

基于深度神经网络权重集成的客户流失预测

利向晴¹, 夏国恩^{1,2}, 张显全¹, 唐琪¹, 叶帅¹

(1. 广西师范大学 计算机科学与信息工程学院, 广西 桂林 541004;

2. 广西财经学院 工商管理学院, 广西 南宁 530003)

摘要:如今许多的企业都面临着客户流失的实质问题,客户流失预测企业的发展也是尤为重要的。对于客户流失预测的问题也有很多解决方案,传统的机器学习方法存在特征工程对模型效果影响较大的缺点,而深度学习则使得算法不会太依赖于领域专业知识和人工特征提取。为提升预测效果,提出一种深度神经网络权重集成方法来对电信客户流失进行预测,其主要思想是在做深度神经网络(DNN)训练的时候,通过随机加权平均(stochastic weight average, SWA)结合同一网络结构下的不同训练阶段的权重获取集成模型,然后对客户流失进行预测。实验结果表明,相比于深度神经网络训练,结合随机加权平均权重集成的深度神经网络训练时间缩短了2.96倍,同时,准确率、精准率、召回率和F1值均有所提升。

关键词:客户流失;深度学习;深度神经网络;随机加权平均;权重集成

中图分类号:TP183

文献标识码:A

文章编号:1673-629X(2021)10-0018-06

doi:10.3969/j.issn.1673-629X.2021.10.004

Customer Churn Prediction Based on Deep Neural Network Weight Ensemble

LI Xiang-qing¹, XIA Guo-en^{1,2}, ZHANG Xian-quan¹, TANG Qi¹, YE Shuai¹

(1. School of Computer Science and Information Engineering, Guangxi Normal University, Guilin 541004, China;

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

Abstract: Now many enterprises are faced with the problem of customer churn, and customer churn prediction is particularly important for the development of enterprises. There are many solutions for customer churn prediction. Traditional machine learning methods have the disadvantage that feature engineering has great influence on model effect, and deep learning makes the algorithm less dependent on domain expertise and artificial feature extraction. In order to improve the prediction effect, a deep neural network combined with stochastic weights average is proposed to predict customer churn. The main idea is to obtain the integration model by combining the stochastic weight average (SWA) with the weights of different training stages under the same network structure when doing DNN training, and then predict customer churn. The experiment shows that compared with the deep neural network training, the training time of the deep neural network combined with the stochastic weighted average weight integration is shortened by 2.96 times, and the accuracy, precision, recall and F1 are improved.

Key words: customer churn; deep learning; deep neural network; stochastic weight average; weighted ensemble

0 引言

随着社会经济的不断发展,各企业在不断涌现,行业的竞争压力也越来越大,许多公司都面临客户流失的实质问题。越来越多的公司开始意识到一个事实:挽留现有的客户是想在行业中生存的最好的市场营销方式,因为吸引新客户成本要比挽留现有客户的成本

大得多^[1]。长期性客户对企业更有利,企业在想办法吸引新客户的同时,也应该避免客户的流失。《Harvard 哈佛商业评论》认为,若能将客户流失率降低5%,则企业的利润将增加25%~85%,而美国著名的财政企业杂志《商业周刊》则认为利润将增加140%^[2]。一般来说,客户流失分为两类:一类是意外

收稿日期:2020-11-20

修回日期:2021-03-22

基金项目:国家自然科学基金资助项目(71862003);2019年度广西高校中青年教师科研基础能力提升项目(2019KY0661)

作者简介:利向晴(1996-),女,硕士研究生,研究方向为数据挖掘与客户流失预测;通讯作者:夏国恩(1977-),男,教授,研究方向为商务智能与客户关系管理。

流失,有时因为生活环境的原因,如客户离开了公司服务的范围,或是客户的财政状况发生了变化,导致客户被迫放弃他们的服务;另一类是主动性流失,由于客户决定将其服务转移到别的公司而导致的流失,这背后的原因或许是别的公司提供了更优的服务^[3]。所以在客户流失之前进行预测,并采取适当营销策略对客户进行挽留,将客户流失的损失降到最低是非常重要的,不仅可以提高企业的利润,也能避免企业在核心竞争中被淘汰。目前对于客户流失预测的方法有很多,如逻辑回归(logistic regression, LR)^[4]、XGBoost 算法^[5]、循环神经网络(recurrent neural network, RNN)^[6]、多层感知机(multi-layer perception, MLP)^[7]等。

研究者们将客户流失预测问题视为二分类问题,如今,客户流失预测方面的研究取得了长足的进展。为提升模型预测的泛化能力,夏国恩等人^[8]提出了一种基于支持向量机的客户流失预测模型,结果表明支持向量机在客户流失预测方面是有效的;周捷等人^[9]提出了基于 LSTM 的模型集成方法,对客户流失进行预测, LSTM 自动学习序列隐含信息以及复杂高阶特征,减少了对特征工程的依赖,同时对其进行集成提高了预测效果;E. Jamalain 等人^[10]提出了一种数据融合和特征提取技术的混合方法,以更准确地预测客户的变动;Wangperawong A 等人^[11]提出深度卷积神经网络和自动编码对客户流失进行分析,使用自动编码无监督学习,以便更好地了解客户流失的原因;Yu R 等人^[12]提出了一种基于客户细分优化的 BP 网络用于客户流失预测,该算法优化了 BP 神经网络的初始权重和阈值,大大提高了客户流失预测的准确性;Mishra A 等人^[13]利用卷积神经网络进行客户流失预测,实验取得了很好的效果,相比于传统机器学习,通过深度学习能更好地发现客户流失的潜在风险。

如今深度学习在图像处理(物体识别、场景识别、人脸身份认证、恢复黑白照片和视频颜色等)、自然语言处理(翻译、文本识别、聊天对话、在野外阅读文本等)、语音处理(语音识别、音乐创作、恢复视频声音等)中都得到了很好的应用^[14]。深度学习算法最大的优点是它尝试以增量形式从数据中逐步学习高级功能。这就使得算法不会太依赖于领域专业知识和人工特征提取。

文中提出一种深度神经网络权重集成方法(stochastic weight average deep neural network, swaDNN)来预测电信客户流失,在做深度神经网络训练的时候,通过随机加权平均(stochastic weight average, SWA)结合相同网络结构的不同训练阶段的权重获得集成模型,对客户流失进行预测。实验结果表明,该方法显著提升了客户流失预测的效果。

1 相关理论基础

1.1 深度神经网络

深度神经网络(deep neural network, DNN)^[15]通过其隐层逐步学习类别,神经网络将多层隐藏层添加到感知器模型中,以增强模型的泛化能力。为使模型更具灵活性,以便更好应用于分类、聚类 and 回归等问题,神经网络层由节点(神经元)组成,其输入函数如下所示:

$$Z = \sum_{i=1}^j w_i x_i + b \quad (1)$$

其中, i 表示第 i 层, j 表示神经网络层数, w 表示权重, b 表示偏置, Z 表示输入函数。单个神经元结构如图 1 所示。

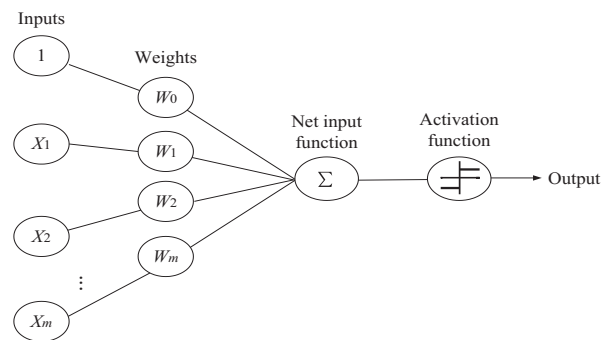


图 1 单个神经元结构

神经元将来自数据的输入与一组系数或权重进行组合,这些系数或权重会放大或衰减该输入,将这些输入权重乘积相加,然后将总和传递给节点的激活函数,激活函数将非线性因素加入到神经网络中,并将神经元的输入映射到输出端,以确定该信号是否应通过网络进一步传输以影响最终结果(如分类),信号通过,则说明神经元已被“激活”。根据各层的位置, DNN 神经网络层分为输入层、隐藏层和输出层,如图 2 所示。

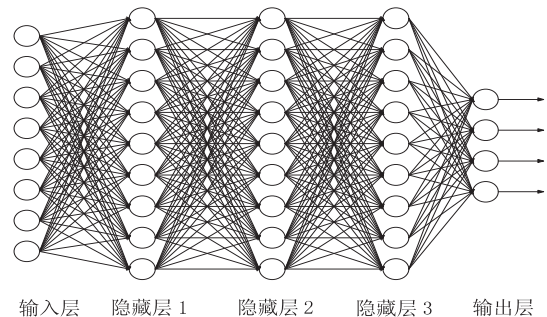


图 2 DNN 模型

DNN 模型图小圆圈代表着神经元,从图中可看出,同一层神经元之间没有连接,层与层之间全连接,神经网络每个连接都有权重,最终输出结果对应的层也成为全连接层,隐藏层和输出层的神经元由输入的数据计算后输出,输出层神经元可以是一个或多个

输出。

1.2 随机加权平均权重集成

随机加权平均 (SWA)^[16] 在随机梯度下降 (stochastic gradient descent, SGD)^[17] 的基础上进行了改进, 随机梯度下降是一种为机器学习算法找到最佳参数配置的方法。迭代地对机器学习网络的参数进行小的调整, 以减少网络的错误。SWA 以周期性或较高的恒定学习率对 SGD 遍历的点进行加权平均。一般的集成方法是对不同的模型进行集成, 然后用相同的输入对模型进行预测, 集成模型的最终预测由某种平均方法来确定。

传统的集成方法有: Snapshot 集成、Fast Geometric Ensembling 集成和 Stacking 集成。Snapshot 集成方法利用循环学习率的策略来训练深度学习模型, Stacking 集成方法通过将训练集分成若干部分, 为同一机器学习算法生成多个参数不同的分类器, Fast Geometric Ensembling 集成使用线性分段的循环学习率, 来取代 Snapshot 集成中的余弦, 这些都是基于模型空间的集成方法, 需要多个训练模型来预测每个模型, 并对最终的预测结果进行平均。而随机加权平均为基于权重空间的新集成方法, 经过训练后的网络为多维权重空间的点, 对于给定的网络结构, 不同权值的组合将产生不同的模型, SWA 的重点是使用改进 SGD 迭代周期性学习率 (lr_schedule 为 cyclic) 或高恒定学习率 (lr_schedule 为 constant), 并利用深度学习训练目标的平坦性来提高泛化能力。其中, 周期性学习率在每周线性地将学习率从 α_1 降到 α_2 , 第 i 次迭代的学习率公式为:

$$\alpha(i) = (1 - t(i))\alpha_1 + t(i)\alpha_2 \quad (2)$$

$$t(i) = \frac{1}{c}(\text{mod}(i - 1, c) + 1) \quad (3)$$

其中, $\alpha_1 \geq \alpha_2$, c 表示周期长度, $t(i)$ 表示第 i 次迭代的时间。SWA 通过加入周期性加权平均来限制权重的变化, 解决了传统 SGD 在反向过程中的权重震荡问题。SWA 局部极小值倾向于累积在损耗值较低的损耗表面区域的边界上, 通过取几个这样的点的平均值, 使得结果具有更低的损耗。SGD 收敛到一个宽平坦损失区域内的解。重量空间是极高维的, 且大部分平面区域的体积集中在边界附近, 因此 SGD 的解总是在平面区域的边界附近找到。另一方面, SWA 有多个 SGD 解决方案, 这使它能够向平坦区域的中心移动。SWA 在训练过程中, 第一个模型用于存储模型权重的平均值, 并在训练结束时作为最终模型用于预测结果; 第二个模型将遍历权重空间, 并用周期性学习率对其进行探索。随机加权平均的权重更新方程如式 (4) 所示:

$$w_{\text{swa}} = \frac{w_{\text{swa}} \cdot n_{\text{models}} + w}{n_{\text{models}} + 1} \quad (4)$$

其中, w_{swa} 表示存储模型权重的平均值, w 表示用于遍历权重空间的值, n 表示模型数。SWA 算法描述如下:

算法: 随机加权平均。

输入: 权重 \hat{w} , 学习率 α_1, α_2 ;

周期长度 c (对于恒定的学习率则 $c = 1$), 迭代次数 n

输出: w_{swa}

```

1.  $w = \hat{w}$ 
2.  $w_{\text{swa}} = w$ 
3. for  $i = 1, 2, \dots, n$  do
4.    $\alpha_i = \alpha(i)$ 
5.    $w = w - \alpha \nabla l_i(w)$ 
6.   if  $\text{mod}(i, c) = 0$  then
7.      $n_{\text{models}} = i/c$ 
8.      $w_{\text{swa}} = \frac{w_{\text{swa}} \cdot n_{\text{models}} + w}{n_{\text{models}} + 1}$ 
9.   end if
10. end for
11. return  $w_{\text{swa}}$ 

```

算法第 1 步初始化权重 \hat{w} , 第 3 ~ 10 步通过一个 for 循环对 w_{swa} 进行更新, 第 5 步进行随机梯度更新, n_{models} 表示模型数, 第 8 步更新平均值, 最终输出 w_{swa} 。

2 深度神经网络权重集成方法的客户流失预测模型

文中提出的基于深度神经网络权重集成的客户流失预测框架如图 3 所示, 主要包含客户数据、数据预处理、模型训练、结果评价几个阶段, 原始的客户数据并不都是可以直接拿来用的, 那会导致某些错误。数据预处理是为了获取更好的数据来进行训练, 接着进行模型训练得到预测结果, 企业可以根据预测制定相应的客户挽留策略, 以免造成更严重的客户流失问题。



图 3 客户流失预测框架

DNN 模型结构如图 4 所示。首先是对客户的数据进行预处理, 进行数据清洗和数据变换, 通过 3 层隐藏层, 最终由输出层输出, 得到预测结果。每一层可以有一个或多个神经元, 文中模型隐层神经元选用 8 个, 输出层只有 1 个神经元。



图 4 DNN 模型结构

神经元常见的激活函数包括 tanh、elu、sigmoid、relu、maxout 等, 在文中实验中, 神经元的激活函数选

用 relu 。 relu 函数能克服梯度消失的问题,使得神经网络具有更快的训练速度, relu 函数的表示如公式(5)所示:

$$f(x) = \max(0, x) \quad (5)$$

输出层设置了 1 个神经元,使用 Sigmoid 作为激活函数,输出在 0 和 1 之间,表示如公式(6)所示:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

客户流失预测是二分类问题,文中实验采用 Binary Cross_entropy 作为损失函数:

$$\text{loss} = - \sum_{i=1}^n \hat{y}_i \log y_i + (1 - \hat{y}_i) \log(1 - \hat{y}_i) \quad (7)$$

其中, \hat{y}_i 表示预测结果, y_i 表示真实值,当 $\hat{y}_i = y_i$ 时, $\text{loss} = 0$,表示无损失。

深度神经网络的层数和神经元数对模型训练的结果具有较大的影响,通过实验得出 3 层隐藏层和隐藏层的神经元设为 8 个的拟合效果较好,输出层只有 1 个神经元。SGD 学习率设为 0.1,SWA 采用周期性学习率,周期长度 c 为 10,学习率 $\alpha_1 = 0.001$, $\alpha_2 = 0.003$,在训练过程中,在模型初始化参数之后,使用 SGD 进行梯度下降,迭代了 c 个 epoch 之后,将模型的参数用加权平均,得到 w_{swa} ,当前模型参数为 w_{swa} ,接着再用 SGD 梯度下降 c 个 epoch,再进行加权平均得到新的 w_{swa} 。训练结束后只会得到一个具有组合权重的集成模型,这将加快后续模型预测的速度。

3 实验

3.1 实验环境

该实验使用 Windows10 操作系统,软件为 Jupyter Notebook,具体硬件配置如下:内存 64 GB;CPU 型号为 Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10 GHz。实验中使用的工具:python3.7、Sklearn、Tensorflow、Keras。

3.2 数据选用及预处理

文中使用的数据集是电信客户流失数据集,选自 kaggle 数据科学大赛。该数据集共包含 7 043 个样本,其中非流失的客户数据为 5 174,流失客户数为 1 869,客户流失率为 26.5%,其中存在 11 个缺失值,对其进行数据清洗,最终获得样本数为 7 032。训练数据大小 5 625,测试数据 1 407,从客户流失数量与非客户流失数量来看,该数据集属于不平衡数据。根据数据的各个属性对流失率的影响,可将流失率相关性低的忽略掉,共获得 17 个属性指标,其中离散属性 13 个,数值属性 4 个,对离散特征进行 one-hot 编码,使用 Scikit-learn 标签编码,将分类的数据变换为整数编码形式。

3.3 对比模型

为了充分验证文中提出的随机加权平均优化深度

神经网络的方法,采用 Random Forest、Logistic Regression、CatBoost、Support Vector Machine、AdaBoost 作为基准对比模型。

(1) Random Forest:随机森林^[18]由多个决策树分类器组成,随机森林的每棵决策树都对类别进行预测,将投票多的决策树作为预测的模型。采用重复 m 次的有放回的采样,并在这些样本上训练树模型,最终取所有单个回归树的预测的平均来实现对未知样本的预测。

$$\hat{f} = \frac{1}{m} \sum_{i=1}^m f_i(x') \quad (8)$$

$$\delta = \sqrt{\frac{\sum_{i=1}^m (f_i(x') - \hat{f})^2}{m - 1}} \quad (9)$$

其中, \hat{f} 表示单个回归树的预测值, m 表示样本或树的数量, δ 表示所有单个回归树的预测的标准差。

(2) Logistic Regression:逻辑回归(LR)是用于分类的模型,逻辑回归将线性回归的输出作为它的输入,逻辑回归通过 Sigmoid 函数映射来解决分类问题,逻辑回归模型和损失函数可表示为:

$$h(x) = \frac{1}{1 + e^{-\theta^T x}} \quad (10)$$

$$\text{cost}(h_\theta(x), y) = \sum_{i=1}^m [-y_i \log(h_\theta(x)) - (1 - y_i) \log(1 - h_\theta(x))] \quad (11)$$

其中, m 表示样本数, y_i 表示真实值, $\theta^T x$ 表示线性回归模型的矩阵形式。一般使用梯度下降法对逻辑回归损失函数进行求解,以减少损失函数的值,最终使得预测效果更准确。

(3) CatBoost:是对传统 Boosting 算法的重大改进,CatBoost 主要思想是在决策树中,先对样本随机排序,节点分裂的标准为类别标签平均值,并通过添加先验分布项来减少噪声和低频率分类数据对于数据分布的影响。

$$\hat{x}_k^i = \frac{\sum_{j=1}^p [x_{\delta_{j,k}} = x_{\delta_{p,k}}] \cdot Y_j + a \cdot p}{\sum_{j=1}^n [x_{\delta_{j,k}} = x_{\delta_{p,k}}] + a} \quad (12)$$

其中, p 是添加的先验项, a 通常是大于 0 的权重系数, $\sigma_{j,k}$ 表示第 j 个数据的第 k 个特征。

(4) Support Vector Machine:支持向量机(SVM)可用于分类和回归,算法的目的是在 n 维空间(n 特征数)中找到可对数据点明确分类的超平面。支持向量机的约束优化问题表示为:

$$\min_{\theta} \frac{1}{m} \left[\sum_{i=1}^m y^{(i)} \text{cost}_1(\theta^T x^{(i)}) + (1 - y^{(i)}) \text{cost}_0(\theta^T x^{(i)}) \right] + \frac{1}{2} \sum_{i=1}^n \theta_j^2$$

$$\text{s. t. } \begin{cases} \theta^T x^{(i)} \geq 1, & \text{if } y^{(i)} = 1 \\ \theta^T x^{(i)} \leq -1, & \text{if } y^{(i)} = 0 \end{cases} \quad (13)$$

支持向量机约束优化问题由损失项和正则项以及约束项组成,其中 n 为样本数,通过上述代价函数来获得一个最优参数 θ ,将优化问题转化为最小参数向量 θ 的范数平方,最终找到能产生最大间距分类现象的 θ 范数。

(5) AdaBoost^[19]:其核心对弱分类器进行迭代训练,弱分类器的权重更新方式如公式(14)所示,在每次迭代中加入一个新的弱分类器,直到加权误差率达到预定的值。

$$w_{t+1,i} = \frac{w_{t,i}}{Z_t} \exp(-\alpha_t y_i G_t(x_i)) \quad (14)$$

$$G(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t G_t(x)\right) \quad (15)$$

其中, $t = 1, 2, \dots, T$, 表示第 t 轮迭代, w 表示样本权重, i 表示第 i 个样本, Z_t 表示归一化因子,第 t 轮迭代的 α_t 可表示为 $0.5 * \ln((1 - \varepsilon_t)/\varepsilon_t)$, 其中 ε_t 为弱分类器 $G_t(x)$ 的加权误差率,当 $\varepsilon_t > 0.5$ 时,达到终止条件, $G(x)$ 为最终的分类器。通过分类结果对权重进行更新可以使被错误分类的样本权重变大,从而在下一轮迭代中得到重视。

3.4 评价指标

为评估模型,采用准确率 (accuracy)、精准率 (precision)、召回率 (recall) 和精准率与召回率的调和平均值 F1 值来定义模型对客户流失预测的效果。

表 1 分类结果混淆矩阵

真实情况	预测结果	
	预测为流失客户	预测为非流失客户
流失客户	TP	FN
非流失客户	FP	TN

准确率、精准率、召回率和 F1 值的计算公式如下:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (16)$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (17)$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (18)$$

$$\text{F1} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (19)$$

其中,TP 表示对客户流失正确预测的样本数;FN 表示实际为流失客户的数据却错误地将其预测为非客户流失的样本数;FP 表示被错误预测为流失客户的非流失客户样本数;TN 为对非客户流失正确预测的样本数。

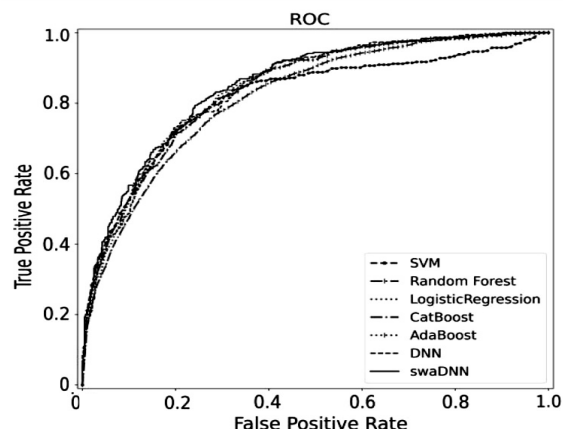
3.5 实验结果以及分析

文中采用 5 折交叉验证,以测试集实验结果的平均值作为模型评判的标准,实验结果如表 2 和图 5 所

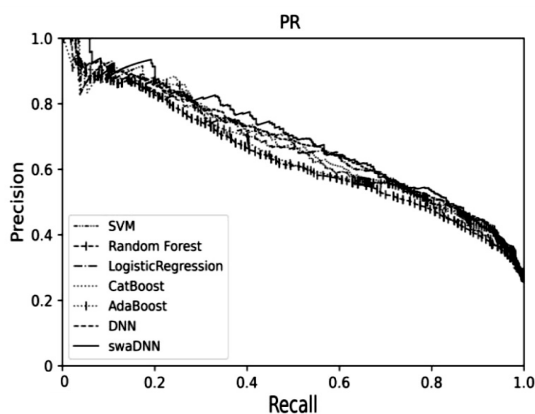
示。随机森林模型在数据集中整体性能的不是很好,说明随机森林模型不太适用于非平衡数据,不能很好地区分流失和非流失数据。支持向量机 (SVM) 模型对数据多的分类的误差比数据少的分类误差小,因为其对不平衡样本数据的预测具有倾向性。逻辑回归模型形式简单,特征对数据结果影响较大,且对正负样本不平衡的数据的分类不是很友好。CatBoost 和 AdaBoost 都是 Boosting 算法的一种实现,相比之下,AdaBoost 的综合性能较好,模型较稳健,但是在处理非平衡数据上精度不是最好的。文中实验 DNN 模型使用 Adam 优化算法迭代更新神经网络的权重,召回率不是很高,表明流失用户被正确预测出来的概率不是很高。swaDNN 中 SGD 的学习率设为 0.1,SWA 的 lr_schedule 采用 cyclic 的形式,相比于 DNN,swaDNN 的准确率、精准率、召回率和 F1 值都分别高出了 0.53%、0.22%、9.26% 和 5.20%,在相同 batch size 情况下,总计训练 100 个 epoch,DNN 训练时间为 52.92 s,swaDNN 的训练时间为 17.84 s,相比之下,swaDNN 的训练时间缩短了 2.96 倍,大大提高了训练的速率。通过与其他模型进行对比,swaDNN 的准确率、精准率、召回率和 F1 值要优于其他模型,实验结果表明文中提出的神经网络权重集成方法可以有效提升预测效果。

表 2 不同模型的预测效果

模型	准确率/%	精准率/%	召回率/%	F1 值/%
RandomForest	78.18	61.75	47.06	53.41
CatBoost	79.95	65.03	53.21	58.85
SVM	79.45	64.81	47.42	54.77
LR	80.02	64.66	52.27	57.99
AdaBoost	80.88	67.21	54.81	60.38
DNN	80.59	67.39	50.41	57.67
swaDNN	81.02	67.54	55.08	60.67



(a) ROC 曲线



(b) PR 曲线

图5 各模型 ROC 曲线与 PR 曲线

4 结束语

文中提出一种深度学习神经网络权重集成方法对电信客户流失进行预测,在做深度神经网络训练的时候,通过随机加权平均(SWA)结合相同网络结构不同训练阶段的权重获得集成模型,对客户流失进行预测。实验结果表明,该方法对客户流失预测具有较好的效果。对模型训练具有更广泛的优化和更好的泛化,提高了神经网络的性能。客户流失预测是一个不断发展的问題,数据的不断庞大,需要新的方法去应对,所以在未来的工作中应不断改进方法以适应复杂性更高的数据。

参考文献:

- [1] 夏国恩. 客户流失预测的现状与发展研究[J]. 计算机应用研究, 2010, 27(2): 413-416.
- [2] AMUDA K A, ADEYEMO A B. Customers churn prediction in financial institution using artificial neural network [EB/OL]. (2019-12-23) [2021-01-26]. <https://arxiv.org/pdf/1912.11346.pdf>.
- [3] SPANOUEDES P, NGUYEN T. Deep learning in customer churn prediction: unsupervised feature learning on abstract company independent feature vectors [EB/OL]. (2017-03-10) [2021-01-26]. <https://arxiv.org/pdf/1703.03869.pdf>.
- [4] CAIGNY A D, COUSSEMENT K, BOCK K W D. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees [J]. European Journal of Operational Research, 2018, 269(2): 760-772.
- [5] 王重仁, 韩冬梅. 基于社交网络分析和 XGBoost 算法的互联网客户流失预测研究 [J]. 信息技术与网络安全, 2017, 36(23): 58-61.
- [6] KASIRAN Z, IBRAHIM Z, MOHD R 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.
- [7] 夏国恩, 唐琪, 张显全. 改进的多层感知机在客户流失预测中的应用 [J]. 计算机工程与应用, 2020, 56(14): 257-263.
- [8] 夏国恩, 金纬东. 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.
- [9] 周捷, 严建峰, 杨璐, 等. LSTM 模型集成方法在客户流失预测中的应用 [J]. 计算机应用与软件, 2019, 36(11): 39-46.
- [10] JAMALIAN E, FOUKERDI R. A hybrid data mining method for customer churn prediction [J]. Engineering, Technology & Applied Science Research, 2018, 8(3): 2991-2997.
- [11] WANGPERAWONG A, BRUN C, LAUDY O, et al. Churn analysis using deep convolutional neural networks and autoencoders [EB/OL]. (2016-04-18) [2021-01-26]. <https://arxiv.org/pdf/1604.05377.pdf>.
- [12] YU R, AN X, JIN B, et al. Particle classification optimization based BP network for telecommunication customer churn prediction [J]. Neural Computing and Application, 2018, 29(3): 707-720.
- [13] MISHRA A, REDDY U S. A novel approach for churn prediction using deep learning [C]//2017 IEEE international conference on computational intelligence and computing research (ICIC). Piscataway, NJ: IEEE, 2017: 1-4.
- [14] 马文斌, 夏国恩. 基于深度神经网络的客户流失预测模型 [J]. 计算机技术与发展, 2019, 29(9): 76-80.
- [15] SZE V, CHEN Y, YANG T, et al. Efficient processing of deep neural networks: a tutorial and survey [J]. Proceedings of the IEEE, 2017, 105(12): 2295-2329.
- [16] IZMAILOV P, PODOPRIKHIN D, GARIPPOV T, et al. Averaging weights leads to wider optima and better generalization [C]//Proceedings of the thirty-fourth conference on uncertainty in artificial intelligence. Monterey: AUAI Press, 2018: 876-885.
- [17] ZHANG Y, SAXE A M, ADVANI M S, et al. Energy - entropy competition and the effectiveness of stochastic gradient descent in machine learning [J]. Molecular Physics, 2018, 116(21): 3214-3223.
- [18] XIE Y, LI X, NGAI E W, et al. Customer churn prediction using improved balanced random forests [J]. Expert Systems with Applications, 2009, 36(3): 5445-5449.
- [19] 武小军, 孟苏芳. 基于客户细分和 AdaBoost 的电子商务客户流失预测研究 [J]. 工业工程, 2017, 20(2): 99-107.