

改进的贝叶斯分类方法在 电信客户流失中的研究与应用

杨 婷,滕少华

(广东工业大学 计算机学院,广东 广州 510006)

摘要: 随着电信市场竞争日益加剧,客户流失成为运营商关注的焦点问题之一. 针对电信数据量庞大且有时间序列的特点,提出一种改进的贝叶斯分类方法来研究电信客户流失问题,通过对不同属性加权改进了朴素贝叶斯分类器默认每个属性对分类结果影响相同的假设,进一步探讨了应用增量学习方法来应对不断增加的数据,以改善分类器的准确率. 实验结果表明,本文的方法有较高的准确率.

关键词: 贝叶斯分类; 电信数据; 增量学习; 客户流失; 预测

中图分类号: TP393

文献标志码: A

文章编号: 1007-7162(2015)03-0067-06

Research and Application of Improved Bayes Algorithm for the Telecommunication Customer Churn

Yang Ting, Teng Shao-hua

(School of Computers, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: With the increasing competition of telecom market, customer churn became one of the focused problems. Because the telecommunication data is huge and has the characteristic of time series, this paper proposes an improved Bayesian classification to study the customer churn problem. The improved Bayesian classification model is designed to make up for the shortcomings of the former Bayes which assumed that each attribute has the same effect on the classification results. Furthermore, by coping with the increasing data, this paper explores the incremental learning method to improve the accuracy of the classifier. The experimental results show that the proposed method has higher accuracy.

Key words: Bayesian classification; telecommunication data; incremental learning method; customer churn; prediction

随着中国电信业的飞速发展和市场竞争加剧,电信产品及服务模式愈加丰富,用户对产品与服务模式的选择更多,客户保持变得越来越困难.此外,获得一个新客户比保持原有客户要有更大的开销且不一定会产生更高的效益.因此保持原有客户,防止客户流失越显重要.

客户流失预测就是通过对客户基本信息与历史行为等数据进行深入分析,提炼出已流失客户在流

失前具有的特征或行为,建立客户流失预测模型.通过模型来预测企业近期内将可能流失的客户,反馈一个流失预测名单给企业,为企业有针对性开展业务来预防或减少客户流失提供参考.

本文基于某电信企业的客户实际通信数据进行深入分析,利用朴素贝叶斯分类器具有计算量少、准确度高、以概率论为理论基础的特点^[1],结合增量学习方法,构建一个改进的贝叶斯分类器,研究客户流

收稿日期: 2014-03-13

基金项目: 教育部重点实验室基金资助项目(110411);广东省自然科学基金资助项目(10451009001004804,9151009001000007);广东省科技计划项目(2012B091000173,2013B090200017,2013B010401029,2013B010401034);广州市科技计划项目(2012J5100054,2013J4500028)

作者简介: 杨 婷(1990-),女,硕士研究生,主要研究方向为数据挖掘.

失问题,探讨并预测客户流失情况.

1 贝叶斯分类器

1.1 贝叶斯分类原理

假设数据集为 D , 其属性集为: $U = \{A_1, A_2, \dots, A_n, C\}$, 其中 $A_1, A_2, \dots, A_n$ 是样本的属性变量, C 是有 m 个值 $C_1, C_2, \dots, C_m$ 的类标号属性变量. 数据集 D 中的每个样本 X 可以表示为 $X = \{x_1, x_2, \dots, x_n, C_j\}$, $x_1, x_2, \dots, x_n$ 分别是 $A_1, A_2, \dots, A_n$ 的 n 个取值, 而 C_j 是所属类的类标号值.

给定一个类标号未知的样本 X , 朴素贝叶斯分类将预测 X 属于具有最大后验概率 $P(C_j|X)$ 的类, 即^[4]:

$$P(C_j|X) = \arg \max P(C_j|X) = \arg \max \frac{P(C_j)P(X|C_j)}{P(X)}, j=1, 2, \dots, m. \quad (1)$$

分母 $P(X)$ 为常数, 根据贝叶斯公式和朴素贝叶斯类条件独立性, 得到朴素贝叶斯分类模型^[4]:

$$V_{BN}(X) = \arg \max P(C_j) \prod_{k=1}^n P(x_k|C_j). \quad (2)$$

1.2 加权朴素贝叶斯

由于各条件属性对分类结果的重要性是不同的, 因此, 在实际应用时需要改正朴素贝叶斯条件独立性的假设, 将其扩展为加权朴素贝叶斯. 对朴素贝叶斯分类法加权, 已有许多研究, Harry 和 Sheng^[5] 提出了根据属性的重要性给不同属性赋权值的加权朴素贝叶斯分类器; 贾娴等^[6] 提出了通过提取特征冗余度判别函数来为不同属性赋权值的方法; 杨敏等^[7] 提出了一种基于属性约简的偏最小二乘回归加权朴素贝叶斯分类算法, 对不同条件属性赋权值; Dong 等^[8] 给出了基于粗糙集的属性加权方法, 文献 [7-8] 也改进了朴素贝叶斯分类算法. 上述方法结合具体应用, 扩展了贝叶斯的应用范围. 有鉴于此, 本文通过信息增益来为分类属性赋权值.

信息增益 (Information Gain, IG) 是信息论中一个重要概念, 在 IG 中, 重要性的度量标准就是看属性能够为分类系统带来的信息量的大小, 值越大, 说明其包含的信息量也越大. 属性 A 给分类 C 带来的信息增益为

$$\text{Gain}(A) = H(C) - H(C|A), \quad (3)$$

其中,

$$H(C) = - \sum_{j=1}^m P(C_j) \log P(C_j), \quad (4)$$

$$H(C|A) = - \sum_{j=1}^m \sum_{i=1}^n P(C_j|x_i) \log(C_j|x_i). \quad (5)$$

式 (4) 中, $H(C)$ 是类 C 的熵, m 是类 C 取值个数, $P(C_j)$ 是样本中属于类 C_j 的概率.

式 (5) 中 $H(C|A)$ 是属性 A 确定时的条件熵, n 为属性 A 取值个数, $P(C_j|x_i)$ 是属性 A 取值 x_i 的条件下分类为 C_j 的概率.

$\text{Gain}(A)$ 就是从特征 A 上获得该划分的 IG, 它越大, 表示该特征在给定记录集中具有的区分度越大, 对分类越重要.

通过计算每个特征属性的信息增益 $\text{Gain}(A_k)$, 得到每个特征属性的权重为

$$W_k = \text{Gain}(A_k) / \sum_{i=1}^n \text{Gain}(A_k), \quad (6)$$

则加权后贝叶斯分类模型变为

$$V_{WBN}(X) = \arg \max P(C_j) \prod_{k=1}^n P(x_k|C_j)^{w_k}. \quad (7)$$

2 电信数据预处理

在数据分析之前, 需要将原始数据集进行预处理, 包括数据清洗、数据集成、数据转换和数据归约, 以删除噪声数据, 解决数据不一致的现象等, 以下是本文针对电信数据的预处理过程.

2.1 客户样本数据整理

客户样本数据集中包含客户的类型、通话级别、月消费额等属性, 如果数据特征的有效值少于总记录数据的 1/5 时, 则删除此类特征; 如果某记录中存在大量空缺值, 而这些空缺值难以以正常方法给予补充, 则删除此类记录.

2.2 客户数据的集成与概化

由于数据集中缺少一些直接体现客户价值和客户流失倾向的特征, 本文从时间维度对数据进行了概化, 构造了以下特征:

(1) 季度总费用: 连续的 12 个月, 分为 4 个季度, 计算每个季度的汇总费用.

(2) 月消费比率: 指下一个月与上一个月的总费用比值. 根据这一原理可构造 11 个月消费比率特征. 用符号可表示为:

$$\text{rate}_i = \text{total_fee}_{i+1} / \text{total_fee}_i, (1 \leq i \leq 11).$$

(3) 根据客户在一个年度内的消费情况可构造未消费月份数, 此特征可反映样本客户消费情况及流失情况. 对于原始数据集中连续 10、11 或 12 个月都没有消费的客户数据, 当作该客户已流失, 在本文中不考虑他们, 这部分数据予以删除.

2.3 数据属性选择

为了便于分析, 通过计算各个属性与类属性的

互信息并结合专家意见删除不相关或冗余的属性以减小数据集,减少样本的维数,降低时间和空间复杂度,简化学习模型.通过数据预处理后,得到构建朴素贝叶斯的特征属性,如表1所示.

表1 特征属性表

Tab.1 Characteristics Attribute Table

连续属性	离散属性
第1季度总消费(A_1)	1年内未消费月数(A_7)
第2季度总消费(A_2)	客户类型(A_8)
第3季度总消费(A_3)	通话级别(A_9)
第4季度总消费(A_4)	类标签(C)
本月总消费(A_5)	
本月/上月总消费(A_6)	

3 改进的贝叶斯分类器

本文首先应用贝叶斯算法计算先验概率与类条件概率;再通过信息增益确定各个特征属性的权重;最后,为避免初始样本覆盖范围不全并考虑到未来的数据随时间不断变化的情况,通过批量加入新数据,采用增量学习来修正分类器的各个参数,使分类器不断改进,具体描述如下.

3.1 分类器体系结构

改进的贝叶斯分类器由数据收集、数据预处理、贝叶斯建模、加权学习、增量学习等过程组成,如图1所示.

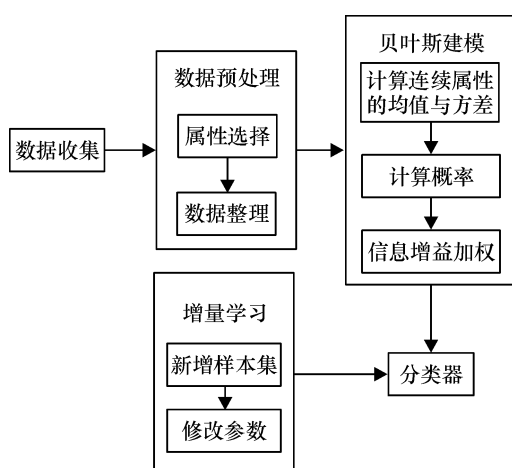


图1 分类器体系结构

Fig.1 Structure of classifier

数据预处理参见第2部分,其他部分描述如下.

3.2 贝叶斯建模

(1) 先验概率.

一般地,对于绝大多数电信企业来说,正常客户

与流失客户的分布是不平衡的,每月流失的客户总是少数,而正常客户占绝大多数.通过数据统计分析发现,正常客户与流失客户的比例大概为4:1,从而得出类先验概率.假设类属性描述为 $C=\{0,1\}$,其中, $C=0$ 表示流失客户, $C=1$ 表示正常客户,则流失客户的先验概率 $P(C=0)=0.2$,正常客户的先验概率 $P(C=1)=0.8$.

(2) 类条件概率.

针对电信数据集,从用户消费行为中可以获知该用户是否流失,如连续若干个月没有消费的用户,很可能已流失;月消费从多到少的用户也可能会流失;而月消费从少到多的用户,则其流失的可能性比较小.由于用户每天的通话情况变化较大、每月电话的使用也会有较大的波动,本文采用按季度汇总数据,获取数据的趋势来进行分析.针对样本中连续属性,通过统计分析发现,其分布类似于正态分布,所以用正态分布来估计其类条件概率.该分布有2个参数,均值 μ 和方差 σ^2 ,类别 C_j 下连续属性 x_k 的类条件概率等于^[4]

$$P(x_k | C_j) = \frac{1}{\sqrt{2\pi}\sigma_{C_j}} \exp\left(-\frac{(x_k - \mu_{C_j})^2}{2\sigma_{C_j}^2}\right) \quad (8)$$

其中, μ_{C_j} 和 $\sigma_{C_j}^2$ 分别是数据集中属于 C_j 类的样本属性 A_k 的平均值和方差.表2中计算了连续属性对应的均值和方差.

表2 连续属性的均值和方差

Tab.2 The mean & variance of continuous attributes

连续属性	均值/方差	$C=0$	$C=1$
第1季度总消费 A_1	均值	329.60	369.62
	方差	322 506.13	282 604.93
第2季度总消费 A_2	均值	252.84	382.44
	方差	150 462.61	303 284.93
第3季度总消费 A_3	均值	162.996	373.220
	方差	107 605.100	229 629.59
第4季度总消费 A_4	均值	73.081	319.848
	方差	76 413.000	135 251.720
本月月总费用 A_5	均值	20.33	115.68
	方差	38 572.02	20 870.87
本月与上月总费用比值 A_6	均值	0.158 9	1.155 0
	方差	0.238 4	0.703 0

对于离散属性,在后验概率计算过程中,当有一个属性的条件概率等于0,则整个类的后验概率都等于0,简单地使用记录比例来估计类条件概率的方法显得太脆弱,为避免这一问题,条件概率为0时一般采用Laplace估计来解决这个问题.类别 C_j 下

离散属性 x_k 的类条件概率计算公式如式(9), 最终得出的类条件概率如表 3、4、5 所示.

$$P(x_k | C_j) = \frac{P(A_k = x_k \wedge C = C_j) + 1}{P(C = C_j) + v}, \quad (9)$$

其中, v 是 A_k 取值的个数.

表 3 未消费月数的类条件概率表

Tab. 3 Class conditional probability of no spending months

A_7	概率	
	$P(A_7 C = 0)$	$P(A_7 C = 1)$
$A_7 = 0$	0.042 148	0.787 899
$0 < A_7 < 4$	0.081 917	0.007 138
$4 \leq A_7 \leq 6$	0.077 158	0.003 739
$A_7 > 6$	0.000 33	0.000 33

表 4 客户类型的类条件概率表

Tab. 4 Class conditional probability of customer type

A_8	概率	
	$P(A_8 C = 0)$	$P(A_8 C = 1)$
CENTREX 单位电话(28 元) (CENT)	0.056 000	0.252 400
普通单机单位电话(PSTN)	0.128 400	0.386 400
普通单机省会城市住宅(PSTN)	0.015 600	0.161 200

表 5 通话级别的类条件概率表

Tab. 5 Class conditional probability of call-level

A_9	概率	
	$P(A_9 C = 0)$	$P(A_9 C = 1)$
市话	0.051 200	0.179 600
市话 + 国际长途(含国际 IP)	0.032 000	0.154 000
市话 + 国内长途(含国内 IP)	0.098 000	0.455 600
市话 + 国际 IP 长途	0.000 400	0.010 400
空缺值	0.018 400	0.000 400

(3) 信息增益加权.

利用 1.2 小节中式(6), 通过计算每个特征属性的信息增益 $\text{Gain}(A_k)$, 得到每个特征属性的权重 W_k , 如表 6 所示.

表 6 各个属性的权值

Tab. 6 Value of the property rights of attributes

属性	权重
A_1	0.101
A_2	0.091
A_3	0.088
A_4	0.116
A_5	0.212
A_6	0.109
A_7	0.221
A_8	0.033
A_9	0.029

3.3 加权分类器构造

在构造分类器前, 将收集到的数据分为训练集 D 、候选增量学习样本集 D_1 、 D_2 、 D_3 及测试集 T 5 部分, 由训练集构造分类模型, 最后由测试集检验模型分类的准确度. 分类器的构造过程如下:

Step1: 数据收集. 本文以某电信公司部分客户一年来的业务数据作为实验数据^[9], 该数据集包含了 156 810 条记录, 每条记录 X 都包含客户基本特征和客户行为特征以及一个类标号特征.

Step2: 数据预处理. 通过数据清洗、数据集成、数据归约后, 形成如表 1 所示的可用于构建朴素贝叶斯分类器的特征属性数据. D 、 D_1 、 D_2 、 D_3 , T 所占比重分别为 30%、10%、10%、10%、40%.

Step3: 对训练集 D , 计算类先验概率和类条件概率.

Step4: 对训练集 D , 根据式(6), 计算每个属性的权重.

Step5: 根据各个参数, 构造加权分类器.

Step6: 对测试集 T , 由式(7), 计算每个测试数据的概率, 将每个测试数据 X 分类为概率 P_i 最大的那个类, 获得测试数据集的分类.

Step7: 比较模型分类结果与测试数据的类属性, 计算分类准确率.

3.4 增量学习

对数据分类, 一般的方法是将所有训练数据记录一次读入内存, 但当数据量非常大或者数据是分批获得时, 这些传统的方法就显示出局限性了, 增量式学习是解决此问题的有效途径, 它能充分利用先验信息和样本信息求解此类问题^[9-14]. 但是, 电信数据量庞大, 很多属性的值因增量数据可能会变动且数据都有一定的时效性, 新的数据随时间不断产生, 原训练集的价值可能会越来越小, 导致构造的分类器准确率会有所下降, 所以本文增加一个增量学习模块, 通过加入新的有价值的数据来修改已构造分类器的部分参数, 从而有效提高分类准确率.

(1) 增量学习.

本文增量学习思路是: 针对候选样本集, 用已构建好的加权贝叶斯分类模型对其进行分类, 针对每一个候选样本, 分类结果分 2 种情况. 第 1 种情况, 此样本分类正确, 则对此类样本不作任何处理. 第 2 种情况, 该样本未流失但分类结果显示流失, 这存在两种可能, 一是分类结果出错, 这时需要用增量数据修正分类模型, 可通过模型重构等方法解决; 二是由

于客户各种原因(诸如较长时间停用等)导致无通信数据,这时仍需要通过增量数据修正模型,以提高分类准确率。

(2) 修正分类器参数.

假设训练集为 D , 样本 X' 为增量数据. 当将样本 $X' = \{x_1, x_2, \dots, x_n, C'\}$ 加入到训练集 D 时, 要重新计算类先验概率和类条件概率, 根据 Dirichlet 先验分布特性, 类先验概率和类条件概率修正方法如式(10)、(11) [8].

$$\begin{cases} P(C_j) = \frac{\delta}{1+\delta} P(C_j), (C_j \neq C'), \\ P(A_{ik} | C_j) = \frac{\delta}{1+\delta} P(A_{ik} | C_j) + \frac{1}{1+\delta}, (C_j = C'), \end{cases} \quad (10)$$

其中, δ 是类标签个数与训练集样本数之和, j 是类的个数, $j = 1, 2, \dots, m$.

$$\begin{cases} P(A_{ik} | C_j) = \frac{\zeta}{1+\zeta} P(A_{ik} | C_j), (C'' = C_j \wedge A_{ik} \neq x_i), \\ P(A_{ik} | C_j) = \frac{\zeta}{1+\zeta} P(A_{ik} | C_j), (C'' = C_j \wedge A_{ik} = x_i), \\ P(A_{ik} | C_j) = C'' \neq C_j, \end{cases} \quad (11)$$

其中, ζ 是 A_i 取值个数与训练集中类 C_j 的个数之和.

由于新样本 X' 的加入, 使训练集中加入了新的样本数据, 类先验概率和类条件概率都发生了变化. 通过分析式(10)和(11)可知, X' 的加入, 只改变了与它相关项的概率, 仅需进行局部修改.

(3) 修正权值.

对新增加的候选样本集用 1.2 节信息增益方法, 求各个属性的权值 $W_{k-2}, k = 1, 2, \dots, n$, 然后和原训练集求得的权值通过整合得到新的权值.

$$W_{k-new} = \alpha W_k + \beta W_{k-2}, \quad (12)$$

其中, $\alpha + \beta = 1$, α 和 β 是两个参数, 用来设置训练集和候选样本集的权重.

(4) 改进的贝叶斯算法.

算法描述如下:

Input: 分类器 C , 候选增量样本集 D_1 , 分类错误的样本集 $S = \emptyset$;

Output: 分类器 C_{-new} .

Step1: 对候选样本集 D_1 中每个样本进行分类, 把分类错误的样本保存入 S ;

Step2: while $S \neq \emptyset$, 对 S 中每个样本利用式(10)、(11)重新计算类先验概率和类条件概率;

Step3: 对分类错误的样本集 S , 利用式(6)计算各个属性的权值, 利用式(12)得到修正后的权值;

Step4: 更新分类器 C 的各个参数, 得到新的分类器 C_{-new} .

Step5: 对测试集 T , 由式(7), 计算每个测试数据的概率, 将每个测试数据 X 分类为概率 P_i 最大的那个类, 获得测试数据集的分类.

Step6: 比较模型分类结果与测试数据的类属性, 重新计算分类准确率.

4 实验及结果分析

为了验证算法的有效性, 利用 UCI 机器学习数据库中 Abalone、Car Evaluation、Heart 数据集 [9] 和电信行业的客户数据对算法进行检验, 分别求出对应的朴素贝叶斯分类准确率、加权贝叶斯分类准确率和改进的增量学习贝叶斯分类准确率, 并进行对比, 结果见表 7.

表 7 几种分类器准确率对比

Tab. 7 Comparison of classifier results

数据集	准确率/%		
	朴素贝叶斯	加权贝叶斯	改进贝叶斯
Abalone	74.33	81.08	89.10
Car	87.94	88.34	90.41
Heart	82.03	85.11	90.20
电信数据	83.23	87.67	89.34

从表 7 可知, 加权贝叶斯分类法优于朴素贝叶斯分类器; 而通过增量学习修正贝叶斯分类器参数后的贝叶斯分类法又优于加权贝叶斯分类法. 它进一步提高了分类准确率. 本文中的 $\alpha = 0.65$ 、 $\beta = 0.35$, 它们是由多次实验获得.

针对 Car 和 Heart 数据集, 本文算法实验结果与文献 [6] 采用方法的实验结果进行对比, 结果如表 8 所示.

表 8 与其他算法的准确率比较

Tab. 8 Compared results with other algorithms

数据集	准确率/%	
	本文方法	文献 [6] 方法
Car	90.41	92.13
Heart	90.20	87.67

另外, 本文增量学习过程中得到的实验结果见表 9.

表9 本文增量学习过程中的实验结果

Tab.9 Incremental learning process results

训练集	候选样本集	测试集	准确率/%
D		T	76.23
D	D_1	T	84.12
D	D_1 D_2	T	87.54
D	D_1 D_2 D_3	T	89.34

从图2中可知,随着对候选样本集进行增量学习,分类准确率越来越高。

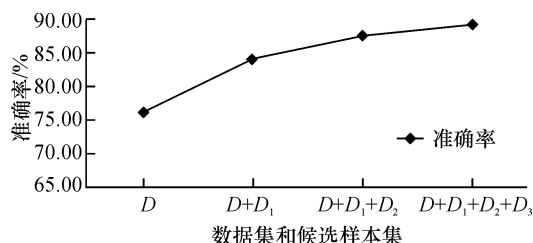


图2 增量学习过程准确率变化图

Fig.2 Accuracy rate trends of incremental learning

5 结束语

以开源数据集及某电信公司部分客户一年来的业务数据作为实验数据^[2],引入按月/季度汇总数据,获取数据的趋势来进行分析,选择合适的特征属性,通过信息增益法对特征属性赋权值,构建一个加权朴素贝叶斯分类器,并对其进行增量学习,提高了预测客户流失的精度。经该电信公司确认,本文的方法能够更好地把握客户流失倾向,及时采取有效的挽留措施,降低客户流失概率,这为运营商提供了有效的决策支持。

参考文献:

- Jiang L X, Zhang H, Cai Z H. A novel bayes model: hidden naïve bayes [J]. IEEE Transactions on Knowledge and Data Engineering (TKDE), 2009, 21(10): 1361-1371.
- 蒋盛益,李霞,郑琪. 数据挖掘原理与实践 [M]. 北京: 电子工业出版社, 2013.
- Harry Z, Sheng S L. Learning weighted naïve bays with accurate ranking [J] // Proceedings of the 4th IEEE International Conference on Data Mining (ICDM 04), Brighton, UK: [s. n.], 2004: 567-570.
- 贾娴,刘培玉,公伟. 基于改进属性加权的朴素贝叶斯入侵取证研究 [J]. 计算机工程与应用, 2013, 49(7): 81-84.
- Jia X, Liu P Y, Gong W. Research of intrusion forensics based on improved attribute weighted naïve Bayes [J]. Computer Engineering and Applications, 2013, 49(7): 81-84.
- 杨敏,贺兴时,刘平丽,等. 基于属性约简的 PLS 加权朴素贝叶斯分类器 [J]. 西安工程大学学报, 2013, 27(1): 118-121.
- Yang M, He X S, Liu P L. Weighted naïve Bayes classifier based on attribute reduction-PLS [J]. Journal of Xi'an Polytechnic University, 2013, 27(1): 118-121.
- Dong W B, Wang G Y, Wang Y. Weighted naïve Bayesian classification algorithm based on rough set [J]. Computer Science, 2007, 34(2): 204-206.
- 宫秀军,孙建平,史忠植. 主动贝叶斯网络分类器 [J]. 计算机研究与发展, 2002, 39(5): 574-579.
- Gong X J, Sun J P, Shi Z Z. An active Bayes network classifier [J]. Computer Research and Development, 2002, 39(5): 574-579.
- Yager R. An extension of the naïve Bayesian classifier [J]. Information Sciences, 2006, 176(5): 577-588.
- 宫秀军,刘少辉,史忠植. 一种增量贝叶斯分类模型 [J]. 计算机学报, 2002, 25(6): 645-650.
- Gong X J, Liu S H, Shi Z Z. An incremental Bayes classification model [J]. Chinese Journal of Computers, 2002, 25(6): 645-650.
- 秦锋,任诗流,程泽凯,等. 基于属性加权的朴素贝叶斯分类算法 [J]. 计算机工程与应用, 2008, 44(6): 107-109.
- Qin F, Ren S L, Cheng Z K, et al. Attribute weighted Naïve Bayes classification [J]. Computer Engineering and Applications, 2008, 44(6): 107-109.
- 陈亮,郑宁,郭艳华,等. 朴素贝叶斯增量学习在病毒上的报分析中的应用 [J]. 计算机应用与软件, 2010, 27(1): 92-95.
- Chen L, Zheng N, Guo Y H, et al. Applying naïve Bayesian incremental learning in virus reporting and analysing [J]. Computer Applications and Software, 2010, 27(1): 92-95.
- 张全新,郑建军,牛振东,等. 贝叶斯分类器集成的增量学习方法 [J]. 北京理工大学学报, 2008, 28(5): 397-400.
- Zhang Q X, Zheng J J, Niu Z D, et al. Increment learning algorithm based on Bayesian classifier integration [J]. Transactions of Beijing Institute of Technology, 2008, 28(5): 397-400.
- 余承依. 基于贝叶斯增量分类的邮件过滤研究 [J]. 科学技术与工程, 2009, 9(9): 2556-2561.
- Yu C Y. Research of mail filtering based on bayesian incremental classification [J]. Science Technology and Engineering, 2009, 9(9): 2556-2561.
- 罗福星. 增量学习朴素贝叶斯中文分类系统的研究 [J]. 长沙: 中南大学软件学院, 2008.
- University of California, Irvine. UCI Machine Learning Repository [DB/OL]. [2014-12-05]. <http://archive.ics.uci.edu/ml/>.