



电力潜在敏感客户预测的大数据方法应用

陈小峰, 赵雅迪, 张利鹏, 朱峰

(北京国网信通埃森哲信息技术有限公司, 北京 100032)

摘要: 随着 95598 业务的不断发展延伸, 人工话务强度增大。为了进一步加深对客户隐性特征以及诉求的认识和理解, 提升 95598 人工精细化客户服务水平, 对投诉倾向等客户服务中的典型应用场景进行了需求细化。基于电力服务工单数据, 选取建模关键指标, 通过熵权法、主成分分析和决策树等数据挖掘算法, 对潜在投诉倾向客户和计划停电敏感客户进行识别, 以便有针对性地进行服务资源调度, 充分做好应对措施, 有效减少投诉压力, 提升服务精度。

关键词: 投诉工单; 数据挖掘; 熵权法; 决策树

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2019271

Application of big data method in forecasting potential sensitive customers of electric power

CHEN Xiaofeng, ZHAO Yadi, ZHANG Lipeng, ZHU Feng

State Grid Information & Telecommunication Accenture Information Technology Co., Ltd., Beijing 100032, China

Abstract: With the continuous development and extension of 95598 business, the intensity of manual telephone traffic increases. In order to further deepen the consciousness and understanding of the hidden features and the demands of the customers, improve the customer service level of 95598, the typical application scenarios in the customer service such as the tendency of the complaint were refined. Based on the data of power service orders, the key index of modeling was selected. Through entropy weight method, principal component analysis and decision tree and other data mining algorithms, potential complaint propensity customers and planned blackout sensitive customers in order to carry out targeted service resource scheduling were identified, fully do a good job of response measures, effectively reduce complaint pressure and improve service accuracy.

Key words: complaint work order, data mining, entropy weight method, decision tree

1 引言

随着现代经济和社会的快速发展, 传统的电力行业机制已经不能满足市场需要。电力体制改

革的不断深化, 使得电网企业的公用性、基础性和服务性等社会特性愈加明显, 一方面承担着服务地方经济发展、提高人们生活品质的重要政治和社会责任; 另一方面, 消费者心理需求从低级



的生理、安全需求逐渐提升至寻求情感归属和尊重的需求^[1-2]。因此,供电企业需要寻求科学依据,拓展服务深度广度,增强面向客户的服务能力建设,不断深化机制建设、持续强化现场管理,实现企业与客户之间的高效沟通,提升客户的满意度^[3-4]。

为了改善服务质量,提高运维效率,企业开始从数据中寻找客户因停电、电费而发生投诉、咨询等行为的影响因素,提前做好预防措施和服务准备工作,提高工作质量和服务水平^[5-6]。随着信息技术的发展,国内开始使用数据挖掘技术分析海量数据并试图寻找引起敏感行为发生的影响因素^[7]。国家电网公司提出构建泛在电力物联网,更充分发挥公司数据价值,实现大数据智能分析、多源大数据治理与跨领域智能分析,主动培育大数据运营、大数据征信等新兴业务^[8]。严宇平等^[9]运用逻辑回归、决策树等数据挖掘技术,建立了客户停电敏感度预测模型,实验证明运用逻辑回归算法的效果比决策树算法好,同时分析了客户特征,为供电企业加强客户停电差异化管理提供了支持,但对指标的建设未过多描述。赵会茹等^[10]在对云南省 150 户农民调研的基础上,采用因子分析法和 Logistic 回归模型,分析了影响电力普遍服务受助个体诉求的主/客观因素,并重点分析了主观影响因素的作用机理。随着服务业务工单量的不断增多,对文本挖掘技术在电力领域中的应用的需求也在不断提升^[11]。吴刚勇等^[12]通过文本分类分析,构建文本分类器模型,实现对“热词频”在不同业务中的分布情况研究,并根据结果开展相应改进措施。邹云峰等^[13]建立 95598 工单文本自动分类的模型,通过验证实现 95598 工单文本快速精准的自动分类,为分析供电服务对客户用电诉求的影响提供依据和数据基础。

本文以客户服务工单中的投诉业务为切入点,运用统计方法和数据挖掘算法对指标进行了梳理并构建投诉倾向客户模型,基于工单内容运用文本挖掘方法得出了停电投诉占总投诉业务的

比率最大的结论,通过决策树算法筛选识别出计划停电敏感潜在客户,这将有助于电力企业准确辨识停电敏感度高的客户,提供差异化停电服务,降低客户停电抱怨的概率^[14-15]。

2 理论方法

2.1 TF-IDF 关键词挖掘算法

词频-逆文档频率方法 (term frequency-inverse document frequency, TF-IDF) 的主要思想是:如果某个词或短语在某分类中出现的频率 TF 高,并且在其他分类中很少出现,则认为此词或者短语具有很好的类别区分能力,适合用来分类,可以标记为关键词汇。设 d 是分类类别, f 是基于工单内容细分的特征词,词频标准化的计算方法如式 (1) 所示。

$$\text{词频}(\text{TF}_{f,d}) = \frac{\text{词汇在所属类别中的出现次数}}{\text{词汇在样本集合的总次数}} \quad (1)$$

其中,TF 表示某一个给定的词语在所属类别中出现的次数。文档频率用 DF 来表示,代表样本集中包含某个词汇的所有类别数目,DF 通常比较大,将其映射到一个较小的取值范围,用逆文档频率 (inverse document frequency, IDF) 表示如式 (2) 所示:

$$\begin{aligned} \text{逆文档频率}(\text{IDF}_f) = \\ \log \left(\frac{\text{数据集总的样本个数}}{\text{包含某个词汇的样本数} + 1} \right) = \\ \log \left(\frac{n}{\text{DF}_f + 1} \right) \end{aligned} \quad (2)$$

其中,取 $\log$ 是为了方便比较,其底数通常取 2 或 10。式 (2) 的意义在于,如果包含该词汇的样本数越少,IDF 越大,则说明该词汇具有很好的类别区分能力。词项的权重用 TF-IDF 来表示,计算如式 (3) 所示:

$$\text{TF-IDF} = \text{词频}(\text{TF}_{f,d}) \cdot \text{逆文档频率}(\text{IDF}_f) \quad (3)$$

使用 TF-IDF 可以计算某个词汇在某个分类中的重要性,因而识别出该词汇是否为关键词,

实现分类规则的构建,精准定位文字描述的基本内容,为后续分类分析工作提供精准数据。

2.2 信息量熵权法

熵权法是指用来判断某个指标离散程度的统计方法,离散程度越大,则该指标对综合评价的影响越大,可列为重要性指标。具体计算方法如下。

步骤 1 给定 n 个样本, m 个指标,则 r_{ij} 为第 i 个样本的第 j 个指标的数值 ($i=1,2,\dots,n$; $j=1,2,\dots,m$);

步骤 2 指标的归一化处理:对于正向指标, $r'_{ij} = \frac{r_{ij} - \min\{r_{1j}, \dots, r_{nj}\}}{\max\{r_{1j}, \dots, r_{nj}\} - \min\{r_{1j}, \dots, r_{nj}\}}$; 对于负向指标, $r'_{ij} = \frac{\max\{r_{1j}, \dots, r_{nj}\} - r_{ij}}{\max\{r_{1j}, \dots, r_{nj}\} - \min\{r_{1j}, \dots, r_{nj}\}}$, 其中, r'_{ij} 为第 i 个样本的第 j 个指标的数值,为了方便处理,归一化的数据仍记为 r_{ij} ,且 $r_{ij} = |r'_{ij}|$;

步骤 3 计算第 j 项指标下第 i 个样本占该指标的比重: $p_{ij} = r_{ij} / \left(\sum_{i=1}^n r_{ij} \right), i=1, \dots, n; j=1, \dots, m$;

步骤 4 计算第 j 项指标的熵值:

$$e_j = -k \sum_{i=1}^n p_{ij} \ln(p_{ij}), k = 1 / \ln(n), e_j \geq 0;$$

步骤 5 计算信息熵冗余度: $d_j = 1 - e_j$;

步骤 6 计算各项指标的权值: $\omega_j = \frac{d_j}{\sum_{j=1}^m d_j}$ 即

为指标重要性得分。

2.3 主成分降维法

主成分分析是利用降维的思想,在损失很少信息的前提下,把多个指标转化为几个综合指标的多元统计方法。其中,每个主成分都是原始变量的线性组合,这些主成分能够反映原始变量的绝大部分信息,且所含的信息互不重叠。

算法通过对原始变量相关矩阵或协方差矩阵内部结构关系的研究,利用原始变量的线性组合形成几个综合指标,在保留原始变量主要信息的

前提下起着降维与简化问题的作用,使得在研究复杂问题时更容易抓住主要矛盾。

2.4 决策树算法

假设给定一个训练集 $D = (x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$, $x_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(n)})^T$ 是输入事例, n 是特征个数, $y_i \in 1, 2, \dots, K$ 是类别个数, N 为样本容量, $i=1, 2, \dots, N$ 。根据给定训练集建立决策树模型,达到正确分类的目的,在文本工单识别与敏感客户筛选中均用到了该算法。具体方法如下:计算每个特征的信息增益值,值最大的特征作为第一个根节点,在子节点上递归调用上述方法来构造决策树,直到所有特征的信息增益值都非常小或没有特征可供选择为止。同时,为了防止过拟合的发生,必要时对生成的决策树进行剪枝。

特征选择的准则是信息增益法,即经验熵 $H(D)$ 表示对数据集 D 进行分类的不确定性,经验条件熵 $H(D|A)$ 表示在特征 A 给定的条件下对数据集 D 进行分类的不确定性, $|D|$ 表示样本容量。假设有 K 个类 C_k , $k=1, 2, \dots, K$, $|C_k|$ 为属于类 C_k 的样本个数, $\sum_{k=1}^K |C_k| = |D|$ 。设特征 A 有 n 个不同的取值 $\{a_1, a_2, \dots, a_n\}$, 根据特征 A 的取值将 D 划分为 n 个子集 $D_1, D_2, \dots, D_n$ 。其中, $|D_i|$ 为 D_i 的样本个数, $\sum_{i=1}^n |D_i| = |D|$ 。记子集 D_i 中属于类 C_k 的样本集合为 D_{ik} , 信息增益值的计算如下。

步骤 1 计算 $H(D) = \sum_{k=1}^K \frac{|C_k|}{|D|} \lg \frac{|C_k|}{|D|}$;

步骤 2 计算 $H(D|A) = \sum_{i=1}^n \frac{|D_i|}{|D|} H(D_i) = - \sum_{i=1}^n \frac{|D_i|}{|D|} \lg \frac{|D_i|}{|D|}$;

$$\sum_{k=1}^K \frac{D_{ik}}{D_i} \lg \frac{D_{ik}}{D_i};$$

步骤 3 计算信息增益值 $g(D, A) = H(D) - H(D|A)$ 。

决策树的剪枝通过极小化决策树整体的损失函数或代价函数来实现。假设树 T 的叶节点个数



为 $|T|$ ， t 是树 T 的叶节点，该叶节点有 N_t 个样本点，其中， k 类的样本点有 N_{tk} ($k=1,2,\dots,K$)个。 $H_t(T)$ 为叶节点 t 上的经验熵， $\alpha \geq 0$ 为参数，则决策树学习的损失函数可以定义为： $C_\alpha(T) = \sum_{t=1}^{|T|} N_t H_t(T) + \alpha |T|$ ，其中，经验熵为 $H_t(T) = -\sum_k \frac{N_{tk}}{N_t} \lg \frac{N_{tk}}{N_t}$ 。输入生成的决策树 T 和参数 α ，输出修剪后的子树 T_α ，剪枝过程如下。

步骤 1 计算每个节点的经验熵。

步骤 2 递归从树的叶节点向上回溯，设一组叶节点回溯到其他父节点之前与之后的整体树分别为 T_B 和 T_A ，其对应的损失函数值分别是 $C_\alpha(T_B)$ 与 $C_\alpha(T_A)$ ，如果 $C_\alpha(T_A) \leq C_\alpha(T_B)$ ，则进行剪枝，将父节点变为新的叶节点。

步骤 3 返回步骤 2，直至不能继续为止，得到损失函数最小的子树 T_α 。

3 场景应用

随着 95598 业务的不断发展延伸，人工话务强度增大，精准服务能力低下等情况日益突出。为了

进一步加深对客户隐性特征以及诉求的认识和理解，提升 95598 人工精细化客户服务水平，实现提质增效，对投诉倾向等客户服务中的典型应用场景进行了需求细化。基于大数据方法进行投诉倾向客户识别，通过对中心业务支持系统、呼叫平台以及营销档案信息进行收集、数据质量验证、数据探索等工作，分析判断数据是否能够支撑业务方面的需求。

基于客户历史 95598 呼叫行为及客户属性，设计因子 143 项，最终纳入模型因子 93 项，指标体系如图 1 所示。

投诉倾向判别模型采用熵权法与主成分分析法相结合的综合权重算法，基于客户历史拨打次数、拨打时长、拨打间隔等呼叫行为，结合客户用电属性，构建客户评分，判断客户投诉倾向高低。选取某电力公司近两年工单数据作为训练集、近 3 个月数据作为验证集进行模型构建与验证。模型构建思路如图 2 所示。

根据各指标相关系数及相关性显著检验 P 值判断变量相关性见表 1，剔除高度线性相关变量，减少变量冗余，剔除相关指标 10 个。

计算指标熵权，筛选剔除无效指标 15 个，确定

呼叫行为			客户属性	
拨打偏好时段	拨打间隔	拨打时长	用电类型	负荷类型
22:00~次日7:00拨打次数 7:00~次日12:00拨打次数 12:00~次日17:00拨打次数 17:00~次日22:00拨打次数 周末拨打次数 元旦假日拨打次数 春节假日拨打次数 清明节假日拨打次数 劳动节假日拨打次数 端午节假日拨打次数 中秋节假日拨打次数 国庆节假日拨打次数 2~3月拨打次数 4~6月拨打次数 7~8月拨打次数 9月拨打次数 10月~次年1月拨打次数	重复拨打平均间隔 单事件首次催办平均时间间隔 单事件首次投诉平均时间间隔 单事件首次催办最短时间间隔 单事件首次投诉最短时间间隔 单事件首次催办最长时间间隔 单事件首次投诉最长时间间隔	95598平均拨打时长 投诉平均拨打时长 催办平均拨打时长 单次投诉事件平均拨打时长 单次催办事件平均拨打时长 重复拨打事件平均拨打时长	大工业用电 居民生活用电 农业生产用电 一般工商业用电	一类负荷 二类负荷 三类负荷
	拨打频数	各供电单位诉求偏好	重要等级	客户地域
	95598拨打总次数 重复拨打率 故障报修拨打次数 咨询拨打次数 投诉拨打次数 建议拨打次数 意见拨打次数 举报拨打次数 表扬拨打次数 服务申请拨打次数 催办拨打次数 单事件最高重复拨打次数 单重复拨打事件平均拨打次数 重复拨打事件数	偏好间隔 偏好时段 偏好频率 偏好时长 偏好拨打原因	一级重要用户 二级重要用户 特级重要用户 临时重要用户	供电单位编码 客户所属台区 客户所属线路
IVR转人工		受理内容关键词	行业分类	
IVR转人工次数 IVR成功转人工业务类别		关键字率 关键词频	城镇居民 乡村居民 农林牧渔业 工业 交通运输、仓储和邮政业 商业、住宿和餐饮业 金融、房地产、商务及居民服务业	
		拨打原因		
		催办各业务主体平均拨打次数 投诉各业务主体平均拨打次数 催办各业务主体平均拨打时长 投诉各业务主体平均拨打时长	客户情绪	
			情绪激动客户	

图 1 指标体系

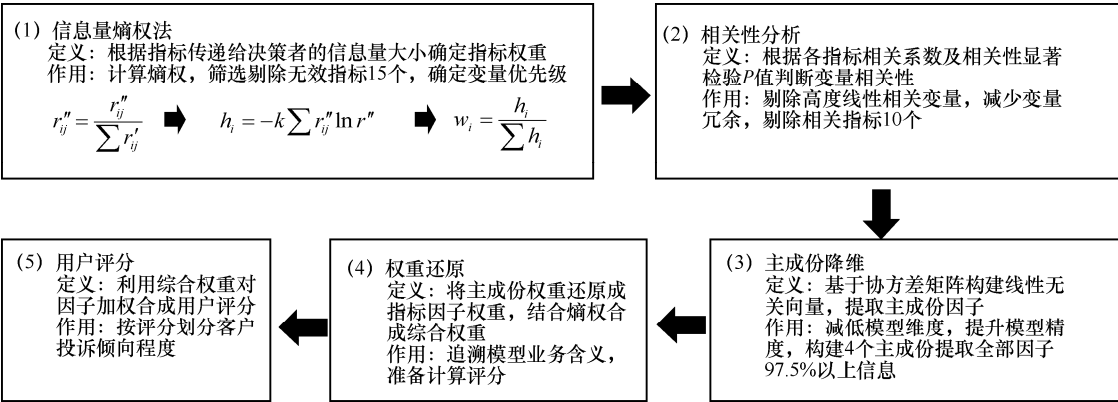


图2 模型构建思路

表1 指标相关性

指标 1	指标 2	相关系数
平均咨询电话时长	最小通话时长	0.98
来电业务类型为投诉下投诉工单的拨打次数	呼入时非投诉最终下投诉工单拨打次数	0.97
工作日拨打次数	拨打总次数	0.97
平均拨打间隔	最小间隔时长	0.95
拨打密集程度	拨打总次数	0.94
平均拨打间隔	最大间隔时长	0.94
呼入时非投诉最终下投诉工单拨打次数	拨打密集程度	0.93
呼入时非投诉最终下投诉工单拨打次数	总拨打次数	0.91
半年内催办拨打次数	半年内拨打次数	0.91
工作日拨打次数	拨打密集程度	0.91
来电业务类型为投诉下投诉工单的拨打次数	拨打密集程度	0.90
平均密集拨打时长	咨询平均通话时长	0.90

变量优先级。基于协方差矩阵构建线性无关向量，提取主成分因子，构建 4 个主成分提取全部因子 97.5%以上信息。从表 2 模型因子权重来看，权重排名前 10 的因子中，通话时长维度的指标占绝大多数，且投诉各子业务类型中通话时长对投诉倾向特征的贡献率较大，投诉倾向客户的服务升级（非投诉转投诉、非投诉转催办）需求强于无投诉倾向客户。

通过自然语言处理技术，对电力客户投诉工单进行深入文本挖掘，利用结巴分词技术分析投诉工单中的受理内容，对分析结果开展特征选取与降维处理后，实现对无关词汇的过滤。利用 TF-IDF 关键词挖掘算法进行词频统计，针对不同

需求类型的客户提供差异化服务策略，从而提高客户满意度和忠诚度。结合实际的电力业务，通过 TF-IDF 算法计算关键词重要性权重值，词语“营业厅、停电、用电、故障”词语出现频率最多，表明客户主要对营业厅、用电、停电等意见较大，因此，可基于投诉工单继续用主成分和决策树方法对停电敏感的客户进行识别。

选取近一年与计划停电相关工单数量 2 373 506 条、能匹配到计划停电信息的工单总量 1 353 101 条、1 118 092 个电话号码，通过决策树算法对模型进行训练，识别计划停电敏感客户。根据业务经验得出影响因子共 126 个，考虑这些影响因子之间



表 2 模型权重排名前 10 的因子

宽表名称	指标权重	指标名称
TIME_MIN	0.481 415 853	最小通话时长
AVG_TIME_003	0.474 726 069	平均咨询通话时长
TIME_UPGRADE_MAX	0.008 128 55	最长服务升级时间
AVG_TIME_005	0.007 388 045	平均投诉通话时长
AVG_TIME_015	0.006 376 885	平均催办通话时长
TIME_UPGRADE_MIN	0.005 700 51	最小服务升级时长
AVG_TIME_POTENTIAL4	0.004 704 907	关联来电为投诉或催办的平均通话时长
AVG_TIME_POTENTIAL3	0.003 650 723	来电业务类型为投诉下投诉工单的平均通话时长
AVG_TIME_COMPLAIN_BUSI	0.002 177 505	营业投诉平均通话时长
AVG_TIME_COMPLAIN_SERV	0.002 137 768	服务投诉平均通话时长

可能存在多重共线性，因此，利用相关分析，筛选相关系数大于 0.9 影响因子，同时利用主成分分析法进行二次筛选，得到最终的影响因子 42 个，保证了指标的质量。运用决策树方法对计划停电敏感客户进行筛选识别，同时给出权重排名前 9 的因子见表 3。

4 实验分析

本文在电力客户服务工单的基础上，运用熵权

法、主成分分析和决策树等算法，综合打分筛选出有潜在投诉倾向的客户，提取近 3 个月工单数据及通话记录数据进行模型验证，结果见表 4。设客户来电率=（来电客户数/客户总数）×100%，客户投诉率=（投诉客户数/来电客户数）×100%，投诉/来电比=（投诉客户数/来电工单数）×100%。

从模型对客户的划分效果来看，有倾向客户占总体客户比例为 5.67%，覆盖了验证期内 73.27%的投诉量、86.39%的客户量；有倾向客户来电比率基

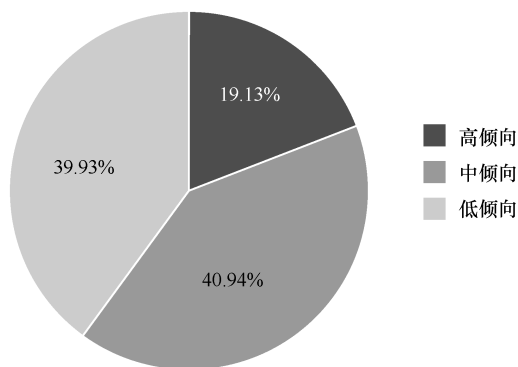
表 3 权重排名前 9 的因子

影响因子	权重
近一个月计划停电事件诉求升级（催办/投诉）拨打事件数	0.056 38
近一个月频繁停电的拨打次数	0.052 10
近一个月停电时间长的拨打次数	0.051 94
近一个月对未按计划停/送电的拨打次数	0.051 51
近一个月对发布渠道有异议的拨打次数	0.049 10
近一个月对停电安排有异议的拨打次数	0.048 02
近 3 个月计划停电事件诉求升级（催办/投诉）拨打事件数	0.046 71
近 3 个月频繁停电的拨打次数	0.046 32
近 3 个月停电时间长的拨打次数	0.042 41

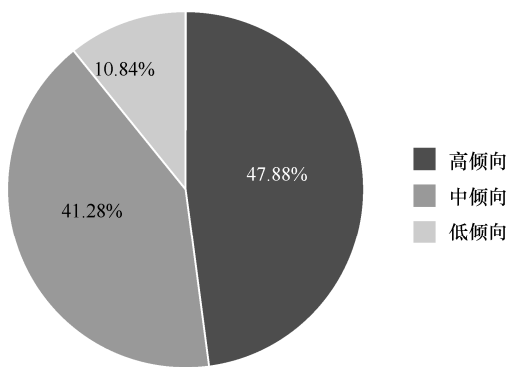
表 4 模型验证结果

对比项	客户总量	来电客户	来电工单	投诉量	投诉客户	客户来电率	客户投诉率	投诉/来电比	覆盖率	正确率
有倾向	2 927 211	369 613	841 482	19 339	19 195	12.63%	5.19%	2.3%	73.27%	92.81%
总体	51 597 624	8 102 828	13 049 487	26 394	22 218	15.70%	0.27%	0.20%	100.00%	—

本持平总体客户,而投诉/来电比、客户投诉率分别为总体客户的11.5倍、19.22倍;有倾向的客户中,高、中、低倾向客户分别占比为19.13%、40.94%、39.93%。其中,实际发生投诉的来电占投诉总量的73.27%,准确度达到92.81%,证明了模型是有效的,各倾向等级客户占比如图3所示。



(a) 客户数量



(b) 投诉客户

图3 各倾向等级客户占比

基于投诉倾向分类结果,运用文本挖掘技术得到停电投诉在整体投诉业务中的占比最高,因此,运用决策树算法筛选计划停电敏感用户,选取近3个月与计划停电相关的工单总量335 533个,涉及289 281个电话号码,能匹配计划停电信息工单总量299 924个,共有187 262个电话号码作为验证集,模型评估如图4所示,模型准确度为81.08%。可将此方法运用在故障停电敏感客户预测等其他场景应用中,降低人工成本,提升客户满意度。

$$\text{准确度: accuracy} = \frac{TP+TN}{TP+TN+FN+FP} = \frac{5\,488+146\,344}{187\,262} = 81.08\%$$

$$\text{灵敏度: sensitivity} = \frac{TP}{TP+FP} = \frac{5\,488}{8\,954} = 61.29\%$$

$$\text{提升度: lift} = \frac{\frac{TP}{TP+FN}}{\frac{TP+FP}{TP+TN+FN+FP}} = \frac{\frac{5\,488}{37\,452}}{\frac{8\,954}{187\,262}} = 3.06$$



对比项		实际情况		合计
		敏感	非敏感	
预测情况	敏感	TP=5 488	FN=31 964	37 452
	非敏感	FP=3 466	TN=146 344	149 810
合计		8 954	178 308	187 262

图4 模型评估

5 结束语

本文提供的电力数据挖掘方法可以将识别的敏感客户集成标签至95598话务人员业务受理界面,为业务人员提供手工标签标记功能,用以标识来电客户。但缺乏模型的比较与选取,后续可增加类似研究以增加模型运用的理论可靠性。在公司泛在电力物联网建设的大背景下,将方法运用至其他电力服务场景,精准定位客户需求,为构建企业中台打下理论方法基础,可不断进行模型的迭代优化,提升服务质量。

参考文献:

- [1] 张哲. 电力市场营销与优质服务的作用[J]. 企业技术开发, 2016(5): 115-116.
ZHENG Z. The role of electric marketing and high quality[J]. Technological Development of Enterprise, 2016(5): 115-116.
- [2] 刘容宇. 探讨供电企业电力营销的现状及其优质服务策略[J]. 通讯世界, 2014(17): 82-83.
LIU R Y. Discussion on the present situation of power marketing in power supply enterprises and its high quality service strategy[J]. Communication World, 2014(17): 82-83.
- [3] 庄雪霞. 现代供电企业电力营销优质服务策略提升分析[J]. 中国高新技术企业, 2013(14): 141-142.
ZHUANG X X. Analysis on the promotion of high quality service strategy of electric power marketing in modern power sup-



- ply enterprises[J]. China High-Tech Enterprises, 2013(14): 141-142.
- [4] 陈丽敏. 剖析电力企业营销服务管理的总体策略[J]. 大科技, 2012(23): 185-186.
- CHEN L N. Analysis of the overall strategy of marketing service management in electric power enterprises[J]. Super Science, 2012(23): 185-186.
- [5] 邹保, 周亮. 电网企业数据服务研究[J]. 电力信息与通信技术, 2010(9): 30-32.
- ZOU B, ZHOU L. Research on data service of power grid enterprises[J]. Electric Power Information and Communication Technology, 2010(9): 30-32.
- [6] 鞠小丽. 基于数据仓库技术的电力营销分析系统[J]. 数字技术与应用, 2012(3): 68-69.
- JU X L. Power marketing analysis system based on data warehouse technology[J]. Digital Technology and Application, 2012(3): 68-69.
- [7] 杨华飞, 李栋华, 程明. 电力大数据关键技术及建设思路的分析和研究[J]. 电力信息与通信技术, 2015, 13(1): 7-10.
- YANG H F, LI D H, CHENG M. Analysis and research on key technologies and construction ideas of electric power big data[J]. Electric Power Information and Communication Technology, 2015, 13(1): 7-10.
- [8] 赵会敏, 雒江涛, 杨军超, 等. 集成 BP 神经网络预测模型的研究与应用[J]. 电信科学, 2016, 32(2): 60-67.
- ZHAO H M, LUO J T, YANG J C, et al. Research and application of prediction model based on ensemble BP neural network[J]. Telecommunications Science, 2016, 32(2): 60-67.
- [9] 严宇平, 吴广财. 基于数据挖掘技术的客户停电敏感度研究与应用[J]. 新技术新工艺, 2015(9): 89-93.
- YAN Y P, WU G C. Research and application of customer blackout sensitivity based on data mining technology[J]. New Technology & New Process, 2015(9): 89-93.
- [10] 赵会茹, 霍慧娟, 李春杰. 基于因子分析和 Logistic 回归模型的电力普遍服务求助个体主观诉求影响因素分析[J]. 陕西电力, 2014, 42(8): 65-71.
- ZHAO H R, HUO H J, LI C J. Analysis on the influencing factors of subjective and objective demands of power universal service recipients based on factor analysis and logistic regression model[J]. Shanxi Electric Power, 2014, 42(8): 65-71.
- [11] STEVEN B, EWAN K, EDWARD L. Python 自然语言处理[M]. 北京: 人民邮电出版社, 2014.
- STEVEN B, EWAN K, EDWARD L. Python natural language processing[M]. Beijing: Posts & Telecom Press, 2014.
- [12] 吴刚勇, 张千斌, 吴恒超, 等. 基于自然语言处理技术的电力客户投诉工单文本挖掘分析[J]. 电力大数据, 2018(10): 68-73.
- WU G Y, ZHANG Q B, WU H C, et al. Text mining analysis of power customer complaints based on natural language processing technology[J]. Power Systems and Big Data, 2018(10):

68-73.

- [13] 邹云峰, 何维民, 赵洪莹, 等. 文本挖掘技术在电力工单数据分析中的应用[J]. 现代电子技术, 2016, 39(17): 149-152.
- ZOU Y F, HE W M, ZHAO H Y, et al. Application of text mining technology in power work order data analysis[J]. Modern Electronics Technique, 2016, 39(17): 149-152.
- [14] REZAC M. Measuring quality of scoring models using information value[J]. Journal of Communication and Computer, 2011(3): 234-239.
- [15] 刘荣, 冯国生, 丁维岱. SAS 统计分析与应用[M]. 北京: 机械工业出版社, 2011.
- LIU R, FENG G S, DING W D. Statistical analysis and application of SAS[M]. Beijing: China Machine Press, 2011.

[作者简介]



陈小峰(1980-), 男, 现就职于北京国网信通埃森哲信息技术有限公司, 主要研究方向为企业管理、精益生产和大数据挖掘。



赵雅迪(1993-), 女, 现就职于北京国网信通埃森哲信息技术有限公司, 主要研究方向为业务数据分析和数据挖掘。



张利鹏(1988-), 男, 现就职于北京国网信通埃森哲信息技术有限公司, 主要研究方向为业务数据分析和数据挖掘。



朱峰(1988-), 男, 现就职于北京国网信通埃森哲信息技术有限公司, 主要研究方向为财金业务分析和咨询。