

基于长短期记忆网络算法的电费回收风险预警

谢林枫¹, 钱立军², 季 聪¹, 江 明², 吕 辉¹

(1. 江苏方天电力技术有限公司, 江苏 南京 211102;

2. 国网江苏省电力有限公司, 江苏 南京 210024)

摘要:电力大客户电费回收风险一直都是电力公司关注的热点,但由于缺乏外部数据支撑和高效预警模型,使电费回收风险预警成为了一个难题。文中首先综合电力客户的电量、电费数据,以及电力客户在工商、税务、法院等部门的风险信息,建立电力大客户的电费回收风险指标体系。其次,基于熵值法得到的风险指标权重系数,过滤弱影响指标,采用相关性分析剔除重叠作用指标,得到客户电费回收风险预警指标。最后,基于深度学习中的长短期记忆(long short-term memory, LSTM)网络算法进行了客户电费回收风险预警。算例结果表明,提出的风险预警模型精确有效,且LSTM在准确率、查准率和查全率3个指标上较Logistic回归更加精准,能够精准定位风险客户,提高电费回收效率。

关键词:电力大客户;电费回收风险;风险评估;风险预警;深度学习;长短期记忆网络

中图分类号:TM71

文献标志码:A

文章编号:2096-3203(2018)05-0098-06

0 引言

一直以来,电费回收率^[1]都是电力公司的重要考核指标,直接关系到电力公司的运营效率和水平。电力大客户合同容量大、用电量,其是否及时缴费,很大程度上影响着电费回收率。因此,近年来,电力公司一直将电费回收风险^[2]作为其运营风险的核心指标之一,也围绕这一主题开展了大量的研究工作,取得了一定的成效。

文献[3—4]采用层次分析法、熵值法建立了电力大客户信用评价体系,用于防范电费回收风险,而本文综合考虑了电力客户电量、电费、工商、司法等数据,建立大客户电费回收风险预警体系。文献[5—6]采用逻辑回归、随机森林等传统数据挖掘算法实现了客户电费回收风险预警,而本文采用长短期记忆(long short-term memory, LSTM)网络算法提取客户风险特征,实现电费回收风险预警。文献[7—8]借助大数据技术、“客户画像”等新技术开展了电费回收风险研究。

电费回收风险受客户生产运营情况、财务经济状况、宏观经济政策等众多因素的影响,相关因素的筛选和量化较为困难。而深度学习^[9-11]算法能够自动选取关键因素,目前在电力负荷预测、变压器故障诊断等方面得到了应用^[11-16]。因此本文研究了循环神经网络(recurrent neural network, RNN)算

法的基本原理,考虑到电费回收风险预警随时间推移存在一定规律,本文进一步研究了RNN算法的特殊类型——LSTM网络算法,用于开展电费回收风险预警工作,取得了较传统算法更精确的预警结果。

1 电费回收风险的指标体系

1.1 指标体系建立

电力大客户电费回收风险受其所在行业背景、自身实力、经营能力等多重因素影响,因此,本文基于电力内部的电量、电费、违约用电等数据,结合外部的工商、税务、法院信息,建立了电力大客户的电费回收风险指标体系,如表1所示。

1.2 指标筛选

在进行客户风险预警模型构建之前,需要对客户风险指标进行筛选,过滤掉弱影响指标,并利用相关性分析剔除作用重复的指标。

首先采用熵值法^[17]计算表1中的各指标权重,结果如表2所示。过滤弱影响指标,剔除表2中权重系数绝对值<1.5%的8个指标,得到27个权重系数较高的指标。对剩余的27个指标进行相关性分析,得到相关性系数较高的指标见表3。剔除实收资本、前年用电量增长情况和最近一年电费回收准时度3个指标,得到最终的24个与客户欠费风险密切相关的指标。

2 深度学习算法的基本原理

2.1 RNN的基本原理

在传统神经网络模型中,输入层、隐含层和输

收稿日期:2018-05-09;修回日期:2018-06-20

基金项目:国家重点研发计划资助项目(2016YFB090-1100);国家电网有限公司指南科技项目(SGTYHT/14-JS-188)

表 1 电费回收风险指标

Tab.1 Tariff recovery risk indices

一级维度	二级维度	三级维度	数据来源
行业背景	行业发展	行业产能增长率	用电数据
		行业产能利用率	用电数据
	行业地位	企业产能利用率在行业中的水平	用电数据
		企业用电量在行业中的水平	用电数据
企业实力	企业股东背景	企业类型	工商登记信息
		股东组成	工商登记信息
	核心股东	核心股东类型	工商登记信息
		核心股东出资比例	工商登记信息
	资本实力	注册资本	工商登记信息
		实缴资本	工商登记信息
商标专利	著名/驰名商标总数	知识产权数据	
	专利总数	知识产权数据	
经营能力	员工及分支机构	员工总人数	工商登记信息
		分支机构数	工商登记信息
	盈利能力	最近一年资产回报率	纳税信息
		净利润率	纳税信息
		企业产能利用率	用电数据
	生产稳定性	每月应收电费金额波动性	用电数据
		每月用电量波动性	用电数据
	经营时长	成立时长	工商登记信息
		企业涉诉被执行记录	法院信息
	异常信息	企业经营异常	工商登记信息
行政处罚			工商登记信息
缴费行为		缴费频率稳定性	用电数据
企业信用	客户余额使用时长	用电数据	
		客户欠费记录	用电数据
违规行为	客户窃电行为	用电数据	
		客户违法用电行为	用电数据
生产状态	电量数据	去年用电量增长情况	用电数据
		前年用电量增长情况	用电数据
	近一年累计拖欠电费数量	年度客户用电量同比增长率	用电数据
		近半年月均电费	用电数据
	电费数据	近一年月电费谷值	用电数据
		近一年电费回收准时度	用电数据

表 2 各电费回收风险指标权重系数

Tab.2 Weight of each tariff recovery risk index

指标名称	权重 / %	指标名称	权重 / %
行业产能增长率	3.872 9	每月用电量波动性	3.886 8
行业产能利用率	3.996 3	成立时长	1.786 3
企业产能利用率在行业中水平	2.815 9	企业涉诉被执行记录	3.976 2
企业用电量在行业中水平	0.022 1	企业经营异常	3.933 6
企业类型	4.099 7	行政处罚	3.824 7
股东组成	3.995 0	缴费频率稳定性	4.270 4
核心股东类型	0.361 2	客户余额使用时长	0.089 9
核心股东企业规模	0.265 1	客户欠费记录	6.558 8
注册资本	1.894 8	客户窃电行为	2.167 5
实缴资本	2.077 7	客户违法用电行为	0.806 7
著名/驰名商标总数	0.621 9	去年用电量增长情况	3.303 1
专利总数	1.840 3	前年用电量增长情况	3.146 7
员工总人数	2.226 4	年度客户用电量同比增长率	3.148 8
分支机构数	0.180 4	近一年累计拖欠电费数量	5.413 6
最近一年资产回报率	3.995 9	近半年月均电费	1.991 0
净利润率	1.303 9	近一年月电费谷值	3.991 8
企业产能利用率	3.887 8	近一年电费回收准时度	6.306 2
每月应收电费金额波动性	3.941 0		

表 3 相关性系数较高的指标

Tab.3 Indices with high correlation coefficient

序号	指标 1	指标 2	相关系数
1	注册资金	实收资本	0.946 2
2	年度用户用电量同比增长率	前年用电量增长情况	0.999 6
3	近一年累计拖欠电费数量	近一年电费回收准时度	0.941 9

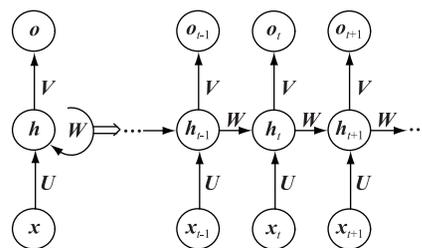


图 1 RNN 的基本结构

Fig.1 The basic structure of RNN

出层之间全连接,而每层的节点相互无连接;序列当前的输出与前一时刻输出相关时,这种连接方式无法体现和处理这种关系。而 RNN 的隐含层之间也有连接,隐含层的输入不仅包括输入层提供的输出,还包括上一时刻隐含层的输出。通过这种方式,RNN 可对前面时刻的信息进行记忆,并应用于当前时刻的计算。RNN 的简单结构如图 1 所示。

$x = [x_1, \dots, x_{t-1}, x_t, x_{t+1}, \dots, x_T]$ 为网络的输入,

T 为 RNN 网络建模的总时刻数, t 时刻的输入即为 x_t ; h_t 为代表 t 时刻的隐含层状态, o_t 为 t 时刻的输出; U, V, W 为输入层到隐含层、隐含层到隐含层、隐含层到输出层的权重系数矩阵。

RNN 的 Forward 过程,在 t 时刻,网络的输出可

以用下式表达:

$$o_t = g(\mathbf{V}h_t) \quad (1)$$

式中:函数 g 一般为 softmax 函数; h_t 为隐含层的状态,其计算公式为:

$$h_t = f(\mathbf{U}x_t + \mathbf{W}h_{t-1}) \quad (2)$$

其中, f 一般为 tanh、relu、logistic 等函数。

在 RNN 的 Backward 过程中,假设输出误差为 $E = \sum_t e_t$,则 t 时刻网络的输出误差可以通过如下的偏微分方程求得:

$$\nabla \mathbf{U} = \frac{\partial E}{\partial \mathbf{U}} = \sum_t \frac{\partial e_t}{\partial \mathbf{U}} \quad (3)$$

$$\nabla \mathbf{V} = \frac{\partial E}{\partial \mathbf{V}} = \sum_t \frac{\partial e_t}{\partial \mathbf{V}} \quad (4)$$

$$\nabla \mathbf{W} = \frac{\partial E}{\partial \mathbf{W}} = \sum_t \frac{\partial e_t}{\partial \mathbf{W}} \quad (5)$$

3.2 LSTM 的基本结构

RNN 很好地利用了历史信息来帮助网络进行当前决策,但当回归/分类信息所需的历史信息间隔较远时,RNN 就会丧失学习能力,这种问题称为长期依赖问题。对长期依赖问题而言,RNN 的 Backward 过程中梯度会呈指数倍衰减,这种衰减导致 RNN 无法处理长期依赖问题。为了克服 RNN 的这种缺陷,国内外专家学者们研究和改进了众多方法,其中 LSTM 表现最为出色。当误差从输出层反向传播回来时,LSTM 可以用记忆元存储下来,以记住较长时间内的信息。

所有的 RNN 都具有重复神经网络模块的链式形式,在标准的 RNN 中,这个重复模块结构简单,一般只有一个 tanh 层;LSTM 同样具有这样的重复结构,但其神经网络层有 4 个,相互直接通过特殊方式进行交互,标准 RNN 模块和 LSTM 模块的链式结构如图 2 所示。

LSTM 一个重要的概念是细胞状态,它可以通过精心设计的“门”结构来更新细胞状态。LSTM 神经网络模块的 4 层结构中:

第一层为遗忘门层,它决定从细胞状态中丢弃什么信息。第二层为输入门层,一般为 sigmoid 函数,决定需要更新的信息。第三层为 tanh 层,通过创建一个新的候选值向量更新细胞状态。第二层和第三层共同作用,更新神经网络模块的细胞状态。第四层为其他相关信息更新层,用于更新由其他因素导致的细胞状态变化。

通过 4 层结构模型,LSTM 很好地解决了标准 RNN 对长期依赖问题处理效果欠佳的问题,目前在语言模型、序列标记中取得一定的成功。

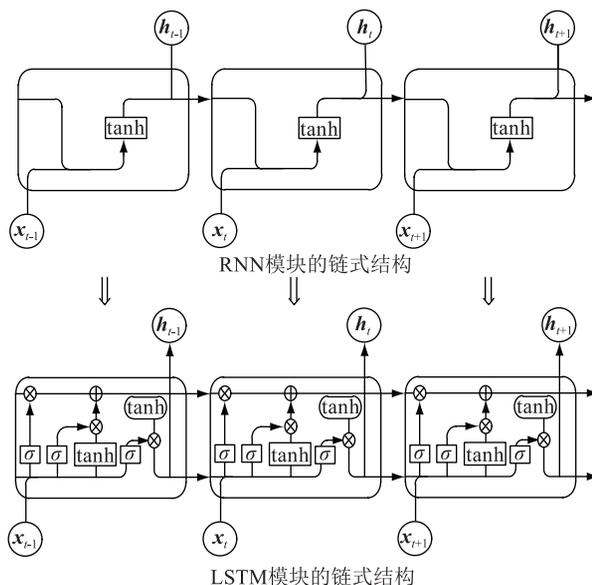


图 2 LSTM 网络模块的链式结构

Fig.2 The chain structure of LSTM network module

3.3 LSTM 的工作原理

LSTM 神经网络模块的四层结构,使得其工作原理比标准 RNN 网络更加复杂,本文以图 4 的 LSTM 模块为参照,详细介绍了 LSTM 的工作原理。假设 LSTM 神经网络模块的各信号定义如图 3 所示。

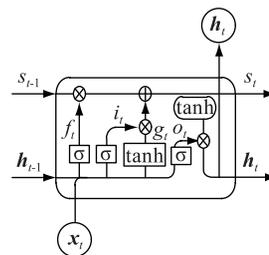


图 3 LSTM 神经网络模块的各信号定义

Fig.3 Definition of signals in LSTM neural network module

LSTM 神经网络模块中的信号传播规则如下式:

$$\begin{cases} f_t = \sigma(\mathbf{W}_{fx}x_t + \mathbf{W}_{fh}h_{t-1}) + b_f \\ i_t = \sigma(\mathbf{W}_{ix}x_t + \mathbf{W}_{ih}h_{t-1}) + b_i \\ g_t = \varphi(\mathbf{W}_{gx}x_t + \mathbf{W}_{gh}h_{t-1}) + b_g \\ o_t = \sigma(\mathbf{W}_{ox}x_t + \mathbf{W}_{oh}h_{t-1}) + b_o \\ s_t = g_t \cdot i_t + s_{t-1} \cdot f_t \\ h_t = o_t \cdot \varphi(s_t) \end{cases} \quad (6)$$

其中: $\varphi(x) = \tanh(x)$, $\sigma(x) = 1/(1 + e^{-x})$,带下标的各 \mathbf{W} 参数代表两个信号之间的权重系数,带下标的各 b 参数代表各信号的偏置值。

令 $x_{ct} = [x_t, h_{t-1}]$,则式(6)可以改写为:

$$\begin{cases} f_t = \sigma(\mathbf{W}_f x_{ct}) + b_f \\ i_t = \sigma(\mathbf{W}_i x_{ct}) + b_i \\ g_t = \varphi(\mathbf{W}_g x_{ct}) + b_g \\ o_t = \sigma(\mathbf{W}_o x_{ct}) + b_o \\ s_t = g_t \cdot i_t + s_{t-1} \cdot f_t \\ \mathbf{h}_t = o_t \cdot \varphi(s_t) \end{cases} \quad (7)$$

式中: f_t 为遗忘门, 决定从细胞状态的上一个状态中丢弃的信息; i_t 和 g_t 构成输入门, 决定新的细胞状态中保存的信息; o_t 为输出层, 决定要输出的信息。

3.4 LSTM 的参数设置

LSTM 虽然学习能力强大, 但也要基于模型要求和人工经验, 设置一些超参数, 使算法的寻优速度更快、分类准确度更高。

(1) Hidden_Size。隐藏层神经元个数, 其个数越多, LSTM 网络越强大, 但计算参数和计算量会因此急剧增加; 而且, 要注意隐藏层神经元个数不能超过训练样本条数, 否则容易出现过拟合。

(2) Keep_Prob。神经元连接不被切断的概率, 在神经元输入端有数据输入时以 Keep_Prob 的概率正常工作, 以 $1 - \text{Keep_Prob}$ 的概率输出为 0, 这种局部连接方式可以降低数据过拟合的概率。

(3) Num_Layers。LSTM 网络的层数, 层数越多, 网络越大, 学习能力越强大, 同时计算量也会大幅增加, 对于电费回收风险这类输入数据类型相对简单的问题, 设置为 2 层一般能够满足需求。

(4) Learning_Rate_Base。学习率初值, 会影响各神经元连接的权值更新速度, 学习率大, 权值更新就快, 到训练后期损失函数可能在最优值附近振荡, 学习率小, 权值更新就慢, 过小的权值可能导致优化损失函数下降速度过慢。因此目前常用的方法是在训练初期设置较高的学习率, 以确保损失函数以较快的速度向最优值靠拢, 随着训练次数的增加, 不断降低学习率, 使损失函数在局部范围内不断逼近最优值。

(5) Learning_Rate_Decay。学习率衰减速度, 与学习率初值搭配, 使 LSTM 网络的学习能力更加强大、智能, 具体操作方法见(4)。

(6) Regularization_Rate。正则化率, 为了防止 LSTM 网络过拟合, 一般的做法是在损失函数后加一项正则化惩罚项, 而正则化率正是该惩罚项的系数, 其值越大, 对 LSTM 网络过拟合的惩罚越大。

(7) Train_Times。训练次数, 随着训练次数的不断增加, LSTM 网络的准确性越高, 但当训练次数达到一定值后, LSTM 网络的准确性将不再提高或提升很小, 而计算量却不断增加。因此在具体操作

时, 应结合研究问题的需要, 选择合适的训练次数。

3 算例分析

为了验证本文所选指标的合理性和算法的先进性, 以某区域电网 3039 户电力大客户为例, 进行电费回收风险预警算例分析。选取其中 2839 户为训练集(其中 2809 户非欠费客户, 30 户欠费客户), 200 户为测试集(其中 191 户非欠费客户, 9 户欠费客户)。LSTM 参数设置见表 4。

表 4 LSTM 参数设置

Tab.4 LSTM parameter setting

参数	说明	取值
Hidden_Size	隐藏层神经元个数	500
Keep_Prob	神经元连接不被切断的概率	0.3
Num_Layers	LSTM 的网络层数	2
Learning_Rate_Base	学习率初值	0.8
Learning_Rate_Decay	学习率衰减速度	0.99
Regularization_Rate	正则化率	0.001
Train_Times	训练次数	1000

采用训练集对 LSTM 网络算法进行训练, 并将训练好的模型用于对 200 户测试样本进行电费回收风险预警, 并与 Logistic 回归、C4.5 决策树、支持向量机(SVM)的预警结果比较, 如表 5 所示(1 表示客户发生欠费, 0 表示客户未发生欠费)。

为了更好地对比算法的性能, 本文在准确率指标的基础上, 引入了预警命中率和覆盖率指标, 其计算规则如下:

$$\begin{cases} H_r = N_{\text{right}} / N_{\text{warn}} \\ C_r = N_{\text{right}} / N_{\text{real}} \end{cases} \quad (8)$$

式中: N_{warn} 为算法计算得到的风险预警客户数; N_{real} 为实际发生电费回收风险的客户数; N_{right} 为算法命中的实际预警客户数; H_r 为算法的命中率, 即算法预警的客户数中实际发生欠费客户数的比例; C_r 为算法的覆盖率, 即实际发生欠费的客户数中被算法预警结果覆盖的比例。

各算法准确率、命中率和覆盖率如表 6 所示。

(1) 准确率而言, 各算法都达到了 95% 以上, 其中 C4.5 决策树准确率最高。

(2) 命中率而言, C4.5 决策树命中率最高, 其计算得到的风险预警客户 8 户, 其中 5 户实际发生了欠费; 而 LSTM 计算得到的风险预警客户 18 户, 其中 9 户实际发生了欠费, 查准率 50.00%。

(3) 覆盖率而言, 实际发生欠费客户 9 户, LSTM 覆盖了全部欠费客户, 覆盖率 100%; Logistic 回归覆盖了其中的 8 户, 覆盖率 88.89%; 在准确率、

表5 风险预警结果比较

Tab.5 Comparison of early warning results of tariff recovery risk

企业序号	真实值	Logistic	C4.5	SVM	LSTM
7	0	1	0	1	0
14	0	1	0	0	1
31	0	1	1	0	1
66	0	1	0	0	1
81	0	0	0	1	0
88	0	1	0	0	1
90	0	1	0	0	1
106	0	0	0	0	1
107	0	0	0	1	0
116	0	1	1	0	1
134	0	1	1	0	1
163	0	1	0	0	1
192	1	1	0	0	1
193	1	1	1	0	1
194	1	1	0	0	1
195	1	1	1	1	1
196	1	1	0	0	1
197	1	0	0	0	1
198	1	1	1	0	1
199	1	1	1	0	1
200	1	1	1	1	1

命中率方面表现最好的 C4.5 决策树算法,仅覆盖到 5 户,覆盖率 55.56%。

表6 风险预警算法效果对比

Tab.6 Comparison of the effect of risk early warning algorithms %

对比参数	Logistic	C4.5	SVM	LSTM
准确率	95.00	96.50	95.00	95.50
命中率	47.06	62.50	40.00	50.00
覆盖率	88.89	55.56	22.22	100.00

从电费回收工作的实际需求看,首要的要求是能够识别所有的欠费客户,也即覆盖率,以便于杜绝电费回收风险转化为实际欠费;其次,命中率的提高,能够将电费催收对象锁定到了更小的客户范围,减少电费催收的工作量。从这两个要求看,LSTM 命中了所有实际欠费客户,而且将催费对象锁定到 18 户,也即在降低工作量的同时,规避了所有的电费回收风险。

4 结语

电费催收工作一直以来都是电力公司的一项重要工作,消耗了大量的人力物力,但由于电费回收风险的潜伏性、突发性,导致电费催收工作中仍

然存在漏网之鱼。因此,本文结合业务人员的关注重点,建立了电费回收风险指标体系,采用熵值法和相关性分析筛选了电费回收风险的关键指标,并采用 LSTM 进行电费回收风险预警,取得了显著的成效,预警准确率、查准率和查全率全面高于传统方法。研究成果可以精准定位高风险客户群体,提高电费催收的针对性。

参考文献:

- [1] 邹晓华. 电力企业绩效评价的现状分析及建议[J]. 电力技术经济, 2001(6): 55-57.
ZOU Xiaohua. Analysis on current status of performance measurement prevailing in power utilities and suggestions [J]. Electric Power Technologic Economics, 2001 (6): 55-57.
- [2] 来恩, 冯蕾. 论防范电费回收风险[J]. 宁夏电力, 2010(5): 54-57.
LAI En, FENG Lei. Discussion on the risk prevention of the electricity charges collection [J]. Ningxia Electric Power, 2010 (5): 54-57.
- [3] 王绵斌, 谭忠富, 张蓉, 等. 供电企业规避电费回收风险的客户信用评价方法[J]. 华东电力, 2007, 35(1): 21-25.
WANG Mianbin, TAN Zhongfu, ZHANG Rong, et al. Consumer credit assessment methods for power supply enterprises to avoid electric toll collection risks [J]. East China Electric Power, 2007, 35(1): 21-25.
- [4] 董玲娟, 赵罡, 郑晨泉, 等. 构建客户信用等级评价和决策的电费风险管理系统[J]. 电力需求侧管理, 2009, 11(6): 61-63.
DONG Lingjuan, ZHAO Gang, ZHENG Chenquan, et al. Construction of electricity tariff risk management system of customer credit evaluation and decision-making [J]. Power Demand Side Management, 2009, 11(6): 61-63.
- [5] 涂莹, 林士勇, 欧阳柳, 等. 基于市场细分的逻辑回归模型在电费回收风险预测中的应用研究[J]. 电力需求侧管理, 2016, 18(4): 46-49.
TU Ying, LIN Shiyong, OUYANG Liu, et al. Research on the risk prediction of electricity fee recovery using logistic regression based on the market segmentation theory [J]. Power Demand Side Management, 2016, 18(4): 46-49.
- [6] 陈羽中, 郭松荣, 陈宏, 等. 基于并行分类算法的电力客户欠费预警[J]. 计算机应用, 2016, 36(6): 1757-1761.
CHEN Yuzhong, GUO Songrong, CHEN Hong, et al. Electricity customers arrears alert based on parallel classification algorithm [J]. Journal of Computer Applications, 2016, 36 (6): 1757-1761.
- [7] 谢骏凯, 徐千, 丁炳森, 等. “客户画像”在电费回收风险管控中的运用[J]. 电力需求侧管理, 2016, 18(S1): 74-76.
XIE Junkai, XU Qian, DING Bingmiao, et al. Application of customer models in the risk control of electricity fees' recovery [J]. Power Demand Side Management, 2016, 18(S1): 74-76.

- [8] 刘东东, 胡少柔, 陈荣腾, 等. 基于大数据技术的电费回收研究[J]. 江苏科技信息, 2016(27): 55-56, 80.
LIU Dongdong, HU Shaorou, CHEN Rongteng, et al. Research on collection of electricity charges based on big data technology [J]. Jiangsu Science & Technology Information, 2016(27): 55-56, 80.
- [9] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553): 436-444.
- [10] 伊恩·古德费洛, 约书亚·本吉奥, 亚伦·库维尔. 深度学习[M]. 北京: 人民邮电出版社, 2017.
IAN G, YOSHUA B, AARON C. Deep learning [M]. Beijing: Posts & Telecom Press, 2017.
- [11] 韩霄汉, 胡小寒, 姚力, 等. 竞争性电力市场背景下电能计量发展趋势的思考[J]. 浙江电力, 2017, 36(3): 30-33.
HAN Xiaohan, HU Xiaohan, YAO Li, et al. Reflection on development trend of electric energy metering in competitive electricity market [J]. Zhejiang Electric Power, 2017, 36(3): 30-33.
- [12] 史俊霞. 基于一体化技术的电力调度系统网络优化[J]. 浙江电力, 2017, 36(5): 66-68.
SHI Junxia. Electric power dispatching system optimization based on integration technology [J]. Zhejiang Electric Power, 2017, 36(5): 66-68.
- [13] 刘洪鑫, 夏通, 徐非非, 等. 针对“峰谷倒置”负荷特性的电价策略研究[J]. 电力系统自动化, 2017, 36(12): 6-10.
LIU Hongxin, XIA Tong, XU Feifei, et al. Study on the price strategy based on ‘peak-valley inversion’ [J]. Electric Power Automation Equipment, 2017, 36(12): 6-10.
- [14] 史静, 周琪, 谈健, 等. 江苏电网夏季空调负荷特性挖掘与温度敏感性辨识[J]. 电力工程技术, 2018, 37(3): 28-32.
SHI Jing, ZHOU Qi, TAN Jian, et al. The load excavation and temperature sensitivity identification of air conditioning in summer of Jiangsu power grid [J]. Electric Power Engineering Technology, 2018, 37(3): 28-32.
- [15] 曹淦昆, 何健伟, 鲍自安. 深度学习在电力领域的研究现状与展望[J]. 上海电力学院学报, 2017(4): 341-345, 361.
CAO Yukun, HE Jianwei, BAO Zian. Status quo and prospect of deep learning in electric power field [J]. Journal of Shanghai University of Electric Power, 2017(4): 341-345, 361.
- [16] JIAO Runhai, HUANG Xujian, MA Xuehai, et al. A model combining stacked auto Encoder and back propagation algorithm for short-term wind power forecasting [J]. IEEE Access, 2018(6): 17851-17858.
- [17] 刘琳, 余顺坤, 吴轶群. 基于熵值法的电力企业二元绩效积分制考核模型研究[J]. 华北电力大学学报, 2012, 39(2): 74-78.
LIU Lin, YU Shunkun, WU Yiqun. The entropy-based dual performance point system assessment model of team in power enterprise [J]. Journal of North China Electric Power University, 2012, 39(2): 74-78.

作者简介:



谢林枫

谢林枫(1972—),男,硕士,高级工程师,从事电力大数据技术、电力营销大数据分析工作(E-mail:java6@js.sgcc.com.cn);

钱立军(1977—),男,硕士,研究员级高级工程师,从事电力营销管理、电力营销大数据分析工作;

季聪(1988—),男,硕士,工程师,从事电力大数据技术、电力营销大数据分析工作。

Application of Long Short-term Memory Network Algorithm in Tariff Recovery Risk Early Warning for Large Power Customers

XIE Linfeng¹, QIAN Lijun², JI Cong¹, JIANG Ming², LYU Hui¹

(1. Jiangsu Frontier Electric Technology Co., Ltd., Nanjing 211102, Jiangsu Province, China;

2. State Grid Jiangsu Electric Power Company, Nanjing 210024, Jiangsu Province, China)

Abstract: For a long time, tariff recovery risk of large power customers is a hot spot for electric power company. But because of the lack of external data support and efficient early warning model, tariff recovery risk early warning has become a difficult problem. Based on electricity consumption and electricity charge data, combining with the relevant business, tax and court information of the enterprise, the paper established a serial of tariff recovery risk indexes for large power customers. Secondly, the entropy method (EM) is adopted to evaluate tariff recovery risk assessment of customers' electricity bills, and the customer risk level is divided according to the distribution of risk score. Weak influence indexes were filtered by weight coefficients and overlapping indexes were dropped by correlation analysis. Tariff recovery risk early warning model was carried out by Long Short-Term Memory (LSTM) network. Numerical example results show that the proposed risk early warning model was accurate and effective, and the result gained by LSTM is better than Logistic regression in accuracy, precision and recall. The tariff recovery risk early warning results can accurately locate high risk customers and improve tariff recovery efficiency.

Key words: large power customer; tariff recovery risk; risk early warning; deep learning; long short-term memory (LSTM)

(编辑 钱悦)