。在这样的环境下，移动终端用户获取有效信息时往往会耗费大量的时间和精力，面对多样化、复杂化的用户需求，如何挖掘移动终端用户的网络行为偏好从而进行用户画像、广告推荐和开展个性化服务具有重要的意义。因此，本文在挖掘移动终端用户网络行为偏好的基础上，将 BP 神经网络与协同过滤推荐算法相融合，优化移动终端用户网络服务的个性化推荐质量，并验证了基于 BP 神经网络的协同过滤推荐算法的个性化推荐效果。下面对本文所做的工作进行总结：

（1）挖掘移动终端用户上网时间的行为偏好，通过验证发现移动终端用户的上网时间在最大上网时段处存在近相似性规律，基于此规律，首先将用户的上网时间分布向量按照最大上网时段划分为 16 个初始分组；其次，根据层次聚类算法得到用户上网时间的分类情况；最后，根据聚类的结果，具体分析各个用户群的上网时间特征，得到用户上网时间的 4 种行为偏好模式：午休上网型用户、上午上网型用户、晚间休闲型用户以及傍晚上网型用户。

（2）挖掘移动终端用户访问内容的行为偏好， 首先利用肘部法则和轮廓系数确定最优聚类数；其次，根据 k-means 聚类算法原理得到用户访问内容的分类情况，根据聚类结果，得到 7 种访问内容的行为偏好模式，分别为社交阅读型用户、游戏社交型用户、娱乐社交型用户、社交新闻型用户、网络购物型用户、学习型用户、新闻关注型用户；最后，将用户上网时间的 4 种行为偏好模式与 7 种访问内容的行为偏好模式相融合，得到用户在不同时间段的上网行为特征。

（3）在挖掘出移动终端用户网络综合行为模式的基础上，结合 BP 神经网络与协同过滤推荐算法提高个性化推荐结果的准确度。首先将用户行为数据按照 4:1 的比例划分训练集和测试集，采用 TF-IDF 算法计算用户对某一类别网络应用的喜好程度，并构建用户—类别评分矩阵；然后采用余弦相似度相关系数法计算相似度，并且通过相似度的高低排序构建目标用户的近邻集合；另外，利用

BP 神经网络根据用户的历史网络行为数据构建评分预测模型，并通过控制变量实验确定较优的神经元个数；最后，通过将本文所使用的个性化推荐算法与改进相似度的协同过滤推荐算法、基于情感分析的协同过滤推荐算法进行对比，研究结果表明本文所使用的推荐算法能够更有效地提高个性化推荐的评分预测准确性。

5.2 展望

本文仅仅是分析了移动终端的用户网络行为较为普遍的一些规律，网络行为分析还有许多值得研究的问题，由于作者的知识储备不全、时间限制等原因，本论文的研究还有许多可以继续探讨的地方。在本文的研究结果基础之上，还可以进一步研究的主要问题有：

（1）用户的网络行为是受多方面因素的影响，因此对于用户行为分析还需要从更多的角度研究，为挖掘用户网络行为模式提供更多的可能性，例如，下一步可以结合用户的年龄、职业、地理位置、上网流量、访问量等因素，利用关联规则算法，从多个角度全面的分析和描述移动终端用户网络行为。

（2）结合移动终端用户对 APP 的权重，设计用户实时个性化服务推荐系统，可以根据用户当前的操作或者搜索的关键词，挖掘用户的网络需求，即时推荐用户感兴趣的内容或者根据网络访问流量进行广告精准投放；或者通过细化移动终端的网络服务类别进一步挖掘用户的行为特征，实现网络信息推荐的高效性与精准性，例如休闲娱乐型网络应用可以细分为短视频、影音视频等。

（3）近年来，许多学者在推荐系统的基础上进行了很多改进和创新，例如提出了矩阵填充、朴素贝叶斯分类和矩阵降维等提高推荐精度和推荐效率，本文在个性化推荐中使用了 BP 神经网络来优化推荐精度，虽然实证中个性化推荐的效果不错，但是模型仍然存在非常大的优化空间。下一步可以通过增加隐含层甚至通过调整输入层和输出层的神经元个数进一步提高个性化推荐精度。

参考文献

[1]Cheng Y, Qiu G, Bu J J,et al Model bloggers' interests based on forgetting mechanism[P]. World Wide Web,2008.

[2]Gerhard Widmer,Miroslav Kubat. Learning in the Presence of Concept Drift and Hidden Contexts[J]. Machine Learning,1996,23(1).

[3]He J,Zhuang F Z,Liu Y C.Bayesian dual neural networks for recommendation[J].Frontiers of Computer Science,2019,13(6):1255-1265.

[4]Humberto T. Marques,Leonardo C. D. Rocha,Pedro H. C. Guerra,Jussara M. Almeida,Wagner Meira,Virgilio A. F. Almeida. Characterizing broadband user behavior[P]. Next-generation residential broadband challenges,2004.

[5]Halvey.A mobile clickstream time zone analysis:implications for real-time mobile collaboration.In Proceedings of KES2004(Volum Ⅱ),volum LNCS 3214 of Lecture Notes in Computer Science,2004:855-861.

[6]I. Barry Crabtree,Stuart J. Soltysiak. Identifying and tracking changing interests[J]. International Journal on Digital Libraries,1998,2(1).

[7]Li L,Yang Z,Wang B.Dynamic adaptation strategies for long-term and short-term user profile to personalize searche[M].Leture Notes in Computer Science,APWeb/WAIM’07,LNCS 4505,Berlin:Springer Berlin.2007:228-240.

[8]Liang M G,Jia K L,Ying Z.Collaborative filtering recommendation based on trust and emotion[J].Journal of Intelligent Information Systems,2018(07):1-23.

[9]LEE S.Using entropy for similarity measures in collaborative filtering[J].Journal of Ambient Intelligence and Humanized Computing,2020,11(1):363-374.

[10]Montjoye Ya D,Quoidbach J,Robic F,etal. Predicting presonality using novel mobile phonebased metrics[M].2013:48-55.

[11]PitkowJE,KehoeCM.Emerging trends in the www user population[J].Communications of the Acm,1996,39(6):106-108.

[12]Pascal Welke,Ionut Andone,Konrad Blaszkiewicz,Alexander Markowetz. Differentiating smartphone users by app usage[J]. Pervasive and Ubiquitous Computing,2016.

[13]Saqib Alam,Nianmin Yao. Big data analytics, text mining and modern english language[J]. Journal of Grid Computing,2019,17(2).

[14]Tan P N,SteinbachM,KumarV. Introduction to Data Mining:Pearson New International Edition PDF eBook[J].Person Schweiz Ag,2013,14(2):279-288.

[15]T.Yamakami. Mobile user drop out ration observation with time zone and Day-of-week Regularity Analysis.In Proceedings of the 2007 International Conference on Multimedia and Ubiquitous Engineering,2007:857-861.

[16]Xu.Q,Erman J,Gerber A,etal.Identifying diverse usage behaviors of smartphone apps[C].Acm Sigcomm Conference on Inernet Measurement Conference,2011:329-344.

[17]Yan J, Liu N, Wang G, et al. How much can behavioral targeting help online advertising[P]. World wide web,2009.

[18]Yan H, Dou Y N, Liu F. Time division based on analyses of network user time span prefence [C].Proceedings of 2009 IEEE International Conference on Network Infrastructure and Digital Content.,2009:213-217.

[19]Zhao S,Ramos J,Tao J,et al. Discovering different kinds of smartphone users through their application usage behaviors[C].In:2016 ACM Interational Joint Conference on Pervasive and Uniquitous Computing,Heidelberg,2016:498-509.

[20]白友东. 基于数据挖掘的网络用户行为分析[D]. 北京邮电大学, 2014.

[21]陈姝彤. 基于用户手机阅读兴趣挖掘的图书采访系统构建[J]. 中国科技信息, 2020(24):32-33.

[22]丛洪杰, 龚安, 李华昱, 帅训波. 基于用户兴趣和项目分类的协同过滤推荐算法[J]. 计算机技术与发展, 2018,28(11):85-88+93.

[23]董富强. 网络用户行为分析研究及其应用[D]. 西安电子科技大学, 2005.

[24]党小超, 郝占军, 王筱娟. 模糊加权 Markov 链的用户行为预测[J]. 兰州大学学报(自然科学版), 2011,47(01):110-115.

[25]符饶. 基于位置服务的潜在好友推荐方法[J]. 软件, 2015,36(01):62-66.

[26]付悦. 基于协同过滤音乐推荐算法的研究[D]. 沈阳理工大学, 2021.

[27]高峰. 基于兴趣分类的用户行为分析系统的研究[D]. 山东大学, 2010.

[28]顾茜. 网络用户特征及行为分析[J]. 数字传媒研究, 2017, 34(10):46-51.

[29]胡明珠. 基于聚类分析的协同过滤推荐算法研究[D]. 长春工业大学, 2021.

[30]黄令贺, 朱庆华, 沈超. 差异与稳定:网络百科用户兴趣动态变化研究[J]. 图书情报知识, 2016(02):101-113.

[31]胡璨. 社交网络用户发布模式分析与兴趣预测研究[D]. 武汉大学, 2019.

[32]黄亮. 网络用户的分析[J]. 科技情报开发与经济, 2003(09):225-226.

[33]互联网实验室. 中国城市居民互联网应用研究报告[J]. 信息空间, 2004(08):61-64.

[34]贺雯静. 校园网用户行为分析系统的设计与实现[D]. 西北大学, 2020.

[35]黄倩, 谢颖华. 一种基于网页浏览行为的用户兴趣度计算方法[J]. 信息技术, 2015(05):184-186+191.

[36]黄良发. 基于移动 APP 行为的用户兴趣度计算[J]. 广东通信技术, 2017,37(05):2-5.

[37]胡旻, 何正宏, 韩伟. 大数据背景下高校网络用户行为分析系统研究[J]. 中国管理信息化, 2020,23(14):178-179.

[38]贾忠涛. 电影个性化推荐系统的研究与实现[D]. 西南科技大学, 2015.

[39]刘建国, 周涛, 汪秉宏. 个性化推荐系统的研究进展[J]. 自然科学进展, 2009,19(01):1-15.

[40]刘丽娟. 网络用户数据挖掘与行为分析[D]. 北京交通大学, 2014.

[41]刘树栋, 孟祥武. 一种基于移动用户位置的网络服务推荐方法[J]. 软件学报, 2014, 25(11):2556-2574.

[42]李旭阳, 邵峰晶. LSTM 与随机森林购买行为预测模型研究[J]. 青岛大学学报(工程技术版), 2018, 33(02):17-20.

[43]李书宁. 网络用户信息行为研究[J]. 图书馆学研究, 2004(07):82-84+101.

[44]李建廷, 郭晔, 汤志军. 基于用户浏览行为分析的用户兴趣度计算[J]. 计算机工程与设计, 2012,33(03):968-972.

[45]李微丽, 罗汝, 颜一鸣. 基于大数据的用户行为分析系统[J]. 科技风, 2020(04):100.

[46]林家民. 基于校园网的用户网络行为分析系统[J]. 电子世界, 2020(08):75-76.

[47]雷鸣, 朱明. 情感分析在电影推荐系统中的应用[J]. 计算机工程与应用, 2016, 52(10):59-63+107.

[48]马彬, 李尚儒, 谢显中. 异构无线网络中基于人工神经网络的自适应垂直切换算法[J]. 电子与信息学报, 2019, 41(05):1210-1216.

[49]彭进香. 大数据背景下互联网用户行为分析[J]. 现代工业经济和信息化, 2019, 9(11):62-63.

[50]齐晶, 刘瀛, 刘艳霞, 胡美振, 乐海丰. 基于标签的协同过滤推荐方法研究[J]. 北京联合大学学报, 2021, 35(02):47-52.

[51]宋丽哲, 牛振东, 余正涛, 宋瀚涛, 董祥军. 一种基于混合模型的用户兴趣漂移方法[J]. 计算机工程, 2006(01):4-6+89.

[52]魏强. 网络用户行为分类的研究[D]. 大连海事大学, 2016.

[53]王飞飞. 移动社交网络微信用户信息共享行为研究[D]. 北京邮电大学, 2018.

[54]王岩, 张杰, 许合利. 结合用户兴趣和改进的协同过滤推荐算法[J]. 小型微型计算机系统, 2020, 41(08):1665-1669.

[55]王锦贵, 王京山. 关于建立网络用户学的思考[J]. 江苏图书馆学报, 2002(03):5-8.

[56]王佳宁. 基于社交网络的用户行为分析[J]. 数码世界, 2017(09):29.

[57]万俨慧, 任晨, 沈敏虎. 基于网络日志的高校用户行为分析[J]. 网络空间安全, 2019, 10(10):49-53.

[58]肖丽媛. 社交网络中的用户行为分析[J]. 中小企业管理与科技(中旬刊), 2019(08):115-116.

[59]杨彬. 基于 Scrapy 的 QQ 空间数据分析研究[D]. 南京理工大学, 2016.

[60]于允飛. 基于用户兴趣划分的推荐算法的研究[D]. 重庆邮电大学, 2017.

[61]中国互联网络信息中心(CNNIC). 2021 年第 48 次中国互联网络发展状况统计报告. 2021.02. http://www.cac.gov.cn.

[62]张玉成, 徐大纹, 王筱娟. 基于加权马尔可夫链的主动用户行为预测模型[J]. 计算机工程与设计, 2011, 32(10):3334-3337+3418.

[63]张雁, 刘才铭. 网络用户的网页访问行为分析架构[J]. 现代信息科技, 2018, 2(09):15-17.

后记

行文至此，落笔为终，百感交集。三载研途，始于 2019 年金秋，终于 2022 年盛夏，感恩遇见。

谆谆告诫，桃李天下。衷心感谢母校提供了舒适整洁的生活环境和营造了浓厚的学习氛围，我喜欢宿舍的上床下桌、喜欢校园里的绿树成荫及百花绽放、喜欢琳琅满目的小超市、喜欢图书馆走廊的朗朗书声、喜欢校园里发生的点点滴滴。

片言之赐，铭记于心。衷心感谢我的导师庞智强教授，庞教授在学术上一丝不苟、精益求精的严谨治学态度让我收获颇多，悉心地指导我在研究生阶段所遇到的问题，在生活中犹如慈父般教我如何做人做事。本篇论文从确定题目、框架至终稿的完成离不开庞老师的细心指导。教诲如春风，日日沐我心，三载师徒缘，十分珍贵，言辞有尽，师恩永记，在此，我要再次向我的导师致以最衷心的感谢和深深的敬意，同时也衷心感谢所有指导过我的老师。

春晖寸草，山高海深。衷心感谢父母多年的养育之恩，他们的悉心培养和教育，使我更加坚强，在我迷茫时指明方向，在我低谷时给予鼓励，家永远是我最温暖的港湾。希望父母在未来的岁月中身体健康、事事顺心，我也会奋力拼搏，不辜负您们的期盼。

山水一程，三生有幸。感谢我的闺蜜王晓清，不管我遇到什么样的烦心事，你都不厌其烦的鼓励和关心我，在春暖花开的日子里，我们一起笑过、哭过、疯过，一切都是那么的淡然美好，这世界上有各种各样的人，恰巧我们成为最好的朋友。感谢 725 的室友们，与我朝夕相处，同宿情谊铭心间。感谢同门在我论文撰写过程中提供的帮助，也让我的学术生活充满乐趣。在这里祝大家前程似锦，未来可期，我们各自努力，他日顶峰相见。

段家滩路 496 号，苍山负雪，明珠天南，有缘再相逢。至此，我的硕士研究生时光全剧终。