

基于深度神经网络的客户流失预测模型

马文斌, 夏国恩

(广西财经学院 工商管理学院, 广西 南宁 530003)

摘要: 客户流失是企业面临的一个重要问题, 为及时发现流失客户, 降低企业损失, 目前已有许多研究对客户流失问题给出解决方案, 但是大部分研究中使用的是浅层学习算法, 预测结果依赖于特征选择, 需要在特征工程上花费大量的时间和精力。随着客户数据的快速增长, 在大数据情况下, 人工特征工程已不能有效地获取高质量特征。深度学习通过模拟人脑多层、逐级地抽取信息特征, 能自动学习到较好的数据特征, 在图像识别、语音识别等领域取得显著成果。为研究深度学习在客户流失预测方面的应用, 构造了基于深度神经网络的流失预测模型, 并在电信客户数据集上, 与经过特征选择的 Logistic 回归、决策树等预测模型作对比, 验证其预测准确度。实验结果表明, 深度神经网络模型取得了较好的预测效果。

关键词: 深度学习; 深度神经网络; 客户流失; 电信

中图分类号: TP31

文献标识码: A

文章编号: 1673-629X(2019)09-0076-05

doi: 10.3969/j.issn.1673-629X.2019.09.015

Customer Churn Prediction Model Based on Deep Neural Network

MA Wen-bin, XIA Guo-en

(School of Business Administration, Guangxi University of Finance and Economics,
Nanning 530003, China)

Abstract: One of the important problem enterprise faced is customer churn. In order to find out the customer loss in time and reduce the loss of enterprises, many researchers have proposed solutions to the problem of customer churn. However, most studies use shallow learning algorithm, whose prediction results depend on feature selection and require a lot of time and energy in feature engineering. With the rapid growth of customer data, in the case of big data, artificial feature engineering has been unable to effectively obtain high-quality features. Deep learning can automatically learn better data features by simulating the human brain to extract information features in multiple layers and step by step, making remarkable achievements in the fields of image recognition and speech recognition. In order to study the application of deep learning in customer churn prediction, a churn prediction model based on deep neural network is constructed and compared with the Logistic regression, decision tree and other models after feature selection in the telecom customer data set to test the prediction accuracy. Experiment shows that deep neural network model has better prediction effect.

Key words: deep learning; deep neural network; customer churn; telecommunications

1 概述

流失客户通常是指在一定时期内终止使用企业的服务或产品的客户。客户流失是企业面临的一个重要问题,也是学术界研究的热点。高流失率代表企业产品的市场份额的减少,客户流失率的降低则意味着企业效益的提高。同时,企业获取新客户的成本也是保留老客户成本的数倍。为及时发现流失客户,减少客户流失量,研究者借助机器学习与数据挖掘算法,构建

了大量的客户流失预测模型。表现好的流失预测模型对于最小化流失率非常重要,因为可以为那些不满意的特定客户提供个性化的促销或优惠活动,以此来挽留将要流失的客户。国内外企业为了深入了解客户行为,寻找影响客户流失的关键因素,通过开展数据挖掘竞赛的形式来发现优秀的客户流失预测解决方案。例如,法国电信运营商 Orange 在 KDD Cup 2009 中提供了大量客户行为数据,供参赛者分析预测;KDD Cup

收稿日期:2018-10-19

修回日期:2019-02-21

网络出版时间:2019-04-24

基金项目: 教育部人文社会科学规划基金项目(17YJA880080);广西跨境电商智能信息处理重点实验室培育基地(广西财经学院)专项资助项目;广西财经学院创新治理与知识产权学科群(政府治理的互联网创新发展)专项资助项目

作者简介: 马文斌(1989-),男,硕士,研究方向为数据挖掘;夏国恩,博士,教授,研究方向为商务智能、智能决策、客户关系管理。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190424.1051.044.html>

2015 使用由学堂在线提供的用户在线学习行为数据,预测用户的流失率;携程也在 2016 年开展了客户流失概率预测竞赛;WSDM Cup 2018 则要求参赛者预测 KKBOX 的订阅用户的流失情况。

经过多年对客户流失预测的研究,取得了较为显著的成果,客户流失中的数据不平衡、预测方法的选择等问题也得到了有效解决。在目前的研究中,研究者将客户流失预测视为一种分类问题,因此有监督学习算法大量地应用于客户流失预测,并取得了不错的效果。根据使用方法的不同,客户流失预测研究主要可分为五个方面。一是基于统计学的方法,具有代表性的方法是聚类算法、回归分析等。姜晓娟等^[1]针对客户数据的类别不平衡、大规模等问题,在聚类算法基础上设置不同权重参数,实验表明该算法具有较好的预测效果。基于统计学方法的流失预测模型的优势是具有较强的可解释性,不足之处在于在大数据背景下,数据往往呈现高维、非线性、非正太分布等特点,此类方法的泛化能力得不到有效的保证。

二是基于人工智能理论的研究。此类研究的代表性方法是人工神经网络。李洋^[2]通过分析客户群特征、服务属性和客户消费数据,对比不同的预测模型,验证了神经网络预测的有效性。Kasiran Z 等^[3]结合增强学习算法与循环神经网络,预测移动手机用户的流失情况。冯鑫等^[4]结合神经网络与自然语言处理,利用客户消费评论信息,预测客户是否会流失,并给出影响客户流失的主要指标。人工神经网络模拟人脑处理信息的结构,能够处理较复杂的数据,但可解释性较低,且容易产生过拟合问题。

三是基于统计学习理论的研究。统计学习理论主要是构建给定数据的概率统计模型,并对未知数据进行预测,朴素贝叶斯算法、决策树、支持向量机等都属于常用的方法。Kirui C 等^[5]利用朴素贝叶斯、贝叶斯网络两种概率模型预测客户流失。尹婷等^[6]结合决策树与贝叶斯分类算法,弥补了决策树算法的缺点。盛昭瀚等^[7]给出一种加权熵的 ID3 算法解决客户流失预测问题。张宇等^[8]使用 C5.0 算法预测邮政短信业务的客户流失情况。夏国恩等^[9]通过与多种预测算法的比较,验证了支持向量机的预测有效性。王观玉等^[10]结合主成分分析与支持向量机,降低数据的冗余性,提高了预测效果。Chen Zhenyu 等^[11]给出一种分层多核支持向量机,融合特征选择过程,在多个数据集上有较好的预测结果。赵琨等^[12]利用双子支持向量机分析信用卡用户的流失情况。支持向量机基于 VC 维理论和结构风险最小化原理,具有较强的泛化能力,但可解释性较低,在小样本的情况下表现优异,但随着数据规模的增大,支持向量机已不能在有效的时间内完成计

算任务。

四是基于集成学习理论的研究。集成学习方法通过集成多种方法的优势,提高预测性能。子算法的选择、子算法预测结果的集成等问题是集成学习方面的研究热点。罗彬等^[13]通过使用聚类算法分组样本集,然后利用不同的算法分别在样本子集上构建预测模型,最后基于成本敏感性,利用人工鱼群算法集成子模型的结果,实验表明提出的集成方法优于单个预测模型的预测性能。Coussement K 等^[14]利用集成学习方法预测在线客户的流失情况。

五是基于社会网络分析的研究。社会网络是一种较为新颖的客户流失预测方法,使用社会网络发现潜在流失客户的假设前提是与流失客户存在于同一社区内或存在关联关系的客户更容易流失。Phadke C 等^[15]基于客户的呼叫网络,给出一个度量客户间社会联系强度的公式,并利用影响扩散模型计算流失客户的净积累影响,最后在真实的移动客户数据上验证了使用社会网络分析预测客户流失的有效性。Verbeke W 等^[16]在关系分类模型中引入非马尔可夫网络,并融合关系分类模型与非关系分类模型,构建了流失预测模型。黄婉秋^[17]基于 RFM 模型和时间序列分析法,结合社区发现、独立级联模型进行客户流失分析,并在零售客户数据上验证了基于社会网络方法的有效性。

上述客户流失预测研究中使用的方法,预测效果依赖于特征处理的好坏,需要花费大量的时间与精力在特征工程上,随着客户数据的快速增长,在大数据情况下,人工特征工程已不能有效地获取高质量特征。但是深度学习通过模拟人脑多层、逐级地抽取信息特征,能够自动学习到可以较好地表示数据集的特征,借助深度学习,构建预测模型时,将不再依赖于特征选择。目前深度学习在客户流失预测方面的研究成果还较少,为探究深度学习在客户流失预测中的应用,文中构建了包含 3 隐层的深度神经网络模型,并在电信客户数据集上与经过特征选择的 Logistic 回归、决策树等预测模型作对比,从而验证深度神经网络模型的预测效果。

2 深度学习简介

人工神经网络是客户流失预测中常用的一种算法,而深度学习是人工神经网络的延伸和发展,是一种拥有多隐层的人工神经网络算法,通过模拟人脑多层、逐级地抽取信息特征,最终获得能够较好地表示输入数据的特征^[18]。2006 年,Hinton 等提出的深度置信网络(DBN)是当前深度学习算法的框架,打破了深层神经网络难以有效训练的僵局^[19]。支持向量机、隐马尔可夫模型、感知机等都是典型的浅层学习算法,与浅层

学习算法相比,深度学习在网络表达复杂目标函数的能力、网络结构的计算复杂度、仿生学角度、信息共享等方面更具有优势^[20]。

根据构造深度学习模型时采用的结构、学习算法等因素,深度学习可分为 3 类:生成深度结构、判别深度结构、混合深层结构^[19]。生成深度结构的代表是深度置信网络;判别深度结构的代表模型是卷积神经网络;混合深层结构则是结合生成深度结构和判别深度结构来实现模式分类的一类深层结构。

目前,借助于大数据,深度学习在许多领域的表现都优于浅层模型。根据数据类型不同,深度学习主要应用在如下领域:一是图像识别,常用的算法是卷积神经网络或改进的卷积神经网络;二是语音识别,常用的算法是循环神经网络(RNN)或改进的循环神经网络;三是自然语言处理,由于自然语言的复杂性,虽然深度学习在自然语言处理上取得了一定的进展,但是并没有在图像、语音上的成果显著。

3 基于深度学习的客户流失预测模型

经典的客户流失预测模型结构见图 1。由图 1 可以看出,经典的客户流失预测模型主要包含数据预处理、属性选择、特征选择、流失预测、结果评价等阶段。属性选择和特征选择主要是为了减小原始数据中存在的主观性,降低数据“噪声”,达到约简数据维度,而不损失或较少损失数据信息的目的。特征选择主要是指从数据集的所有特征中,利用某种度量方法,筛选出分类预测效果最好的一组特征子集,常用的特征选择方法有互信息、Fisher 比率、ReliefF 等。当数据维度较大时,组合筛选出最优特征子集,需要花费大量的时间。

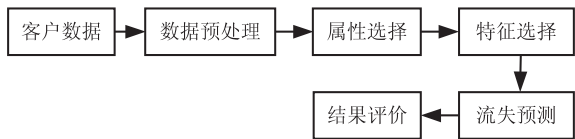


图 1 经典客户流失预测模型结构

基于深度学习的客户流失预测模型如图 2 所示。由图 2 可知,经典客户流失预测模型与基于深度学习的客户流失预测模型最大的区别是在特征处理方面。特征工程需要一定的领域知识,且费时费力,最后选择的特征子集也不一定具有较好的预测结果。在基于深度学习的客户流失预测模型中,深度学习算法可以自主逐层地进行特征处理,没有属性选择、特征选择等特征工程阶段,节省了时间成本,且能够获得更为准确刻画数据信息的特征子集。



图 2 基于深度学习的客户流失预测模型结构

基于深度学习的预测模型结构的预测过程是:多来源收集客户行为数据,确定初始属性集;对数据进行缺失值处理、异常值处理、峰度转换、标准化等预处理工作;将准备好的数据集输入深度学习算法,逐层学习数据特征,训练预测模型;评价预测结果,采用常用的精确率、召回率等评价指标,评价预测模型的性能。

目前,常用的深度学习框架包括 TensorFlow、Caffe、Keras、PyTorch、CNTK 等。其中,Caffe 采用配置文件定义网络结构,容易使用,且支持 python 接口,仅需要少量的代码构建预测模型,训练速度较快。因此,文中基于 Caffe 框架,研究深度学习算法在网络客户流失预测中的应用,通过参考现有深度学习算法模型,调整隐层以及各层的参数,构建了包含 3 个隐层的深度神经网络模型,如图 3 所示。Caffe 中每一个网络模块都是一个层,文中构建的深度神经网络模型使用了数据层、全连接层、DropOut 层、损失层等。这里对 layers 进行描述。

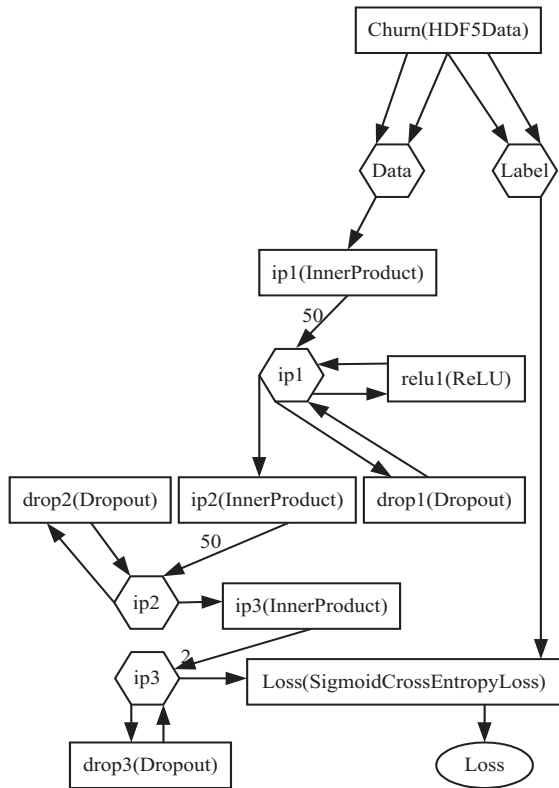


图 3 深度神经网络模型

数据层:Caffe 不直接处理原始数据,需要由处理程序转换为 Caffe 支持的数据格式。目前,Caffe 支持 HDF5、LMDB 等多种数据格式,文中构建的深度神经网络使用 HDF5 格式。数据层定义 4D 的输入(1,1,1,87),表示一次输入一个数据,数据大小是(1,87)。

全连接层:全连接层的每个节点与相邻层的所有节点都有连接。文中构建的深度神经网络的隐层是三个全连接层的堆叠,可看作是对输入数据逐层地提取

信息,最后学习到较好的数据特征。全连接层的神经元数目分别是 87、50、50,损失层的神经元数目则是 2 个。为加快收敛速度,全连接层的激活函数采用 ReLU(rectified linear unit)。ReLU 函数(式 1)是一种非饱和激活函数,Sigmoid、Tanh 等饱和激活函数存在严重的梯度消失问题,训练收敛速度较慢。

$$f(x)=\begin{cases}0 & \text{for } x < 0 \\ x & \text{for } x \geqslant 0\end{cases}\quad (1)$$

DropOut 层:为了防止训练网络时产生过拟合现象,提高模型泛化能力,文中构建的网络中使用了 DropOut。DropOut 是一种参数正则化方法,在训练网络过程中,按照一定的概率从网络中暂时丢弃部分节点,减少特征之间的相互作用,能够有效防止过拟合,提高模型健壮性。文中构建的网络中全连接层的丢弃率分别是 0.5、0.4、0.3。

损失层:损失函数度量网络输出的好坏,通过最小化损失,训练得到较好的网络。Caffe 中定义了多种损失函数,如 EuclideanLoss、HingeLoss、SoftmaxLoss 等,由于客户流失预测是一种二类分类问题,因此采用 SigmoidCrossEntropyLoss。

4 实验结果与分析

4.1 数据集

客户流失预测是在客户的历史行为数据上提取、选择客户特征,并运用分类预测算法建立预测模型,预测客户未来的状态。文中实验所用的电信客户行为数据来源于美国 DUKE 大学,其中训练集共 100 000 个样本,包含流失客户 49 562 个,非流失客户 50 438 个,两类客户的比例基本为 1 : 1;测试集共 51 306 个样本,包含流失客户 924 个,非流失客户 49 514 个,客户流失率为 1.8%,数据类别严重不平衡。原始数据中

部分属性存在缺失的情况,通过删除缺失率过高的属性以及填充缺失率较低的属性,共取得 87 个初始属性指标。

4.2 预测算法和模型评价

实验分别采用 Logistic 回归、朴素贝叶斯和决策树 3 种常用算法构建预测模型,与深度学习预测模型进行对比,并从精确率、召回率、准确率、提升系数和 F_1 值 5 个方面评价模型预测结果。由表 1 可知,精确率 = $A/(A+C)$;召回率 = $A/(A+B)$;准确率 = $(A+D)/(A+B+C+D)$;提升系数 = 精确度/测试集的客户流失率; $F_1=(2 * \text{精确率} * \text{召回率})/(\text{精确率}+\text{召回率})$ 。

表 1 混淆矩阵

客户实际状态	预测流失	预测非流失
流失	A	B
非流失	C	D

4.3 实验环境

实验所用的 Logistic 回归、朴素贝叶斯和决策树等算法的实现主要使用基于 Python 的机器学习库 Scikit-Learn。数据预处理主要使用 Pandas 数据分析库。实验所用电脑的内存是 16 G,处理器是 Intel(R) Xeon(R) CPU E5-1603 v3,操作系统为 Win7 64 位。支持向量机也是客户流失预测中常用的方法,但是在现有的硬件条件下,在实验所用的数据集上,支持向量机不能在有效时间内计算出结果,因此没有选择支持向量机作为对比算法。

4.4 实验结果分析

深度神经网络的预测效果与网络的学习率相关,实验通过设定步长和搜索范围,经过多次对比,确定了预测效果较好的学习率为 0.002。不同模型的预测结果如表 2 所示。

表 2 不同模型的预测结果

预测算法	精确率	召回率	准确率	提升系数	F_1
DNN	0.022 2	0.419 9	0.657 2	1.235 3	0.042 3
Logistic 回归	0.021 2	0.510 8	0.567 4	1.179 8	0.040 8
朴素贝叶斯	0.018 9	0.826 8	0.224 7	1.050 5	0.037 0
决策树	0.020 7	0.534 6	0.535 7	1.148 5	0.039 8

由表 2 可知,深度神经网络(DNN)具有较好的预测结果。对比数据发现:在精确率上,DNN 的结果相对较好,分别比 Logistic 回归等三种算法高出 0.1%、0.33%、0.15%。精确率表示预测为流失客户的样本中的正确率,DNN 的精确率最高,表明在预测为流失客户的样本集中,DNN 预测正确的比例相对更高;在召回率上,DNN 的结果低于其他三种算法,说明 DNN 在实际流失的样本集中,预测正确的比例较低;在准确

率上,DNN 的表现也优于其他三种算法,说明 DNN 预测正确的流失样本与非流失样本的数量更多;在提升系数上,DNN 的表现同样优于其他三种算法,提升效果明显;在 F_1 值上,DNN 的结果同样优于其他三种算法, F_1 值是精确率和召回率的一种加权平均,DNN 的精确率比其他算法高,召回率比其他算法低,但 F_1 值最高,同时测试数据具有严重的类别不平衡性,说明 DNN 的综合性能更优。

朴素贝叶斯模型的召回率高达0.8268,但精确度、 F_1 值在四个预测模型中最低,说明朴素贝叶斯模型预测错误的非流失客户更多,模型的整体性能不高。整体而言,与经过特征选择的Logistic回归等模型相比,DNN具有较好的预测效果。

5 结束语

客户流失预测是一个不断发展的课题,过去的研究成果解决了客户流失预测领域的一些重要问题,但随着大数据时代的来临,客户流失预测出现了新的特点,例如数据的超大规模、更高的复杂性等,对经典的预测方法提出了挑战,需要新的方法来应对变化。深度学习在处理大数据方面具有很大的优势,在图像、语音、自然语言处理等领域取得了较为显著的成果,但在客户流失预测方面的研究较少。为探究深度学习在客户流失预测上的效果,构造了包含3个隐层的深度神经网络,并在某电信客户数据集上与Logistic回归、决策树等常用预测算法进行对比,实验结果表明,与经过特征选择的Logistic回归等模型相比,构造的深度神经网络模型拥有较好的预测效果。由于条件所限,未能构建拥有更多隐层的深度神经网络模型,也未能在更大规模的数据集上验证深度神经网络的有效性。下一步,将探究更深层神经网络的性能以及卷积神经网络等经典模型在网络客户流失预测上的应用,并搜集更大规模的数据用于分析预测大数据环境下的客户流失问题。

参考文献:

- [1] 姜晓娟,郭一娜. 基于改进聚类的电信客户流失预测分析[J]. 太原理工大学学报,2014,45(4):532-536.
- [2] 李洋. 基于神经网络的客户流失数据挖掘预测模型[J]. 计算机应用,2013,33(S1):48-51.
- [3] KASIRAN Z, IBRAHIM Z, MOHD RIBUAN M S. Customer churn prediction using recurrent neural network with reinforcement learning algorithm in mobile phone users[J]. International Journal of Intelligent Information Processing, 2014,5(1):1-11.
- [4] 冯鑫,王晨,刘苑,等. 基于评论情感倾向和神经网络的客户流失预测研究[J]. 中国电子科学研究院学报,2018,13(3):340-345.
- [5] KIRUI C, HONG L, CHERUIYOT W, et al. Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining[J]. International Journal of Computer Science Issues, 2013,10(1):165-172.
- [6] 尹婷,马军,覃锡忠,等. 贝叶斯决策树在客户流失预测中的应用[J]. 计算机工程与应用,2014,50(7):125-128.
- [7] 盛昭瀚,柳炳祥. 客户流失危机分析的决策树方法[J]. 管理科学学报,2005,8(2):20-25.
- [8] 张宇,张之明. 一种基于C5.0决策树的客户流失预测模型研究[J]. 统计与信息论坛,2015,30(1):89-94.
- [9] 夏国恩,金炜东. 基于支持向量机的客户流失预测模型[J]. 系统工程理论与实践,2008,28(1):71-77.
- [10] 王观玉,郭勇. 支持向量机在电信客户流失预测中的应用研究[J]. 计算机仿真,2011,28(4):115-118.
- [11] CHEN Zhenyu, SHU Peng, SUN Minghe. A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data[J]. European Journal of Operational Research, 2012,223(2):461-472.
- [12] 赵琨,许洪贵,田英杰. 基于双子支持向量机的信用卡流失分析[J]. 数学的实践与认识,2015,45(17):85-92.
- [13] 罗彬,邵培基,夏国恩. 基于多分类器动态选择与成本敏感优化集成的电信客户流失预测研究[J]. 管理学报,2012,9(9):1373-1381.
- [14] COUSSEMENT K, BOCK K W D, MIZERSKI D. Customer churn prediction in the online gambling industry: the beneficial effect of ensemble learning[J]. Journal of Business Research, 2013,66(9):1629-1636.
- [15] PHADKE C, UZUNALIOGLU H, MENDIRATTA V B, et al. Prediction of subscriber churn using social network analysis[J]. Bell Labs Technical Journal, 2013,17(4):63-76.
- [16] VERBEKE W, MARTENS D, BAESSENS B. Social network analysis for customer churn prediction[J]. Applied Soft Computing Journal, 2014,14(1):431-446.
- [17] 黄婉秋. 一种基于社交网络的潜在流失客户发现方法[J]. 北京交通大学学报,2014,38(3):123-127.
- [18] 余凯,贾磊,陈雨强,等. 深度学习的昨天、今天和明天[J]. 计算机研究与发展,2013,50(9):1799-1804.
- [19] 孙志远,鲁成祥,史忠植,等. 深度学习研究与进展[J]. 计算机科学,2016,43(2):1-8.
- [20] 刘建伟,刘媛,罗雄麟. 深度学习研究进展[J]. 计算机应用研究,2014,31(7):1921-1930.