

基于实例和属性加权朴素贝叶斯的 电气故障分类研究

舒一飞¹, 郭汶昇¹, 樊博¹, 康洁滢¹, 许诗雨², 杨林²

(1. 国网宁夏电力有限公司营销服务中心, 银川 750000; 2. 四川大学电气工程学院, 成都 610065)

摘要: 为实现对电气故障快速、准确和动态的分类, 提出一种有机结合实例和属性加权的朴素贝叶斯电气故障分类方法(AIWNB); 朴素贝叶斯分类方法中的先验概率和条件概率采用两种实例加权方式加以改进, 积极实例权值取决于各属性值频度的统计值, 而消极实例权值通过逐条计算训练实例与测试实例间的相关性加以确定; 属性权值则基于互信息定义为属性-属性相关性和属性-类相关性之间的残差; 所提出的 AIWNB 方法将属性加权和实例加权有机结合在朴素贝叶斯统一框架内, 利用高低压用户的电气实测数据进行验证, 实验结果表明, 与朴素贝叶斯相比, 加权后的朴素贝叶斯方法更具竞争性, 准确率和 F1 分数可提升 3.09% 和 9.39%, 证明所提的 AIWNB 算法在电气故障分类的实用性及有效性; 并与其他电气故障分类方法进行对比, 验证算法的优越性。

关键词: 属性加权; 积极实例加权; 消极实例加权; 朴素贝叶斯框架; 电气故障分类

Research on Electrical Fault Classification Based on Instance and Attribute Weighted Naive Bayes

SHU Yifei¹, GUO Wensheng¹, FAN Bo¹, KANG Jieying¹, XU Shiyu², YANG Lin²

(1. Marketing Service Center, State Grid Ningxia Electric Power Co., Ltd., Yinchuan 750000, China;

2. College of Electrical Engineering, Sichuan University, Chengdu 610065, China)

Abstract: In order to achieve rapid, accurate and dynamic classification of electrical faults, an electrical fault classification model based on attribute and instance weighted naive bayes (AIWNB) is proposed. The prior probability and conditional probability in the naive Bayes classification method are improved by using two instance weighting methods. The eager instance weight depends on the statistics of the frequency of each attribute value, and the lazy instance weight is determined by calculating the correlation between the training instance and the test instance one by one. Attribute weight is defined as the residual between attribute-attribute correlation and attribute-class correlation based on mutual information. The proposed AIWNB method organically combines attribute weighting and instance weighting in a unified framework of Naive Bayes, the electrical measurement data of high and low voltage users is used to verify. Experimental results show that compared with pure Naive Bayes, the weighted Naive Bayes is more competitive, and the accuracy and F1 score can be increased by 3.09% and 9.39%, the practicality and effectiveness of algorithm are proved in the classification of electrical faults. Compared with other electrical fault classification methods, the superiority of algorithm is verified.

Keywords: attribute weighted; eager instance weighted; lazy instance weight; naive bayes framework; electrical faults classification

0 引言

我国电气急速发展, 电气化进程显著改善人民生活, 但由诸多因素引起的电气故障时有发生, 给用电安全、经济建设及电气市场发展带来直接影响^[1]。因此及时、快速、准确地对电气故障进行分类, 对后续检修、防治, 用电线路升级都具有重要意义。居民用电过程中的电气故障主要有漏电, 短路, 过压, 断相等, 而电气故障诊断分类的传

统方法是基于知识驱动的, 该方法由故障诊断知识库、专家知识库、故障诊断推理机等组成, 通过将监视的物理征兆导入推理机, 并由故障诊断推理机与专家知识库交互, 求解出物理征兆下的故障类型, 从而完成分类^[2]。该方法简单易行, 但存在较大的主观性, 不能做到对电网监测数据的实时充分利用和准确的故障分析分类。

近年来, 诸如神经网络、支持向量机、K 近邻、决策树等快速发展的数据驱动方式为故障分类提供了新的手段,

收稿日期: 2021-11-22; 修回日期: 2022-01-04。

基金项目: 国网宁夏电力有限公司营销服务中心(国网宁夏电力有限公司计量中心)科研项目(JG29YX210027)。

作者简介: 舒一飞(1989-), 宁夏银川人, 大学本科, 工程师, 主要从事采集主站建设运维、线损管理、反窃电大数据应用方向的研究。

通讯作者: 杨林(1997-), 四川成都人, 硕士研究生, 主要从事机器学习, 深度学习, 信号处理方向的研究。

引用格式: 舒一飞, 郭汶昇, 樊博, 等. 基于实例和属性加权朴素贝叶斯的电气故障分类研究[J]. 计算机测量与控制, 2022, 30(5): 169-174, 180.

由于具备坚实的数学基础, 无需先验知识, 可仅利用原始的数据特征就完成故障分类任务^[3]。但目前这种方法面临挑战, 数据驱动方式对数据的数量及质量都有着较高的要求。不充分的数据量将导致网络架构无法拟合, 过差的数据质量将使得网络的推广能力下降, 对于居民用电故障分类来说, 缺乏大量真实故障数据, 需通过仿真机产生大量的模拟故障数据片段, 不具有较强的推广能力^[2]。同时这种方法无法明确地表征数据特征, 仅在架构内部进行学习, 没有充分完整地数据属性及特征进行分析和利用, 造成了数据资源的浪费。

而朴素贝叶斯将贝叶斯概率、统计学理论和特征条件独立假设相结合, 对样本的数据信息进行了严密的推理计算, 形成了稳固的理论体系与方法, 可以解决数据分析不确定性和不完整性问题^[4]。同时为降低属性条件独立假设所带来的负面影响, 改进朴素贝叶斯网络分类算法引起了越来越多学者的广泛关注, 并由此产生了许多改进的分类模型和学习算法; 其中包括通过增加有向边限制条件独立假设的结构扩展、根据待测实例领域选择的实例选择、选择最合理属性子集的属性选择以及属性加权、实例加权五种改进方式^[5]。

其中效果较优的为属性加权朴素贝叶斯和实例加权朴素贝叶斯^[5]。属性加权是根据属性对分类预测的影响力来为不同的属性分配不同的权值, 然后在加权以后的实例样本集上构建贝叶斯分类器; 实例加权则根据实例分布情况对训练实例赋予不同的权重, 然后在加权的训练实例集上构建分类器^[6]; 加权方式可以科学严谨地表征数据特征, 并且能够充分对其进行计算与分析, 使得输入算法框架内的特征信息浓度更高, 较充分地提升算法性能表现^[7-8]。

但目前的研究局限于仅采用一种加权方法, 未将属性加权与实例加权有机结合在朴素贝叶斯框架内, 文章充分考虑不同属性应具有不同重要性, 不同实例应具有不同可靠性, 在朴素贝叶斯框架内采用双重加权, 解决数据分析不确定性和数据不完整性问题, 增加对数据片段特征的捕捉和利用, 提高算法的实用性和鲁棒性, 完成对电气故障高效准确的分类, 并可推广至其他分类情形; 并与其他基于数据驱动方式的电力故障分类方法进行对比, 验证文中算法的有效性。

1 朴素贝叶斯分类

朴素贝叶斯分类是依据贝叶斯定理计算条件概率从而实现分类预测^[9], 而贝叶斯定理的前提条件是各条件属性相互独立且对分类结果影响相同, 但这在现实中常常难以满足。

如果 $D = \{A_1, A_2, \dots, A_m, C\}$ 为训练数据集, 其中 $\{A_1, A_2, \dots, A_m\}$ 为 m 个属性变量, a_i 为是属性 A_i 的取值, $C = \{c_1, c_2, \dots, c_q\}$ 是类别总数为 q 的类别变量。一个待分类实例 $x_i = [a_1, a_2, \dots, a_m]$ 属于 c_j 类的条件概率, 根据贝叶斯定理有:

$$P(c_j | a_1, a_2, \dots, a_m) = \frac{P(a_1, a_2, \dots, a_m | c_j)P(c_j)}{P(a_1, a_2, \dots, a_m)} \quad (1)$$

其中: 为进行分类预测的后验概率, $P(a_1, a_2, \dots, a_m |$

$c_j)$ 为给定第 c_j 类下属性值为 $[a_1, a_2, \dots, a_m]$ 的条件概率, $P(c_j)$ 是第 c_j 类的先验概率, $P(a_1, a_2, \dots, a_m)$ 是属性的先验概率, 对于各类别都相同。

基于各属性相互独立的前提条件, 条件概率可表示为:

$$P(a_1, a_2, \dots, a_m | c_j) = \prod_{i=1}^m P(a_i | c_j) \quad (2)$$

其中: $P(a_i | c_j)$ 表示第 c_j 类下属性 a_i 的条件概率。对于待分类实例 x_i , 基于后验概率最大的原则, 朴素贝叶斯分类可表示为:

$$c(x) = \arg \max_{c_j \in C} P(c_j) \prod_{i=1}^m P(a_i | c_j) \quad (3)$$

式中, $P(c_j)$ 和 $P(a_i | c_j)$ 可以通过训练样本而求得, 其计算方式如下:

$$P(c_j) = \frac{\sum_{i=1}^n \delta(c_i, c_j) + \frac{1}{q}}{n + 1} \quad (4)$$

$$P(a_j | c_j) = \frac{\sum_{i=1}^n \delta(a_{ij}, a_j) + 1 \delta(c_i, c_j) + \frac{1}{n_j}}{\sum_{i=1}^n \delta(c_i, c_j) + 1} \quad (5)$$

其中: n 为训练样本总数, a_{ij} 为第 i 个训练样本的第 j 个属性值, n_j 为第 j 个属性值在训练样本中的个数, $\delta(r, t)$ 为二值函数, 定义为:

$$\delta(r, t) = \begin{cases} 1, & r = t \\ 0, & r \neq t \end{cases} \quad (6)$$

2 实例和属性加权的朴素贝叶斯

朴素贝叶斯分类基于各条件属性相互独立且对分类结果影响相同的假设条件, 这在现实中难以满足^[10]。事实上, 属性集中的一些属性比其他属性对预测分类更加重要, 因此, 可以通过属性加权为每个属性分配一个不同的权重, 通常, 对预测分类越重要的属性为其分配越大的权重, 反之亦然。此外, 一些训练实例比其他实例对预测分类更可靠, 它们对最终模型的影响应该也更大^[11], 因此实例加权也是另一种缓解朴素贝叶斯劣势的有效方法。

2.1 双重加权朴素贝叶斯框架

鉴于此, 本文融合属性加权和实例加权的优势, 提出属性和实例加权的朴素贝叶斯分类 (AIWNB)。改进后的 AIWNB, 一方面采用属性加权方法对不同的属性分别赋予不同的权重, 另一方面, 采用实例加权方法更准确地估计先验概率和条件概率。因此, AIWNB 可以同时继承属性加权的优点和实例加权的优点, 从而获得比单一属性加权或单一实例加权更好的性能。

在 AIWNB 中, 将属性权重纳入朴素贝叶斯分类公式中, 并将实例权重纳入先验概率和条件概率中。因此, AIWNB 的预测分类可表示为:

$$c(x) = \arg \max_{c_j \in C} P^{w_i} (c_j) \prod_{i=1}^m P^{w_i} (a_i | a_j)^{w_i} \quad (7)$$

其中: w_i^{at} 为第 i 个属性的权重, w_i^{ms} 为第 i 个训练实例的权重, 先验概率 $P^{w_i} (c_j)$ 和条件概率 $P^{w_i} (a_i | c_j)$ 可以计

算为:

$$P^{w_i} (c_j) = \frac{\sum_{i=1}^n \tau_i^{w_i} \delta(c_i, c_j) + \frac{1}{q}}{\sum_{i=1}^n \tau_i^{w_i} + 1} \quad (8)$$

$$P^{w_i} (c_j | c_i) = \frac{\sum_{i=1}^n \tau_i^{w_i} \delta(a_{ij}, a_j) \delta(c_i, c_j) + \frac{1}{n_j}}{\sum_{i=1}^n \tau_i^{w_i} \delta(c_i, c_j) + 1} \quad (9)$$

AIWNB 的整体流程如图 1 所示, 权重 τ_i^{att} 和 τ_i^{ms} 的计算见 2.2 和 2.3。

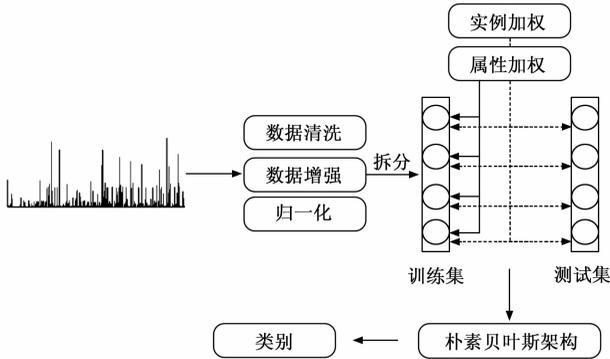


图1 AIWNB算法架构

2.2 实例权重 τ_i^{ms} 的设置

在实例加权中, 按照训练实例与测试实例之间的距离为训练实例来分配不同的权值; 与测试实例最相近的给予最大的权值, 反之, 最远的给予最小的权值。当实例权值 τ_i^{ms} 为 0 时, 该实例被删除; 实例权值 τ_i^{ms} 为 1 时, 即可认为该实例被选择, 因此实例加权较实例选择更为灵活。

寻找实例权值的方法通常采用积极学习方法和消极学习方法^[12], 文中采用了这两种实例加权方法, 并通过实验比较两种实例加权算法的差异性; 并据此提出二者的较为适宜的使用环境, 发挥更大的算法优势。

2.2.1 积极实例加权

为了保持朴素贝叶斯的计算简单性, 采用一种简单、高效、有效的方法——基于属性值频度的实例加权滤波器来学习每个实例的权重^[16]。这种方法是基于以下观察结果: 1) 每个属性值的频率包含一些重要的信息, 这些信息可以用来定义训练实例的权重; 2) 每个训练实例的权重与其属性值频率向量和整个训练数据集的属性值个数向量呈正相关。

首先, 计算每个属性值的频率,

$$f_{ij} = \frac{\sum_{r=1}^n \delta(a_{rj}, a_{ij})}{n} \quad (10)$$

f_{ij} 是 a_{ij} (第 i 个实例第 j 个属性值) 的频率, n 是训练实例的数量, a_{rj} 是第 r 个实例第 j 个属性值。

令 n_j 为第 j 个属性的值的数量, 并且可表示为 $[n_1, n_2, \dots, n_m]$; 则第 i 个实例的权重定义为属性值频率向量与属性值个数向量的内积,

$$\tau_i^{ms} = \langle f_{i1}, f_{i2}, \dots, f_{im} \rangle \cdot \langle n_1, n_1, \dots, n_m \rangle = \sum_{j=1}^m (f_{ij} * n_j) \quad (11)$$

2.2.2 消极实例加权

不同于积极实例加权, 消极实例加权集中注意力于计算并统计训练实例与测试实例间的相关性, 虽然时间复杂度较积极加权方式更高, 但是可以在分类准确率上得到提升^[16]。 x 为测试实例, y_i 为第 i 个训练实例, 它们间的相似性定义为:

$$s(x, y_i) = \sum_{j=1}^m \delta(a_j(x), a_j(y_i)) \quad (12)$$

从式 (16) 可以看出, $s(x, y_i)$ 简单地对 x 和 y_i 的相同属性的数量进行统计的函数, 进而粗略地衡量两个实例间的相似程度, 定义第 i 个训练实例的权重为:

$$\tau_i^{ms} = 1 + s(x, y_i) \quad (13)$$

2.3 属性权重 τ_i^{att} 的设置

利用基于相关性的属性权重方法为每个属性指定不同的权重, 具有强预测性的属性应该与类别标签相关, 与其他属性低相关或无关^[13]。属性权重 τ_i^{att} 为 0 时, 该属性被删除; 属性权重 τ_i^{att} 为 1 时, 该属性保留且在分类中占据主导地位, 较属性选择更加具有鲁棒性与有效性。

鉴于互信息可以度量每对随机离散变量之间的相关性^[15], 因此, 文中利用互信息来度量属性一类之间、属性一属性之间的相关性, 从而计算得到类关联度和属性间的相互关联度。每个属性的权重^[14]可以通过对归一化后的类关联度与平均相互关联度之间的差值进行 Sigmoid 变换后获取, 具体的计算步骤如下所示:

1) 属性 (A_i) — 类 (C) 的相关性、属性 (A_i) — 属性 (A_k) 的相关性采用互信息来度量, 分别定义为:

$$I(A_i; C) = \sum_{a_i} \sum_c P(a_i, c) \log \frac{P(a_i, c)}{P(a_i) P^{w_i}(c_j)} \quad (14)$$

$$I(A_i; A_k) = \sum_{a_i} \sum_{a_k} P(a_i, a_k) \log \frac{P(a_i, a_k)}{P(a_i) P(a_k)} \quad (15)$$

其中: $P^{w_i}(c_j)$ 为式 (4) 中的实例加权后的先验概率。

2) 对 $I(A_i; C)$ 和 $I(A_i; A_k)$ 进行归一化操作:

$$NI(A_i; C) = \frac{I(A_i; C)}{\frac{1}{m} \sum_{j=1}^m I(A; C)} \quad (16)$$

$$NI(A_i; A_k) = \frac{I(A_i; A_k)}{\frac{1}{m(m-1)} \sum_{j=1}^m \sum_{k=1 \wedge k \neq j}^m I(A_i; A_k)} \quad (17)$$

其中: $NI(A_i; C)$ 和 $NI(A_i; A_k)$ 分别为归一化类关联度和相互关联度; 归一化后的各相关性可直接用于运算, 不受相关属性的单位限制。

3) 将属性权重定义为类关联度与平均相互关联度的差值。

$$\Delta \tau_i^{att} = \underbrace{NI(A_i; C)}_{\text{relevance}} - \underbrace{\frac{1}{m-1} \sum_{k=1 \wedge k \neq i}^m NI(A_i; A_k)}_{\text{average redundancy}} \quad (18)$$

$\Delta \tau_i^{att}$ 即为属性 A_i 的权值。

根据上式计算得到的权值可能为负数, 应使权值范围

落入 $[0, 1]$, 利用 sigmoid 函数对其进行变换, 可得到各属性权值,

$$\omega_i^{att} = \frac{1}{1 + e^{-\Delta\omega_i^{att}}} \quad (19)$$

进一步地, 将属性加权和两种实例加权相结合, 由此得到如下两种用于电气设备故障分类的算法, AIWNB^E (属性一积极实例加权朴素贝叶斯) 和 AIWNB^L (属性一消极实例加权朴素贝叶斯)。

算法 1: AIWNB^E

输入: 原始训练集

输出: 属性权值向量 ω^{att} , 实例权值向量 ω^{ins} , 类别 c

1. 计算训练集各实例的属性值频率 f_{ij} 及属性值个数 n_j ;
2. 计算实例权值 ω_j^{ins} ;
3. 计算各属性 A_j 的 $I(A_j; C)$ 与 $I(A_j; A_k)$;
4. 对各属性的 $I(A_j; C)$ 与 $I(A_j; A_k)$ 归一化得到 $NI(A_j; C)$, $NI(A_j; A_k)$;
5. 计算各属性的 $\Delta\omega_j^{att}$ 和 ω_j^{att} ;
6. 返回属性权值 ω_j^{att} 和实例权值 ω_j^{ins} ;
7. 计算先验概率 $P(c)$, 条件概率 $P(a_j | c)$, 类别概率 $c(x)$;
8. 返回类别。

算法 2: AIWNB^L

输入: 原始训练集

输出: 属性权值向量 ω^{att} , 实例权值向量 ω^{ins} , 类别 c

1. 计算训练实例 y_i 和待测实例 x 的相似性 $s(x, y_i)$;
2. 计算训练实例权值 ω_i^{ins} ;
3. 计算各属性 A_j 的 $I(A_j; C)$ 与 $I(A_j; A_k)$;
4. 对各属性的 $I(A_j; C)$ 与 $I(A_j; A_k)$ 归一化得到 $NI(A_j; C)$, $NI(A_j; A_k)$;
5. 计算各属性的 $\Delta\omega_j^{att}$ 和 ω_j^{att} ;
6. 计算先验概率 $P(c)$, 条件概率 $P(a_j | c)$, 类别概率 $c(x)$;
7. 返回类别。

3 实例验证

3.1 数据预处理

用电数据采集于国网宁夏电气有限公司计量中心, 包含高低压用户 783 户为期 31 天的三相电压, 三相电流, 功率及功率因数 8 项数据。获得数据后对异常数据进行清洗, 因电气故障的发生将导致各项数据发生相应变化, 仅去除掉空值、数值偏差极大、重复值的数据段。

根据已统计到的用电情况, 并结合故障发生时电压电流变化情况, 经由相关文献[17-20]及专家标注, 可总结为如表 1 的故障类别及判定依据。如果设备负载是家用热电偶、灯、加热器等, 供电线路出现以下 4 种故障时设备电流、电压和功率变化趋势如表 1 所示, 正常运行数据段标签为“1”。

由于数据集内包含的故障数据段过少, 通过在故障数据段内的每个属性叠加随机高斯噪声, 增加少量故障数据片段, 提升加权朴素贝叶斯对其特征的捕捉能力; 随机噪声与真实情况接近, 保证数据可用, 较大程度模拟真实故障数据。公式如下:

表 1 电气故障判据

故障	故障判据	标签
漏电	电流增大, 电压降低, 功率因数降低	2
短路	故障相电压为 0, 非故障相电压增加	3
	电流不变, 三相不平衡	
过压	电压超过额定电压 20% 以上	4
	电流增加, 功率增加, 功率因数基本不变	
断相	故障相无电压	5
	非故障相电压为正常低压相的 0.866 倍	
	电流、功率和功率因数降低	

$$Gauss(x) = \frac{1}{\sqrt{2\pi\delta}} e^{-\frac{(x-\mu)^2}{2\delta^2}} \quad (20)$$

其中: μ 为高斯噪声的均值, 设置为 0; δ 为高斯噪声的标准差, 设置为 0.1; $Gauss(x)$ 为生成的噪声, 将其与原始数据叠加, 完成数据增强。仅对 4 种电气故障类型进行数据增强, 正常运作片段不作处理。原始数据及数据增强后的各电气故障类型数量对比如表 2 所示。

表 2 数据增强前后故障类型数量

故障标签	原始数量	增强后数量
1	23 159	23 159
2	41	8 241
3	88	8 888
4	53	8 003
5	522	9 918

经由数据增强后总共获得 58 209 条数据段, 随机取得 10 000 条数据作为测试集, 并对训练集与测试集进行归一化, 归一化可提高 AIWNB 的收敛速度以及分类精度, 文章采取 min-max 归一化方法, 其公式如下,

$$A_j^* = \frac{A_j - \min}{\max - \min} \quad (21)$$

其中: \min 为所有数据段的当前属性 A_j 的最小值, \max 为所有数据段的当前属性 A_j 的最大值。

3.2 评价指标

为准确评估文章提出模型在分类中的表现, 选择准确率 (Accuracy) 和 F1 分数也作为衡量模型性能和有效性的指标, 其计算公式如下:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (22)$$

$$F1 = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (23)$$

其中: 精准率 (Precision) 及召回率 (Recall) 计算公式如下,

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (24)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (25)$$

式中, TP 表示正类别中准确分类样本的数量; TN 表示负类别中准确分类样本的数量; FP 表示正类别中错误分类样本

的数量; FN 是负类别中错误分类样本的数量。

3.3 结果及分析

按前文所述, 对属性及实例分别求解其权值, 为使实验更严谨、更具说服力, 将分别使用单独属性加权, 单独积极实例加权, 单独消极实例加权, 属性一积极实例加权, 属性一消极实例加权与原始朴素贝叶斯实验结果进行对比, 获取到效果最佳的加权朴素贝叶斯方式。

3.3.1 属性及实例权值

积极实例加权依赖于对各属性的属性值频率计算, 而观察采集到的数据得知实际运行数据的各属性值分散, 导致属性值频率 f_{ij} 极小, 并不利于计算; 因此借用模糊化思想, 文中按照表 1 所述的各属性值变化界定范围, 按照类别进行划分, 进而计算得到各类实例的权值, 如表 3 所示。

表 3 积极实例加权各类实例权值

类别	1	2	3	4	5
权值	0.295 5	0.114 4	0.026 7	0.394 4	0.169 0

消极实例加权更侧重于训练实例与测试实例的相似性, 通过计算每条训练实例与每条测试实例的相似性对实例进行权值的划分, 进而改善朴素贝叶斯对不同实例的注意程度, 提升分类精度。因数据片段过多, 部分实例权值如表 4 所示。

表 4 消极实例加权一部分实例权值

数据段序号	权值	数据段序号	权值
1	0.873 2	131	0.733 6
3 645	0.877 5	7 603	0.904 0

通过对属性一类相关性和属性一属性相关性的残差进行 sigmoid 变化得到各属性权值, 基于互信息的各属性权值如表 5 所示。

表 5 基于互信息的各属性权值

属性	电压-A 相	电压-B 相	电压-C 相	电流-A 相
权值	0.614 6	0.632 7	0.580 8	0.416 2
属性	电流-B 相	电流-C 相	功率	功率因数
权值	0.416 9	0.416 4	0.423 1	0.496 2

3.3.2 算法性能分析

为全面评价算法的整体性能, 文章对各算法的准确率及 F1 分数进行计算, 具体值如表 6~7 所示, 各算法总体准确率如图 2 所示。

表 6 算法准确率对比

算法/类别	NB	AW NB	IW NB ^E	IW NB ^L	AIW NB ^E	AIW NB ^L
1	94.97	95.72	96.24	96.41	97.18	98.06
2	96.47	96.72	97.24	97.3	98.23	98.91
3	99.67	99.78	99.76	99.78	99.88	99.78
4	98.92	99.22	99.22	99.33	99.07	99.37
5	99.89	100	99.98	100	100	100

表 7 算法 F1 分数对比

算法/类别	NB	AW NB	IW NB ^E	IW NB ^L	AIW NB ^E	AIW NB ^L
1	90.88	92.45	93.25	93.54	94.99	96.61
2	86.06	86.50	89.09	89.42	92.80	95.45
3	94.18	96.04	95.70	96.04	97.80	96.04
4	98.65	99.02	99.02	99.16	98.83	99.21
5	99.68	100	99.94	100	100	100

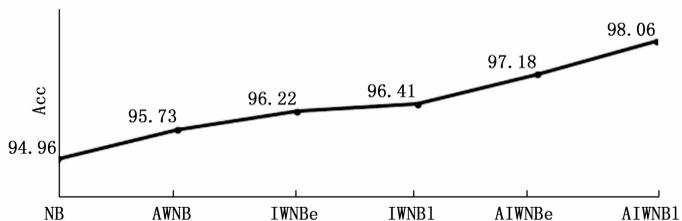


图 2 算法总准确率曲线图

可以看出, 加权朴素贝叶斯在纳入不同属性、实例的权值后, 相较于单一朴素贝叶斯提高了对电气故障分类的准确率, 同时获得较高的 F1 分数, 证明在加权朴素贝叶斯能够提取到更多的全局特征, 可较大程度提高分类准确率。积极实例加权与消极实例加权的准确率相差较小, 但在实验过程中, 消极实例加权因需要对训练实例及测试实例逐条计算相关性, 耗费更多的运算时间, 实际使用中偏向于实时性, 可考虑使用积极实例加权。而在电气数据分析和诊断中, 消极实例加权的优势更大, 可以获得最高的分类准确率。

双重加权朴素贝叶斯较单一朴素贝叶斯的准确率提升为 0.11%~3.09%, F1 分数提升为 0.32%~9.39%。属性一实例双重加权朴素贝叶斯对于待分类的电气故障的高准确率来源于对属性和实例两者权值的同时分析, 完整又准确地提取各电气故障的特征分布; 并且在第 4 类与第 5 类取得接近完全准确的分类效果。属性一积极实例加权朴素贝叶斯 (AIWNB^E) 及属性一消极实例加权朴素贝叶斯 (AIWNB^L) 的混淆矩阵如图 3 所示。

AIWNB^E 与 AIWNB^L 在面对第 2, 3, 4, 5 类时几乎可以取得完全准确的分类结果, 且在 1 类的偏差较小; 由于消极实例加权更侧重于对测试实例的分析, 进而会取得更加优良的分类效果, 而代价则是增加运算时间, 积极实例加权的分类效果稍弱, 但运算时间较少, 仅需根据属性范围确定权值。

3.4 其他电气故障分类方法性能对比

文中基于属性一消极实例加权朴素贝叶斯的电气故障分类在实际数据中取得了 98.06% 的准确率, 为进一步验证算法有效性, 选取 K 近邻 (KNN, k-nearest neighbor)^[21], 决策树 (DT, decision tree)^[22]、支持向量机 (SVM, support vector machine)^[23]、极端梯度提升 (XGBoost)^[24] 等基于数据驱动方式作为电气故障分类方法进行对比; 上述机器学习方

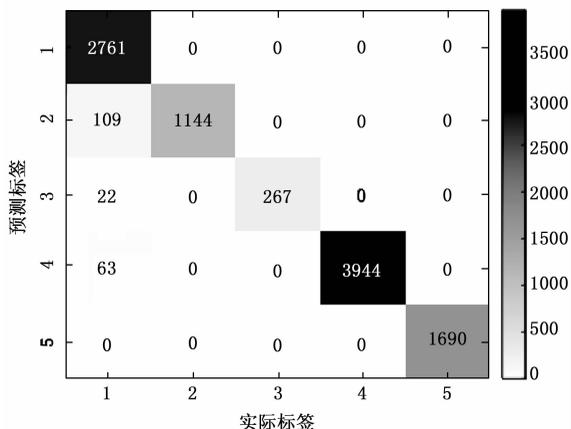
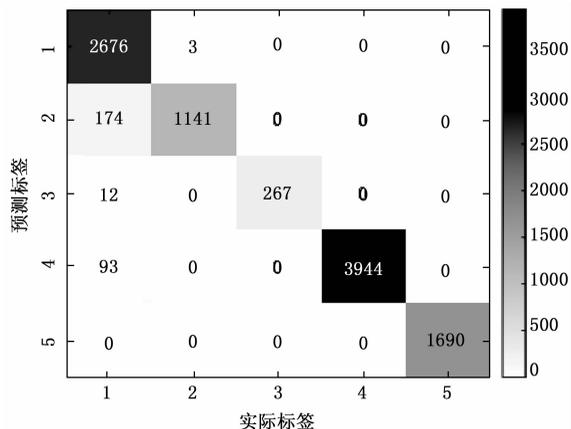


图 3 双重加权朴素贝叶斯分类混淆矩阵

法使用广泛, 分类效果较好, 可形成较好的对比组。

为保证对比实验的可信度, 在完全相同的训练集与测试集下收集各电气故障分类方法的分类准确率, 如表 8 所示。

表 8 其他电气故障分类方法准确率对比

算法/类别	AIWNB ^E	AIWNB ^L	KNN	DT	SVM	XGBoost
1	94.99	96.61	95.43	95.52	94.54	95.70
2	92.80	95.45	85.12	94.09	89.44	94.49
3	97.80	96.04	96.32	93.69	92.39	96.81
4	98.83	99.21	95.98	95.38	96.36	99.62
5	100	100	96.71	98.32	98.64	99.72

对各方法指标数据的进一步观察可知, 双重加权朴素贝叶斯在获得属性及实例的权值后, 能够更加准确地对电气故障进行分类, 相较于 KNN、DT 与 SVM 取得了更好的分类效果; 而 XGBoost 在第 4 类取得了更好的分类效果, 但 XGBoost 依赖于梯度的下降, 计算量较大, 根据实验数据显示运算时间为 1.18 s; 而 AIWNB^L 的运算时间仅为 0.536 s, 具有更好的实时性。

4 结束语

文章采用朴素贝叶斯架构, 提出了基于属性加权和实

例加权的电气故障分类算法。该算法通过互信息统计属性一类间依赖性和属性一属性冗余度获取各属性权值, 提升算法对各属性特征的提取与利用能力; 分别计算实例属性值频率和实例间相关性完成积极、消极实例加权, 进一步提升算法对数据全局特征的把控能力。在实验过程中, 双重加权朴素贝叶斯在准确率、F1 分数优于单独加权朴素贝叶斯、单一朴素贝叶斯与其他常用分类方法, 较显著降低误分数量, 提高分类实时性。文章算法是通用算法, 可以推广至其他待分类情形, 譬如心电信号、岩爆等。

参考文献:

- [1] 李 晖, 刘 栋, 姚丹阳. 面向碳达峰碳中和目标的我国电气系统发展研判 [J]. 中国电机工程学报, 2021, 41 (18): 6245 - 6259.
- [2] 齐 奔, 梁金刚, 张立国, 等. 基于贝叶斯分类器的核电厂故障诊断方法研究 [J]. 原子能科学技术, 2022, 56 (3): 512 - 519.
- [3] 阎 馨, 朱永浩, 屠乃威, 等. 基于 PCA 与权重贝叶斯的工作面煤与瓦斯突出预测 [J]. 计算机工程, 2021, 47 (8): 315 - 320.
- [4] 束美艳, 魏家玺, 周也莹, 等. 基于朴素贝叶斯分类的柑橘叶片溃疡病诊断 [J]. 浙江大学学报 (农业与生命科学版), 2021, 47 (4): 429 - 438.
- [5] JIANG L, CAI Z, ZHANG H, et al. Improving Naive Bayes for Classification [J]. International Journal of Computers and Applications, 2010, 32 (3): 327 - 332.
- [6] 郭树强, 黄 蕊, 李 卿. 改进加权朴素贝叶斯的软件缺陷预测算法 [J]. 控制工程, 2021, 28 (3): 600 - 605.
- [7] 赵海霞, 李 赟, 石洪波. 基于高维数据的加权朴素贝叶斯算法研究 [J]. 统计与决策, 2020, 36, 548 (8): 7 - 11.
- [8] 周颜军, 王双成, 王 辉. 基于贝叶斯网络的分类器研究 [J]. 东北师大学报 (自然科学版), 2003 (2): 21 - 27.
- [9] 余良俊. 属性加权的贝叶斯网络分类算法及其应用研究 [D]. 武汉: 中国地质大学, 2018.
- [10] WU J, CAI Z. Attribute weighting via differential evolution algorithm for attribute weighted naive bayes [J]. Journal of Computational Information Systems, 2011, 7 (5): 1672 - 1679.
- [11] 伍杰华, 沈 静, 周 蓓. 改进朴素贝叶斯模型的复杂网络关系预测 [J]. 计算机工程与科学, 2017, 39 (10): 1825 - 1831.
- [12] MCCANN S, LOWE D G. Local naive bayes nearest neighbor for image classification [C] //2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012: 3650 - 3656.
- [13] ZAIDI N A, CERQIDES J, CARMAN M J, et al. Alleviating naive bayes attribute independence assumption by attribute weighting [J]. Journal of Machine Learning Research, 2013, 14: 1947 - 1988.
- [14] HALL M. A decision tree-based attribute weighting filter for naive bayes [J]. Knowledge-Based Systems, 2007, 20 (2), 120 - 126.

(下转第 180 页)

- sing microwave interconnects [C] // 2012 IEEE 21st Conference on Electrical Performance of Electronic Packaging and Systems, Tempe: IEEE, 2012; 3-6.
- [2] CHUANG H H, HSU C J, HONG J, et al. A broadband chip-level power-bus model feasible for power integrity chip-package codesign in high-speed memory circuits [J]. IEEE Transactions on Electromagnetic Compatibility, 2009, 52 (1): 235-239.
- [3] FADEYI J, MARKUS E D, KANZUMBA K. A Review of the Impact of Reflection of Transient Signals on TEM Transmission Lines [C] // 2018 Open Innovations Conference (OI), Johannesburg: IEEE, 2018; 26-33.
- [4] CHEN B, PAN S, WANG J, et al. Differential Crosstalk Mitigation in the Pin Field Area of SerDes Channel With Trace Routing Guidance [J]. IEEE Transactions on Electromagnetic Compatibility, 2019, 61 (4): 1385-1394.
- [5] HUREMOVIC A, HADZIALIC M. Novel approach to analytical jitter modeling [J]. Journal of Communications and Networks, 2015, 17 (5): 534-540.
- [6] SONG T, LIU C, PENG Y, et al. Full-chip signal integrity analysis and optimization of 3-D ICs [J]. IEEE Transactions on Very Large Scale Integration Systems, 2015, 24 (5): 1636-1648.
- [7] SJIARIEL R. Power integrity simulation of power delivery network system [C] // Microwave & Optoelectronics Conference, Berlin: IEEE, 2016; 30-35.
- [8] 马雪松, 姚静波, 邓茂林. 新一代运载火箭时序仿真系统信号完整性分析 [J]. 计算机测量与控制, 2015 (1): 30-35.
- [9] 李俊杰, 曹旭东, 梁华庆. 嵌入式应用系统中高速 PCB 设计技术的研究及实现 [J]. 计算机测量与控制, 2016 (6): 25-29.
- [10] KINGER X C, STEVEN Y J, DAKROUB M, et al. Signaling margin oriented LPDDR PDN design with SIPI synthesis methodology [C] // International Symposium on Electromagnetic Compatibility (EMC), Ottawa, IEEE, 2016; 50-53.
- [11] PANDEY A K. Power-aware signal integrity analysis of DDR4 data bus in onboard memory module [C] // Signal & Power Integrity, Turin: IEEE, 2016; 35-40.
- [12] 王海三. DDR 总线信号完整性分析技术研究 [D]. 上海: 华东理工大学, 2015.
- [13] HUANG K H, HSIEH C Y, TSENG B C, et al. Comprehensive signal and power co-investigation on DDR4 simulation and measurement [C] // IEEE International Symposium on Electromagnetic Compatibility and IEEE Asia-Pacific Symposium on Electromagnetic Compatibility (EMC/APEMC), Singapore: IEEE, 2018; 1041-1044.
- [14] SEO D, LEE H, PARK M, et al. Enhancement of Differential Signal Integrity by Employing a Novel Face Via Structure [J]. IEEE Transactions on Electromagnetic Compatibility, 2017, 60 (1): 1-8.
- [15] CHENG W, SARKAR A, et al. Worst Case Switching Pattern for Core Noise Analysis [C] // DesignCon, Santa Clara, 2009; 1-16.
- [16] 马晓剑, 李晓春, 王宁, 等. 一种新型混合型超宽带 EBG 结构 [J]. 微波学报, 2016, 32 (2): 30-34.
- [17] 鱼鹏. 高速电路中 PDN 电源噪声分析及去耦网络设计 [D]. 西安: 西安电子科技大学, 2018.
- [18] KHO J, LOH C I, KRSTNIK B, et al. Effect of SSN-induced PDN noise on a LVDS output buffer [C] // 2007 Asia-Pacific Conference on Applied Electromagnetics, Melaka: IEEE, 2008; 50-59.
- [19] SHI H, LIU G, LIU A, et al. Simultaneous switching noise in FPGA and structure ASIC devices, methodologies for analysis, modeling, and validation [C] // 56th Electronic Components and Technology Conference 2006, San Diego: IEEE, 2006; 229-236.
- [20] MENG H, YIN L, TAN A, et al. Power and Signal Integrity Analysis of High-speed Mixed-signal Backplanes Based on VPX [C] // IEEE Symposium on Electromagnetic Compatibility, Signal Integrity and Power Integrity (EMC, SI&PI), Long Beach: IEEE, 2018; 577-581.
- [21] SHIUE G H, YE H C L, HUANG P W, et al. Ground bounce noise induced by crosstalk noise for two parallel ground planes with a narrow open-stub line and adjacent signal traces in multilayer package structure [J]. IEEE Transactions on components, packaging and manufacturing technology, 2014, 4 (5): 870-881.
- [22] 岳根霞, 刘金花, 刘峰. 基于决策树算法的医疗大数据填补及分类仿真 [J]. 计算机仿真, 2021, 38 (1): 451-454, 459.
- [23] 梁亮, 杨敏华, 李英芳. 基于 ICA 与 SVM 算法的高光谱遥感影像分类 [J]. 光谱学与光谱分析, 2010, 30 (10): 2724-2728.
- [24] 陈凯, 南东亮, 孙永辉, 等. 基于 XGBoost 算法的电网二次设备缺陷分类研究 [J]. 南京信息工程大学学报 (自然科学版), 2019, 11 (4): 483-489.
- [15] JIANG L, ZHANG H, CAI Z, et al. Weighted average of one-dependence estimators. Journal of Experimental Theoretical Artificial Intelligence [J]. 2012, 24 (2): 219-230.
- [16] ZHANG H, JIANG L, YU L. Attribute, and instance weighted naive Bayes [J]. Pattern Recognition, 2021, 111: 107674.
- [17] 刘湖云, 梁海莲, 顾晓峰, 等. ESD 应力下改进型 SCR 器件设计与漏电特性优化 [J]. 西安电子科技大学学报, 2018, 45 (6): 118-122.
- [18] 欧阳金鑫, 叶俊君, 张真, 等. 电网故障下多馈入直流输电系统相继换相失败机理与特性 [J]. 电气系统自动化: 2021 (20): 93-102.
- [19] 孙家豪, 王笑雪, 李光辉, 等. 换相失败引起送端电网过电
- 压的定量计算及影响因素 [J]. 电气系统及其自动化学报, 2020, 32 (12): 62-68.
- [20] 齐宏伟, 孙逸民, 王帆, 等. 核电厂高压电气断相故障仿真分析 [J]. 计算机仿真, 2021, 38 (1): 380-386.
- [21] 卞则康, 张进, 王士同. 基于类标感知的 KNN 分类算法 [J]. 模式识别与人工智能, 2021, 34 (10): 873-884.