计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200800025

一种基于标签相关度的 Relief 特征选择算法

丁思凡　王　锋　魏　巍
山西大学计算机与信息技术学院　太原 030006
(sifan_ding_0718@163.com)

摘　要　特征选择在机器学习和数据挖掘中起到了至关重要的作用。Relief 作为一种高效的过滤式特征选择算法,能处理多种类型的数据,且对噪声的容忍力较强,因此被广泛应用。然而,经典的 Relief 算法对离散特征的评价较为简单,在实际进行特征选择时并未充分挖掘特征与类标签之间的潜在关系,具有很大的改进空间。针对经典的 Relief 算法对离散特征的评价方式较为简单这一不足,提出了一种基于标签相关度的离散特征评价方法。该算法充分考虑了不同特征的特性,给出了一种面向混合特征的距离度量方式,同时从离散特征与标签之间的相关度出发,重新定义了 Relief 算法对离散特征的评价体系。实验结果表明,改进后的 Relief 算法与经典的 Relief 算法和现有的一些面向混合数据的特征选择算法相比,其分类精度均有不同程度的提升,具有良好的性能。

关键词:特征选择;Relief;标签相关度;VDM;决策树
中图法分类号　TP181

Relief Feature Selection Algorithm Based on Label Correlation

DING Si-fan, WANG Feng and WEI Wei
School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China

Abstract　Feature selection plays a vital role in machine learning and data mining. Relief, as an efficient filtering feature selection algorithm, is widely used because it can process multiple types of data and has a strong tolerance for noise. However, classic Relief algorithm provides a relatively simple evaluation to discrete features. In actual feature selection, the potential relationship between features and class labels is not fully explored, and there is a lot of room for improvement. Aiming at the shortcomings of classic Relief algorithm's simple evaluation method for discrete features, a discrete feature evaluation method based on label correlation is proposed. The algorithm fully considers the characteristics of different features and gives a distance measurement method for mixed features. At the same time, starting from the correlation between discrete features and tags, it redefines the Relief algorithm's evaluation system for discrete features. Experimental results show that, compared with the classic Relief algorithm and some existing feature selection algorithms for mixed data, the classification accuracy of the improved Relief algorithm has been improved to varying degrees and has a good performance.

Keywords　Feature selection, Relief, Label correlation, VDM, Decision tree

1　引言

特征选择是数据挖掘和机器学习中一类重要的数据预处理技术,特征选择算法旨在从原始样本空间中去除与分类不相关或冗余的特征,找到一个最佳特征子集[1-3]来描述整个数据集。根据其与学习算法的关系,特征选择算法可分为过滤式(Filter)方法[4]、封装式(Wrapper)方法[5]和嵌入式(Embedded)方法[6]。过滤式特征选择方法按照一定的准则直接从原始的特征集合中得到特征子集,独立于学习算法之外,如信息增益[7]、互信息[8]以及本文的 Relief 算法[9]等;其优点是简便快捷,适合于较大规模的数据集[10-11],但其缺乏与分类模型的交互性,后续的学习算法偏差较大。封装式特征选择方法在特征选择过程中与学习算法紧密结合,使用分类模型来评价特征子集,如 k 近邻算法[12]、贝叶斯分类器[13]等;其分类精度比过滤式方法更好,但是容易过拟合,且计算量大,不适合大规模的数据集。嵌入式特征选择方法独立于以上两种

到稿日期:2020-08-04　返修日期:2020-09-22　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61772323);山西省应用基础研究项目(201801D221170)
This work was supported by the National Natural Science Foundation of China(61772323) and Basic Applied Research Project of Shanxi Province of China(201801D221170).
通信作者:王锋(sxuwangfeng@126.com)

方法，其最大的特点就是将特征选择的过程和学习模型的训练合二为一，并且在同一个优化过程中完成，使用一个目标函数来进行特征选择，常见的有决策树算法[14]、基于支持 SVM 的特征选择算法[15]等；这类方法可快速得到特征子集，但对参数敏感，因此如何构造目标函数也是一大难点。

Relief 系列算法是一类效果较好的过滤式特征选择算法。Relief 是由 Kira 等[16]于 1992 年提出的用于二分类的多变量过滤式特征选择算法，其也只能适用于二分类问题。因此，Kononenko 等[17]在 Relief 算法的基础上进行扩展，提出了能解决多类问题的 ReliefF 算法，其可以很好地去除不相关特征。但当样本空间内存在大量无关特征或错误标签时，Relief 和 ReliefF 算法的性能会受到很大的影响。为此，Sun 等[18]于 2007 年提出了一种迭代式特征选择算法 I-Relief，其放弃对近邻的确定，考虑了整个样本空间中的所有样本对权重的影响，相比 Relief 算法来说有着更高的分类精度。Iterative Relief[19]指定目标实例周围的半径，并在该半径范围内选择近邻样本，即最接近的近邻样本对特征权重的影响大于边缘样本。与 Iterative Relief 类似，SURF[20]使用距离阈值 T 将实例定义为邻居(其中 T 等于数据中所有实例对之间的平均距离)。上述研究都很好地扩展了 Relief 系列算法[21]。

但是，Relief 系列算法还有很多待研究的问题。在属性之间的距离度量方面，Relief 算法需要找到计算实例在某类样本集中最近的样本，这便涉及到了距离的度量。经典的 Relief 算法使用的是欧氏距离，即属性值之间的差值会被平方化，而文献[22]的实验表明，属性间的差值是否被平方化对结果没有显著的影响。同时，有研究人员提出任何有效的距离度量方式都可以被 Relief 使用[23]。因此，确定一种优异的距离度量方式是一个亟待解决的问题。

此外，当样本空间中只存在单一类型的特征时，无论该特征是离散的还是连续的，Relief 算法的性能都良好，但当数据集中同时包含离散特征和连续取值特征时，Relief 算法的性能会大打折扣[24]。对此，一种解决方法是使用 ramp 函数，当特征的取值为离散型取值时，该特征会被分配用户预先定义的最大值和最小值两个离散值，如 0 或 1 这样完整的差异值；当特征的取值为连续型取值时，该特征则被分配实例样本与近邻样本的距离函数值[25-26]。但是，由于这种方法会增加额外的自定义参数，需要依赖于问题的优化，因此在实际应用中可能会遇到诸多挑战。而经典 Relief 算法对于离散特征的评价是 0 和 1 两个离散值，显然这种简单的评价方式并没有考虑到实际样本，因此如何确定一种通用却简单高效的离散特征评价方法显得至关重要。

针对以上两个突出问题，本文提出了一种基于标签相关度的 Relief 算法。该算法考虑了不同特征自身的特点，面向不同类型的特征时采用不同的度量方式[27-28]，即一种面向混合特征的度量方法。同时，针对离散特征，考虑其与标签的相关性[29-30]，不同的离散特征会有不同的参数用于样本特征加权。该算法充分考虑了不同问题的实际情况，能在不增加过多过程的同时，有效解决 Relief 面对不同问题的适应性，提高了算法的分类精度和自适应性。

本文第 2 节介绍了 Relief 算法和特征选择的基本概念[31-32]；第 3 节介绍了面向混合特征的度量方式；第 4 节介绍了离散特征的评价方法以及改进的 Relief 算法；第 5 节选取了 UCI 中常用的 4 组数据集进行实验分析，结果进一步证明了该改进算法的高效性，特别是当面对混合特征的数据集时，其对特征子集的求解精度得到明显提高，从而为 Relief 算法对离散特征、混合特征数据集的特征选择提供了新途径；最后总结全文。

2 基本概念

设 $D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$ 是一个含有 n 个样本 m 个特征的数据集。其中，x_i 为样本，$y_i\in\{1,2,\cdots,d\}$ 为样本标签，$C=\{c_1,c_2,\cdots,c_m\}$ 是含 m 个特征的集合。x_{ij} $(1\leqslant i\leqslant n, 1\leqslant j\leqslant m)$ 表示第 i 个样本在第 j 个特征上的值。$|g|$ 为集合中含有元素的个数。当 c_j 为离散特征时，c_j 的取值集合 C_j 为 $\{c_{j1},\cdots,c_{jk},\cdots,c_{j|c_j|}\}$ $(1\leqslant j\leqslant m, 1\leqslant k\leqslant|c_j|)$，而 $|C_j|$ 为特征集合 c_j 内的样本个数，此时在 c_j 上有 $\forall i, x_{ij}\in C_j$。

特征选择是按照一定的评价准则从原始数据集中选择部分特征来构造一个最佳特征子集，使其能够描述原始样本空间[33]。

即在一个含 m 个特征 $c_1,c_2,\cdots,c_n$ 的数据集 D 中，给定一种方法 *Method*，按照一定的准则评价每个特征，筛选出部分特征 $c_i,c_j,\cdots,c_k$ $(1\leqslant i,j,k\leqslant m)$ 以得到最终的特征子集 $F\subseteq D$。

Relief 算法的主要思想是：对于一个二分类问题，Relief 算法会从数据集中选择一个样本 x_i，然后从与 x_i 同标签的样本集中寻找 x_i 的最近邻样本 x_i^{NH}，从与 x_i 不同标签的样本集中寻找 x_i 的最近邻样本 x_i^{NM}，再根据以下策略更新每个特征权重。

如果 x_i 和 x_i^{NH} 在某个特征上的距离小于 x_i 和 x_i^{NM} 之间的距离，说明该特征对区分同类和异类的最近邻起着正面作用，则增加该特征的权重。反之，如果 x_i 和 x_i^{NH} 在某个特征上的距离大于 x_i 和 x_i^{NM} 之间的距离，说明该特征对区分同类和异类的最近邻起着负面作用，则降低该特征的权重。多次重复以上过程，最后得到各个特征的平均权重。特征的权重越大，该特征的分类能力就越强；反之，该特征的分类能力就越弱。

3 面向混合特征的度量方式

在 Relief 算法的研究过程中，确定最佳的度量方式仍是一个值得深入思考的问题，针对不同类型的特征使用不同的度量方式，最后汇总一个面向混合特征的度量方式。

特征主要可分为两类：1)连续特征，该特征在定义域上的取值有无穷多个，$|c_j|\approx n$，特征的取值个数接近于样本数，连续特征能够直接参与数值计算；2)离散特征，该特征在定义域

上的取值数量有限，$|c_j|=n$，特征的取值个数远小于样本数，且不能直接参与数值计算。离散特征也可以进一步细分为有序特征和无序特征，有序特征的特点是特征值间存在次序关系，能够与连续特征一样参与数值计算，例如调查问卷中的打分项 1－5；而无序特征是无法直接参与数值计算的，如颜色、性别等特征。

对于无序离散特征，使用 VDM[34] 进行度量。

定义 1 令 $m_{c_j,c_{j1},i}$ 表示特征 c_j 取 c_{j1} 值时的样本个数，$m_{c_j,c_{j1},i}$ 表示在第 i 样本簇中在特征 c_j 上取值为 c_{j1} 的样本数，$|c_j|$ 为样本簇数，则特征 c_j 上两个特征值 c_{j1} 与 c_{j2} 之间的 VDM 距离为：

$$VDM(c_{j1},c_{j2})=\sum_{k=1}^{|c_j|}\left|\frac{m_{c_j,c_{j1},k}}{m_{c_j,c_{j1}}}-\frac{m_{c_j,c_{j2},k}}{m_{c_j,c_{j2}}}\right| \tag{1}$$

其中，c_{j1} 与 c_{j2} 为特征 c_j 上的两个特征值。

对于连续特征和有序离散特征，为了与离散特征的 VDM 度量的维度相对应，这里我们使用的是哈曼顿距离。

$$dist_{mk}(x_1,x_2,j)=\left(\sum_{j=1}^{m}|x_{1j}-x_{2j}|\right) \tag{2}$$

其中，x_1,x_2 表示两个不同的样本。

针对混合特征，使用 VDM 和哈曼顿距离相结合的方式。对于一个含 n 个特征的数据集，设有 n_{mk} 个有序特征，$n-n_{mk}$ 个无序特征，令有序属性排列在无序属性之前，则：

$$dist(x_1,x_2)=\left(\sum_{j=1}^{n_{mk}}|x_{1j}-x_{2j}|+\sum_{j=n_{mk}+1}^{n}VDM(x_{1j},x_{2j})\right) \tag{3}$$

这便是改进 Relief 算法对于混合特征的度量方式。在之后的实验中，经典 Relief 算法和改进 Relief 算法都将使用这种度量方式。

4 基于标签相关度的特征选择算法

Relief 是一种高效的多变量过滤式特征选择算法，能处理多种类型的数据，简单、快捷且高效，可以很好地去除无关特征。但经典的 Relief 算法对离散特征的评价方式较为简单，没有很好地考虑特征与标签之间的潜在联系。本文从离散特征与标签的相关性出发，重新定义了离散特征的评价方法，提出了一种基于标签相关度的 Relief 算法。

4.1 面向离散特征的评价体系

对于一个离散特征 c_j，$|c_j|$ 为其含有的元素个数，对于 $\forall j\in\{1,2,\cdots,|c_j|\}$，均有 $c_{jk}\in\{c_{j1},c_{j2},\cdots,c_{j|c_j|}\}$，同时，$\forall i\in\{1,2,\cdots,n\}$，均有 $x_{ij}\in\{c_{j1},c_{j2},\cdots,c_{j|c_j|}\}$。

首先考虑一个特征值 c_{jk}，给出以下定义。

定义 2 $\forall i\in\{1,2,\cdots,n\}$，$x_{ij}$ 为样本 x_i 在特征 c_j 上的特征值，则：

$$\sigma(x_{ij},c_{jk})=\begin{cases}0, & x_{ij}=c_{jk}\\ 1, & x_{ij}\neq c_{jk}\end{cases} \tag{4}$$

定义 3 当 y 为样本标签集，d 为样本标签类别数，且 $y_{all}=\{y_0,y_1,\cdots,y_d\}$，对于样本 x_i 有：

$$\sigma(y_i,y)=\begin{cases}0, & y_i\notin y\\ 1, & y_i\in y\end{cases} \tag{5}$$

下面将给出一个特征值 c_{jk} 对标签相关度的推导过程：

$$\varphi(c_{jk},y)=\frac{\sum_{i=1}^{n}[\sigma(x_{ij},c_{jk})\cdot\sigma(y_i,y)]}{\sum_{i=1}^{n}[\sigma(x_{ij},c_{jk})\cdot\sigma(y_i,y_{all})]} \tag{6}$$

考虑该特征值占总样本数的比例 $\theta(c_{jk})$：

$$\theta(c_{jk})=\frac{\sum_{i=1}^{n}[\sigma(x_{ij},c_{jk})]}{n} \tag{7}$$

因此，该特征值相对于标签的相关度可以表示为：

$$p(c_{jk},y)=\varphi(c_{jk},y)\cdot\theta(c_{jk}) \tag{8}$$

$$p(c_{jk},y)=\frac{\sum_{i=1}^{n}[\sigma(x_{ij},c_{jk})\cdot\sigma(y_i,y)]}{n} \tag{9}$$

综合考虑离散特征的所有特征值，可以构成一个多维矩阵。

$$\begin{bmatrix}p(c_{j1},y_0) & p(c_{j2},y_0) & \cdots & p(c_{j|c_j|},y_0)\\ p(c_{j1},y_1) & p(c_{j2},y_1) & \cdots & p(c_{j|c_j|},y_1)\\ \vdots & \vdots & & \vdots\\ p(c_{j1},y_d) & p(c_{j2},y_d) & \cdots & p(c_{j|c_j|},y_d)\end{bmatrix} \tag{10}$$

特别地，在二分类数据集上，式(10)可改写为：

$$\begin{bmatrix}p(c_{j1},y_0) & p(c_{j2},y_0) & \cdots & p(c_{j|c_j|},y_0)\\ p(c_{j1},y_1) & p(c_{j2},y_1) & \cdots & p(c_{j|c_j|},y_1)\end{bmatrix} \tag{11}$$

在该二维矩阵中，每一行表示该标签下所有特征值占总样本的权重，而每一列表示一个特征在两个标签上所占的权重。式(11)表示一个特征值在不同标签上的分布情况，若一个特征值在不同标签上的分布差异大，说明该特征值可能在一定程度上影响着标签。综合考虑该特征上所有的特征值，则可以得到该离散特征的评价系数。

$$weight(c_j)=\sum_{k=1}^{|c_j|}|p(c_{jk},y_0)-p(c_{jk},y_1)| \tag{12}$$

通过分析可发现，该系数 $weight(c_j)$ 的范围介于 0～1 之间。若 $weight(c_j)$ 趋于 0，则其与标签的相关性低，对分类起到的作用可能很小；反之，该特征与标签的相关性高，可能具有较强的分类能力。

需要注意的是，仅仅依靠该方式来选择特征子集显然是不合理、欠考虑的。该评价方式仅仅适用于离散特征，更重要的是，该方法只是描述了离散特征相对于标签的分布情况，并不能真正反映该特征对标签的区分能力，因此我们需要将其与 Relief 算法相结合。

4.2 基于标签相关度的 Relief 算法

上文介绍了面向混合特征的度量方式以及离散特征的评价体系，该方法考虑了每个离散特征值与标签之间的相关性，其不依赖于问题本身，而是宏观考虑了离散特征与标签的联系，具有良好的自适应性和普适性。

在此基础上，本节将给出基于标签相关度的 Relief 特征选择算法，如算法 1 所示。

算法 1 改进的 Relief 算法

输入：训练样本 $D=\{(x_i,y_i)\}_{i=1}^{n}$，$x_i\in R^m$，抽样次数 num

输出：特征子集 T

1. 初始化：特征权重 $w_j=0$，$j=1,\cdots,m$

2. for i=1 to num do

随机选择一个样本 T；

$\min\limits_{y_i} dist(x_i,x_i^{NH})$，从与 x_i 同标签样本集中找到 x_i 的最近邻样本 x_i^{NH}；

$\min\limits_{y_i} dist(x_i,x_i^{NM})$，从与 x_i 不同标签样本集中找到 x_i 的最近邻样本 x_i^{NH}；

for j=1 to m do

$w_j=w_j-diff(j,x_i,x_i^{NH})/num+diff(j,x_i,x_i^{NM})/num$

3. 按照一定的评价准则筛选特征子集 T。

算法 1 中，$diff(j,x_1,x_2)$表示样本 x_1 和 x_2 在第 j 个特征上的距离差。

对于连续特征：

$$diff(j,x_1,x_2)=\frac{|x_{1j}-x_{2j}|}{\max(c_j)-\min(c_j)} \tag{13}$$

对于离散特征：

$$diff(j,x_1,x_2)=\begin{cases}0, & x_{1j}=x_{2j}\\ weight(c_j), & x_{1j}\neq x_{2j}\end{cases} \tag{14}$$

以上即为改进后的 Relief 算法，其使用了哈曼顿距离和 VDM 进行混合度量，针对不同特征采用不同的度量方式，保证了在每次迭代过程中都选择了与 x_i 最近邻的样本。在此基础上，使用了一种新的离散特征评价方式，充分考虑了不同特征相对于标签的划分情况，依据不同特征赋予不同的权重。由于该方式是从特征与标签的相关性出发，很好地契合了 Relief 算法的思想，同时又考虑了问题的实际情况，在不改变算法的时间复杂度的情况下，优化了离散特征的评价体系，能更好地筛选出特征子集。

5 实验分析

本文选择了 UCI 中 4 组常用的数据集。由于本文提出的新算法是面向混合数据的特征选择算法，为有效验证新算法的性能，除了经典的 Relief 算法，本节还选取了一种面向混合数据的基于信息熵的启发式粗糙特征选择算法[35]作为对比算法。为了方便表示，将基于信息熵的启发式前向搜索算法标记为 shang，经典 Relief 算法标记为 Relief，本文提出的改进 Relief 算法标记为 ReliefM。实验中使用的数据集如表 1 所列。

表 1 实验中使用的数据集

Table 1 Datasets in experimental

数据集	样本数	属性个数	类别
Australian	690	14	2
Credit	690	15	2
Hepatitis	155	19	2
Somerville Happiness Survey	143	6	2

学习算法使用的是决策树 J48[36]，在每个数据集上运行 10 次，每次运行时都将数据集进行随机划分，将数据集的 2/3 作为训练集，将剩下的 1/3 作为测试集。Relief 和 ReliefM 算法会在划分好的训练集和测试集上进行 10 次实验，并取 10 次实验结果的平均值作为该训练集上的每个特征得分，在测试集上使用决策树 J48 得到分类精度，最后会汇总 10 次实验的结果。

实验中分别比较了 shang 算法、Relief 算法和 ReliefM 算法在表 1 所列的 4 个数据集上的分类精度，结果如图 1—图 4 所示，其中 X 轴表示第 i 次实验，Y 轴表示分类精度。

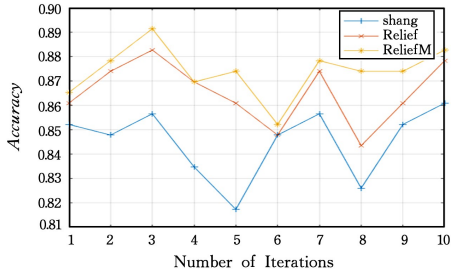

图 1 Australian 数据集上的分类精度

Fig. 1 Classification accuracy on Australian dataset

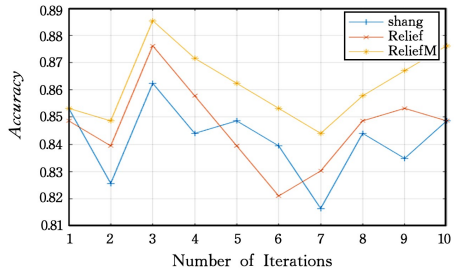

图 2 Credit 数据集上的分类精度

Fig. 2 Classification accuracy on Credit dataset

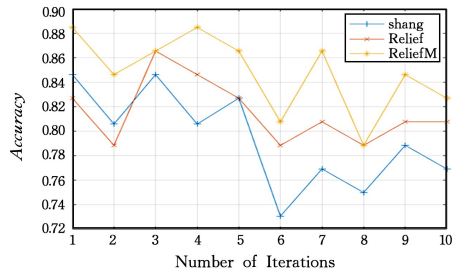

图 3 Hepatitis 数据集上的分类精度

Fig. 3 Classification accuracy on Hepatitis dataset

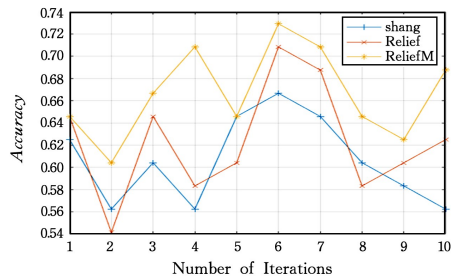

图 4 Somerville Happiness-Survey 数据集上的分类精度

Fig. 4 Classification accuracy on Somerville Happiness-Survey dataset

表 2 列出了 Relief 和 ReliefM 在 4 组数据集上得到的特征子集的分类精度的均值和标准差。表 2 中，在 4 个数据集

上，改进 Relief 算法所得特征子集的分类精度明显优于经典 Relief 算法和基于信息熵的启发式特征选择算法。

表 2 各数据集上的结果

Table 2 Results on each datasets

数据集	shang	Relief	ReliefM
Australian	0.8478±0.0107	0.8652±0.0128	0.8739±0.0105
Credit	0.8417±0.0134	0.8463±0.0152	0.8619±0.0131
Hepatitis	0.7939±0.0394	0.8154±0.0260	0.8481±0.0320
Somerville Happiness Survey	0.6062±0.0386	0.6229±0.0505	0.6667±0.0405

值得说明的是，改进 Relief 算法所得特征的权重低于 Relief 算法，这是因为在改进 Relief 算法的离散特征评价方式中，每个特征被赋予的权重最大为 1，使得 ReliefM 算法得到的特征权重低于 Relief 算法，但由于离散特征加权的系数考虑了特征与标签之间的关联，因此改进 Relief 算法能很好地区分出不同的特征，有效地提高了分类精度。在部分数据集上出现了两者分类结果一样的情况，这与 Relief 算法会以一定比例抽取样本进行迭代有关，但 ReliefM 算法的总体结果优于 Relief。

实验结果表明，使用面向不同特征的度量方式和离散特征的评价方法，ReliefM 的分类性能有了明显的改善。在部分实验中，实验会受到迭代次数的影响，从而导致两者的效果一致，但总体上 ReliefM 的精度优于 Relief，有效地提高了分类精度。

综上分析，与经典 Relief 算法相比，ReliefM 算法具有如下几点优势：

(1)面向不同特征采用不同的度量方式，能正确地选择出与实例最近邻的样本。

(2)离散特征的评价体系考虑了特征与标签的相关性，具有良好的自适应性，同时该方法并没有过多提升问题的复杂性。

(3)多个样本的多次实验结果表明，改进后的 Relief 算法的分类精度相比经典 Relief 算法有着不同程度的提升。

结束语 针对经典 Relief 面对离散特征评价方式较为简单的不足，本文提出了一种面向混合特征的距离度量方式，在此基础上，通过分析每个离散特征相对于标签的相关度，提出了一种基于标签相关度的离散特征加权机制，该加权机制综合考虑了不同特征的实际情况，能更好地区分不同特征，优化了特征子集的求解过程。进一步，我们选取了 UCI 中 4 组常用的数据集，结果验证了改进的 Relief 算法的有效性。该算法很好地考虑了不同特征的具体情况，在含混合特征的数据集中效果会更明显，该算法可有效应用于离散特征和混合特征数据集的特征选择中。

下一步的研究工作将重点关注以下两个方面：1)进一步改进离散特征评价方法，将其扩展到多分类的 Relief-F 上；2)进一步优化基于标签相关度的相似度度量，以提高算法的性能。同时对比多种混合特征选择算法，不断研究、改进，以提升 Relief 系列算法对离散特征和混合特征的特征选择效果。

参 考 文 献

[1] LIU H, YU L. Toward integrating feature selection algorithms for classification and clustering[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(4): 491-502.

[2] WANG S, LI T R, LUO C, et al. Domain-wise approaches for updating approximations with multi-dimensional variation of ordered information systems[J]. Information Sciences, 2019, 478: 100-124.

[3] ZENG A P, LI T R, HU J, et al. Dynamical updating fuzzy rough approximations for hybrid data under the variation of attribute values[J]. Information Sciences, 2017, 378: 363-388.

[4] DASH M, CHOI K, SCHEUERMANN P, et al. Feature selection for clustering - a filter solution[C]// 2002 IEEE International Conference on Data Mining. Maebashi City, Japan, 2002: 115-122.

[5] ZHU Z, ONG Y, DASH M. Wrapper-Filter Feature Selection Algorithm Using a Memetic Framework[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2007, 37(1): 70-76.

[6] LIU Y F, YE D Y, LI W B, et al. Robust neighborhood embedding for unsupervised feature selection[J]. Knowledge-Based Systems, 2020, 193: 105462.

[7] HUANG D, CHOW T W S. Effective feature selection scheme using mutual information[J]. Neurocomputing, 2005, 63(Jan): 325-343.

[8] XU J L, ZHOU Y M, CHEN L, et al. Unsupervised feature selection based on mutual information[J]. Journal of Computer Research and Development, 2012, 49(2): 372-382.

[9] HUANG X J. Research on Relief Algorithm for Feature Selection[D]. Suzhou: Suzhou University, 2018.

[10] WANG F, LIANG J Y, QIAN Y H. Attribute reduction: a dimension incremental strategy[J]. Knowledge-Based Systems, 2013, 39(2): 95-108.

[11] LIANG J Y, WANG F, DANG C Y, et al. A group incremental approach to feature selection applying rough set technique[J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 26(2): 294-308.

[12] ISLAM M J, WU Q M J, AHMADI M, et al. Investigating the Performance of Naive-Bayes Classifiers and K-Nearest Neighbor Classifiers[J]. Journal of Convergence Information Technology, 2010, 5(2): 133-137.

[13] WANG G C. Research and Application of Naive Bayes Classifier [D]. Chongqing: Chongqing Jiaotong University, 2010.

[14] SAFAVIAN S R, LANDGREBE D. A survey of decision tree classifier methodology[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1991, 21(3): 660-674.

[15] ZHOU X, TUCK D P. MSVM-RFE: extensions of SVM-RFE for multiclass gene selection on DNA microarray data[J]. Bioinformatics, 2007, 23(9): 1106-1114.

[16] KIRA K, RENDELL L A. The feature selection problem: Traditional methods and a new algorithm[C]// AAAI. 1992: 129-134.

[17] KONONENKO I. Estimating attributes: analysis and extensions of Relief[C]// Maching Learning: ECML-94. 1994: 171-182.
[18] SUN Y. Iterative RELIEF for Feature Weighting: Algorithms, Theories, and Applications[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1035-1051.
[19] DRAPER B, KAITO C, BINS J. Iterative Relief[C]// 2003 Conference on Computer Vision and Pattern Recognition Workshop, Madison, Wisconsin. USA, 2003: 62-62.
[20] GREENE C S, PENROD N M, KIRALIS J, et al. Spatially uniform relieff (SURF) for computationally-efficient filtering of gene-gene interactions[J]. BioData Mining, 2009, 2(1): 5.
[21] URBANOWICZ R J, MEEKER M, LA CAVA W, et al. Relief-based feature selection: Introduction and review[J]. Journal of biomedical informatics, 2018, 85: 189-203.
[22] KONONENKO I, ŠIMEC E, ROBNIK-ŠIKONJA M. Overcoming the myopia of inductive learning algorithms with RELIEFF [J]. Applied Intelligence, 1997, 7(1): 39-55.
[23] TODOROV A. Statistical Approaches to Gene X Environment Interactions for Complex Phenotypes[M]. MIT Press, 2016: 95-116.
[24] KONONENKO I, ŠIKONJA M R. Non-myopic feature quality evaluation with (R) ReliefF[J]. Computational methods of feature selection, 2008, 7(10): 169-191.
[25] HONG S J. Use of contextual information for feature ranking and discretization[J]. IEEE transactions on knowledge and data engineering, 1997, 9(5): 718-730.
[26] ROBNIK-ŠIKONJA M, KONONENKO I. Theoretical and empirical analysis of ReliefF and RReliefF[J]. Machine learning, 2003, 53(1/2): 23-69.
[27] WANG J, WANG S T. Double exponential fuzzy C-means algorithm based on mixed distance learning [J]. Journal of Software, 2010, 21(8): 1878-1888.
[28] LI H L, GUO C H. Review of feature representation and similarity measurement in time series data mining[J]. Computer Application Research, 2013, 30(5): 1285-1291.
[29] XIE M X, GUO J Z, ZHANG H B, et al. Research on similarity measurement method of high-dimensional data[J]. Computer Engineering and Science, 2010, 32(5): 92-96.
[30] LIU J, JIN D, DU H J, et al. A new hybrid feature selection method RRK[J]. Journal of Jilin University (Engineering Science Edition), 2009, 39(2): 419-423.
[31] ZHANG L X, WANG J Y, ZHAO Y N, et al. Combined feature selection based on Relief[J]. Fudan Journal (Natural Science Edition), 2004(5): 893-898.
[32] WANG J, CI L L, YAO K Z. A Summary of Feature Selection Methods[J]. Computer Engineering and Science, 2005(12): 72-75.
[33] DING X M, WANG H J, WANG Y G, et al. Unsupervised feature selection method based on improved ReliefF[J]. Application of Computer Systems, 2018, 27(3): 149-155.
[34] STANFILL C, WALTZ D. Toward memory-based reasoning [J]. Communications of the ACM, 1986, 29(12): 1213-1228.
[35] WANG F, LIANG J Y. An efficient feature selection algorithm for hybrid data[J]. Neurocomputing, 2016, 193: 33-41.
[36] BHARGAVA N, SHARMA G, BHARGAVA R, et al. Decision tree analysis on j48 algorithm for data mining[J]. Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering, 2013, 3(6): 1114-1119.

DING Si-fan, born in 1999, bachelor, is a student member of China Computer Federation. His main research interests include data mining and machine learning.

WANG Feng, born in 1984, Ph. D. Her mian research interests include the areas of feature selection, rough set theory, granular computing and artificial intelligence.