

doi:10.3969/j.issn.1672-4348.2018.03.009

基于交易数据的券商个人客户风险偏好识别

肖琳^{1,2}

(1.福建工程学院 信息科学与工程学院, 福建 福州 350118;

2.福建省大数据重点实验室, 福建 福州 350118)

摘要: 针对个人客户,基于客户的交易标的,应用主成分分析法、改进的 K-means 聚类算法为某证券公司的个人客户建立投资风险偏好模型。研究表明:该模型可为券商的客户细分提供精细化数据支撑,提升券商对客户风险偏好的认识,促进券商营销的精准度和成功率,为券商制订科学的营销策略提供决策支持。

关键词: 客户分类; 风险属性; 聚类分析; 券商

中图分类号: TP39; F830.9

文献标志码: A

文章编号: 1672-4348(2018)03-0248-05

Research on the risk preference recognition model of individual customers of brokerage firms based on transaction data

XIAO Lin^{1,2}

(1.School of Information Science and Engineering, Fujian University of Technology, Fuzhou 350118, China;

2.Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fuzhou 350118, China)

Abstract: An investment risk preference model was established for individual customers of a brokerage firm, which was based on the customers' transaction targets, the principal component analysis method and the modified K-means clustering algorithm. Results show that the model can provide detailed data support for brokerage firms to classify individual customers and improve their understanding of customers' risk preferences, and promote their marketing accuracy and success rate, thus providing decision support for the firms to formulate scientific marketing strategies.

Keywords: customer classification; risk attributes; clustering analysis; brokerage firms

证券投资者不断地成长、成熟,逐渐认识到了证券市场难以预测与把握,并反思自己的投资行为和理念。尤其是个人客户开始倾向于思考自己的风险特征,关注与其风险属性相匹配的产品和服务。另一方面,在实际营销中,券商可提供的产品众多,如何体现客户对产品的偏爱和找到匹配的客户,无疑是增加营销成功率的关键。券商必须通过调研,了解客户的实际状况与需求,才能正确地评价客户并为之提供合适的产品或服务。券

商可以从客户外在行为中挖掘出他们内在的特征和需求,并据此对客户进行分类管理。这也是证券行业监管规定和要求^[1]。

已经有许多学者将数据挖掘技术应用于证券行业的相关研究。如吴峰用概念聚类的方法分析客户的交易行为,针对客户买卖行为,得出影响客户盈亏的一般性规律^[2]; R.J. Kuo 对数据进行新二阶段法的处理,以此开展客户分类的研究^[3]; 梁敏君描述了分形聚类方法对客户在资金、贡献

收稿日期: 2018-04-18

基金项目: 福建工程学院青年基金项目(CY-Z15092); 福建省教育厅中青年教师科研项目(JAS160355); 福建省教育厅教育规划重点项目(FJJKCG15-051)

作者简介: 肖琳(1980-),男,福建周宁人,讲师,博士研究生,研究方向:大数据应用与知识管理。

度、交易频率方面的分类研究^[4]。钱维佳提出了证券客户细分的多维模型,从客户贡献度、忠诚度、活跃度、收益率、资产、响应度、依赖度、套牢度、风险度 9 个方面来划分客户^[5]。尔古打机等提出基于用户行为分析的移动终端偏好模型研究^[6]。以上研究多数基于客户价值、客户生命周期、客户忠诚度等角度来进行客户分类管理,对证券公司客户基于投资标的风险偏好研究较少。少量涉及到投资偏好的研究也比较局限,不能突出客户的风险偏好特征。因此本文提出基于个人客户风险偏好识别的模型非常有意义。

1 问题描述和模型构建

本文是对证券公司个人客户的风险偏好识别提出 1 种模型。其主要思路是对券商的个人客户交易数据进行清洗、提取主成分;再运用数据挖掘的算法,开展深层次的聚类与分析,从而发现隐藏的客户风险偏好的相关规律。图 1 为模型框架。

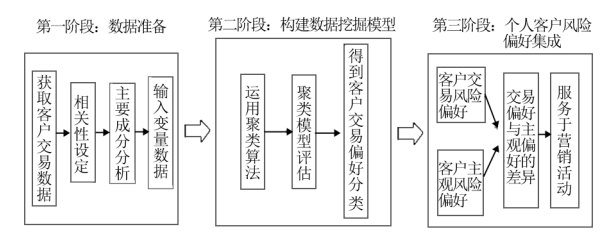


图 1 个人客户风险偏好识别模型

Fig.1 Individual customers' risk preference recognition model

第 1 阶段:数据准备。本文使用主成分分析来简化数据。原始数据是某证券公司的真实交易数据。主成分矩阵数据是输入变量。

第 2 阶段:建立数据挖掘模型。建立数据挖掘的循环迭代模型。利用改进的 K-means 算法得到客户的交易类别划分结果。

第 3 阶段:客户风险偏好集成。一方面基于上述的聚类结果,即客户的交易类别划分,结合不同产品的风险值,可确定客户的交易风险偏好。另一方面,通过证券交易系统的在线问卷调查可以获得客户的主观风险偏好。二者相结合,可以得到客户的风险偏好的最终结果。

1.1 数据准备步骤

应用主成分分析的目的是减少数据集的维

数,先将原始数据样本进行 z 标准化变换消除量纲影响,再求其相关系数矩阵,最后做出成分分析得到成分矩阵。

定义 1 对于有 n 个样本的原始矩阵 X , n 为样本数, p 为变量数。

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$$

对 X 进行 z 标准化变换的公式如公式(1)所示。

$$x_{aj}^* = \frac{x_{aj} - \bar{x}_j}{\sigma_j} \tag{1}$$

其中 $\bar{x}_j = \frac{1}{N} \sum_a x_{aj}$, $\sigma_j^2 = \frac{1}{n} \sum_{a=1}^n (x_{aj} - \bar{x}_j)^2$, $j \in 1, 2, \dots, p$ 。 x_{aj} 为矩阵 X 的原始数据。需要对数据进行消除量纲操作,应用公式(2)计算可以得到相关系数矩阵,即 X 的协方差阵。

$$R_{ij} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}} \tag{2}$$

对协方差阵 R 计算特征值和正交化单位特征向量 a_i , $i = 1, 2, \dots, p$ 。由特征向量 a_i 按公式(3)计算可得各综合向量指标 F_i 。

$$F_i = a_{1i}x_1 + a_{2i}x_2 + \cdots + a_{pi}x_p, i = 1, 2, \dots, p. \tag{3}$$

由各综合向量指标可计算其方差贡献率,再根据累计的贡献度来筛选出主成分。从而得出样本的主成分矩阵。

1.2 改进的 K-means 聚类算法

本文采用改进的 K-means 算法对客户样本主成分矩阵进行聚类分析。K-means 算法是数据挖掘中经典的聚类分析法^[7]。由于证券公司客户样本数据庞大,K-means 比较适合应用于证券公司的客户分类。

传统 K-means 算法步骤:

- 1)利用随机函数,在样本范围内选取 k 个点作为初始聚类中心。
- 2)计算各个样本到聚类中心的距离,把样本归到离它最近的那个聚类中心所在的簇。

- 3) 对各簇重新按均值计算新的聚类中心。
- 4) 回到 2) 步骤, 循环迭代直到各类聚心稳定。

本文算法改进的部分是初始聚心的选择策略和删除临近大簇的小簇 2 个方面。

(1) 改进的初始聚心选择策略

本文选择初始聚类中心的方法为: 先取每 1 种成分中最大值的对象作为 m 个初始聚心, 接着从其余的样本对象中再找出新的聚类中心, 依据是与现有的全部聚类中心的距离之和最大, 循环选取, 直到聚类中心总数满足 k 为止。这样得到的初始聚类中心在样本中的分布相对均匀且受样本的输入顺序影响较小, 缺点是牺牲了一点时间, 但这对于现代计算机来说, 可以忽略不计。

(2) 聚类数目 K 的删减

原始的 K-means 算法是没有删减聚类中心的^[2]。本文的改进是每轮聚类结束对每个簇中对象数目进行统计, 对任意最邻近的两个簇, 如果满足(4)式则删除簇 i 的簇心。

$$\text{num_}C_i < \text{num_}C_j \text{ 且 } d_C_{ij} < 2 * \max d_C_i \tag{4}$$

式(4)解释为编号 i 的簇中对象数目小于其最邻近的标号 j 的簇的对象数, 且两个簇的簇心距离要小于 2 倍的第 i 簇簇心到其最远对象的距离。

若簇 i 被删除, 这个簇的对象在下次聚类中按距离远近原则重新被分配到最临近的其他簇中, 这样的目的是使簇数量按邻近原则合并达到一个相对合理值。如图 2 所示的簇 C_1 和 C_2 将有一个被删除, 在下次聚类时对象将合并为一个簇。这样考虑是为了合并紧密的簇。而有些簇的对象数目较少但与其他簇距离偏远, 具有独立特征的, 不易满足上述距离的不等式, 因此不会被删除。

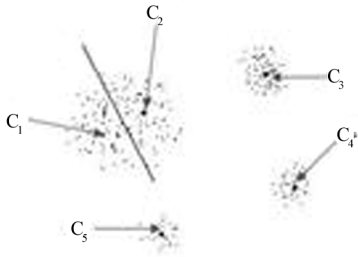


图 2 C_1 与 C_2 合并
Fig.2 Merge C_1 and C_2

为了易于实现算法, 本文设计的一个簇结点。其中节点的信息包含簇编号、簇心位置、该簇对象数目、最邻近簇心编号、距簇心最远的对象距离值、下一个簇节点地址。

改进后的 K-means 算法流程为:

- 1) 确定 K 个簇的初始聚心。
- 2) 找出每个簇的最邻近簇, 并计算其与最邻近簇簇心之间距离。
- 3) 将所有的样本对象分配到最邻近簇, 同时更新每个簇对象数目及簇中与簇心最远的对象距离值。
- 4) 对象分配结束, 按均值调整每个簇簇心, 若所有簇心位置无调整, 算法结束。

5) 按照式(4) 计算每个簇, 删除满足条件的簇心, 每删除一个簇心, K 数目减 1。修改相应的链表结点指针域保持链表完整。回到步骤 2)。

为了避免特殊的噪声点在上述算法中引起簇不断合并为一个簇。可以在步骤 4) 中增加阈值 t , 满足簇的总数目小于 t 时, 停止算法。通过几次训练, 也可以得到合理的 t 值设定。

2 实证分析

2.1 数据准备

以中国 X 证券公司为例, 使用了该公司 2015 年 4 月到 6 月的数据。原始样本矩阵经过公式(1)–(3)处理后, 得到各成分的方差累积贡献度如表 1 所示。表 1 显示的是前 10 项综合指标解释的总方差可以代表原有信息的 95.23%。本文依据这 10 个综合指标建立样本的主成份矩阵。

2.2 聚类结果

按上文所述算法进行聚类。经过多次的训练, 样本客户群体最终在 15 个分类的时候达到一个相对稳定的结果。15 个类别特征描述如表 2 所示。

2.3 模型稳定性测试

使用 2015 年度 4–6 月份和 7–9 月份的数据验证模型的稳定性。验证结果如表 3 所示。结果显示两组数据分类结果相差不大, 各个类别占比只有千分以下的差距。也可以看出中国个人客户偏爱 A 股占大多数。

表 1 各成分累积方差贡献度

Tab.1 Cumulative variance contribution of each component

成份	初始特征值			成份	初始特征值		
	合计	方差/%	累积/%		合计	方差/%	累积/%
1	3.361	14.615	14.615	13	0.075	0.325	98.587
2	2.517	10.945	25.559	14	0.052	0.225	98.812
3	2.128	9.252	34.811	15	0.051	0.220	99.032
4	2.054	8.930	43.741	16	0.049	0.212	99.244
5	2.014	8.754	52.495	17	0.046	0.201	99.445
6	1.990	8.652	61.148	18	0.031	0.134	99.578
7	1.987	8.639	69.786	19	0.030	0.130	99.708
8	1.958	8.511	78.298	20	0.027	0.119	99.827
9	1.950	8.479	86.777	21	0.021	0.090	99.917
10	1.945	8.458	95.234	22	0.011	0.046	99.963
11	0.453	1.969	97.203	23	0.008	0.037	100.00
12	0.244	1.059	98.262				

表 2 X 公司客户按产品偏好分群聚类结果

Tab.2 Clustering results of X company's customers according to product preference

群体序号	群名	产品配置特征
1	主攻新股申购客户	比例约为 75%以上资产专门用于申购新股
2	AB 股混合客户	比例约为 45%的 A 股,38%的 B 股
3	A 股与 ST 股票混合客户	比例约为 57%A 股及 31%ST 类股票
4	专攻 B 股客户	比例约为 83%的仓位持有 B 股
5	偏爱场内基金客户	比例约为 70%资产投在场内基金,30%其他
6	主攻 ST 类股票的客户	比例约为 81%资产购买 ST 类股票
7	理性配置客户	比例约为 35%A 股,36%场内基金,其中 12%股票型基金
8	主攻创业板投资客	比例约为 73%的仓位投资创业板股票
9	专投股票型基金客户	比例约为 93%资产投资于股票型基金
10	混合投资客户	比例约为 A 股占 35%,B 股约 12%,开放式基金 16%,货币 8%
11	A 股客户	比例约为 92%的 A 股市值占比
12	主攻场内基金客户	比例约为 86%资产投资于场内基金
13	偏 A 股客户	比例约为 57%A 股配比,26%创业板股票投资,17%其他
14	A 股客户+配置基金	比例约为 62%的 A 股,15%的场内基金配比,其中 7%股票型基金
15	偏爱债券与基金客户	比例约为债券型基金 86%,股票型基金 8%

表 3 X 公司客户以 3 个月份为时间窗口的分群聚类验证结果

Tab.3 Verification results of the clustering of X company's customers using 3 months as a time window

群体 序号	4-6 月份使用 模型分类	占比	5-7 月份使用 模型分类	占比
1	18 452	3.888 21	17 650	3.726 93
2	1 650	0.347 68	1 541	0.325 39
3	13 898	2.928 59	12 792	2.701 12
4	2 281	0.480 65	2 713	0.572 87
5	3 098	0.652 81	3 107	0.656 06
6	7 629	1.607 58	7 615	1.607 96
7	4 050	0.853 41	4 072	0.859 83
8	1 102	0.232 21	1 090	0.230 16
9	49 076	10.341 32	48 994	10.345 45
10	31 920	6.726 20	32 220	6.803 49
11	320 847	67.609 07	321 308	67.846 61
12	3 226	0.679 78	3 259	0.688 16
13	4 131	0.870 48	4 335	0.915 36
14	6 472	1.363 78	6 399	1.351 19
15	1 524	0.321 13	1 549	0.327 08

3 营销策略

客户风险类别会发生变化.因此风险偏好识别模型对客户的风险属性应持续跟踪、动态评估,

不断更新才能更客观真实的反映客户风险属性.根据以上的模型分析结果,本文为证券公司理财顾问开展专业服务和专项营销活动提出以下建议。

(1)依据市场规律,证券公司为客户提供适当性服务,可依据客户风险属性,提供适合于客户风险偏好的资产配置或投资组合建议服务,帮助客户将高风险资产的配置比例控制在与其自身能力相适应的范围内。这样更容易获得客户的认可,建立和维系忠实客户群。

(2)在专项营销中,证券公司可以先评估营销产品的风险特征,匹配与产品相适应的客户的风险偏好和类别,在此基础上从客户数据库中筛选出目标客户名单,供理财顾问更有针对性的开展一对一营销。这样易于锁定可能的目标客户群体,提高产品销售的适当性和成功率。

4 结论

综上所述,本文提出了客户风险偏好分析模型。对客户的交易数据,使用主成分分析法对冗余变量进行剔除,应用改进的 K-means 算法建立客户细分模型,达到客户分组目的,并对每一群组客户进行特征描述,进而提出相应的营销策略。本文的创新在于用侧重交易标的来划分客户群组,动态修正来体现客户的成长和偏好变化,提出针对不同标的群类提供相匹配产品的营销策略,更具针对性和提高营销成功率,是帮助证券公司有的放矢地进行服务和产品的推荐。

参考文献:

[1] 何海鹰,朱建平,谢帮昌.证券投资意识调查分析[J].统计研究,2008,25(9):49-54.
[2] 吴峰,施鹏飞.概念聚类挖掘方法的客户交易行为分析[J].微型电脑应用,2000,16(5):26-28.
[3] KUO R J, HOL M, HUC M. Cluster analysis in industrial market segmentation through artificial neural network[J]. Computers and Industrial Engineering, 2002, 4(2): 391-399.
[4] 梁敏君.分形聚类分析在证券客户细分中的应用研究[D].合肥:合肥工业大学,2009.
[5] 钱维佳,王延清.基于动态聚类的证券业客户细分实证研究[J].计算机应用,2010,30(2):495-498.
[6] 尔古打机.基于用户行为分析的移动终端偏好模型研究[C].第八届(2013)中国管理学年会论文集(选编),2013.
[7] 王学民.应用多元分析[M].3版.上海:上海财经大学出版社,2009:209-211.

(特约编辑:黄家瑜)