

基于学习排序的电信行业潜力用户推荐策略

刘梦迪¹ 刘佳¹ 王越¹ 张清伟¹ 高宗宝¹ 程永明¹

LIU Mengdi LIU Jia WANG Yue ZHANG Qingwei GAO Zongbao CHENG Yongming

摘要

随着市场态势的愈发激烈以及技术层面的不断更新迭代,消费者对服务质量和个性化体验的要求越来越高,鉴于电信领域在精准营销中尚缺乏有效的智能化用户推荐方案,文章提出一种基于排序学习的电信行业潜力用户推荐策略,旨在为电信运营商推荐最有可能响应特定营销活动的目标用户。首先获取海量真实的电信用户数据,对电信用户行为进行深度分析,提取目标电信用户的重要特征,构造复合特征并提出复合特征度量函数。基于机器学习技术 Ranking SVM 算法,将排序问题转化为分类问题,训练特征,学习智能电信用户推荐模型,为电信运营商推荐最具潜力的目标用户群,提高精准营销效率及投资回报率。通过对比不同策略对不同规模的电信用户样本集的推荐结果,基于排序学习的电信行业潜力用户推荐策略具有更高的准确率。

关键词

电信领域;学习排序;潜在用户推荐;精准营销;智能化

doi: 10.3969/j.issn.1672-9528.2024.12.014

0 引言

在数字化时代,电信行业面临着前所未有的挑战与机遇。随着市场竞争的加剧和技术的快速迭代,传统的一刀切营销策略已难以满足用户日益增长的个性化需求,同时也无法有效应对成本控制的压力。在此背景下,精细化营销成为电信运营商提升竞争力的关键。然而,如何在海量用户数据中智能且精准地识别并优先接触最有潜力的客户,成为了电信行业亟待解决的难题^[1-3]。

传统的营销方式,如大众邮件、电话营销等,往往采取广撒网的策略,缺乏针对性,不仅效率低下,而且容易引起用户反感,导致营销成本高昂而回报率低。近年来,随着大数据分析和机器学习技术的成熟,精细化营销逐渐成为电信行业的主流趋势。通过对用户通信电信行为、偏好和消费模式等的深度洞察,运营商能够识别出最具潜力的目标客户群体,从而实现资源的有效配置,提高营销活动的效率及投资回报率。

在精细化营销领域,学习排序(learning to rank, LTR)^[4]技术展现出了巨大的潜力。LTR 是一种机器学习方法,其目标是学习一个模型,用于预测和排序一系列项目的相关性或用户偏好。在电信行业,LTR 可以被用来分析用户数据,预测哪些用户对特定营销活动的响应度最高,从而帮助运营商

制定更为精准的营销策略。然而,尽管 LTR 在多个领域已成功应用,但在电信行业中的实践仍然处于探索阶段,存在许多未被充分挖掘的机会和挑战。鉴于此,本文致力于探索如何将电信行业的大数据资源与先进的 LTR 技术,特别是针对排序问题的支持向量机算法 Ranking SVM^[5-6],进行整合,应用于电信行业的潜在价值用户推荐策略中,通过构建智能化的分析模型,为电信运营商推荐最有潜力的目标客户,以实现营销效率和投资回报率的最大化。接下来的章节将具体讲述 Ranking SVM 在潜在价值用户推荐策略中的应用,以及不同规模的电信用户样本集下预测的准确度。

1 基于学习排序的潜在用户推荐策略

1.1 潜在用户特征分析与复合特征构建

本文中所使用的真实电信用户数据均来自某电信运营商,由于运营商将套餐超支用户及套餐升级用户视为他们的潜在用户,因此本文获取到的潜在电信用户信息包含近半年来套餐超支用户、套餐升级用户的号码及一系列固定用户特征、用户兴趣特征、用户行为特征、消费行为特征等,共获取到 236 个潜在电信用户样本,包含年龄、性别、职业、上网时长、地理位置信息、每月消费费用、订阅服务类型、订阅频率、信用评分、月主叫时长、月被叫时长、月通用流量使用量,月通用通话分钟数、月短信使用量、投诉记录等 47 个特征。

为提高电信用户推荐的效率和准确度,本文对获取的多

1. 中国移动通信集团设计院有限公司山东分公司
山东济南 250000

维特征的电信用户数据进行深度分析,提取与潜在用户推荐有决定性关系的显著特征,创新性地对上述 47 个特征进行提取并融合构建了 5 个复合特征:用户年龄、用户职业、附加服务订阅活跃度、电信行为指数及用户服务体验值,并结合数据分析,定义符合特征度量函数。

用户年龄:根据真实数据分析, $24 \leq \text{年龄} \leq 45$ 的用户经济独立,收入相对较高,对电信服务的需求多样且稳定,包括大量的数据流量、国际漫游、商业通讯等。他们更可能订阅高端套餐和服务,对电信公司的价值贡献较大,因此该年龄段的用户对于运营商来说潜力更大。

用户职业:用户职业分为管理层、专业技术人员、销售人员、教育医疗工作者、自由职业者、工人等。不同职业的用户往往有不同的需求、兴趣和消费习惯,职业种类直接影响他们对特定产品或服务的潜在需求。如企业管理层及专业技术人员可能需经常出差、远程办公、参加视频会议、接收文件等,对大流量数据包、国际漫游服务等有较高的需求;销售与市场人员经常需要电话联系及外出拜访客户,对语音通话、移动数据流量包需求较高;教育工作者在线教学活动较多,对于流量相关的教育优惠套餐,家庭网络计划有较多兴趣。因此,不同职业的用户平均消费水平,对服务多样性、复杂性的需求,对电信运营商的忠诚度,服务续约率等也都存在差异,综合以上各种因素,本文用职业指数 δ_p 来量化职业对电信运营商精准营销的潜在价值,职业指数越大,用户潜在价值越大。经过数据分析,管理层 $\delta_p(0.9) >$ 专业技术人员 $\delta_p(0.8) >$ 销售人员 $\delta_p(0.6) >$ 自由职业者 $\delta_p(0.5) >$ 教育医疗工作者 $\delta_p(0.3) >$ 工人 $\delta_p(0.2)$ 。并采用独热编码对用户职业数据进行处理,将分类变量转化为数值型数据。

附加服务订阅活跃度:不同的订阅种类可能指向不同的用户群体,电信运营商通过用户订阅的附加服务多样性,可以了解用户的偏好和需求。例如:视频流媒体订阅表明用户对娱乐内容感兴趣;在线教育课程订阅表明用户有自我提升的需求;高频次服务订阅者可能是该服务的忠实粉丝,需要投入更多资源维护并挖掘更大潜力;非长期订阅但隔一段时间又会重复订阅的用户可能对该服务有兴趣,但对服务费用存在顾虑。如果为其推荐个性化或者更优惠的服务,极大可能性将其发展成为新的潜力用户;附加服务订阅活跃度结合了附加服务订阅种类及附加服务订阅频次两个特征,用以下评分函数量化附加服务订阅活跃度在电信用户潜在价值判定中的评分。基于现网历史数据来看,运营商更希望为其推荐乐于尝试更多附加服务种类的用户,以便为用户推荐更多相似的服务,所以附加服务订阅多样性的权重要比附加服务订阅频次权重更大。

$$V_s = \omega_n \cdot \text{DS} + \omega_f \cdot \text{SA} \quad (1)$$

式中:DS 是月平均附加服务订阅种类数;SA 是月平均附加服务订阅次数; $\omega_n = 0.7$ 是月平均附加服务订阅种类数的权重; $\omega_f = 0.3$ 是月平均附加服务订阅次数的权重。 $V_s \geq 2.1$ 基本可将该用户推荐给运营商作为精准营销对象。

电信行为指数:该特征结合了月平均主叫通话时长、月平均通用流量使用量、月平均短彩信使用量及月消费费用 4 个特征。高通话时长表明用户对语音服务有较高依赖性,如果运营商为其提供性价比更高的语音服务,这项服务很可能被用户接受并愿意支付更多费用,比如销售人员或者商务人士;流量使用量是衡量用户对数据服务需求的重要指标,高流量可能是移动互联网的重度使用者,他们对于数据套餐升级或额外的服务感兴趣,比如视频流媒体爱好者或者远程办公人员;随着微信、QQ 等即时通讯软件的普及,短彩信的使用频率不断下降,但仍存在特殊需求用户需使用短彩信进行交流,如果使用量较高,也很可能是运营商需要发展的潜力用户;月消费费用是可直接用于衡量用户价值的指标,高消费用户是对运营商收入贡献最大的用户,他们可能更倾向于使用 VIP 服务或者订阅附加功能,需要运营商尽力维持并为其提供专属客服支持或推荐更丰富的套餐选项。为方便月平均主叫通话时长、月平均通用流量使用量、月平均短彩信使用量及月消费费用 4 个特征的计算,用 z-score 方法对数据进行标准化,并用以下评分函数量化电信行为指数在电信用户潜在价值判定中的评分。

$$V_d = \alpha_t \cdot \text{TS} + \alpha_g \cdot \text{FL} + \alpha_m \cdot \text{MS} + \alpha_c \cdot \text{CT} \quad (2)$$

式中:TS、FL、MS、CT 分别为标准化后的月平均主叫通话时长、月平均通用流量使用量、月平均短彩信使用量及月消费费用; $\alpha_t = 0.2$ 、 $\alpha_g = 0.3$ 、 $\alpha_m = 0.1$ 、 $\alpha_c = 0.4$ 分别为其权重。 $V_d \geq 2$ 基本可判定该用户潜在价值较高,可推荐给电信运营商作为目标用户。

用户服务体验值:投诉情况及投诉后满意度是衡量运营商服务质量的标准,高频投诉及低满意度既表明用户对服务的不满,又反映出用户对运营商本身的忠诚度和期望。用户服务体验值结合了月平均投诉次数及平均投诉结果满意度两个特征。运营商可以通过改善服务质量、为其提供个性化服务或者提供优惠套餐升级的方式将其转变为潜力用户。本文定义了以下评分函数,用于评定用户服务体验值在电信用户潜在价值上的贡献。

$$V_e = \gamma_c \cdot \frac{1}{\text{CN}} + \gamma_s \cdot \text{SF} \quad (3)$$

式中:CN、SF 分别为投诉次数和投诉满意度,取值区间为 (0~1); $\gamma_c = 0.4$ 、 $\gamma_s = 0.6$ 分别为其权重。 $V_e \geq 0.5$ 基本可判定该用户为潜力用户。

1.2 学习排序驱动的潜在电信用户推荐

1.2.1 学习排序

学习排序属于监督学习技术，核心是机器学习，其目标是学习一个评分函数，该函数能够对输入的项目（或项目组）进行评分，从而产生一个按重要性或相关性排序的列表。排序学习广泛应用于信息检索、推荐系统和广告排名等领域。学习排序有 3 种主要的训练方式：PointWise、PairWise、ListWise。PairWise 的方法是将同一个查询中两个不同的项目作为一个样本，主要思想是把排序问题转换为二元分类^[7-10]问题。对于同一查询的相关文档集中，对任何两个标签的文档，都可以得到一个训练实例 (d_i, d_j) ，如果 $d_i > d_j$ 则赋值+1，否则-1，这样就能获得二元分类器所需的训练样本，PairWise 排序方法公式为：

$$\begin{pmatrix} d_1^{(i)}, 5 \\ d_2^{(i)}, 3 \\ \vdots \\ d_{n^{(i)}}^{(i)}, 2 \end{pmatrix} \xrightarrow{\text{transform}} \left\{ (d_1^{(i)}, d_2^{(i)}), (d_1^{(i)}, d_{n^{(i)}}^{(i)}), \dots (d_2^{(i)}, d_{n^{(i)}}^{(i)}) \right\} \quad (4)$$

$$5 > 3, 5 > 2, \dots 3 > 2$$

常用的 PairWise 实现算法有 RankNet、LambdaRank、RankingSVM 等。

(1) RankNet 不同于传统的监督学习，通过比较文档对来训练模型，预测两个文档的相对排序。RankNet 的架构包含输入层、隐藏层及输出层 3 层，每个文档表示一个特征向量，通过网络计算其得分。为了训练模型，RankNet 使用了交叉熵损失函数最小化预测排序与 groudtruth 之间的差异。该算法能够处理非线性的关系，并且可以用于处理大规模数据集。

(2) LambdaRank 是基于梯度提升树（GBDT）的学习排序算法，它直接优化排序指标，LambdaRank 引入了“lambda”权重，用于衡量每次排名变动对排序指标的影响。LambdaRank 在每次迭代中都会更新叶子节点的预测值，以最大化排序指标的改进。这意味着模型会更加关注那些能显著提高排序质量的样本。该算法在提高最终排序质量方面往往表现优异。

(3) RankRanking SVM 的基本理念是把排序任务转变成一系列两两比较的分类问题，这种方法的核心在于识别出哪些项目应该排在其他项目之前。具体来说，算法会构建许多由训练集中元素组成的配对，其中每个配对代表了一个“偏好”——即一个元素应该比另一个元素排名更靠前。接下来，Ranking SVM 运用支持向量机（SVM）这一强大的分类工具来学习这些偏好关系，从而确定一个能够正确反映所有训练配对中隐含排序顺序的决策边界。该算法的数学目标是最大化分类间隔的同时最小化配对间的分类错误，这通常通过一个特定的优化函数来实现，旨在找到一个最优的权重向量，以确保正例（即正确的排序配对）和负例（即错误的排序配对）

被正确分类。该算法适用于线性可分的数据，并且可以高效地处理高维特征空间。

相比传统的分类方法，Ranking SVM 具有以下优势：试图使高潜力用户评分低于低潜力用户，直接针对排序性能进行优化；更关注于样本之间的相对顺序，而不是绝对的分类边界，在处理噪声数据时表现更好；更适用于中等大小的数据集，效率更高；允许采用不同的损失函数，以适应特定场景下的需求等。

1.2.2 Ranking SVM 在电信行业潜力用户推荐策略中的应用

电信行业潜力用户的推荐策略本质是一类二元分类问题，即从电信用户群体中辨识出具有高潜力的用户，同时排除那些不符合推荐标准的用户。鉴于目前电信用户数据集大小适中且数据不平衡，加之用户行为特征较多，我们决定采用 Ranking SVM 算法来处理电信行业潜力用户推荐问题。本文借鉴了 Ranking SVM 算法的优势，利用复合特征对用户行为进行量化评估，为每位用户分配一个权重得分。这一得分反映了用户在潜在价值或参与度上的估计水平。

基于学习排序模型的电信行业潜力用户推荐策略的首要且重点工作是在电信大数据中提取潜力用户特征，本文显著特征提取及复合特征构建过程已在 1.1 节详细介绍。紧接着，基于获取的电信用户样本数据，构建训练集及测试集。训练集作为输入，构建初始模型。将训练集分为 k 个子集，轮流将其中一个子集作为验证集输入到初始模型中，重新训练模型并进行模型参数调优。将测试集输入到排序学习模型中，对模型进行评估，确保模型的泛化能力。本文排序学习模型训练过程如图 1。

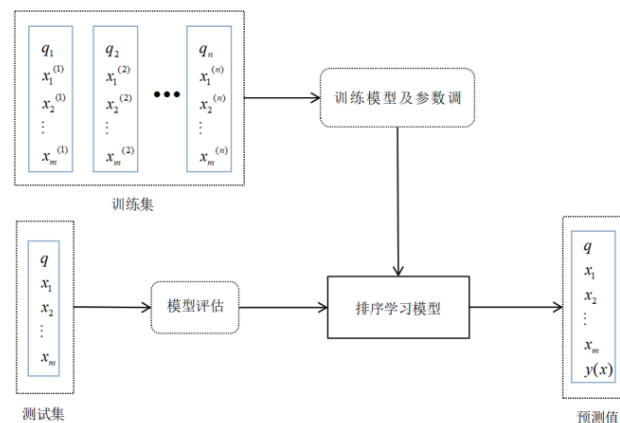


图 1 排序学习训练过程

图 1 中， $q_1, q_2, \dots, q_n$ 代表每个电信用户，在文档中代表每个用户的电话号码， $x_1, x_2, \dots, x_m$ 为每个用户行为特征组合，基于获取到的数据，提取了 10 个单个显著的特征，并构建复合特征。 $y_1, y_2, \dots, y_n$ 代表每个用户的潜力评分，分值越高表示用户的营销潜力越大。

2 实验对比分析

在本文中,电信用户数据集总计包含 1 685 744 条记录,其中 3261 条被标识为潜力用户,738 条作为测试集,20 000 条作为训练集。为验证 Ranking SVM 相对于其他算法的有效性,再次进行多轮实验,对比 Ranking SVM、RankNet 和 LambdaRank 在不同数据规模下的推荐准确性。通过专注于分析排名前 500 名的结果,以检验这些算法识别潜力用户的能力。实验结果显示,Ranking SVM 算法在前 500 名的推荐列表中,491 条数据是潜力用户;RankNet 和 LambdaRank 算法前 500 名的推荐列表中分别只有 54 条和 130 条是潜力用户。这凸显了 Ranking SVM 在潜力用户推荐方面的准确率。当缩减训练集至 15 000 条记录时,观察到 Ranking SVM 依然保持了其优势,前 500 名推荐中有 482 条是潜力用户,而 RankNet 和 LambdaRank 则分别有 49 条和 127 条。因此,即便在数据量减少的情况下,Ranking SVM 算法依旧能够高效且准确地筛选出高价值的电信用户,实验结果如图 2 所示。

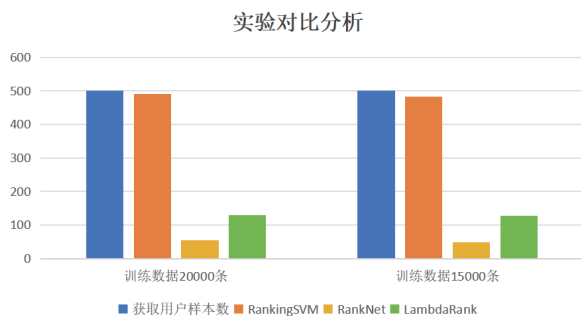


图 2 Ranking SVM、RankNet、LambdaRank 实验对比结果

3 结语

面对电信行业的动态竞争格局与消费者对卓越服务及定制化体验的追求,本文引入基于排序学习的电信用户推荐策略,为精准营销开辟了新路径。通过对电信用户大数据集进行精细剖析,精确定位了驱动用户响应的核心属性,为提升模型训练效率及预测精度,本文对核心属性进行深入分析,创新性地构造复合特征并提出复合特征度量函数。并借助 Ranking SVM 算法构建了先进的预测模型,精准锁定高潜力客户群体。实证研究表明,不论是在小型还是大型数据集上,这一策略均能显著优化用户推荐的精确度,从而极大增强电信企业的营销效能。面对未来市场的不确定性,持续的技术创新和以用户为中心的服务理念将是电信企业保持竞争力的关键,将致力于融合新兴技术和优化现有模型,以适应瞬息万变的趋势,为电信行业者打造更智能、更贴合用户需求的推荐体系。

参考文献:

- [1] 范斌鑫. 电信行业用户个性化套餐智能推荐模型及应用研究 [D]. 天津: 河北工业大学, 2019.
- [2] 兰慧. 基于消费行为的电信套餐推荐模型研究 [D]. 成都: 电子科技大学, 2013.
- [3] 王婷. 基于电信客户消费行为的手机套餐质量研究 [D]. 哈尔滨: 哈尔滨工业大学, 2020.
- [4] 黄震华, 张佳雯, 田春岐, 等. 基于排序学习的推荐算法研究综述 [J]. 软件学报, 2016, 27(3): 691-713.
- [5] CHEN F Q, YU Z T, WU Z J, et al. Expert ranking method based on ListNet with multiple features[J]. Journal of beijing institute of technology, 2014,23(2):240-247.
- [6] LIU T Y. Learning to rank for information retrieval[J]. Foundations and trends , 2009,3(3): 225-331.
- [7] 程学旗, 兰艳艳. 网络大数据的文本内容分析 [J]. 大数据, 2015, 1(3):62-71.
- [8] FANG X, HU P J, LI Z P, et al. Predicting adoption probabilities in social networks[DB/OL]. (2013-09-24)[2024-01-11].<https://doi.org/10.48550/arXiv.1309.6369>.
- [9] MA H, ZHOU T C, LYU M R, et al. Improving recommender systems by incorporating social contextual information[J]. ACM transactions on information systems (TOIS), 2011, 29(2): 1-23.
- [10] HESS C, SCHLIEDER C. Trust-based recommendations for documents[J]. AI communications, 2008, 21(2-3):145-153.

【作者简介】

刘梦迪(1993—), 女, 山东济南人, 硕士研究生, 研发工程师, 研究方向: 人工智能、模式识别。

刘佳(1992—), 男, 山东威海人, 硕士研究生, 研发工程师, 研究方向: 人工智能、模式识别。

王越(1992—), 男, 山东淄博人, 硕士研究生, 研发工程师, 研究方向: 人工智能、模式识别。

张清伟(1990—), 男, 山东聊城人, 硕士研究生, 研发工程师, 研究方向: 数据挖掘、算法优化。

高宗宝(1990—), 男, 山东诸城人, 硕士研究生, 研发工程师, 研究方向: 数据挖掘、算法优化。

程永明(1986—), 男, 山东济南人, 硕士研究生, 研发工程师, 研究方向: 数据挖掘、算法优化。

(收稿日期: 2024-09-03)