

基于 Filter 模型和随机森林的电信客户投诉分类方法

岳丹阳 方 帅

(合肥工业大学计算机与信息学院 安徽 合肥 230009)

摘要 电信客户投诉记录中包含大量非结构化文本信息和客户通信行为信息。为了进行客户投诉的自动分析和精准分类,针对这些具有高维混合特征的数据,提出一种基于 Filter 模型和随机森林的客户投诉分类方法。采用 Filter 模型进行特征选择,以信息增益比来衡量特征的重要性,从而删除无关冗余的特征,降低维数;构建基于随机森林的客户投诉分类模型。采用某电信运营商3个月的投诉数据进行实验,结果表明:该方法相比其他集成学习方法,能够显著提高投诉处理的准确性,从而验证了其有效性。

关键词 客户投诉处理 文本分类 随机森林 特征选择

中图分类号 TP181

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2020.07.037

CLASSIFICATION METHOD OF TELECOM CUSTOMER COMPLAINTS BASED ON FILTER MODEL AND RANDOM FOREST

Yue Danyang Fang Shuai

(School of Computer and Information, Hefei University of Technology, Hefei 230009, Anhui, China)

Abstract Telecom customer complaint records contain a lot of unstructured text information and customer communication behavior information. For the purpose of automatic analysis and accurate classification of customer complaints, this paper proposes a classification method of customer complaints based on filter model and random forest. Filter model was used to select features, and information gain ratio was used to measure the importance of features, so as to delete irrelevant redundant features and reduce dimension. A customer complaints classification model based on random forest was constructed, and the experiment was carried out with the complaint data of a telecom operator for 3 months. The results show that compared with other integrated learning methods, this method can significantly improve the accuracy of complaint processing, thus verifying its effectiveness.

Keywords Customer complaints processing Text classification Random forest Feature selection

0 引言

通信服务种类的增多使用户对运营商的服务质量要求越来越高,随之而来的就是呈爆炸式增长的用户投诉量。传统的客户投诉分类方法主要是依靠客户关系管理系统和领域专家的经验判断,根据客户投诉文本中的关键词进行自动标识,再分配给相应技术支撑部门的领域专家进行处理^[1]。其中最重要的环节便是对投诉进行分类^[2]。例如,客户关于上网速度慢的投诉中有相当一部分不是因为运营商的通信设备或网络

问题,而是由于客户自身的手机终端配置或感知问题而导致的。若对投诉原因定位不准,则无法及时分配到相应的支撑部门,提供适当的解决方案,就会降低客户满意度,还可能造成投诉升级。客户投诉数据中包含了大量的非结构文本信息和用户的通信行为信息,因此,对这些高维混合特征数据进行自动分析和精准分类就成为有效处理电信客户投诉的关键。

近年来,集成学习方法逐渐被应用到高维混合数据的分类中^[3-4]。集成学习将多个学习器进行结合,往往能够获得比单一学习器更为显著的泛化性能。常见的集成学习方法根据个体学习器的生成方式可以分

为两种^[5]。一种是个体学习器之间存在强依赖关系必须串行生成的序列化方法,如 Boosting^[6];另一种是个体学习器不存在依赖关系,可以同时生成的并行化方法,如 Bagging^[7-8]和随机森林^[9]。随机森林简单易实现,在现实任务中具有强大的分类性能,被广泛应用于数据分类问题中,如垃圾网页检测^[10]、恶意软件检测等^[11]。然而,面对维度高、噪声大的数据,传统随机森林算法也面临着计算复杂度高和分类效果较差的问题。姚立等^[12]基于主题模型提取文本特征并对该算法进行了改进,将原始文本映射到主题空间中,从而降低文本噪声对分类结果的影响,然后再采用随机森林算法进行分类。卢晓勇等^[10]则提出一种基于随机森林和欠采样集成的二元分类器算法,以解决垃圾网页检测过程中的不平衡分类和“维数灾难”问题。由于电信客户投诉信息中包含大量的文本信息和通信行为信息,这些信息中存在大量冗余无关的特征,因此本文通过 Filter 模型进行特征选择,提出了一种基于 Filter 模型和随机森林的客户投诉分类方法。该方法首先将客户投诉文本和客户通信行为信息进行预处理,采用词频-逆文档频率(Term Frequency-Inverse Document Frequency, TF-IDF)进行特征提取,然后采用 Filter 模型进行特征选择,并构建基于随机森林的客户投诉分类方法,最后以某通信运营商的投诉数据进行实验,验证本文方法的有效性。

1 客户投诉的特征提取

用于描述客户投诉的特征包含客户行为特征和投诉文本特征。客户行为特征主要指运营业务支持系统中记录的关于客户手机通信网络制式、通信状态和所占用基站等信息,客户投诉文本的内容通常包含客户对通信服务质量的感知与描述以及关于地点、时间及其信号强度和呼叫质量的信息,这些信息都可作为判断客户投诉原因的依据。因此,本文首先提取客户投诉的文本特征,然后选取其中最重要的若干个特征与客户行为特征相结合构成客户投诉的特征向量。

常见的文本特征提取方法就是 TF-IDF。该方法是一种基于词频与逆文档频率的统计方法,其基本思想为:若一个词或一个短语在某个文档分类中出现的次数很多,而包含该词的文档分类却很少,则认为此词对分类的贡献程度较大,有较好的分类能力。 $tf(w, d)$ 为词频,表示词 w 在文档 d 中出现的频率, idf_w 为逆文档频率,表示了词 w 的类别区分能力,包含词 w 的文档越少则该值越大。TF-IDF 的计算公式如下:

$$tf(w, d) = \frac{f(w, d)}{\sum_k f(i_k, j)} \quad (1)$$

$$idf_w = \log\left(\frac{N}{1 + df_w}\right) \quad (2)$$

式中: $f(w, d)$ 表示词 w 在文档 d 中出现的频数; df_w 表示特征词 w 出现的文档数量; N 表示所有文档的数量。则得到特征词 w 的权重公式为:

$$tfidf_w = tf(w, d) \times idf_w \quad (3)$$

2 基于 Filter 模型的特征选择

特征选择是根据特征的相关性和冗余性,从原始特征集中进行搜索寻优,直到选择出满足特定评价指标的特征子集的过程。特征选择的方法主要有三种:滤波法(Filter)、封装法(Wrapper)和嵌入法(Embedded)^[13]。基于 Filter 模型的特征选择方法是直接利用数据的统计特性将特征的重要性进行排序,从而去除得分较低的特征。常用的度量重要性的方法有距离度量、相关性度量、一致性度量和信息度量。基于 Wrapper 模型的特征选择方法是将分类器学习模型和特征子集相结合,根据最终的分类结果错误率来评价特征子集的好坏。该方法会枚举出所有可能的特征组合,准确率高但是计算速度慢,泛化能力较弱,不适合大数据集。基于 Embedded 模型的特征选择方法是将特征选择作为学习算法的一部分,同时解决特征选择和分类问题,如基于决策树的特征选择法等。它同时具有了 Filter 和 Wrapper 的优点,但是损失函数的选择和缩放参数的设定较难给出。

本文选择 Filter 模型中的信息度量方式进行特征选择。采用特征的信息增益比(information gain ratio, IGR)进行特征选择,以校正信息增益计算过程中取值较多的特征对选择结果的影响。假设一个特征变量 X 与一个类别变量 Y 相关,IGR 值较高的特征则被认为与该分类变量 Y 相关度更高。IGR 是信息增益(information gain, IG)与特征 X 的内在熵值(intrinsic value, IV)的比率,即:

$$IGR(Y, X) = \frac{IG(Y, X)}{IV(X)} = \frac{H(Y) - H(Y|X)}{IV(X)} \quad (4)$$

式中: $H(Y) = -\sum_{y \in Y} p(y) \log_2 p(y)$ 是变量 Y 的熵值, $p(y)$ 是 Y 的边际概率密度函数。熵是一种常用的随机变量不确定性的度量,在特征提取中,常用来表示对 Y 进行分类的不确定性,熵越小,不确定性就越小。 $H(Y|X)$ 是条件熵,即:

$$H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2 p(y|x) \quad (5)$$

式中: $p(y|x)$ 是 y 给定 x 的条件概率。信息增益 $IG(Y, X)$ 定义的是一种度量,以反映添加特征 X 时变量 Y 减少的熵。加入特征 X 后,通过计算 Y 的 IG 值来度量特征 X 的贡献。特征 X 的内在值表示为:

$$IV(X) = - \sum_{x \in X} p(x) \log_2 p(x) \quad (6)$$

根据上述方法,在选择一个特征 X 来预测变量 Y 时,信息增益比可以通过考虑多值特征的数目和大小来减少对多值特征的偏差。因此采用的特征信息增益比越大,表明该特征对分类结果的影响越大,重要性程度也越高。

3 基于随机森林的客户投诉分类方法

由 Breiman 在 2001 年提出的随机森林算法是 Bagging 算法的一个变体,是一种基于决策树的组合分类器算法。随机森林算法在决策树的训练过程中引入了随机属性选择:传统决策树在选择划分属性时是在当前结点属性集合中选一个最优集合,随机森林则是先使用 Bagging 方法生成不同训练集,对基决策树的每个结点,在该结点的属性集合中得到一个包含数个属性的子集,再从子集中选择最优属性用于划分。即每棵决策树的训练样本由随机采样获得,且生成决策树时各节点分裂时所选择的属性也是随机的,两个随机性的结合,降低了决策树相似的可能性,这使得该算法面对噪声时的鲁棒性更好,对非平衡数据处理得到的结果也更加稳健,进而提升了随机森林的分类精度。基于随机森林的客户投诉分类方法共包含以下三个步骤,如图 1 所示。

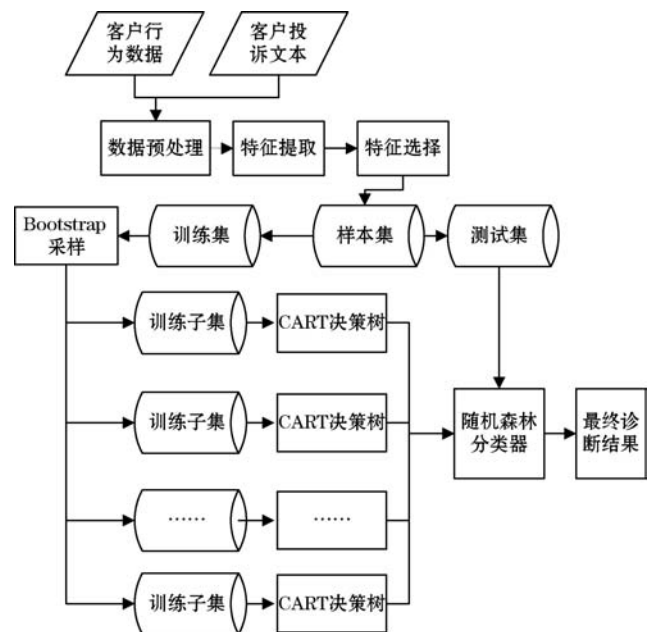


图 1 基于随机森林的客户投诉分类方法

(1) 训练集的采样生成。首先将样本分为训练集、验证集和测试集。对于训练集采用重采样的方法,在每一轮的学习中根据样本分布对训练集重新进行采样,从含有 D 个样本的原训练集中有放回地随机抽取 n 个新的样本集合,建立 n 棵决策树,每次采样中未被抽到的样本则会形成 n 个袋外数据(out of bag, OOB)。

(2) 决策树基分类器的训练。决策树算法的主要理论基础为信息熵,算法将分裂属性的取值划分成为两个子集,从两个子集出发,采用递归的方法在每个子集中再产生两个分支。随机森林中采用分类回归树(classification and regression tree, CART)算法建立没有剪枝的决策树作为基分类器。

分类回归树以 Gini 指数作为分裂标准来选取划分属性。Gini 指数是一种不等性度量,可用于度量任何不均匀分布,数值介于 0(表示完全相等)到 1(表示完全不等)之间。此处用于描述一个节点的不纯度,用以衡量从数据集中随机抽取两个样本,其类别标记不一致的概率。假设数据集 $T\{X, Y\}$ 中包含 J 个类别的样本, Gini 指数的定义如下:

$$Gini(t) = 1 - \sum_{j=1}^J [P(j|t)]^2 \quad (7)$$

式中: $P(j|t)$ 为类别 $j(j=1, 2, \dots, J)$ 在节点 t 处的概率。如果将样本集合分成 m 个部分,则进行这个划分的 Gini 指数为:

$$Gini(T) = \sum_{i=1}^m \frac{n_i}{n} Gini(i) \quad (8)$$

式中: m 为子节点的数目; n_i 为子节点 i 处的样本数; n 为母节点的样本数。

在候选属性集合中选择使得划分后 Gini 指数最小的属性作为最优划分属性,并根据其属性创建分支,每棵树都不进行任何剪枝操作。

(3) 随机森林的生成。随机森林是由大量的基分类器(决策树)组合而成的强分类器。重复步骤(2),遍历已建好的决策树,由 k 棵决策树组成随机森林。对于随机森林算法的分类结果,一般用投票机制进行决策,表示如下:

$$y^* = \operatorname{argmax}_{t \in T} I(P(j|t) = y) \quad (9)$$

式中: $I(\cdot)$ 是示性函数,取值是 0 或 1。 Y 为可能的类别。随机森林在输出待测样本的分类结果时就是根据 k 棵决策树投票最多的类别作为样本的最终类别。

4 实验结果与分析

4.1 数据收集与预处理

我们收集了某电信公司的技术支持部门在 2015

年8月至10月期间的客户投诉工单文本共3 000条,消除一些无效投诉后共获得数据1 433条。样本数据包括了关于客户投诉文本、客户通信行为信息、投诉处理结果和客户反馈的详细信息。客户投诉文本是公司客服部门在接到客户投诉信息后在客户管理系统中记录下来的工单信息。某客户的投诉文本如表1所示。

表1 客户投诉工单实例

工单内容
1. 时间: 2015-08-31 14:46。 2. 地点: ***小区2期。 3. 投诉描述: 室内没有信号,室外只有1~2格信号。换机测试无效。 4. 用户要求:在早上8:30之前晚上7:00点以后联系用户
1. 故障发生时间: 2015-08-31 07:38。 2. 发生地点: ***世纪城。 3. 客户投诉描述: 用户表示无法主叫,网络忙,表示在每天早上7:00左右都不行,可以被叫,满格。 4. 开关机测试情况:其他。 5. 其他地点是否正常:正常

客户通信行为信息来自运营商业支持系统,本文选择六个客户行为特征,包括:(1)干扰,用来判断投诉是否来自信号干扰地区;(2)热点区域,指投诉是否来自人群聚集区,如学校工厂等附近;(3)热点时段,指投诉是否发生在人群聚集期间,如学校放学和工厂下班等时刻;(4)基站状况,判断客户手机所占信号基站的运行状况是否良好;(5)弱信号区域,指投诉客户所在地区是否来自规划拟建设区域;(6)手机制式匹配,指投诉客户手机终端类型的网络制式与所占网络制式是否匹配。该六个客户行为特征变量取值均为0或1,1表示是,0表示否。

4.2 特征提取和选择

分类方法中主要使用文本的权重和语义,故对于一段文本,首先要进行分词。本文使用的分词工具是Python库中的jieba包。由于文本来自于移动通信行业的客户投诉工单,所以我们将移动通信业的特有名词如“大王卡”、“哔哩哔哩元卡”等定义为用户词典加入分词库中。

为提高文本分类的准确性,减少无意义词语造成的噪音干扰,还需要去除停用词。停用词本身没有实际意义,仅用于连接完整句子。本文选取1 803个停用词,作为停用词表放在记事本文件里,分词完毕后再进一步去除停用词。

使用TF-IDF方法对投诉工单的文本信息进行提取,共提取了118个文本特征。通过信息增益比计算,

筛选出前十个对分类贡献最大的特征,与另外六个客户行为特征组合构建客户投诉的特征向量。根据信息增益比值,选择出前十个重要的特征,如表2所示。

表2 选择出的十个重要特征

编号	1	2	3	4	5
特征	晚上	功能	网页	打开	只能
IGR	0.100 7	0.094 8	0.094 0	0.093 1	0.092 3
编号	6	7	8	9	10
特征	开通	打不开	速度	数据	查询
IGR	0.091 7	0.090 6	0.090 0	0.089 5	0.089 4

4.3 衡量指标的确定

使用混淆矩阵来衡量本文方法的有效性。混淆矩阵是一个由两行两列构成的表,表示真阳性(True positive, TP)、假阳性(False positive, FP)、假阴性(False negative, FN)和真阴性(True negative, TN)的数目。如果实际属于通信质量问题的投诉被诊断为1类,则为真阳性,如被诊断为2类则为假阴性。同样地,实际属于客户终端问题的投诉被诊断为2类为真阴性,被诊断为1类则为假阳性。采用准确度(Accuracy)、召回率(Recall)、精度(Precision)、F-值(F-measure)和AUC(Area under the Curve of ROC)值等五个性能指标。各值计算公式如下:

$$Accuracy = \frac{TP + TN}{Total}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F\text{-measure} = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$$

准确率是衡量分类器分类正确的概率。召回率又称敏感性或真阳性率,为衡量分类器在正确预测客户终端类投诉的概率;精度又称正向预测值,为通信质量问题的投诉个数除以标记为1类的总数;F-值同时考虑精度和召回率来衡量分类器的绩效。AUC值是指ROC曲线下方的面积,是一个用来评价二分类模型优劣的常用指标,AUC值越高通常表明模型的效果越好。

4.4 结果分析

对所提出的分类模型进行了十折交叉验证,以检验其可靠性。该模型分别进行十次训练和测试,计算上述五项性能指标十折交叉验证的平均值。考虑到分类器的分类性能可能会随特征数的变化而变化,为提

高分类模型的性能,分析具有不同文本特征数的分类模型的性能,结果如表 3 所示。

表 3 含不同特征的分类器性能比较

分类器	准确率	精度	召回率	F 值	AUC
F2	0.588	0.591	0.884	0.709	0.621
F1	0.846	0.810	0.953	0.876	0.928
(F1 + F2)_10	0.854	0.806	0.978	0.884	0.926
(F1 + F2)_20	0.854	0.807	0.977	0.884	0.927
(F1 + F2)_50	0.851	0.806	0.970	0.881	0.924

下面比较纯客户行为特征(F1)、纯文本特征(F2)和分别加入 10 个、20 个、50 个文本特征时的情况。结果表明选择 10 个文本特征与行为特征结合时分类性能整体较高,与选择 20 个文本特征时的结果相近,但是比纯文本特征和纯客户行为特征的分类准确度明显提高,也表明了将客户投诉文本和行为特征综合考虑进行投诉诊断的必要性。常见的分类器还有贝叶斯网络、朴素贝叶斯、支持向量机、Bagging、Adaboost 和随机子空间等方法。为了比较它们与本文提出的基于随机森林的分类器的性能,使用相同的数据集对六个分类器进行了训练,采用分类算法中的默认参数,进行十折交叉验证后得到不同分类器的性能指标,如表 4 所示。

表 4 不同分类器下的分类性能指标比较

分类器	准确率	精度	召回率	F 值	AUC
本文方法	0.854 0	0.806 0	0.978 0	0.884 0	0.926 0
Adaboost	0.830 0	0.820 0	0.897 0	0.857 0	0.906 0
Bagging	0.846 0	0.813 0	0.947 0	0.875 0	0.919 0
随机子空间	0.843 0	0.808 0	0.948 0	0.873 0	0.920 0
贝叶斯网络	0.849 0	0.810 0	0.958 0	0.878 0	0.921 0
朴素贝叶斯	0.849 0	0.808 0	0.963 0	0.879 0	0.921 0
支持向量机	0.820 0	0.783 0	0.945 0	0.856 0	0.801 0

结果表明,本文方法在进行客户投诉诊断时,与其他分类器相比精度更高,更有竞争力。由于本文方法考虑到在结合客户文本特征和行为特征时使用信息增益比进行特征选择,筛选出了十个贡献较大的文本特征,故本文还对比了在三种不同特征组合下,六种分类器性能指标上的绩效,以验证文本特征的选择对分类器性能的影响。① 未进行特征选择的全部混合特征(F1 + F2);② 纯文本特征(F2);③ 包含 10 个文本特征的混合特征(F1 + F2)_10。图 2 - 图 5 展示了选取不同特征时性能指标值的比较。结果表明,本文方法构建的分类器(F1 + F2)_10 性能远好于使用全部混合特征(F1 + F2)和只使用文本特征(F2)的分类器。

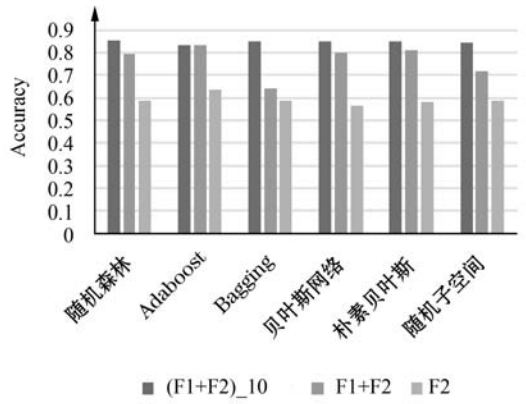


图 2 不同分类器的准确率

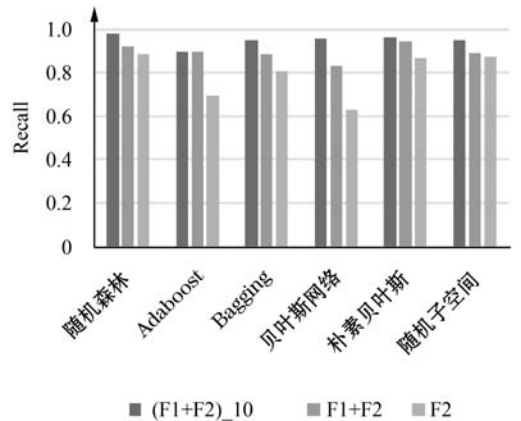


图 3 不同分类器的召回率

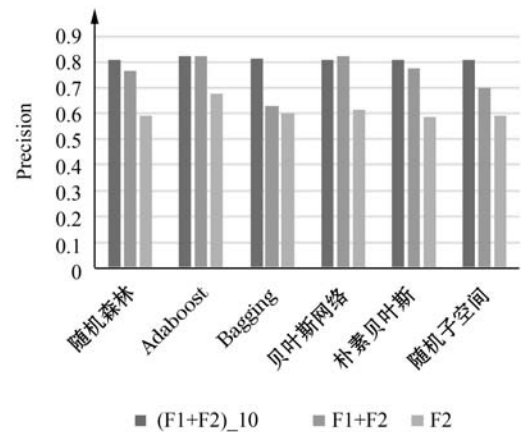


图 4 不同分类器的精度

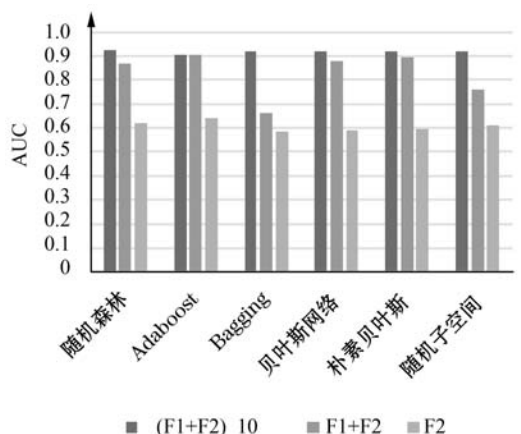


图 5 不同分类器 AUC 值

从比较分析可以看出,本文提出的基于随机森林的客户投诉分类器具有以下优点。首先,在文本特征的基础上融合了客户行为特征,而不是仅仅依据客户投诉的文本内容。其次,对提取出的文本特征利用信息增益比进行选择,筛选出其中对分类贡献最大的10个重要特征,使选择的特征更具解释性。最后,使用集成分类器,在单分类器的基础上增加属性与样本子集的随机性抽取,提高了分类器的分类准确率。

5 结 语

本文提出了一种基于 Filter 模型和随机森林的客户投诉分类方法,将来源于业务支持系统的客户行为特征和来自于客户投诉的文本特征数据相结合构建特征向量。客户行为特征的选择反映了领域专家的知识,文本特征则是经信息增益比进行特征选择后得到对分类贡献最大的特征,且具有较好解释性。使用了集成学习方法中的随机森林分类器,有效防止数据过拟合,提高了分类精度和泛化性能。本文使用一家知名的电信运营商数据进行实例研究。实验结果表明,该方法具有较高的分类精度,与经典贝叶斯网和朴素贝叶斯概率模型等概率模型相比具有较高分类精度,与 Bagging、Adaboost 等集成分类器相比,本文构建的分类器性能同样具有竞争力。本文提出的方法有助于电信技术人员高效地处理客户投诉,从而进一步提高客户满意度。

参 考 文 献

- [1] Faed A, Chang E, Saberi M, et al. Intelligent customer complaint handling utilising principal component and data envelopment analysis (PDA)[J]. *Applied Soft Computing*, 2016, 47(C):614-630.
 - [2] 梁昕露, 李美娟. 电信业投诉分类方法及其应用研究[J]. *中国管理科学*, 2015(S1):188-192.
 - [3] Yu Z, Luo P, You J, et al. Incremental semi-supervised clustering ensemble for high dimensional data clustering[J]. *IEEE Transactions on Knowledge & Data Engineering*, 2016, 28(3):701-714.
 - [4] Raychev B, Katamoto Y, Koujiba M, et al. Ensemble-based local learning for high-dimensional data regression [C]//2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2017.
 - [5] 周志华. 机器学习[M]. 北京:清华大学出版社,2016.
 - [6] Menezes F S D, Liska G R, Cirillo M A, et al. Data classification with binary response through the boosting algorithm and logistic regression [J]. *Expert Systems with Applications*, 2016, 69:62-73.
 - [7] Bauer E, Kohavi R. An empirical comparison of voting classification algorithms: bagging, boosting, and variants[J]. *Machine Learning*, 1999, 36(1-2):105-139.
 - [8] Spanakis G, Weiss G, Roefs A. Enhancing classification of ecological momentary assessment data using bagging and boosting[C]//2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, 2017.
 - [9] Onan A, Korukoğlu S, Bulut H. Ensemble of keyword extraction methods and classifiers in text classification[J]. *Expert Systems with Applications*, 2016, 57(C):232-247.
 - [10] 卢晓勇, 陈木生. 基于随机森林和欠采样集成的垃圾网页检测[J]. *计算机应用*, 2016, 36(3):731-734.
 - [11] 杨宏宇, 徐晋. 基于改进随机森林算法的 Android 恶意软件检测[J]. *通信学报*, 2017, 38(4):8-16.
 - [12] 姚立, 张曦煌. 基于主题模型的改进随机森林算法在文本分类中的应用[J]. *计算机应用与软件*, 2017, 34(8):173-178, 212.
 - [13] 黄铨. 特征降维技术的研究与进展[J]. *计算机科学*, 2018, 45(S1):16-21.
-
- (上接第 157 页)
- [7] McCartney E J. Optics of the atmosphere: scattering by molecules and particles[J]. *Journal of Modern Optics*, 1977, 14(7):521.
 - [8] Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms [J]. *Physica D Nonlinear Phenomena*, 1992, 60(1/4):259-268.
 - [9] 何人杰, 樊养余. 基于非局部全变分正则化优化的单幅雾天图像恢复新方法[J]. *电子与信息学报*, 2016, 38(10):2509-2514.
 - [10] 薛文丹, 赵凤群. 涉及景深的雾天图像增强的偏微分方程模型[J]. *计算机工程与应用*, 2017, 53(19):192-197.
 - [11] 赵胜楠, 魏伟波, 潘振宽, 等. 暗原色先验与 NL-CTV 模型相结合的图像去雾方法[J]. *计算机工程与应用*, 2018, 54(16):182-186.
 - [12] 魏伟波, 李帅, 潘振宽, 等. 基于 H-NL-MTV 模型的单幅彩色图像去雾[J]. *华南理工大学学报(自然科学版)*, 2018, 46(12):34-41, 50.
 - [13] 庞志峰, 张慧丽, 史宝丽. 基于 $e \sim p$ 的 DTU 图像去噪模型[J]. *北京航空航天大学学报*, 2019, 45(3):464-471.
 - [14] Gong Y, Sbalzarini I F. Curvature filters efficiently reduce certain variational energies[J]. *IEEE Transactions on Image Processing*, 2017, 26(4):1786-1798.
 - [15] Tarel J P, Hautiere N. Fast visibility restoration from a single color or gray level image[C]//IEEE 12th International Conference on Computer Vision. IEEE, 2009:2201-2208.
 - [16] Hautiere N, Tarel J P, Aubert D, et al. Blind contrast enhancement assessment by gradient ratioing at visible edges [J]. *Image Analysis & Stereology*, 2008, 27(2):87-95.