

# 基于多属性联合的朴素贝叶斯分类算法

谢小军<sup>1</sup>, 陈光喜<sup>2</sup>

(1. 桂林电子科技大学 数学与计算科学学院, 广西 桂林 541004;  
2. 桂林电子科技大学 广西高校图像处理实验室, 广西 桂林 541004)

**摘要:**朴素贝叶斯分类算法由于条件独立性假设对属性施加了一定的限制,这可能会降低分类性能。针对此问题,为了削弱条件独立性假设对分类结果带来的负面影响,从结构扩展的角度提出了一种基于多属性联合的朴素贝叶斯分类算法。该算法通过计算条件属性组合相对于决策属性依赖度的大小,选择最大相对属性依赖度的属性组合进行联合作为新的条件属性。通过实验仿真,结果表明该方法可行而且有效,特别是对一些属性之间关联性比较强的数据集,分类效果提高尤为明显。

**关键词:**朴素贝叶斯;分类;粗糙集理论;相对属性依赖度;属性联合

**中图分类号:**TP181

**文献标识码:**A

**文章编号:**1673-629X(2016)12-0077-05

doi:10.3969/j.issn.1673-629X.2016.12.017

## Naive Bayes Classification Algorithm Based on United Multi-attribute

XIE Xiao-jun<sup>1</sup>, CHEN Guang-xi<sup>2</sup>

(1. School of Mathematics and Computer Science, Guilin University of Electronic Technology, Guilin 541004, China;  
2. Guangxi University Image Processing Laboratory, Guilin University of Electronic Technology,  
Guilin 541004, China)

**Abstract:** Naive Bayes classification algorithm imposes certain restrictions on the properties due to the conditional independence assumption, which may reduce the classification performance. Aiming at this problem, in order to weaken the negative influence of conditional independence assumption on the classification results, a new Naive Bayesian classification algorithm based on multi attributes is proposed from the point of view of structure expansion. In this algorithm, the combination of the maximum relative attribute dependency is selected as a new condition attribute by computing the size of the combination of the conditional attributes with respect to the decision attribute. The experiment results show that it is feasible and effective, especially for some of the properties between the correlation stronger data sets, classification effect is particularly obvious.

**Key words:** Naive Bayes; classification; rough set theory; relative attribute dependency; attribute combination

## 0 引言

朴素贝叶斯分类器是一种基于贝叶斯统计理论的有监督学习方法,对已标记训练样本进行分类时,基于一个强有力的假设,即训练样本中的所有属性均独立于训练样本。然而此假设在现实中往往不成立,因此,研究人员做了大量工作来削弱朴素贝叶斯独立性假设。Jiang等<sup>[1]</sup>对改进朴素贝叶斯的方法做了综述,这些改进方法粗略分为:结构扩展、特征选择、数据扩展、局部学习和属性加权五大类。基于属性加权的思想,国内许多研究人员做了大量研究工作。文献[2-5]分别使用粗糙集属性重要度、属性之间互信息、属性之间

相关系数、分类概率等建立了加权朴素贝叶斯分类模型。Wu Jia等<sup>[6]</sup>提出了一种自适应属性加权的朴素贝叶斯算法(AISWNB),该算法通过使用人工免疫系统里的免疫理论来搜索最优权重值,并能够自我调整权重值,从而得到更精确的条件概率。Lee C H等<sup>[7]</sup>提出了一种梯度下降的特征值加权的朴素贝叶斯分类学习方法(VWNB),该算法通过梯度下降法计算特征值的最优权重,为每一个特征值分配不同的权重。Tütüncü G Y等<sup>[8]</sup>提出一种聚合的模糊朴素贝叶斯分类。从结构扩展的角度,Kononenko<sup>[9]</sup>于1991年提出的半朴素贝叶斯分类模型就是一种经典的通过结构扩

收稿日期:2016-02-26

修回日期:2016-06-15

网络出版时间:2016-11-22

基金项目:广西壮族自治区自然科学基金(2013GXNSFC019330);广西壮族自治区高校科研资助项目(2013YB086)

作者简介:谢小军(1990-),男,硕士研究生,研究方向为数值计算与软件应用;陈光喜,教授,研究方向为可信计算、图像处理。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20161122.1228.038.html>

展改进朴素贝叶斯算法,该算法将独立性放宽到了属性的子集之间,从而有效减少了属性的独立性假设对分类性能的不良影响。Friedman<sup>[10]</sup>于1997年提出树增强朴素贝叶斯分类模型(TAN),它要求属性节点除了类节点可以作为父节点外,至多只能拥有一个其他的非类属性作为其父节点,故保留了其结构的特点,并放松了独立性假设,从而使属性之间有着简单的依赖关系。Webb等<sup>[11]</sup>提出了一种平均单依赖估计(AODE)的方法来削弱属性独立性假设,该方法为分类器的所有属性平均分配同一个约束类,弱化了独立性假设,扩充了算法结构,而且大多数情况下具有较好的综合性能。但是存在如下缺陷:

(1)在AODE模型中,所有的结构扩展的朴素贝叶斯分类模型中,所有的属性节点对分类的影响是相同的,这往往并不成立,使用不同的属性节点为父节点的扩展朴素贝叶斯分类模型在分类测试中的影响程度应该是不同的。

(2)AODE算法属于组合学习的分类算法,在进行训练时需要训练出多个模型,然后将训练得到的多个模型对测试实例进行分类。故该算法在时间和简洁性方面要逊色于单个模型分类算法。

针对问题(1),文献[12]提出一种加权平均的单依赖估计模型(WAODE),并设计了四种加权的方法;针对问题(2),文中提出一种多属性联合的朴素贝叶斯分类算法,该算法也是从结构上扩充了朴素贝叶斯分类算法,削弱了条件独立性假设,并且结构扩展更加灵活,最终只要学习一个模型进行分类测试。

基于粗糙集理论<sup>[13-14]</sup>,通过计算条件属性组合相对于决策属性的依赖度大小,给出了一种基于属性最大依赖度的联合属性组的方法。通过实验证明,该方法在一定程度上能够有效提高分类效果。

## 1 朴素贝叶斯分类模型研究

朴素贝叶斯分类模型是贝叶斯分类模型中结构最为简单的模型。由一个父节点和多个子节点构成的树状结构,如图1所示。

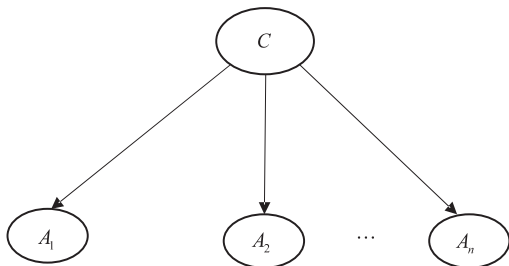


图1 朴素贝叶斯分类模型

它假设在类属性变量已知时,各条件属性变量是相互独立的。然而现实生活中,此假设在很多情况下

并不成立,即当属性之间有很强的依赖时,这个假设会对朴素贝叶斯分类模型的分类准确率产生负面影响,但另一方面也大大简化了贝叶斯分类模型构建的复杂性,故朴素贝叶斯分类模型具有简单和高效等特点,从而得到了广泛应用<sup>[15-18]</sup>。

假定输入训练数据集  $D = \{X_1, X_2, \dots, X_p\}$ , 其中  $A_1, A_2, \dots, A_n$  表示  $n$  个属性,  $C$  表示  $m$  个类别  $C_1, C_2, \dots, C_m$ 。待分类样本  $X = \langle x_1, x_2, \dots, x_n \rangle$ , 其中  $x_i$  为属性  $A_i$  的值,则  $X$  属于类别  $C_k$  的概率为:

$$P(C_k | X) = \frac{P(X | C_k) P(C_k)}{P(X)} \quad (1)$$

由于计算  $P(X | C_k)$  过于复杂,所以假定  $n$  个属性变量是相互独立的,得:

$$P(X | C_k) = \prod_{i=1}^n P(x_i | C_k) \quad (2)$$

由于  $P(X)$  为常数,因此只要最大化  $P(X | C_k) P(C_k)$ ,则根据极大后验假设得朴素贝叶斯分类模型为:

$$C(X) = \operatorname{argmax}_{C_k \in C} P(C_k) \prod_{i=1}^n P(x_i | C_k) \quad (3)$$

其中,概率  $P(C_k)$  和  $P(x_i | C_k)$  的计算公式如下:

$$P(C_k) = \frac{\sum_{j=1}^n \delta(C_j, C_k) + 1}{n + C |} \quad (4)$$

$$P(x_i | C_k) = \frac{\sum_{j=1}^n \delta(x_{ji}, x_i) \delta(C_j, C_k) + 1}{\sum_{j=1}^m \delta(C_j, C_k) + |A_i|} \quad (5)$$

其中,  $n$  为训练实例个数;  $C_j$  为实例  $j$  的类标记;  $x_{ji}$  表示第  $j$  个训练实例的第  $i$  个属性的值,其中:

$$\delta(C_j, C_k) = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases} \quad (6)$$

整个朴素贝叶斯分类模型的实现主要分三个步骤:

(1)数据预处理。此过程的目的是为NBC的分类做铺垫,将获取的数据全部转化为数值型,并将缺失数据进行补齐,若是连续数据则需要对其离散化。最后将数据分为两部分:训练数据与测试数据。

(2)构造分类模型。此过程的任务就是学习分类模型,统计各类别在训练实例中出现频率和条件属性变量出现在各类别的条件概率值。其输入是训练样本,输出是分类模型。

(3)测试分类模型。此过程的任务是利用分类模型对测试数据集进行分类,其输入是分类模型和测试数据集,输出是测试数据集的分类结果。

## 2 结构扩展的朴素贝叶斯分类模型

结构扩展的朴素贝叶斯分类模型(Augmented Naive Bayesian, ANB),是基于朴素贝叶斯分类模型中的

全部属性变量都是类属性的马尔可夫覆盖,对最初的朴素贝叶斯分类模型的结构进行扩展。基本原理:属性变量之间的关联关系通过有限的有向边表示,从而从结构上扩展了朴素贝叶斯分类模型的结构。该思想放松了朴素贝叶斯分类模型的条件独立性假设,并且扩展了朴素贝叶斯分类模型的网络结构。ANB 的关键是如何设计一个高效的扩展算法。关键点是要无约束地构造出所有属性节点的父节点,必须要学习由条件属性变量  $A_1, A_2, \dots, A_n$  组成的结构。然而,学习无限制的贝叶斯网络结构已被证明是一个 NP-难问题。于是,比较现实的方法是学习有限制条件贝叶斯网络结构。为此,广大研究者提出许多经典的模型和算法。这里主要研究分析了 AODE。基本思想是除类属性节点外为所有的其他属性平均分配同一个约束类。图 2 给出了一个 AODE 结构的例子。

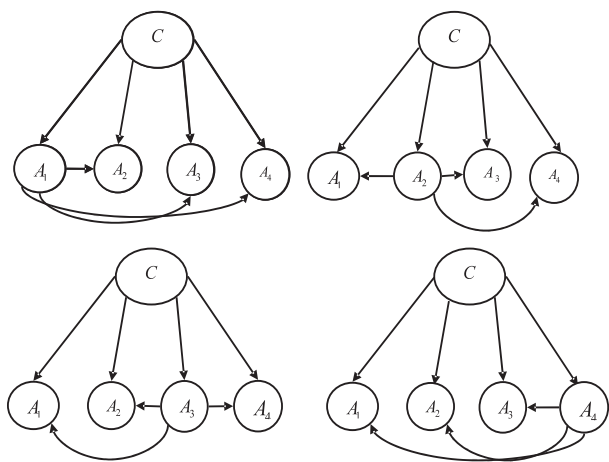


图 2 AODE 结构实例

对于测试实例  $X = \langle x_1, x_2, \dots, x_n \rangle$ , ANB 使用式 (7) 进行分类测试:

$$C(X) = \operatorname{argmax}_{C_k \in C} P(x_1, x_2, \dots, x_n | C_k) P(C_k) \quad (7)$$

应用乘法规则有:

$$C(X) = \operatorname{argmax}_{C_k \in C} P(C_k) \prod_{i=1}^n P(x_i | x_{i+1}, \dots, x_n, C_k) \quad (8)$$

应用马尔可夫原理可得到:

$$C(X) = \operatorname{argmax}_{C_k \in C} P(C_k) \prod_{i=1}^n P(x_i | \Pi_{x_i}, C_k) \quad (9)$$

ANB 的构造步骤如下:

算法: ANB( $D, X$ )。

输入: 训练实例集  $D$  和测试实例  $X$ ;

输出:  $C(X)$ 。

Step1: 对于训练实例集  $D$ , 应用结构学习算法找到所有属性节点的父节点;

Step2: 根据式 (9) 预测测试实例  $X$  的类标记  $C(X)$ ;

Step3: 返回  $C(X)$ 。

### 3 基于多属性联合的朴素贝叶斯分类算法

#### 3.1 粗糙集相关概念

(1) 在信息系统  $S = (U, Q, D, V, f)$  中, 任何属性  $P \subseteq Q$ , 知识  $P$  的不可分辨关系定义为:

$$\operatorname{IND}(P) = \{(x, y) | f(x, a) = f(y, a), a \in P\} \quad (10)$$

其中,  $U$  为论域。

记  $I$  为  $U$  中一组等价关系。对于  $X \subseteq U$ , 集合  $X$  关于  $I$  的下近似是根据已有知识断定, 那些一定是属于  $X$  的对象所组成的最大集合, 有时也称为  $X$  的正区 (positive region), 记做  $\operatorname{POS}(X)$  [16-17]:

$$I_*(x) = \{x \in U | I(x) \subseteq X\} \quad (11)$$

集合  $X$  关于  $I$  的上近似 (Upper approximation) 是指可能属于  $X$  的一些元素所组成的最小集合:

$$I_*(x) = \{x \in U | I(x) \cap X \neq \emptyset\} \quad (12)$$

(2) 属性依赖度: 对于任意属性  $q_i \in Q$ , 类属性变量集合  $D$  与条件属性变量  $q_i$  的重要性是由  $D$  集合与  $q_i$  之间的相互依赖程度所决定。 $D$  集合与  $q_i$  之间的相互依赖程度定义为:

$$\gamma_{q_i}(D) = \frac{|\operatorname{POS}_{q_i}(D)|}{|U|} \quad (13)$$

#### 3.2 基于粗糙集属性依赖度的多属性联合

文中基于结构扩展的思想, 其中半朴素贝叶斯分类模型主要考虑如何有效而快速构成“组合属性”, 当目标数据集过于庞大, 或者数据集中的属性太多, 那么进行属性组合的时间将呈指数级增长, 对运行环境有一定的要求, 否则可能会造成系统的崩溃, 因此半朴素贝叶斯分类模型在使用上有一定的局限性, 最好是对规模较小的数据集; 而 AODE 算法则需要构造多个模型进行学习训练, 故在时间上和简洁性上要逊色于单个模型分类算法。总结以上两种算法的不足, 文中基于粗糙集中属性依赖度的概念, 提出了一种能够快速对属性进行组合的方法, 实际上也是从结构上扩展了朴素贝叶斯算法, 并且最终只要通过学习一个模型对测试实例进行预测。

对于信息系统  $S = (U, Q, D, V, f)$ ,  $Q = \{q_1, q_2, \dots, q_n\}$ ,  $n$  为属性个数,  $k$  为选择进行联合的属性个数,  $1 \leq k \leq n$ , 记  $A$  为经过联合后得到的新属性组集合。当  $k = 1$  时即为朴素贝叶斯分类模型。文中主要是研究当  $k = 2, 3$  时的属性联合算法, 即基于双属性联合与三属性联合的朴素贝叶斯分类算法。在选择联合属性时, 基于联合条件属性相对于决策属性依赖度的大小, 选择最大相对依赖度作为联合条件属性。这是有意义的。因为属性的相对依赖度的大小反映了联合条件属性相对于决策属性的一致程度, 相对依赖度越大, 在一定程度上说明该联合条件属性相对于决策属性一致性

就越强,也就说明该联合条件属性相对于决策属性越重要。

首先给出  $k = 2$  时属性联合后属性组的值的定义,  $k = 3$  时可以类比得到。

定义 1:对于属性集合  $Q = \{q_1, q_2, \cdots, q_n\}$ , 属性  $q_i$  对应的属性值记为  $V_i = \{v_1, v_2, \cdots, v_{|q_i|}\}$ , 对任意的  $q_{n_0}, q_{m_0} \in Q, n_0 \neq m_0$ , 记  $q_{n_0}, q_{m_0}$  联合后的属性为  $a_{n_0}$ ,  $1 \leq n_0 \leq n$ , 则联合后新的属性  $a_{n_0}$  的属性值为  $V_{n_0} \times V_{m_0}$ , 记为  $V_{n_0}$ :

$$V_{n_0} \times V_{m_0} = \{(v_i, v_j) \mid v_i \in V_{n_0}, v_j \in V_{m_0}\}$$

由式(2)可知:

$$P(V_1, V_2, \cdots, V_n / C_i) = \prod_{k=1}^n P(V_k / C_i) = \prod_{k=1}^n P((V_k, V_j) / C_i), (1 \leq j \leq n, j \neq k)$$

其中

$$P((V_k, V_j) / C_i) = \frac{N_{C_i, V_i} + 1}{N_{C_i} + |V_k|} \tag{14}$$

下面给出  $k = 2$  时属性联合实现的过程。

算法 1:双属性联合的朴素贝叶斯分类算法(Two Properties Combined Naive Bayesian Classification, Two-PCNBC)。

输入:决策表  $S = (U, Q, D, V)$ ;

输出:新的属性组集合  $A = \{a_1, a_2, \cdots, a_n\}$  以及对应的属性值集合  $V = \{V_1, V_2, \cdots, V_n\}$ 。

Step1:令  $j = 1, q_j \in Q, A = \emptyset$ ;

Step2:对  $\forall q_i \in Q, i = 1, 2, \cdots, n$ , 利用式(13) 计算所有的  $\gamma_{(q_i, q_j)}(D)$ , 选择满足  $\gamma_{a_j}(D) = \max\{\gamma_{(q_i, q_j)}(D)\}$  的联合属性组合  $a_j = (q_j, q_i)$  加入到  $A$ , 即令  $A = A \cup \{a_j\}$ ;

Step3:令  $j = j + 1$ ;

Step4:如果  $j > n$ , 则终止, 否则转 Step2。

通过上述过程最终得到新的属性组集合  $A = \{a_1, a_2, \cdots, a_n\}, 1 \leq n' \leq n$ , 可以发现, 此方法对两个属性进行联合后的属性个数最多为  $n$ , 对  $\forall a_i$  都有  $a_i = (q_i, q_0)$ 。其中  $q_0 \in Q$  时, 即为 AODE 算法, 此时的  $q_0$  即为所有条件属性除去决策属性另一个父节点, 即约束类。

当  $k = 3$  时, 属性联合的实现过程如下:

算法 2:三个属性联合的朴素贝叶斯分类算法(Three Properties Combined Naive Bayesian Classification, Three-PCNBC)。

输入:决策表  $S = (U, Q, D, V)$ ;

输出:新的属性组集合  $A = \{a_1, a_2, \cdots, a_n\}$  以及对应的属性值集合  $V = \{V_1, V_2, \cdots, V_n\}$ 。

Step1:令  $j = 1, j' = 1, q_j \in Q, A = \emptyset$ ;

Step2:对  $\forall q_i, q_t \in Q, i = j + 1, \cdots, n, t = 1, 2, \cdots, n$ ,

利用式(13) 计算  $\gamma_{(q_j, q_i, q_t)}(D)$ , 选择满足  $\gamma_{(q_j, q_i, q_t)}(D) = \max\{\gamma_{(q_j, q_i, q_t)}(D)\}$  的联合属性组合  $(q_j, q_i, q_{t_0})$  已在  $A$  中存在, 则直接转 Step3, 否则将  $a_j$  加入到新的属性集合中, 令  $A = A \cup \{a_i\}, j' = j' + 1$ ;

Step3:令  $i = i + 1$ ;

Step4:如果  $i < n$ , 转 Step2, 否则转 Step5;

Step5:令  $j = j + 1$ ;

Step6:如果  $j \geq n - 1$ , 则终止, 否则转向 Step2。

通过上述过程, 三个属性联合后得到新的属性组合记为  $A = \{a_1, a_2, \cdots, a_n\}, 1 \leq n' \leq C_n^2$ , 故联合后最多有  $n * (n - 1) / 2$  个属性。

下面给出基于粗糙集属性依赖度的多属性联合朴素贝叶斯算法的步骤:

Step1:数据预处理。将数据全部转化为数值型, 并将缺失数据进行补齐, 若是连续数据则将数据进行离散化。

Step2:确定  $k$  值, 若  $k = 2$  则使用算法 1 进行条件属性联合; 若  $k = 3$  则使用算法 2 进行属性联合。通过联合后得到新信息系统  $S' = (U', A, D, V', f)$ 。

Step3:将得到的新信息系统  $S'$  中数据集分成训练样本和待分类样本, 对训练样本的每一个样本对象  $X$  进行遍历, 利用式(14) 计算所有的先验概率值, 即  $P(x_i / C_k)$  以及在类别  $C_i$  下各个条件属性  $x_i$  的取值概率  $P(x_i / C_k)$ 。

Step4:根据式(1) 计算待分类样本中待分类对象  $X$  属于其他各类别的后验概率, 并通过式(3) 得出最终分类结果。

## 4 实验仿真及分析

为了验证算法的可行性和有效性, 下面将比较传统的朴素贝叶斯分类算法(NB)、树增强型朴素贝叶斯分类模型(TAN)、加权的平均单依赖估计(WAODE)以及文中提出的属性联合算法(Two-PCNBC 和 Three-PCNBC)在分类精度之间的差别。选用 UCI 机器学习库中的 8 个数据集进行实验仿真测试, 数据集见表 1, 所有数据都经过预处理。

表 1 预处理后的数据集

数据集	属性数	类别	实例数
Tic-tac-toe	9	2	958
Iris	4	3	150
Balance scale	4	3	625
Wine	13	3	178
Diabetes	8	2	768
Breast-cancer	10	2	286
Vote	16	2	435
Zoo	16	7	101



首先对数据集的顺序进行随机打乱,采用分割数据集的方法进行测试,其中训练集为 70%,测试集为 30%,以十折交叉验证的分类精度来评价这 7 种算法的分类效果。实验仿真结果见表 2。

表 2 实验结果

数据集	NB	TAN	WAODE	Two-PCNB	Three-PCNB
Tic-tac-toe	71.498	72.31	69.35	72.25	98.96
Iris	92.26	94.07	95.33	97.11	96.40
Balance scale	91.44	85.97	89.44	93.21	91.46
Wine	96.78	94.30	95.62	96.50	89.13
Diabetes	75.68	76.04	76.14	80.02	71.37
Breast-cancer	72.94	69.53	72.01	76.95	86.43
Vote	90.21	94.43	94.53	90.12	96.12
Zoo	93.34	96.63	97.13	96.53	97.84
Average	85.52	85.41	86.19	87.84	90.96

从表 2 可以看出,文中提出的基于粗糙集属性依赖度的属性联合朴素贝叶斯算法是可行的,可以从数据中挖掘更多的信息,特别是相对于数据集 Tic-tac-toe 和 Iris 中属性之间关联性比较强的数据集,分类效果提高尤为明显。从算法 Two-PCNB 和 Three-PCNB 的比较发现,并不是联合的属性个数越多就越好,有的数据集选择两个属性进行联合比选择三个属性进行联合的效果反而要更好些。而对数据集 Wine,发现对属性联合时反而降低了它的分类精度,这是因为在离散化时得到的单个属性的属性值个数比较多,导致属性联合时对应的属性值域大大增加,从而增加了属性的复杂度,在一定程度上对分类效果产生了负面影响。所以文中算法更适合属性个数相对较小、属性对应的属性值域比较集中的数据集。

总体上,通过图 3 可以看出,文中算法的改进效果还是相当不错的。

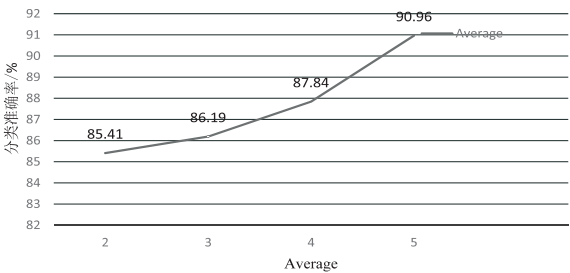


图 3 8 个数据集的平均分类准确率比较

5 结束语

文中基于粗糙集理论,结合粗糙集中属性相对依赖度的概念,提出了一种基于最大相对属性依赖度的属性联合朴素贝叶斯分类算法。通过平均选择适当的属性个数进行联合,在 8 个数据集进行了仿真测试实验,并与最新提出的相关算法进行比较,表明了该方法的有效性和可行性。如何改进在进行属性联合时属性

值增加的问题,以及将算法应用到现实中一些特定的数据都将作为下一步的研究工作。

参考文献:

[1] Jiang L,Zhang H,Cai Z. A novel Bayes model:hidden Naive Bayes[J]. IEEE Transactions on Knowledge & Data Engineering,2008,21(10):1361-1371.

[2] 邓维斌,王国胤,王 燕. 基于 Rough Set 的加权朴素贝叶斯分类算法[J]. 计算机科学,2007,34(2):204-206.

[3] 张明卫,王 波,张 斌,等. 基于相关系数的加权朴素贝叶斯分类算法[J]. 东北大学学报:自然科学版,2008,29(7):952-955.

[4] 郑 默,刘琮荪. 一种属性相关性的加权贝叶斯分类算法研究[J]. 微型机与应用,2011,30(7):96-98.

[5] 张步良. 基于分类概率加权的朴素贝叶斯分类方法[J]. 重庆理工大学学报:自然科学版,2012,26(7):81-83.

[6] Wu J,Pan S,Zhu X,et al. Self-adaptive attribute weighting for Naive Bayes classification[J]. Expert Systems with Applications,2015,42(3):1487-1502.

[7] Lee C H. A gradient approach for value weighted classification learning in naive Bayes [J]. Knowledge - Based Systems, 2015,85(C):71-79.

[8] Tütüncü G Y,Kayaalp N. An aggregated fuzzy naive Bayes data classifier[J]. Journal of Computational & Applied Mathematics,2015,286(C):17-27.

[9] Kononenko I. Semi-naive Bayesian classifier[C]//EWSL-91. Berlin:Springer,1991:206-219.

[10] Friedman N,Geiger D,Goldszmidt M. Bayesian network classifiers[J]. Machine Learning,1997,29(2):131-163.

[11] Webb G I,Boughton J R,Wang Z. Not so naive Bayes:aggregating one-dependence estimators [J]. Machine Learning, 2005,58(1):5-24.

[12] Jiang Liangxiao,Zhang H,Cai Zhihua,et al. Weighted average of one-dependence estimators[J]. Journal of Experimental & Theoretical Artificial Intelligence,2012,24(2):219-230.

[13] Grzymała-Busse J W,Pawlak Z,Słowiński R,et al. Rough set [J]. Communications of the ACM,1995,38(11):800-805.

[14] 王国胤,姚一豫,于 洪. 粗糙集理论与应用研究综述[J]. 计算机学报,2009,32(7):1229-1246.

[15] 马小龙. 一种改进的贝叶斯算法在垃圾邮件过滤中的研究[J]. 计算机应用研究,2012,29(3):1091-1094.

[16] 张 轮,杨文臣,刘 拓,等. 基于朴素贝叶斯分类的高速公路交通事件检测[J]. 同济大学学报:自然科学版,2014,42(4):558-563.

[17] 朱克楠,尹宝林,冒亚明,等. 基于有效窗口和朴素贝叶斯的恶意代码分类[J]. 计算机研究与发展,2014,51(2):373-381.

[18] 苏 中,张宏江,马少平. 基于贝叶斯分类器的图像检索相关反馈算法[J]. 软件学报,2002,13(10):2001-2006.