

FinTech 时代基于大数据挖掘的 银行产品推荐模型研究

中国民生银行、对外经济贸易大学统计学院、云南大学大数据研究院 王彦博

首都经济贸易大学信息学院 陈喆辉

对外经济贸易大学统计学院 胡丽君 高璐 文会党

FinTech 时代,产品推荐模型研究是商业银行实现精准营销的关键所在。针对此问题,人们在大数据挖掘技术发展基础上,提出了不同的推荐算法。本文着重探索利用 K-means 聚类算法结合关联规则挖掘中支持度和置信度的思想,形成基于客户细分的产品推荐模型。实证分析结果表明该模型具有较好的应用效果。

一、大数据挖掘与产品推荐

随着国内银行业的竞争压力日益增大,各商业银行为了满足不同类型的客户需求,不断推出各种银行产品。在此背景下,银行需要一套行之有效的模型来为客户精准、快捷地推荐相关产品,实现银行与客户的共赢。基于大数据挖掘技术发展,目前应用较为广泛的产品推荐模型主要基于关联规则挖掘技术和协同过滤技术。关联规则挖掘技术于 1993 年被提出,其最为经典的算法是 Apriori 算法。协同过滤技术于 1992 年被提出,随后逐步发展完善。本文结合商业银行应用实际,探索将大数据挖掘聚类技术与关联规则挖掘技术相结合,提出基于客户细分的产品推荐模型。实证分析结果表明,基于客户细分的产品推荐模型的应用效果是随机推荐的两倍以上,模型收效显著。

二、产品推荐模型及应用

产品推荐的主要目的是给客户推荐他们偏好并且有

较高概率会购买的商品。随着数据量的增多和产品营销需求的加大,产品推荐模型也逐步成为业界和学界所关注的焦点问题之一,人们渴望寻找一种易于实现且推荐精准度高的模型来实现产品的精准营销。

1. 基于关联规则挖掘的产品推荐

关联规则挖掘技术通过经典的“尿布·啤酒”规则阐述了产品推荐的概念,其经典算法为 1994 年提出的 Apriori 算法,此后相关研究工作向负关联规则挖掘、加权关联规则挖掘、数量关联规则挖掘、分配规律挖掘、序列模式挖掘、分类关联规则挖掘等领域不断深化演进。有研究表明可以通过客户过往购买产品的兴趣度来向其推荐未来可能感兴趣的产品;而个性化推荐应考虑运用强关联规则的概念,以有效避免冗杂的噪音规则。

基于关联规则挖掘的产品推荐模型主要包含两个阶段的计算,一是寻找过往全部客户中购买产品组合的频繁项集,所谓频繁项集是指某一产品组合在事务数据库中出现的频率较高;二是从频繁项集中提取产品间的关联规则,若某一关联规则的置信度较高则认为此关联规则是可信的,可用于客户产品推荐。然而在实际应用中,该技术往往存在产品推荐规模较大、需要生成的频繁项集较多、整体推荐效率不高等问题。

2. 基于协同过滤的产品推荐

基于协同过滤的产品推荐一般采用最近邻技术,利用客户的历史喜好信息计算客户之间的距离,然后选取

目标客户其最近邻客户对产品的加权评价值来预测目标客户对特定产品的偏好程度,从而支持对目标客户进行产品推荐。基于协同过滤的产品推荐可以进一步分为基于用户的协同过滤推荐以及基于项目的协同过滤推荐。

基于用户的协同过滤推荐其基本思想是基于客户相似度分析,根据目标客户的最近邻客户对某一产品的选择,从而逼近目标客户的偏好。首先需要找到与目标客户兴趣相似的客群,需要计算客户之间的相似度,这里的相似度是指两个客户的兴趣相似度,一般通过 Jaccard 公式或余弦公式来计算。其次根据计算出的客户相似度,为客户 A 推荐与他最为相似的客户 B 所持有的差异化产品,以及与客户 A 次为相似的客户 C 所持有的差异化产品。基于用户的协同过滤推荐主要适用于客户数量稳定、业务实时性较强且产品更新换代较快的业务场景。然而该技术存在的不足主要是当客户数量过多、客户集更新较快时,将造成计算量过大、客户相似性矩阵不易于维护等问题。

基于项目的协同过滤推荐其基本思想是基于物品相似度分析,根据客户对目标产品的最近邻产品进行估计,从而逼近客户对目标产品的偏好。首先计算物品之间的相似度,一般也是通过 Jaccard 公式或余弦公式来计算。其次根据物品相似度分析以及客户的历史行为向客户推荐相似度最高的一个或几个产品。基于项目的协同过滤推荐适用于产品数量稳定且产品数量低于客户数量的应用场景。然而该技术存在的不足主要是很可能会来自于不同领域的产品误判为具有较高相似性。

三、基于客户细分的银行产品推荐模型

本文基于大数据挖掘技术发展,探索使用聚类技术作为关联规则挖掘技术数据预处理方法,以提高产品推荐的精准度。尤其当客户样本量较大时,聚类技术可以将客户根据相似性原则聚为不同的客群,然后在此基础上使用推荐技术,将有助于提高产品推荐的响应速度和

精准程度。

本研究采用 K-means 算法来实现聚类分析。K-means 算法是一种经典的聚类技术,它将一个空间中的若干个数据点聚成 K 个群。K-means 算法是一个重复移动聚类中心点的过程,通过迭代的方法,逐次更新各群的中心值,直至得到最终聚类结果。K-means 算法一般以欧几里得距离作为样本间相似度的评价指标,其基本思想如下:在给定数据集中任意选择若干(K)个样本作为初始聚类中心点,计算数据集中各样本与这些数据中心点之间的距离,将各样本归属到距离最小的聚类中;再通过计算已经归属到各个聚类中的样本平均值来更新各个聚类的中心点;再次计算数据集中各样本与这些新的数据中心点之间的距离,将各样本归属到距离最小的聚类中;再通过计算已经归属到各个聚类中的样本平均值来更新各个聚类的中心点;不断重复,直到聚类中心点不再发生变化,或变化程度已经小于预设指标或重复轮数已经达到预设指标。

在关联规则挖掘技术应用方面,本研究做了简化处理。传统基于关联规则挖掘的推荐技术是在全部客户中采用支持度和置信度指标来挖掘产品之间的关联关系,从而向客户进行产品推荐,其中关键的一步是生成所有的频繁项集,当支持度设定较高时将难以找寻到新颖的规则,而当支持度设定较低时很可能产生海量的噪音规则,而从中寻找出强关联规则的效率是很低的。由于关联规则挖掘技术的局限性,考虑采用更具有针对性的关联规则推荐,即若一个客户持有 A 和 B 两种产品,则寻找覆盖 A、B 两种产品的客户群体,将客群内除 A、B 以外的 N 个众数产品推荐给该客户(本研究将 N 选为 2)。有针对性的关联规则推荐是对经典关联规则挖掘技术的简化处理,但仍没有利用客户的其他信息,因此建议考虑与 K-means 聚类技术相结合以引入客户其他信息参与计算。所谓“物以类聚,人以群分”,本研究先利用 K-means 聚类技术进行客群细分,而后

表 1 产品推荐胜率

客群 ID	产品 1	产品 2	1 个产品的推荐胜率	2 个产品的推荐胜率
客群 1	i6	i5	0.1553	0.3204
客群 2	i3	i6	0.2332	0.3281
客群 3	i6	i11	0.2674	0.3919
客群 4	i11	i5	0.1528	0.2986
客群 5	i11	i12	0.3115	0.3279
客群 6	i11	i13	0.3158	0.3421
客群 7	i6	i3	0.0909	0.3387
客群 8	i6	i7	0.0957	0.2589
客群 9	i6	i3	0.1224	0.3184
客群 10	i6	i3	0.1988	0.2733
客群 11	i3	i5	0.1939	0.2727
客群 12	i6	i3	0.1204	0.3148
客群 13	i11	i6	0.1221	0.4854
客群 14	i9	i8	0.0211	0.3263
客群 15	i11	i6	0.1187	0.3288
平均推荐胜率			0.1680	0.3284

在不同客群中采用有针对性的关联规则推荐来实现产品推荐。

在实证分析方面，本研究以国内某商业银行为例，从该银行的企业级数据仓库中提取了公司业务全量客户数据，并针对 15 种主流产品进行基于大数据挖掘的产品推荐建模。用以描述客户的数据涉及 141 个字段属性，利用 K-means 算法将全部客户样本聚类为 15 个客群。针对每个客群，将群内客户较为普遍持有的两项产品向群内还未持有该两项产品的客户进行推荐。以形成产品推荐策略后的三个月数据作为验证数据集进行评估，将所推荐的产品与真实的产品持有变化情况进行对比，由此获得产品推荐的准确率。表 1 以向客户推荐一个产品及两个产品后相关客户至少响应持有一个产品来计算产品推荐胜率。

通过上述分析可见，基于客户细分的产品推荐模型能够有效提升推荐的精准程度，从平均推荐胜率来看，15 个客群的两个产品推荐胜率为 32.84%，且针对各类客群的推荐胜率较为稳定，是随机推荐胜率 13.33%（即从 15 个产品中随机挑选 2 个产品向客户推荐，并最终有 1 个产品被客户响应）的 2.46 倍。

在 FinTech 时代背景下，本文基于大数据挖掘技术发展，针对银行产品推荐模型进行了研究，并结合银行业务实际提出了基于客户细分的产品推荐模型，利用 K-means 聚类技术将客户进行细分，而后利用关联规则挖掘中支持度和置信度的思想向相关客户推荐产品。实证分析结果表明，基于客户细分的产品推荐模型其产品推荐精准度是随机推荐的两倍以上，支持以客户精准营销实现银行与客户的价值共赢。FCC

栏目编辑：焦卢玲 jiaoluling@fcc.com.cn