

基于交互信息的数据集特征结构研究^{*}

刘 娟 朱翔鸥 刘文斌

(温州大学 物理与电子信息工程学院 温州 325035)

摘 要 机器学习分类领域提出大量的分类算法,如何为数据集找到合适的分类算法成为研究的重要内容之一.文献[8]提出一种新的数据集离散化方法用来刻画数据集的特征,且在推荐方法方面取得较好的结果.本文在此基础上利用交互信息理论刻画数据集的属性与属性及属性与类标签之间协作关系,提出基于二变量和基于三变量的交互信息特征结构.通过 12 种分类算法在 UCI 数据库中的 98 个数据集上的性能实验,结果表明与文献[8]的方法相比,两种方法都能明显提高推荐方法的精度和命中率,且对于适应性较差的数据集,基于三变量的交互信息方法更为有效.

关键词 分类算法,交互信息,数据集特征结构

中图法分类号 TP 391.4

Research on Dataset Feature Structure Based on Interaction Information

LIU Juan, ZHU Xiang-Ou, LIU Wen-Bing

(College of Physics & Electronic Information Engineering, Wenzhou University, Wenzhou 325035)

ABSTRACT

In machine learning area, classification algorithms are widely studied and a large number of different types of algorithms are proposed. How to select appropriate ones from so many classification algorithms for the datasets becomes a crucial problem. Recently, a new method in reference [8] is proposed to characterize datasets and achieve better results in algorithm recommendation. In this paper, two methods are presented to characterize datasets under the theory of interaction information. The performance of 12 different types of classification algorithms on the 98 UCI datasets illustrates that both two-variable and three-variable interaction information methods can improve the precision and the hit rate of recommended algorithms. Furthermore, the latter performs even better under datasets with poor adaptability.

Key Words Classification Algorithm, Interaction Information, Dataset Feature Structure

^{*} 国家自然科学基金项目(No. 60970065, 61272018)、浙江省自然科学基金项目(No. R1110261)、温州大学研究生创新基金项目(No. 31606036010138)资助

收稿日期: 2012-11-22; 修回日期: 2013-04-23

作者简介: 刘娟(通讯作者),女,1988年生,硕士研究生,主要研究方向为数据分析、模式识别. E-mail: liujuan555ok@163.com. 朱翔鸥,男,1969年生,硕士,教授,主要研究方向为低压电器及智能化、低压电器检测. 刘文斌,男,1969年生,博士,教授,主要研究方向为数据挖掘和智能信息处理.

1 引言

过去几十年,机器学习领域分类的研究者提出大量的分类算法,如基于树的 C4.5、CART 分类算法,基于概率的 Naive Bayes、AODE 分类算法,基于规则的 RIPPER、CN2 和 PART 分类算法,支持向量机(SVM)分类算法,BP 神经网络算法,及近邻法、Boosting 方法和遗传算法等。然而,各种分类算法在不同的数据集上的性能表现往往差异较大,如 Weiss 等发现 BP 神经网络在 Appendicitis 和 Iris 数据集上的精度优于决策树,而在 Breast 和 Thyroid 数据集上则不如决策树^[1]; Shavlik 等在数据集 Soybean, Audiology, Heart Disease、NETtalk-A 与 Chess 上也得到类似的结论^[2]。Brazdil 和 Duin 的研究表明,没有哪一个分类算法在所有数据集上具有一致的性能表现^[3-4]。因此,如何为给定数据集在众多分类算法中快速有效地挑选一个合适的算法就成为机器学习领域的重要研究问题。

除实验方法不同,大量研究表明算法性能的不一致性与数据集本身的结构有密切关系。Gama 和 Brazdil^[5]指出通过研究数据集的特征将有助于挑选合适的算法。他们利用统计和信息理论的方法描述数据集的特征,并用 C4.5 方法产生推荐规则,结果证实统计相似的数据集可推荐相同的分类算法。尽管不能保证推荐到最佳的分类算法,但却缩小了用户的选择范围^[4]。Shawkat 等^[6]用基于规则的方法研究分类问题与分类算法种类的关系。Kalousis 等^[7]首次研究分类算法的相似性和数据集的相似性问题。根据算法在各种数据集上的误差分布及性能等级进行聚类,得到等性能组分类算法、等性能组数据集和多层感知器(MLP)性能组数据集等。

现有研究表明,相似的数据集其可用分类算法相似。然而数据集在属性数、样本规模及属性取值方面千差万别,如何有效地用上述特征来刻画一个数据集,并以此识别相似数据集就成为算法推荐的关键。最近, Song 等^[8]提出一种对数据集进行二进制离散化的方法,并构造了数据集的统计结构特征向量,然后通过 k -近邻识别相似数据集,将相似数据集的分类算法推荐给新数据集。这种方法虽然取得较好的效果,但对数据集结构的刻画并未考虑属性与类标签之间的内在联系。Pritam 等^[9]指出统计交互信息有助于深入分析数据集的潜在结构,并能解释属性之间的依赖、协同、冗余等关系。对于分类问题,通常与类标签有密切关系的是属性的一个子集,属性之间的交互信息能捕获属性、类标签之间的多

元依赖性。Jakulin 的研究表明可通过属性间的交互来分析数据集^[10-13]。本文在 Song 等的基础上,利用统计和交互信息理论提出一种刻画数据集的方法,并将基于这两种方法的推荐算法性能与 Song 的方法进行比较。

2 相关概念

香农信息论使用熵测量随机变量所包含的信息量,熵表征随机变量所包含的不确定性;条件熵是指在已知某一条件下原随机变量剩余的不确定性;两个随机变量之间的交互信息是指两个随机变量所共同包含的信息量,它与两个随机变量之间的相关性成正比^[14]。

2.1 二变量交互信息

二变量交互信息其实是常说的互信息,主要用来度量两个变量之间的共有信息量,给定两个离散随机变量 X 与 Y , X 有 m 种可能值, Y 有 n 种可能值, X 的概率空间为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ p(x_1) & p(x_2) & \cdots & p(x_m) \end{bmatrix},$$

Y 的概率空间为

$$\begin{bmatrix} Y \\ P \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \\ p(y_1) & p(y_2) & \cdots & p(y_n) \end{bmatrix},$$

则它们的交互信息可表示为

$$I(X; Y) = H(X) - H(X|Y),$$

其中 X 的信息熵为

$$H(X) = - \sum_i p(x_i) \lg(p(x_i)),$$

条件熵 $H(X|Y)$ 为

$$H(X|Y) = - \sum_{i,j} p(x_i, y_j) \lg(p(x_i, y_j)).$$

2.2 三变量交互信息

给定 3 个变量 X, Y, Z , 它们之间的交互信息 $I(X; Y; Z)$ 可表示为

$$\begin{aligned} I(X; Y; Z) &= I(XY; Z) - I(X; Z) - I(Y; Z) \\ &= I(Y; Z|X) - I(Y; Z) \\ &= I(X; Z|Y) - I(X; Z), \end{aligned}$$

其中,

$$I(X; Y|Z) = H(X|Z) + H(Y|Z) - H(X, Y|Z)$$

表示给定 Z 的情况下, X 与 Y 之间的条件互信息。三变量的交互信息 $I(X; Y; Z)$ 实质上是度量给定 Z 的情况下, $\{Y, X\}$ 共享信息的增益。当变量 Z 促进或加强 $\{Y, X\}$ 之间的相关性时, $I(X; Y; Z)$ 为正; 当 Z 对 $\{Y, X\}$ 之间的相关性没有影响时, $I(X; Y; Z)$ 为 0;

当 Z 抑制 $\{Y, X\}$ 之间的相关性时 $I(X; Y; Z)$ 为负.

3 特征向量构造方法及评价标准

3.1 数据集属性的离散化

下面简要介绍 Song 等的离散化方法^[8]. 给定数据集 D 其属性集合为

$$A = \{A_1, A_2, \dots, A_n\},$$

类标签 C 集合表示为

$$V_{Class} = \{C_1, C_2, C_3\}.$$

假定 A_1 属性值集合为

$$V_{A_1} = \{A_{11}, A_{12}, A_{13}\},$$

A_2 属性值集合为

$$V_{A_2} = \{A_{21}, A_{22}\},$$

将每个实例的属性或类别用一个二进制向量表示, 如第一个实例的 A_1 属性值可表示为 $V_{A_1}[1] = [1, 0, 0]$. 这样数据集 D 就转化为一个二进制数据集 D_B . 最后统计每个属性值出现的百分比, 就可得到一个向量:

$$V_I = [0.4, 0.4, 0.2, 0.4, \dots, 0.2, 0.3, 0.4, 0.3].$$

主要过程如图 1 所示.

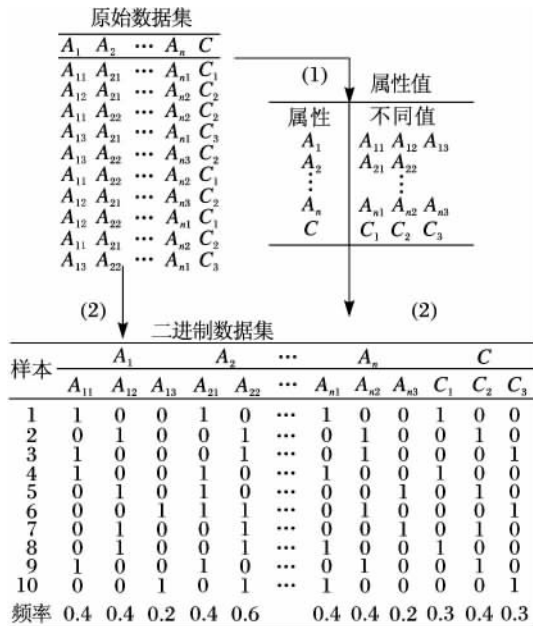


图 1 数据集转换

Fig. 1 Dataset transformation

对两个不同属性 A_i 和 A_j 取异或操作可得二项集合向量 $V_{II} = \{V_{A_i} \oplus V_{A_j} \mid 1 \leq i < j \leq n\}$, 则 $V_{I \& II} = \{V_I, V_{II}\}$ 就表示数据集的特征结构向量. 如前所述, 该方法的缺点是没有考虑属性与属性及属性与类标

签之间的依赖关系. 下面将结合二变量互信息及三变量交互信息, 进一步拓展上述数据集的特征向量的构造方法.

3.2 二变量互信息向量

本文定义属性与类标签的互信息向量 V_{fc} 及属性之间的互信息向量 V_{ff} 分别为

$$V_{fc} =$$

$$\{I(A_i; C) \mid A_i = \{V_{A_i}\}, C = \{V_{class}\}, i = 1, \dots, n\},$$

$$V_{ff} =$$

$$\{I(A_i; A_j) \mid A_i = \{V_{A_i}\}, A_j = \{V_{A_j}\}, 1 \leq i < j \leq n\},$$

特征向量 $V_{fc \& ff} = \{V_{fc}, V_{ff}\}$ 即为数据集的二变量互信息特征结构向量.

3.3 三变量交互信息向量

与二变量互信息特征结构向量类似, A_i 和 A_j 表示数据集的两个不同属性, C 表示类标签, 三变量交互信息向量可表示为

$$V_{Interaction} = \{I(A_i; A_j; C) \mid A_i = \{V_{A_i}\}, A_j = \{V_{A_j}\}, 1 \leq i < j \leq n, C = \{V_{class}\}\}.$$

3.4 向量归一化

由于数据集的属性数及其取值数和类标签数各不相同, 用上述方法所得的特征向量的长度也将不同, 这样无法直接比较它们的相似性. 为解决问题, Song 等采用一种简单的统计方法——五数概括法^[8]. 对于计算出来的互信息和交互信息值按升序排列后, 分别取最小值、1/8 分位数、2/8 分位数, \dots , 7/8 分位数、最大值、平均值和标准方差, 这样就能获得统一长度的数据集特征向量.

3.5 评价标准

为衡量推荐方法的性能, 文献[8]中定义 3 种指标: 分类精度、推荐精度及命中率来验证提出的数据集特征结构方法的有效性.

1) 分类精度. 一个算法 Alg 在数据集 D 上的性能应该综合考虑其分类精度和时间复杂度, 定义为

$$P_{Alg, D} = \frac{Acc_{Alg, D}}{1 + \alpha \times \lg(Runtime_{Alg, D})},$$

其中 $Acc_{Alg, D}$ 和 $Runtime_{Alg, D}$ 分别是算法 Alg 的精度和运行时间. α 是表示精度与时间重要性的参数, 如 $\alpha = 10\%$ 代表用 10% 的精度来换取 10 倍的运行时间加速. 对于数据集 D , 设 $P_{BestAlg, D}$ 为其最佳算法 BestAlg 的性能, $Stdev$ 表示所有候选算法的标准方差, 则在区间 $[P_{BestAlg, D} - \beta \times Stdev, P_{BestAlg, D}]$ 的算法定义为数据集的可用分类算法, 参数 β 决定可用算法区间的大小.

2) 推荐精度. 推荐精度主要用来衡量推荐算法逼近最优算法的程度, 定义为

$$RA = \frac{P_{Alg R D} - P_{Alg W D}}{P_{Alg A D} - P_{Alg W D}} \times 100\%$$

其中 $P_{Alg R}$ 表示推荐分类算法的性能, $P_{Alg A}$ 表示数据集最佳分类算法的性能, $P_{Alg W}$ 表示最差候选算法的性能. 平均推荐精度 ARA 表示在所有数据集上的推荐精度, 定义为

$$ARA = \frac{\sum_{i=1}^N RA_i}{N} \times 100\%$$

3) 命中率. 对于指定数据集 D , 如果推荐算法的性能在其可用分类算法性能子集中, 则表示该推荐算法命中. 命中率可衡量推荐方法的有效性, 定义为

$$HR = \frac{\sum_{i=1}^N HCount_i}{N} \times 100\%$$

如果第 i 个数据集的推荐算法命中, 则 $HCount_i$ 的值为 1, $\sum_{i=1}^N HCount_i$ 表示所有数据集中命中次数之和, N 表示数据集总数.

4 实验与结果分析

4.1 实验准备

本文选取 12 种分类算法和 98 个数据集进行验证实验, 数据集主要来自 UCI 数据仓库 (<http://archive.ics.uci.edu/ml/datasets.html>). 数据集的离散化采用 WEKA 中的 MDL 的方法. 这 12 种分类算法主要分为: 基于概率的 Naive Bayes、Bayes Net; 基于树的 C4.5, 基于规则的 PART、JRIP、ONER 算法; 懒惰学习算法 IB1; 支持向量机算法

SMO 及 Bagging 算法中分别以 C4.5、Naive Bayes 作为基分类器的合成算法和 Boosting 算法中分别以 C4.5、Naive Bayes 作为基分类器的合成算法. 在推荐的过程中, 近邻数设置为验证数据集总数的 10%. 参数 α 的值分别设置为 0%、0.05%、0.1%, 参数 β 的值设为 0.5.

4.2 推荐分类算法的性能

为直观描述推荐方法的性能, 采用散点图表示推荐算法精度与实际最佳算法精度的关系. 如果数据点在对角线上, 表明推荐算法的分类精度与最佳算法的分类精度一样. 如果在对角线下面, 表明推荐算法的分类精度低于最佳算法的分类精度. 为衡量推荐算法分类精度的分布, 本文设置一条与对角线平行且包括 90% 数据点的斜线 $y = x - b$. 显然, 截距 b 的大小恰好反映它与对角线之间的宽度, 本文称之为推荐精度间隙. b 越小, 表明推荐方法效果越好. 图 2、图 3 和图 4 分别是 3 种特征结构方法 $V_{I\&II}$ 、 $V_{fc\&ff}$ 、 $V_{Interaction}$ 的推荐精度. 可看出: 1) 无论哪种推荐方法, 都有相当部分的点在对角线附近. 这表明通过数据集的相似性, 可为大多数数据集推荐到接近最佳分类算法性能的算法; 2) 随着 α 增大, 特征结构 $V_{I\&II}$ 和 $V_{fc\&ff}$ 分类精度的 b 逐渐增大, 这是由于 α 的增大导致算法以部分精度来换取较短时间, 从而引起推荐精度的下降. 而 $V_{Interaction}$ 分类精度的 b 则基本保持不变, 这说明这种特征向量对 α 值不敏感. 3) 在 3 种 α 取值情况下, 特征结构 $V_{fc\&ff}$ 和 $V_{Interaction}$ 分类精度的 b 比 $V_{I\&II}$ 大约分别减少 15%、7%、7% 和 2%、5.5%、5.5%. 因此, 基于交互信息的特征结构方法能更好地刻画数据集.

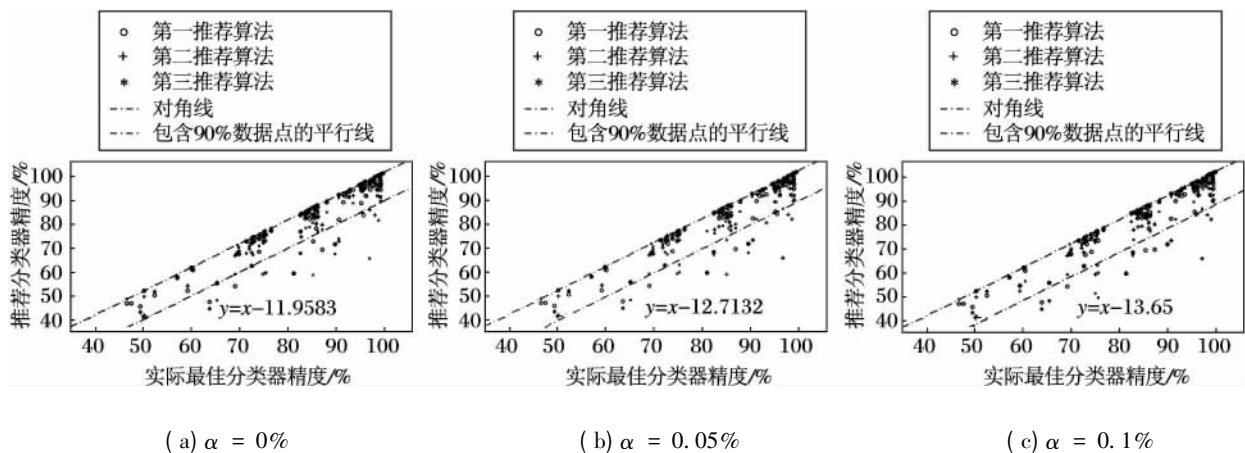


图 2 实际最佳算法的分类精度和 $V_{I\&II}$ 推荐算法的分类精度比较

Fig. 2 Comparison of classification accuracy of real optimal algorithm and recommended algorithms under $V_{I\&II}$

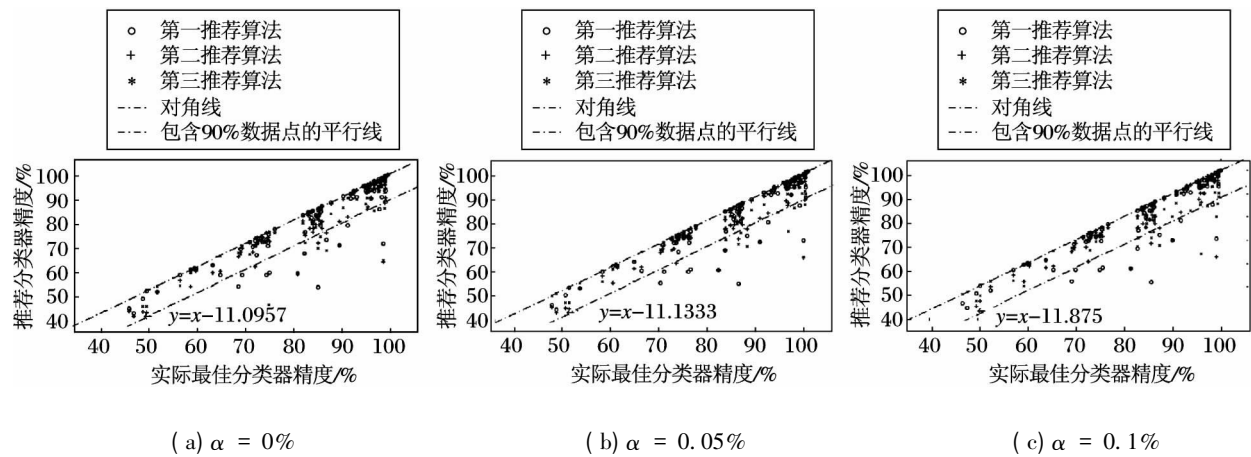


图3 实际最佳算法的分类精度和 $V_{fc\&ff}$ 推荐算法的分类精度比较

Fig.3 Comparison of classification accuracy of real optimal algorithm and recommended algorithms under $V_{fc\&ff}$

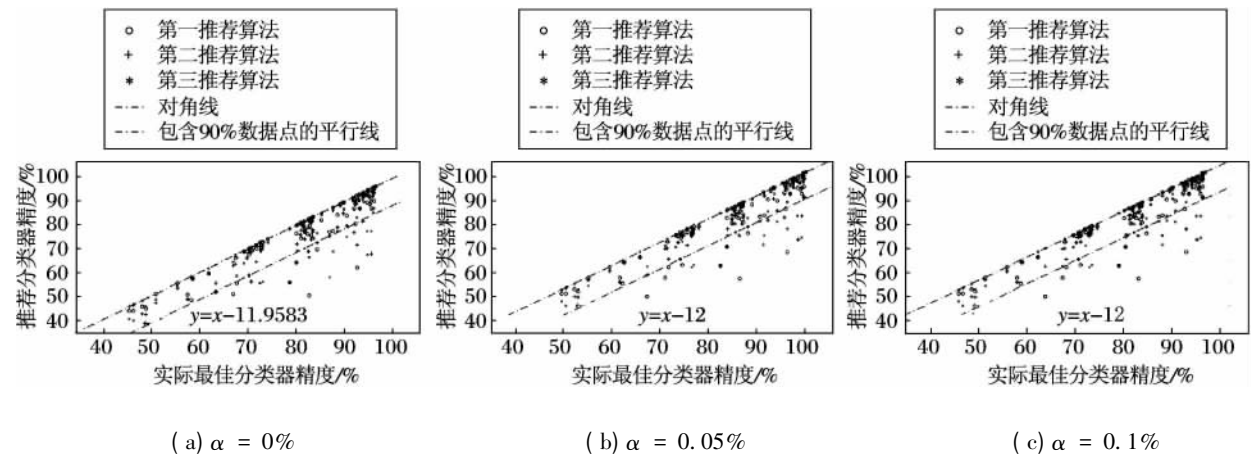


图4 实际最佳算法的分类精度和 $V_{Interaction}$ 推荐算法的分类精度比较

Fig.4 Comparison of classification accuracy of real optimal algorithm and recommended algorithms under $V_{Interaction}$

无论是哪种数据集特征结构方法,都存在一定的偏差,因此,实际推荐过程总是给出几个候选算法.表1总结在最好算法的平均精度是80.03%时,98个数据集在3个 α 取值下前3个推荐算法的精度,可看出:1) $V_{fc\&ff}$ 的前3个推荐分类算法的平均分类精度都比 $V_{I\&II}$ 高,3个推荐算法的平均精度提高

约0.7%,这表明二变量互信息特征结构方法比 $V_{I\&II}$ 更有效;2) $V_{Interaction}$ 除第1推荐分类算法的分类精度低于 $V_{I\&II}$ 方法外,第2、3推荐算法的分类精度都高于 $V_{I\&II}$ 方法,3个推荐算法的平均精度提高约0.5%,这表明基于三变量交互信息特征结构方法也比 $V_{I\&II}$ 有效.

表1 3种特征结构推荐算法的平均分类精度比较

Table 1 Comparison of average classification accuracy of 3 feature recommendation algorithms

推荐分类算法	%								
	$\alpha = 0$			$\alpha = 0.05$			$\alpha = 0.1$		
	$V_{I\&II}$	$V_{fc\&ff}$	$V_{Interaction}$	$V_{I\&II}$	$V_{fc\&ff}$	$V_{Interaction}$	$V_{I\&II}$	$V_{fc\&ff}$	$V_{Interaction}$
第一推荐分类算法	75.55	75.94	75.43	75.89	75.97	75.41	75.89	75.97	75.41
第二推荐分类算法	74.59	75.67	75.72	74.04	75.56	75.36	74.04	75.56	75.36
第三推荐分类算法	74.19	75.00	74.71	74.04	75.00	75.12	74.44	75.00	75.20
3个推荐算法的平均值	74.78	75.54	75.3	74.66	75.51	75.53	74.79	75.51	75.32

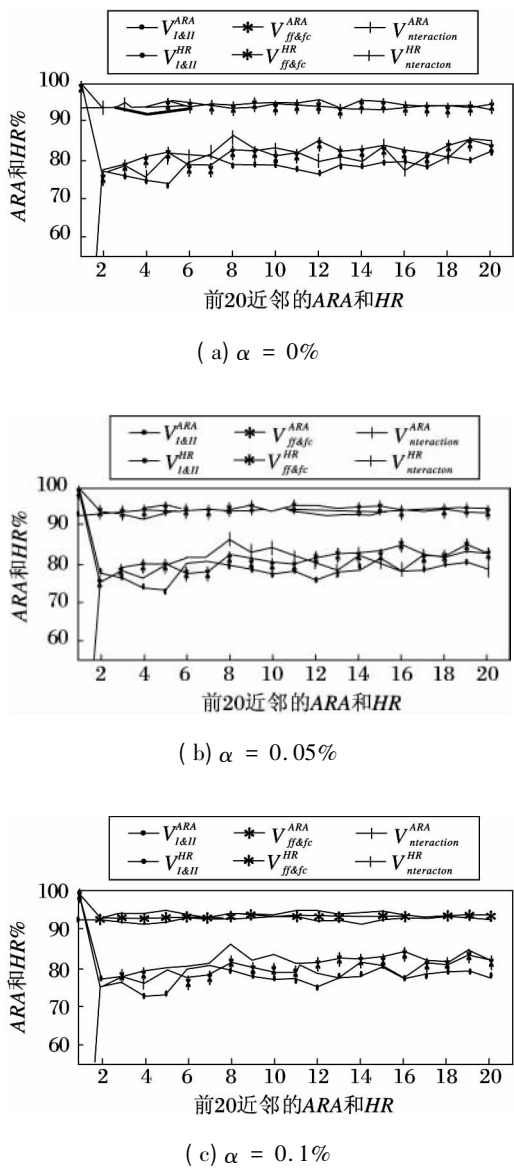


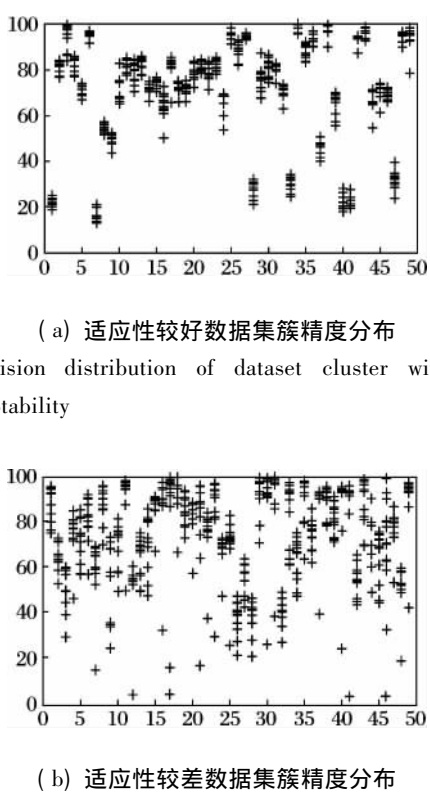
图5 近邻数与平均推荐精度 ARA 和命中率 HR 的关系
Fig.5 Relationship between the number of the nearest neighbors and ARA and HR

4.3 近邻数对精度和命中率的影响

在推荐过程,近邻数的多少将影响推荐方法的精度及命中率.图5描述从1~20近邻下的平均推荐精度 ARA 和命中率 HR .从图可看出:1)在3种特征结构方法下,近邻数的分布对推荐算法平均精度的影响不大,大约在93%~95%之间.与 $V_{I\&II}$ 相比, $V_{fc\&ff}$ 和 $V_{Interaction}$ 的平均推荐精度大约能提高0.5%~1%.2)在命中率方面,基于 $V_{fc\&ff}$ 和 $V_{Interaction}$ 方法推荐的算法几乎总是优于 $V_{I\&II}$.因此,基于交互信息特征结构 $V_{fc\&ff}$ 和 $V_{Interaction}$ 的推荐方法比 $V_{I\&II}$ 更精确更稳定.

4.4 数据集的适用性对推荐性能的影响

在实际中,有些数据集在大部分分类算法下都具有较好的分类结果,即数据集对算法的适应性好;而有些数据集在不同算法下的结果相差很大,即数据集的适应性较差.本文按照98个数据集在12种分类算法的精度进行简单的聚类,将其等分为两组.图6给出这两组数据集的最优分类精度的分布图:(a)分类精度相差不大,(b)分类精度相差较大.表2与表3分别列出在3种 α 取值下(a)数据集与(b)数据集推荐精度散点图中的平行线间隙值.



(a) 适应性较好数据集簇精度分布
(a) Precision distribution of dataset cluster with better adaptability
(b) 适应性较差数据集簇精度分布
(b) Precision distribution of dataset cluster with poor adaptability

图6 98个数据集在12种分类算法下的精度分布
Fig.6 Precision distribution for 98 datasets under 12 kinds of classification algorithms

表2 图6(a)适应性较好数据集簇推荐精度间隙 b
Table 2 Gap b of recommendation precision for dataset cluster with better adaptability in Fig.6(a)

	$\alpha = 0\%$	$\alpha = 0.05\%$	$\alpha = 0.1\%$
$V_{I\&II}$	7.4277	7.6835	7.6835
$V_{fc\&ff}$	6.6186	7.1502	7.1502
$V_{Interaction}$	7.4277	7.4277	7.4277

表3 图6(b) 适应性较差数据集簇推荐精度间隙 b
Table 3 Gap b of recommendation precision for dataset cluster with poor adaptability in Fig. 6(b)

	$\alpha = 0\%$	$\alpha = 0.05\%$	$\alpha = 0.1\%$
$V_{I\&II}$	14.7308	14.7308	14.7308
$V_{fc\&ff}$	14.119	14.119	14.119
$V_{Interaction}$	13.3344	13.3344	13.3344

由表可看出,这两种数据集的精度平行线的间隙大约分别为7和14.这也说明适应性较差数据集分类算法推荐的难度远远大于适应性较好数据集.此外, $V_{fc\&ff}$ 和 $V_{Interaction}$ 方法在适应性较好数据集上的精度平行线间隙比 $V_{I\&II}$ 分别减少约11%、6.5%、6.5%和0%、3%、3%,而在适应性较差数据集上的精度平行线间隙比 $V_{I\&II}$ 分别减少约4%、4%、4%和9.5%、9.5%、9.5%.说明对于适应性好的数据集, $V_{fc\&ff}$ 方法的推荐性能比 $V_{Interaction}$ 明显提高;而对于适应性差的数据集, $V_{Interaction}$ 方法的推荐性能比 $V_{fc\&ff}$ 明显提高更.后者尤为重要,因为这才是分类算法推荐的主要应用领域. $V_{Interaction}$ 推荐方法之所以在适应性差的数据集上更有效,可能是其不仅考虑属性与类别之间的相关性,还涉及属性与属性之间的协作关系.

5 结束语

在模式识别领域,随着各种分类算法的提出,如何为新的数据集推荐一个合适的分类算法成为一个重要的研究问题.由于已有的研究表明相似的数据集可适用相同的分类算法,因此如何刻画数据集之间的相似性就成为一个最基本的问题.本文在文献[8]的基础上,提出两种基于交互信息理论的特征结构方法.首先,与文献[8]的方法相比,本文提出的数据集特征结构方法能将平均推荐精度提高大约0.5%~1%,命中率提高大约5%~7%,这说明基于交互信息的数据集特征结构方法能较好地刻画数据集,并作分类算法的自动推荐.其次,本文研究了数据集适应性对推荐算法精度的影响,结果表明对于适应性差的数据集,同时考虑数据集属性与类别相关性和属性与属性之间协作关系的特征结构 $V_{Interaction}$ 方法比仅考虑属性与类别相关性的特征结构 $V_{fc\&ff}$ 方法的推荐性能更好,约提高5%.最后,本文提出的两种方法虽然更好地刻画了数据集的特征,但其缺点是计算复杂度高.因此,将来的工作是进一步研究如何精简这种方法计算量,如将特征选

择与本文的方法结合起来,或是更加简便的交互信息计算方法.

参考文献

- [1] Weiss M S, Kapouleas L. An Empirical Comparison of Pattern Recognition, Neural Nets, and Machine Learning Classification Methods // Proc of the 11th International Joint Conference on Artificial Intelligence. Detroit, USA, 1989: 781-787
 - [2] Shavlik J W, Mooney R J, Towell G G. Symbolic and Neural Learning Algorithms: An Experimental Comparison. Machine Learning, 1991, 6(2): 111-143
 - [3] Duin R P W. A Note on Comparing Classifiers. Pattern Recognition Letters, 1996, 17(5): 529-536
 - [4] Brazdil P, Gama J, Henery B. Characterizing the Applicability of Classification Algorithms Using Meta-Level Learning // Proc of the European Conference on Machine Learning. Catania, Italy, 1994: 83-102
 - [5] Gama J, Brazdil P. Characterization of Classification Algorithms // Proc of the 7th Portuguese Conference on Artificial Intelligence. Funchal, Portugal, 1995: 189-200
 - [6] Ali S, Smith K A. On Learning Algorithm Selection for Classification. Applied Soft Computing, 2006, 6(2): 119-138
 - [7] Kalousis A, Gama J, Hilario M. On Data and Algorithms: Understanding Inductive Performance. Machine Learning, 2004, 54(3): 275-312
 - [8] Song Qinbao, Wang Guangtao, Wang Chao. Automatic Recommendation of Classification Algorithms Based on Data Set Characteristics. Pattern Recognition, 2012, 45(7): 2672-2689
 - [9] Chanda P, Cho Y R, Zhang Aidong, et al. Mining of Attribute Interactions Using Information Theoretic Metrics // Proc of the IEEE International Conference on Data Mining. Miami, USA, 2009: 350-355
 - [10] Jakulin A, Bratko I. Testing the Significance of Attribute Interactions // Proc of the 21st International Conference on Machine Learning. Banff, Canada, 2004: 409-416
 - [11] Jakulin A. Machine Learning Based on Attribute Interactions. Master Dissertation. Ljubljana, The Republic of Slovenia: University of Ljubljana, 2005
 - [12] Jakulin A, Bratko I. Analyzing Attribute Dependencies // Proc of the 7th European Conference on Principles and Practice of Knowledge Discovery in Databases. Dubrovnik, Croatia, 2003: 229-240
 - [13] Jakulin A, Bratko I, Smrke D, et al. Attribute Interactions in Medical Data Analysis // Proc of the 9th Conference on Artificial Intelligence in Medicine in Europe. Protaras, Cyprus, 2003: 229-238
 - [14] Xie Jingbo, Wang Xizhao. An Extended Heuristic Algorithm to ID3 Based on the Mutual Information between Attributes. Computer Engineering and Applications, 2004, 40(3): 93-94 (in Chinese)
- (谢竞博,王熙熙.基于属性间交互信息的ID3算法.计算机工程与应用,2004,40(30):93-94)