

基于神经网络的卷烟零售户销售假烟行为的预警模型研究

朱卫东^{1,2,3}, 孙宜博³, 吴勇^{2,3}, 李矿⁴, 杜承勇³

1 合肥工业大学, 经济学院, 安徽省合肥市屯溪路 193 号 230009;

2 过程优化与智能决策教育部重点实验室, 合肥, 230009;

3 合肥工业大学管理学院, 安徽省合肥市屯溪路 193 号 230009;

4 安徽省烟草公司淮北市公司, 淮北, 235100

摘要: 在卷烟零售户销售假烟驱动因素分析的基础上, 通过对安徽某市烟草专卖局的实地调研以及所获得历史数据的系统分析, 构建了基于 BP 神经网络的零售户销售假烟行为的预警模型, 该模型综合考虑了零售户的销售行为、零售户自身的特征以及时间因素等十二个预警输入指标。实证表明所构建的模型稳健性较好, 预测准确率较高。

关键词: 烟草行业; BP 神经网络; 假烟销售; 预警模型

doi:10.3969/j.issn.1004-5708.2014.06.020

中图分类号: TS4-06

文献标志码: A

文章编号: 1004-5708 (2014) 06-0127-06

A neural network-based early warning model for counterfeit cigarette traded by retailers

ZHU Weidong^{1,2,3}, SUN Yibo³, WU Yong³, LI Kuang⁴, DU Chengyong³

1 School of Economics, Hefei University of Technology, Hefei 230009, China;

2 Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei 230009, China;

3 School of Management, Hefei University of Technology, Hefei 230009, China;

4 Huaibei Municipal Tobacco Company, Huaibei 235100, China

Abstract: Field study and data analysis were conducted in a city of Anhui province to establish a early warning model of counterfeit cigarette sales based on BP neural network by analyzing retailer's motivation for counterfeit selling. This model took into account twelve factors such as retailers selling practice, their individual characteristics and time factors. It was proved to have good stability and improved accuracy of early warning.

Keywords: tobacco industry; BP neural network; sales of counterfeit cigarettes; early warning model

烟草行业是一个特殊的行业, 它既与政府财政收入关系密切, 又与政府监管行为紧密相关, 同时又会对消费者健康造成影响。2012 年, 烟草行业实现工商税利 8649.39 亿元, 上缴国家财政 7166.62 亿元, 占全国财政总收入 6.1%^[1]。少数卷烟零售商通过销售假烟来追求自身利益的最大化, 这种行为严重的损害

了烟草行业的健康发展, 也对国家利益和消费者利益造成了危害。然而通过对安徽省某市烟草销售行业进行调研, 发现该市的烟草专卖局内管系统现有的预警准确率仅为 4% 左右, 不能为烟草专卖局执法部门提供科学准确的执法对象, 不能改变当前“地毯式”的执法方式, 同时也对卷烟零售户的日常经营造成了较大干扰。

市场预测是根据历史数据资料, 运用科学的理论方法通过对市场的各类行为进行研究分析来判断市场的发展趋势, 预测未来市场的趋势变化, 为管理者的经营决策提供科学合理的依据。许多学者对管理预测理论的实际应用进行研究: 王增民^[2]利用灰色加权

基金项目: 国家自然科学基金项目 (NO. 71071048); 安徽省烟草公司科技立项项目 (NO. HBYC201101)

作者简介: 朱卫东, 博士, 博士生导师, 教授, 研究方向: 决策科学、决策支持系统、会计信息与决策, Email: zhuwd@hfut.edu.cn

收稿日期: 2014-01-12

马尔可夫链对移动通讯用户数进行预测;赵辉和王辉^[3]基于河南某市的卷烟历史销售数据构建了三个基于时间序列的卷烟需求模型,并对比分析了三种模型的适用情况;仲东亭等^[4]基于修正的神经网络预测模型利用江苏某地的销售数据建立了一维时间数据销售总量的预测模型。然而目前的市场预测研究中,学者大都采用定量的历史数据构建预测模型,很少考虑到预测对象的自身特征,而预测对象的差异会导致预测结果的差别。

本文通过对卷烟零售户销售假烟行为的特征以及影响因素进行分析,并对安徽省某市烟草专卖局进行实地调研,对传统的销量指标进行细化,并增加了刻画零售商自身特征的定性指标和控制时间效应的月份指标。在此基础上,通过相关试验确定预警模型的最优参数,并通过BP神经网络对数据样本进行自适应学习完善并确定预警模型,最后利用所构建的预警模型对剩余样本进行预警检验,对预警结果进行分析得到相关管理启示。

1 卷烟零售户销售假烟行为的相关分析

1.1 违规零售户历史数据的特征分析

通过对安徽某市查处的违规销售假烟的历史数据进行分析,发现卷烟零售户的经营规模、经营业态、以及所处的市场位置等特征的不同,会使零售户的销售行为产生较大差异:

在经营业态上,违规零售户主要为批零兼营杂货店、一般杂货店以及中小超市;在所处商圈内,其主要分布在居民区(村)和商业(集贸)区;所处的市场细分类型上,销售假烟的零售户主要分布在农村、乡镇、市区和城乡结合部。

对于经营规模来说,违规零售户主要分布在大型和中型经营规模的零售商家内,主要原因是小型零售户的销售卷烟数量和种类较少,客户较为固定,出售假烟会对其在消费者中的声誉造成严重损害,而对于大中型零售户来说,其销售烟的种类繁多、数量较大,并且固定客户较少,客户流动性较大,其出售假烟的声誉损伤较小。

1.2 零售户销售假烟行为的影响因素分析

(1) 利益因素

首先,销售假烟可以带来非常丰厚的利润,由于烟草行业实行烟草专卖制度,零售户所销售的卷烟是从当地烟草专卖局获得,烟草专卖局对每种品牌的卷烟均有相应的指导价格,而零售户销售的假烟从非法途径获得,进货成本较低;其次,烟草专卖局对卖假

烟的零售户处罚较轻,零售户即使被抓到后的预期损失远远小于非法销售假烟所带来的巨额利润。

(2) 环境因素

实地调研发现零售户销售假烟与所处的地理环境有着很大的关系,当周围的商家销售假烟较多时,在从众心理的影响下零售户往往会“随大流”,进入销售假烟的行列,出现假烟销售的“扎堆现象”。尤其在零售户发现别的商家违法销售假烟很少被查处时,就会产生一定的侥幸心理,认为自己销售假烟也不会被查处,这更进一步的促使零售户违规销售假烟。

2 基于BP神经网络的零售户违规销售的预警模型

BP(Back Propagation)神经网络^[5-6]也称误差反向传播神经网络,其学习过程为一个反复迭代的过程,是应用得最为广泛的神经网络预测模型^[7]。人工神经网络具有以下特征:(1)具有自学习和自适应性;(2)容错性以及较强的鲁棒性;(3)并行性;(4)复杂的非线性。由于BP神经网络以上的优良特性,其被广泛应用于图像识别^[8],上市公司财务预警^[9]、预测^[10-11]、顾客满意度以及风险评价^[12-15]等领域。模型的输入是利用神经网络构建预测模型的前提基础,它与模型预测结果准确度密切相关,本文在此重点对预警模型的输入进行梳理和选择。

2.1 零售户销售假烟行为的预测指标的遴选

由于零售户销售假烟时会导致自身合法的销售量发生变化,可以通过考察零售户销售量的变化来对零售户销售行为进行研究^[16]。从1.1节违规零售户历史数据的特征分析中发现不同特征的零售户的销售假烟的行为是不同的,基于此,本文在传统的销售数据的基础上增加了零售户的特征指标以及时间指标。

2.1.1 销售类指标

通过调研发现仅仅通过销售总量的变化有时不能准确的反映其销售行为的规律,本文通过在销售总量的基础上来增加更精细的指标。烟草专卖局根据卷烟的不含税调拨价把卷烟分为五个价类,由于各地区之间的经济发展水平和人们生活习惯的不同,各地区五类烟的销售情况不尽相同,本文在考虑销售总量的基础上,同时考虑每类烟的销售状况,由此本文预警指标的销售类指标确定如下表所示:

(1)月销量变化率(A_1)表示卷烟零售户当月销量与上月销量之间的差异,用来衡量销量在一个相对长的时间内的变化规律;

(2)旬销量变化率(A_2)表示卷烟零售户旬销

量与前三旬平均销量之间的差异，用来衡量零售户当期总销量的变化；

（3）五类烟销量变化率（ A_3-A_7 ）表示各价类烟的旬销量与前三旬平均销量之间的差异，用来衡量各价类卷烟的销量变化。

2.1.2 零售户的特征指标

通过对安徽某市实地调研以及违规销售的历史数据的分析，本文选取经营规模、所处的市场细分类型、经营业态以及所处的商圈类型四个指标来对零售户的特征进行衡量，具体指标及说明如下表所示。

表 1 销售类指标及说明
Tab. 1 Sales index and its expalnation

销售类指标 A	计算公式	备注
月销量变化率 A_1	$R_1=[Q_1(t)-Q_1(t-1)]/Q_1(t-1)$	$Q_1(t)$ 为当期月销售总量
旬销量变化率 A_2	$R_2=[Q_2(t)-Q_2]/Q_2, Q_2=[Q_2(t-1)+Q_2(t-2)+Q_2(t-3)]/3$	$Q_2(t)$ 为当期旬销售总量
第一类烟销量变化率 A_3	$R_3=[Q_3(t)-Q_3]/Q_3, Q_3=[Q_3(t-1)+Q_3(t-2)+Q_3(t-3)]/3$	$Q_3(t)$ 为当期一类烟旬销量
第二类烟销量变化率 A_4	$R_4=[Q_4(t)-Q_4]/Q_4, Q_4=[Q_4(t-1)+Q_4(t-2)+Q_4(t-3)]/3$	$Q_4(t)$ 为当期二类烟旬销量
第三类烟销量变化率 A_5	$R_5=[Q_5(t)-Q_5]/Q_5, Q_5=[Q_5(t-1)+Q_5(t-2)+Q_5(t-3)]/3$	$Q_5(t)$ 为当期三类烟旬销量
第四类烟销量变化率 A_6	$R_6=[Q_6(t)-Q_6]/Q_6, Q_6=[Q_6(t-1)+Q_6(t-2)+Q_6(t-3)]/3$	$Q_6(t)$ 为当期四类烟旬销量
第五类烟销量变化率 A_7	$R_7=[Q_7(t)-Q_7]/Q_7, Q_7=[Q_7(t-1)+Q_7(t-2)+Q_7(t-3)]/3$	$Q_7(t)$ 为当期五类烟旬销量

表 2 零售户的特征指标及说明
Tab. 2 Retail clients' characteristics and explanation

零售户的特征指标 B	备注
经营规模 B_1	经营规模分大、中、小三类
所处的市场细分类型 B_2	市场细分类型有：城乡结合部、农村、市区、县城城区、乡镇
经营业态 B_3	经营业态分为：百货店、餐饮（宾馆）、大型超市（大型商场）、流动摊点（小推车）、批零兼营杂货店、小卖部、一般杂货店、杂货店、中小超市（中小超市、百货店）、专营店、娱乐场所（酒店娱乐类）
所处的商圈类型 B_4	商圈类型有：工业区、居民区（村）、旅客中转区、商业（集贸）区、学区、娱乐（旅游）区、政务（商务）区、其他

2.1.3 时间因素

通过对安徽某市卷烟零售户的实地调研发现卷烟的销售量会因时间的变化产生波动，比如春节期间的销售量远高于其他月份，而且各月份之间的销售量也有一定的差异，因此本文选择“月份（ C_1 ）”作为预警模型的一个控制时间因素的输入。

综上所述，通过理论分析和实地调研梳理出了预测零售户违规销售假烟的十二个相关因素，其中主要包括零售户销售行为、零售户自身特征以及时间因素三个方面，具体如表 3 所示。为了方便数据处理，本文需将定性指标进行数值化，标准见表 4。

表 3 零售户销售假烟预测模型的输入指标
Tab. 3 Indicators of prediction model of counterfeit cigarette sales

零售户违规销售假烟预测相关指标	
销售类指标 A	零售户的特征指标 B
月销量变化率 A_1	经营规模 B_1
旬销量变化率 A_2	所处的市场细分类型 B_2
第一类烟销量变化率 A_3	经营业态 B_3
第二类烟销量变化率 A_4	所处的商圈类型 B_4
第三类烟销量变化率 A_5	时间因素 C
第四类烟销量变化率 A_6	月份 C_1
第五类烟销量变化率 A_7	

表 4 定性变量的数值化表
Tab. 4 Qualitative indicators

经营规模 B ₁		经营业态 B ₃		商圈类型 B ₄	
属性	数值化	属性	数值化	属性	数值化
大	1	百货店	1	工业区	1
中	2	餐饮或宾馆	2	居民区(村)	2
小	3	大型超市或商场	3	旅客中转区	3
市场类型细分 B ₂		流动摊点	4	商业(集贸)区	4
属性	数值化	批零兼营杂货店	5	学区	5
城乡结合部	1	小卖部	6	娱乐(旅游)	6
农村	2	一般杂货店	7	政务(商务)区	7
市区	3	杂货店	8	其他	8
县城城区	4	中小超市	9		
乡镇	5	专营店	10		
		娱乐场所	11		

注：月份(C₁)指标数值化按照：一月数值化为1，二月数值化为2，依次类推。

2.2 参数选择及预警模型的构建

(1) 数据样本的选择

以安徽某市烟草专卖局从2010年1月上旬至2012年9月中旬所有零售户的销售数据，剔除属性信息缺失和销量数据缺失的样本，最后得到用户销售假烟的样本为328个，正常经营的用户样本为539809个。由于样本的比例不平衡，借鉴石晓军^[17]的方法使用1:3的样本配比，所以本文从正常数据样本中随机抽取984个样本，与328个违规用户合在一起作为初始样本，本文设置训练集、验证集和测试集的样本比例分别为6:3:1。

(2) 隐含层数及隐含层节点数的确定

Carroll 和 Dickinson^[18]证明了三层BP神经网络只要隐含层节点数足够多，并拥有足够的训练数据和训练时间，若其输出层采用线性转换函数，隐含层采用Sigmoid转换函数，可以模拟任意复杂系统。基于此本文所构建的模型为三层BP神经网络（即只含有一个隐含层），输入层到隐含层的转移函数为双曲正切函数，隐含层到输出层的转移函数为线性传递函数。

通过2.1节对预测指标的梳理，本文共选取十二

个影响因素对零售户销售假烟行为进行预警，模型的输出为零售户销售假烟的可能性，所以本文构建的预警模型有十二个输入、一个输出。然而，对于网络隐含层接点数(m)的确定目前尚无完整的理论指导，主要根据经验和试凑来确定，根据经验公式

$m = \sqrt{n+l} + \alpha$ ，其中模型的输入 $n=12$ ，输出 $l=1$ ， α 为1到10之间的常数，因此本文所构建的预测模型初始隐含层节点数从5开始，逐渐增加隐含层节点数，经过多次训练得到本文模型的误判率在隐含层节点数 $m=6$ 时最小，因此本文所构建的模型隐含层节点数为6。

(3) 其他参数的确定

初始权值决定了神经网络的初始训练从误差曲面哪一点开始，会对神经网络各神经元的学习速度产生影响，本文借鉴通常的做法，取初始权值为(-1, 1)之间的随机数。在学习过程中，神经网络系统的学习率与模型稳定性关系密切，在选取隐含层节点数的训练中分别设置0.002、0.004和0.006，经过比较发现，隐含层节点数为6、学习率为0.004的网络模型误判率最低，所以预测模型学习率设置为0.004。另外，

本预测模型将冲量系数设置为 0.001，模型的训练迭代次数设置为 1500。

综上所述，该预警模型主要参数：十二个输入、一个输出；一个隐含层，六个隐含层节点；学习率为 0.004，输入层到隐含层的转移函数为双曲正切函数，隐含层到输出层的转移函数为线性传递函数，初始权值为 (-1, 1) 之间的随机数，在此基础上借助 SAS 软件构建本文所需的 BP 神经网络模型（如图 1 所示）。

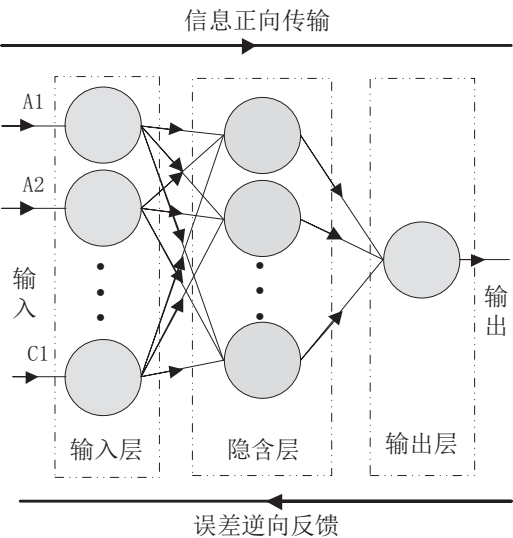


图 1 本文的神经网络结构图

Fig. 1 Neural network structure of this paper

通过对初始样本自适应学习得到相应参数，该模型对训练数据集进行判别的错误分类率为 21.71%，对检验数据集进行判别的错误分类率为 22.80%，两者相差 1.09%，说明所建立的模型稳健性较好；测试集误判率为 23.68%，准确率为 76.32%，说明该模型预测准确率比较高。

2.3 模拟预警与结果分析

为了方便烟草专卖局员工可以直观的观察用户违规销售的概率，同时为了便于烟草公司针对模型的预警结果采取相应的管理措施，本文按照违规可能性的大小将零售户分为四类，如下表所示。

表 5 模型输出结果与违规可能性对应表

Tab. 5 Corresponding results of model outputs and possibility of irregularities

模型输出结果	(0, 0.5)	(0.5, 0.6)	(0.6, 0.7)	(0.7, 1)
违规可能性	较小	一般	较大	大

注：此处假设违规可能性一般、较大或大的零售户为可能违规的零售户

利用所构建的预警模型对构建模型剩余的 538825 个数据样本进行模拟预警，得到其中有 29253 组数据样本可能为违规样本，这些违规样本分布在 1860 个用户中，其中 13 个违规可能性大，512 个违规可能性较大，1335 个违规可能性一般，通过模型的模拟预警所得到的 1860 个可能销售假烟的零售户特征如图 2 所示。通过对模型的预测结果进行分析发现：对于商圈类型来说，1860 个可能销售假烟的零售户中 57.7% 分布在居民区，34.6% 分布在商业集贸区，而烟草专卖局所查处的历史数据中 68.9% 分布在居民区，24.39% 分布在商业集贸区；对于经营规模来说，预测结果中 54.6% 为大型零售户，42.8% 为中型零售户，而已查处销售假烟的零售户 48.78% 为大型零售户，46.04% 为中型零售户。1860 家可能违规的零售户的特征与已经查处的违规销售假烟零售户的特征较为类似，这进一步验证了本文将零售户的自身特征加入到预警指标的合理性。

在模拟预测结果分析的基础上并结合历史违规用户的特征分析，可以发现由于不同特征的用户在销售假烟的行为上有着较大差异，烟草专卖局应该对零售户进行分类管理，针对不同类型的用户制定具体的有针对性的监督管理措施。在具体的专卖管理工作中，烟草专卖局可以采取如下措施来遏制零售户销售假烟的行为：在经营规模上，烟草专卖局的重点监管对象应该主要集中在大型或者中型的零售户中；在市场细分类型上，适当的将监督力度由市区向农村、乡镇等地区转移，加大对农村、乡镇以及县城城区的监管水平；在经营业态上，要对中小超市和批零兼营杂货店进行重点检查；在商圈类型上，要经常对处于居民区和商业集贸区零售户进行检查。

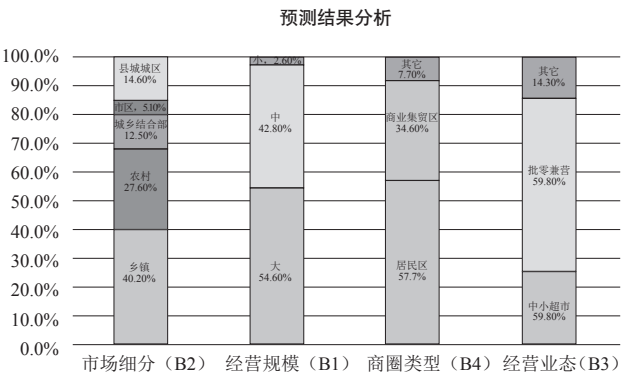


图 2 可能违规零售户的分布特征

Fig. 2 Distribution characteristics of possible illegal retailers

3 结论

本模型在对零售户的历史数据分析的基础上,通过对零售户违规行为驱动因素进行分析,综合考虑零售户的销售行为、零售户所处的地理位置和时间因素以及零售户自身的特征等多个因素来对违规行为进行预测,改变了以往单纯使用销售数据预测零售户销售违规行为的方法。对模型的相关分析结果显示本文所构建的预警模型稳健性比较理想,预测准确率比较高。

烟草销售中零售户的违规行为会对烟草行业产生严重的不利影响,而烟草行业的健康发展对国家财政收入有重要影响,构建预警模型对卷烟零售户违规销售假烟进行预测,首先,可以为烟草专卖的检查工作提供重要的科学依据。通过本模型可以得到每个零售户的违规概率,为烟草专卖检查工作提供了依据,可以根据违规概率的大小进行针对性检查,可以减少撒网式、地毯式的检查工作,同时还可以提高烟草公司的精准化管理水平,降低管理成本,提高经济效益;其次,可以减少执法人员对零售户的正常经营活动带来的干扰,提高烟草专卖的客户满意度;另外,对零售户有一定的约束作用,由于本文所构建的预警模型的准确性较高,对卷烟零售户的侥幸心理有一定的影响,从而减少零售户卖假烟的行为。

由于该市烟草专卖局的数据库中用户的部分属性信息缺失,模型中衡量零售户个人特征的部分指标(如教育程度)未加入预测模型,以后可以在本模型的基础上增加输入项,对所缺失的部分属性指标进行的完善,进一步提高模型的预测精准度。基于本文梳理出的零售户销售假烟预警指标,后续可以研究各指标对预警结果的具体影响路径及相关系数,进而有针对性的制定卷烟零售户的监督管理措施,进一步提高烟草专卖的精准化水平。

参考文献

- [1] 数据来源: http://www.tobacco.gov.cn/html/10/1005/100501/10050103/4288217_n.html.
- [2] 王增民,王开珏. 基于灰色加权马尔可夫链的移动通信

- 市场预测[J]. 数学的实践与认识, 2012(22):8-15.
- [3] 赵辉,王辉. 基于多种时序模型的河北省某市卷烟需求预测比较[J]. 经济研究导刊, 2011(08):142-145.
- [4] 仲东亭,张玥. BP神经网络对烟草销售量预测方法的改进研究[J]. 工业技术经济, 2007(09):115-118.
- [5] Rumelhart D E, Hinton G E, McClelland J L, et al. A general framework for parallel distributed processing [J]. Parallel distributed processing, 1986, 1(2).
- [6] Syed M N, Pardalos P M. Neural network models in combinatorial optimization [M] Springer, 2013:2027-2093.
- [7] Vellido A, Lisboa P J, Vaughan J. Neural networks in business: a survey of applications (1992--1998) [J]. Expert Systems with Applications, 1999, 17(1):51-70.
- [8] 孙亚. 基于粒子群 BP神经网络人脸识别算法[J]. 计算机仿真, 2008(08):201-204.
- [9] 杨淑娥,黄礼. 基于 BP神经网络的上市公司财务预警模型[J]. 系统工程理论与实践, 2005(01):12-18.
- [10] 邵惠芳,许自成,李东亮,等. 基于 BP神经网络建立烤烟感官质量的预测模型[J]. 中国烟草学报, 2011(01):19-25.
- [11] 王文波,费浦生,羿旭明. 基于 EMD 与神经网络的中国股票市场预测[J]. 系统工程理论与实践, 2010(06):1027-1033.
- [12] 郭岚,张祥建. 基于 BP神经网络的企业核心竞争力评价[J]. 科学学研究, 2007(01):132-137.
- [13] 彭黔荣,蔡元青,王东山,等. 根据常规化学指标识别烟叶品质的 BP神经网络模型[J]. 中国烟草学报, 2005(05):19-25.
- [14] 岑成德,权净. 服务属性对顾客满意感影响程度研究——人工神经网络方法[J]. 南开管理评论, 2005(02):16-22.
- [15] Wang R, Zeng J, Feng X, et al. Evaluation of effect of plastic injection molding process parameters on shrinkage based on neural network simulation [J]. Journal of Macromolecular Science, Part B, 2013, 52(1):206-221.
- [16] 谢星峰. 基于神经网络的卷烟销售预测模型的研究及应用[D]. 昆明理工大学, 2005.
- [17] 石晓军. Logistic 违约率模型最优样本配比与分界点的模拟分析[J]. 数理统计与管理, 2006(06):675-682.
- [18] Carroll S M, Dickinson B W. Construction of neural nets using the radon transform[C]. International Joint Conference on Neural Networks. IEEE, 1989: 607-611.