DOI：10.7652/xjtuxb201602003

采用互补信息熵的分类器集成差异性度量方法

赵军阳[1,2]，韩崇昭[2]，韩德强[2]，张春霞[3]
（1.第二炮兵工程大学 202 教研室，710025，西安；2.西安交通大学电子与信息工程学院，710049，西安；
3.西安交通大学数学与统计学院，710049，西安）

摘要：针对多分类器系统差异性评价中无法直接处理模糊数据的问题，提出了一种采用互补信息熵的分类器集成差异性度量（CIE）方法。首先利用训练数据生成一系列基分类器，并对测试数据进行分类，将分类结果依次组合生成分类数据空间；然后采用模糊关系条件下的互补信息熵度量分类数据空间蕴含的不确定信息量，据此信息量判断基分类器间的差异性；最后以加入基分类器后数据空间差异性增加为选择分类器的基本准则，构建集成分类器系统，用于验证 CIE 差异性度量与集成分类精度之间的关系。实验结果表明，与 Q 统计方法相比，利用 CIE 方法进行分类器集成，平均集成分类精度提高了 2.03%，分类器系统集成规模降低约 17%，而且提高了集成系统处理多样化数据的能力。

关键词：分类器集成；差异性；互补信息熵；模糊关系
中图分类号：TP391.4 **文献标志码**：A **文章编号**：0253-987X(2016)02-0013-07

A Novel Measure Method for Diversity of Classifier Integrations Using Complement Information Entropy

ZHAO Junyang[1,2], HAN Chongzhao[2], HAN Deqiang[2], ZHANG Chunxia[3]
(1. Staff Room 202, The Second Artillery Engineering University, Xi'an 710025, China; 2. School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China; 3. School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: A novel diversity measure method using complement information entropy (CIE) is proposed to solve the problem that the diversity estimation of multiple classifier systems is unable to deal directly with fuzzy data. A set of base classifiers is generated by using training data, and then is used to label test data. The outputs of the classifiers are reorganized into a new classification data space. Then the complement information entropy model is introduced under fuzzy relation to measure uncertainty information of the new space and the uncertainty information is used to estimate the diversity of base classifiers. Finally, an ensemble system is constructed based on the criterion that the ensemble diversity of the classifier set increases when a base classifier is added, and the ensemble system is used to validate the performance of CIE. Experimental results and a comparison with the Q-statistic method show that the average classification accuracy of CIE increases by 2.03%, and the number of ensemble classifiers reduces by 17%. Moreover, CIE also improves the ability of ensemble systems to process diverse data.

Keywords: classifier ensemble; diversity; complement information entropy; fuzzy relation

收稿日期：2015-06-21。 作者简介：赵军阳(1981—)，男，讲师，博士后；韩崇昭(通信作者)，男，教授，博士生导师。 基金项目：国家自然科学基金资助项目(61074176,41174162)；中国博士后科学基金资助项目(2013M532048)。
网络出版时间：2015-11-27 网络出版地址：http://www.cnki.net/kcms/detail/61.1069.T.20151127.2115.002.html

分类器集成是指针对某一问题，将一系列基分类器进行组合，来提高分类的精度和泛化性能的方法。目前，多分类器集成已得到广泛而深入的研究，并成为机器学习、模式识别等领域的主要研究方向之一。很显然，如果进行组合的是相同且无差异的分类器，集成系统并不能提高整体分类效果。因此，要提高多分类器系统的性能，基分类器必须具有一定的差异性，即要求 B 分类器能将 A 分类器错误分类的样本重新划分到正确的类别。

分类器差异性的研究主要涉及分类器差异性生成模式、差异性度量方法、差异性与集成分类性能关系以及如何利用差异性度量优化分类器集成系统等方面的研究[1-2]。其中，分类器差异性生成模式的研究是提高集成系统性能的基础[3]，也是众多文献的研究热点。差异性的获得可通过采用不同类型的分类器、设置分类器的不同参数配置和采用不同的训练数据集来实现[4]。如何度量分类器间的差异性是研究者需要关注的另一个重要问题。分类器差异性的正确度量和分析对于设计性能优良的分类器系统至关重要。目前，国内外学者已经提出一些度量分类器差异性的方法，以期对分类器系统差异特性进行统计分析[5-9]，如 Kuncheva 总结的 Q 统计、双错度量和熵度量等[5]，Windeatt 提出的基于模式的度量方法[6]；或者指导分类器集成系统的优化设计与实现[10-14]，以提高分类器的集成性能。现有的一些方法虽然能在一定程度上表示分类器之间的差异性，但主要是从分类器正确分类和错误分类的一致性角度出发进行定义，必须根据标准类别信息首先对分类器输出结果的正确性进行判别，无法直接度量分类器本身蕴含的分类信息。为此本文从信息熵角度出发研究如何直接度量分类器的差异性，提出一种基于互补信息熵的分类器差异性度量(CIE)方法，根据不同分类器所蕴含不确定信息量的差别来实现分类器的差异性评价，并分析差异性度量方法与系统集成性能之间的联系。数据实验表明，本文方法能有效度量分类器差异性，在降低分类器集成规模的同时，提高或保持集成系统的集成分类精度。

1 常用的差异性度量方法

目前比较常用的差异性度量方法主要可以分为两类：成对度量方法[5]和非成对度量方法[15]。

1.1 成对度量方法

成对差异度量方法首先计算分类器系统中每一对分类器之间的差异性度量值，L 个分类器对应 $L(L-1)/2$ 对差异值，然后对各差异值求取平均值得到系统的差异度。以下介绍几种常见的成对度量方法。

(1)相关系数(Correlation Coefficient，ρ)

$$\rho_{i,j}=(N^{11}N^{00}-N^{01}N^{10})/[((N^{11}+N^{10})(N^{01}+N^{00})(N^{11}+N^{01})(N^{10}+N^{00}))^{1/2}] \tag{1}$$

式中：N^{01} 表示分类器 D_i 和 D_j 的联合分类输出概率，0 表示 D_i 分类错误，1 表示 D_i 分类正确；其余定义类似。

(2)Q 统计(Q-statistic，Q)

$$Q_{i,j}=\frac{N^{11}N^{00}-N^{01}N^{10}}{N^{11}N^{00}+N^{01}N^{10}} \tag{2}$$

(3)不一致度量(Disagreement Measure，D_{is})

$$D_{\mathrm{is}}=\frac{N^{01}+N^{10}}{N^{11}+N^{10}+N^{01}+N^{00}} \tag{3}$$

(4)双错度量(Double-Fault Measure，D_{DF})

$$D_{\mathrm{DF}}=\frac{N^{00}}{N^{11}+N^{10}+N^{01}+N^{00}} \tag{4}$$

1.2 非成对度量方法

非成对差异性度量方法不强调分类器两两之间的关系，而是对整个分类器集合进行计算得到系统的差异度。

(1)熵度量(Entropy，E)

$$E=\frac{1}{N}\frac{2}{L-1}\sum_{i=1}^{N}\min\{l(x_i),(L-l(x_i))\} \tag{5}$$

式中：$l(x_i)$ 表示在一组 L 个分类器中，将样本 x_i 正确分类的分类器个数；N 为样本数。

(2)KW 方差(Kohavi-Wolpert variance，D_{KW})

$$D_{\mathrm{KW}}=\frac{1}{NL^2}\sum_{i=1}^{N}l(x_i)(L-l(x_i)) \tag{6}$$

(3)Kappa 度量(Interrater agreement，κ)

$$\kappa=1-\frac{(1/L)\sum_{i=1}^{N}l(x_i)(L-l(x_i))}{N(L-1)\bar{P}(1-\bar{P})} \tag{7}$$

式中：$\bar{P}$ 表示平均样本分类精度。

(4)难点度量(Difficulty，θ)

$$\theta=\mathrm{var}(Z) \tag{8}$$

式中：Z 表示对于随机给定的输入 x，分类正确的分类器在所有集成分类器中的比率。

(5)广义差异性度量(Generalised Diversity，D_{G})

$$D_{\mathrm{G}}=1-\frac{p(2)}{p(1)} \tag{9}$$

式中：$p(1)$ 表示 1 个分类器的出错概率；$p(2)$ 表示 2 个分类器的出错概率。

(6)一致错误差异性度量(Coincident Failure

$Diveristy, D_{CF}$)

$$D_{CF}=\begin{cases}0, & p_0=1.0\\ \dfrac{1}{1-p_0}\displaystyle\sum_{i=1}^{L}\dfrac{L-i}{L-1}p_i, & p_0<1\end{cases} \tag{10}$$

式中：p_0 表示所有个体部分类正确；p_i 表示 L 个分类器中有 i 个得出错误分类结果的概率。

2　互补信息熵差异性度量与集成

2.1　模糊近似空间中的互补信息熵

为了度量数据空间蕴含的不确定信息，目前已提出多种信息熵度量方法，但无论是 Shannon 信息熵[16]还是梁吉业提出的粗糙集中的信息熵模型[17]均要求数据空间满足一定的等价关系，只能处理离散数据。然而，实际的数据未必存在明确的边界区分，需利用连续特征函数进行描述，通过模糊隶属函数进行处理。为了适应任意模糊关系下的信息度量，文献[18]对 Shannon 熵进行改进，提出了模糊关系下的信息熵模型；作者则在文献[19]中考虑类别划分的补集，提出了任意模糊关系下的互补信息熵模型，可以直接处理连续或模糊数据。

定义 1　设 $U=\{x_1,x_2,\cdots,x_n\}$ 为有限非空论域，R 是 U 上的任意模糊关系，则模糊近似空间 (U,R) 的互补信息熵[19]定义为

$$H(R)=\sum_{i=1}^{n}\frac{1}{|U|}\left(1-\frac{|[x_i]_R|}{|U|}\right) \tag{11}$$

式中：$|[x_i]_R|=\displaystyle\sum_{j=1}^{n}r_{ij}$ 表示对象 x_i 在模糊关系 R 下的势，其中 r_{ij} 表示 2 个对象的相似度。

2.2　互补信息熵差异性度量方法

上节介绍的差异性度量方法不仅要求分类器的输出结果为 0/1 模式，而且需要预先判断分类器输出的正确性，无法直接度量分类器输出信息，不能适应连续或模糊数据的集成处理。互补信息熵则不仅能应用于模糊信息系统的信息处理，而且无需预先离散化，也可以度量分类器系统所蕴含的信息量。为此，本文将其用于分析分类器的差异性，提出一种采用互补信息熵的差异性度量方法。

假设分类器系统中基分类器 c_i 的分类输出结果为 $O_i=\{o_{i1},o_{i2},\cdots,o_{iN}\}$，将各基分类器的输出组合起来构成一个新的分类数据空间，即 $U=\{o_{ij}\mid i=1,\cdots,L;j=1,\cdots,N\}$，其中，$L$ 表示分类器个数，N 表示样本个数，每一个分类器的输出即为数据空间 U 中的一个数据对象，各个分类器间的差异性越大，则蕴含的互补信息熵也越大，由此得到一种新的差异性度量方法。

定义 2　设 $O=\{o_1,o_2,\cdots,o_L\}$ 为有限非空论域，R 是 O 上的任意模糊关系，则基于互补信息熵的差异性度量方法（Complement Information Entropy，CIE）定义为

$$D_{CIE}=\frac{1}{L}\sum_{i=1}^{L}\frac{|[o_i]_R|}{|U|}\left(1-\frac{|[o_i]_R|}{|U|}\right) \tag{12}$$

式中：$|[o_i]_R|$ 表示在第 i 个分类器输出的各样本对象结果在模糊关系 R 下的势。

定义 2 基于不同分类器间的相似关系，综合度量基分类器对各个原始样本数据的分类效果及其互补信息，给出基分类器集合的差异性，省略了对分类器输出结果的正确性判别过程，具有更好的适应性。D_{CIE} 值越大，则差异性越大，可用于指导基分类器的评价和选择。为此，本文依据互补信息熵差异性度量方法提出增量式的基分类器差异重要性评价方法，其定义如下。

定义 3　给定一个基分类器集成系统 (O,C)，O 为有限非空论域，C 为所有分类器集合，$B\subseteq C$，$\forall c_i\in C-B$，则分类器 c_i 关于分类器集合 B 中的差异重要性定义为

$$S(c_i,B)=D_{CIE}(B\cup\{c_i\})-D_{CIE}(B) \tag{13}$$

该定义以基分类器集成系统差异性增加为基本准则。若加入一个基分类器后，集成系统的差异性增加，则保留该分类器；若集成系统的差异性降低，则舍去该分类器。基于该准则可实现基分类器的自动选择，有利于降低集成规模。

2.3　基于互补信息熵差异重要性评价的选择性集成方法

为了验证 CIE 差异性度量方法与集成分类精度之间的关系，设计了一种基于互补信息熵分类器差异性评价的集成方法（简称 CIE 集成方法），即首先将原始数据集划分为训练集和测试集，然后采用 Bootstrap 采样方法在训练集上生成 N 个数据子集，再基于这些数据子集对基分类器进行训练得到每个数据对象的分类输出结果。在基分类器训练结束后，基于定义 3 评价当前分类器对基分类器集合的重要程度。如果重要性大于 0，则保留该分类器；若重要性小于等于 0，则舍去该分类器，继续评价下一个分类器。将选择的基分类器输出结果通过多数投票法进行组合，可得到最终的分类结果。

CIE 集成方法步骤如下。

步骤 1　初始条件。令 $U\leftarrow$有限数据集，$C\leftarrow$初始空分类集成系统。

步骤 2 生成训练子集。利用 Bootstrap 采样方法生成 N 个训练子集。

步骤 3 训练基分类器。在每个训练子集上训练单一分类器，得到 N 个分类器集合$\{C_i\}_{i=1,\cdots,N}$。

步骤 4 基分类器性能评价与选择。根据式(13)分类器差异重要性评价结果自动选择分类器加入集成系统 C。

步骤 5 生成分类器集成系统。将加入的各基分类器组合得到最终的分类器集成系统 C^*，利用多数投票方法组合输出结果。

步骤 6 集成系统分类精度评价。基于 10 折交叉验证方法评价集成系统 C^* 的分类精度。

CIE 集成方法在运行过程中无需重复进行类别标记，利用差异性评价方法对在样本采样后的训练子集中生成的基分类器进行选择，不仅能够提高分类器间的差异性，也有助于降低分类器系统的集成规模和复杂度，提高系统的识别效果。

3 数据实验分析

3.1 实验数据

本文利用机器学习领域常用的加州大学 Irvine 分校 UCI(University of California Irvine)数据库[20]中的 12 种数据集对 CIE 集成方法的性能进行验证实验，涉及医学诊断、客户分类、污水处理、车辆分析和葡萄酒识别等方面，详细信息如表 1 所示。12 种数据集的类别数为 2～13 类，特征值均为数值类型，特征既有连续型，也有离散型，特征维数在 4～56 之间，样本数在 32 到 1 000 之间。

表 1 UCI 实验数据

数据集缩写	数据集全称	样本数	特征维数	类别数
BC	breast cancer data	286	9	2
WBC	wisconsin breast cancer	699	9	2
Cre	german credit	1 000	20	2
Cle	cleveland clinic	303	13	5
Der	dermatology database	366	34	6
LC	lung cancer	32	56	3
Iris	iris plants	150	4	3
Veh	vehicle silhouettes	946	18	4
Wat	water treatment	527	38	13
Win	wine recognition	178	13	3
Ion	Ionosphere	351	34	2
Wdbc	wisconsin diagnostic breast cancer	569	30	2

3.2 CIE 集成方法分类性能比较实验

在开始算法性能实验前，需首先设置基分类器的训练个数 N，各方法的分类精度为 P。从表 1 中随机选取 Wbc、Cre、Wat 和 Wdbc 4 个数据集，并选择常用的决策树(decision tree, DT)和支持向量机(support vector machine, SVM)作为基分类器，其中 SVM 核函数采用径向基函数。在此基础上，分析集成系统训练的基分类器数量对 CIE 方法集成分类性能的影响，实验结果如图 1 所示。

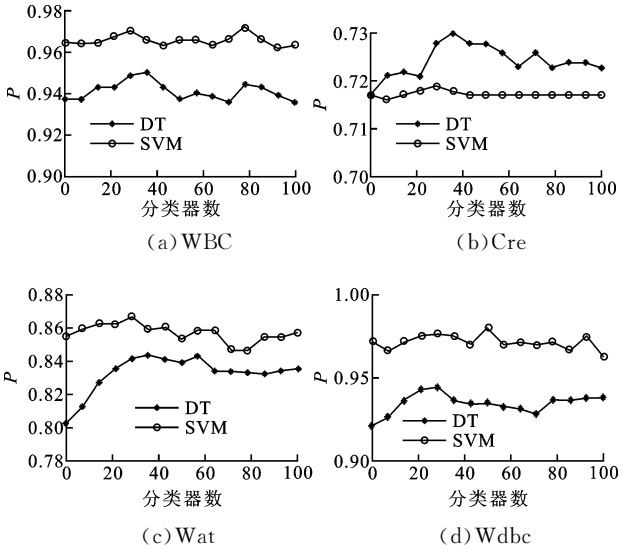

图 1 不同基分类器数对算法分类性能的影响

由图 1 可知，随着分类器数量的增加，集成分类精度存在先升后稳的趋势，甚至还会降低，表明分类器数量并非越多越好，满足集成系统的选择需要即可。为提高集成系统的训练效率，以下统一设置 $N=10$。

进行 CIE 集成方法的分类性能验证实验，并将结果与 Bagging(Bag)、Adaboost(Ada)和 RSM 等主要集成算法进行分析比较。首先将数据集样本随机划分为 20 份，循环将其中 9 份组合作为训练集，剩余 1 份作为测试集，并在每个循环中生成 10 个基分类器作为候选集合，然后根据重要性评价方法自动选择合适的分类器组合得到分类器集成系统。Bagging、Adaboost 和 RSM 等集成算法采用新西兰 Waikato 大学开发的 WEKA 机器学习软件对数据集进行分类实验。所有算法的参数设置均为 WEKA 的默认设置。

图 2 和图 3 分别为采用决策树和 SVM 为基分类器时，上述方法在这些数据集上的分类性能比较结果。从图中结果可以得出：

(1)CIE 集成方法的分类性能在多数数据集上

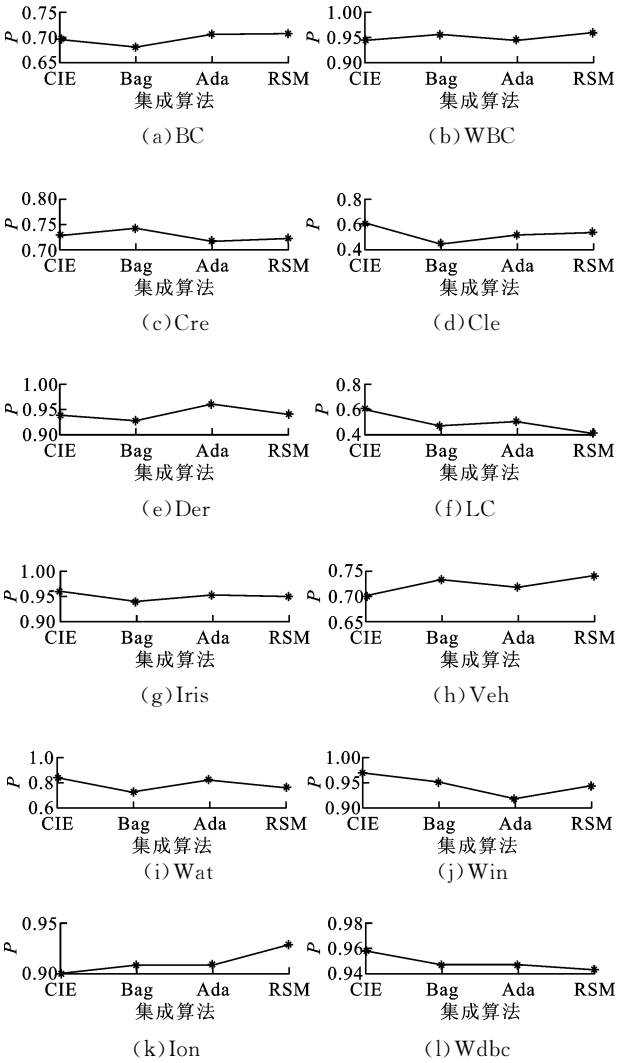

图 2　以决策树为基分类器时几种集成算法的分类性能比较

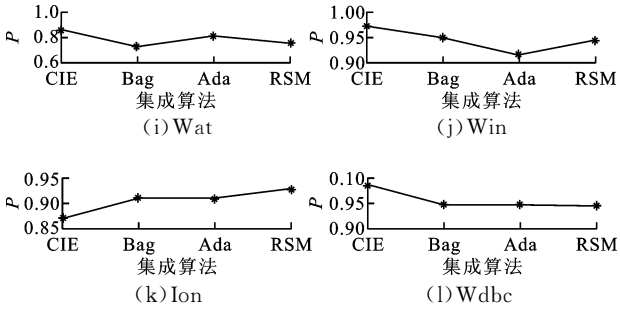

图 3　以 SVM 为基分类器时几种集成算法的分类性能比较

接近或超过 Bagging、Adaboost 和 RSM 算法，表明以差异性评价作为选择分类器的标准是可行的；

(2)当采用决策树为基分类器时，CIE 集成方法在半数数据集上获得最优性能，而当采用 SVM 为基分类器时，CIE 集成方法在 8 个数据集上性能表现突出，在 Cle、LC、Iris、Wat 和 Wdbc 这 5 个数据集上表现更为明显，如在 Cle 上的分类精度相比 Bagging 算法提高了 38.5%。

3.3　CIE 差异性度量方法性能分析实验

在 CIE 集成方法框架下，为了比较 CIE 度量方法与其他差异性度量方法的性能差异，引入 Q 统计、熵度量、KW 方差和双错度量等常用方法替换 CIE 差异性度量方法，并与原始 CIE 集成方法进行比较。图 4 和图 5 是分别以决策树和 SVM 为基分类器时的精度对比结果。对图 4、图 5 的结果分析可得如下结果。

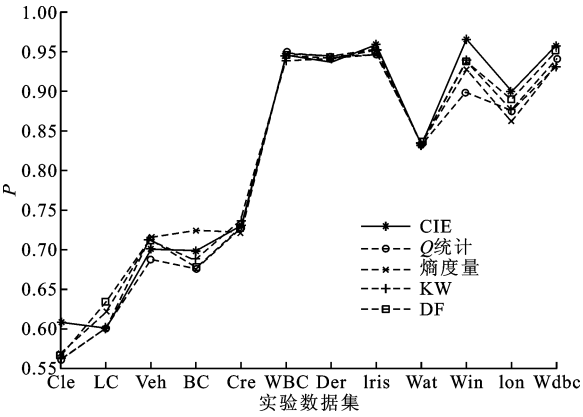

图 4　以决策树为基分类器时几种差异性度量方法的集成分类性能比较

(1)采用决策树作为基分类器时，基于 CIE 度量集成后的系统分类精度与基于其他 4 种差异性度量方法相比，在 6 个数据集上获得最佳分类效果；采用 SVM 作为基分类器时也在 4 个数据集上获得最高精度；在其余数据集上的分类性能则与其他方法相近，表明 CIE 差异性度量方法可有效应用于分类

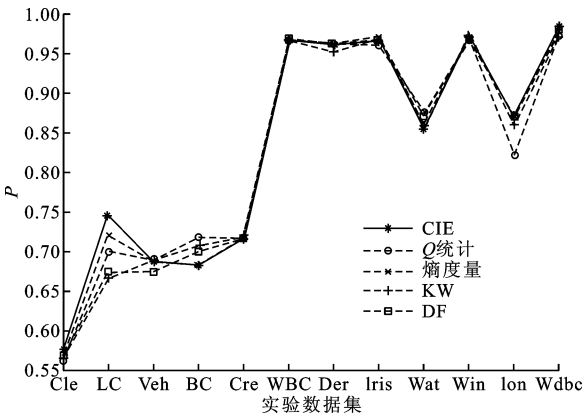

图 5 以 SVM 为基分类器时几种差异性度量方法的集成分类性能比较

器集成系统的差异性评价，并指导分类器集成系统设计和优化。

(2)通过对图 4 和图 5 实验结果的统计分析可以看出，CIE 度量方法综合性能最优，在不同基分类器条件下均取得了最高平均分类精度，如表 2 所示。其次是熵度量和双错度量方法。熵度量和双错度量在文献[5]中也指出其具有较好的差异性度量能力，整体性能表现要优于 Q 统计和 KW 方差。

表 2 几种差异性度量方法下 CIE 集成方法的平均分类精度

基分类器	P/%				
	CIE	Q 统计	熵度量	KW	DF
决策树	81.97	80.34	81.12	80.88	81.43
SVM	83.20	82.60	82.93	82.44	82.64

图 6 和图 7 给出了本节实验过程中，基于上述差异性度量方法的集成系统最终选择的基分类器个数。由图中可知，无论采用何种差异性度量方法，多分类器系统集成的平均分类器个数在 2.1～4.0 之间。与传统上多达几十甚至上百个分类器的复杂集成系统相比，CIE 方法可在选择少量基分类器的同时，获得较优的分类性能，平均比 Q 统计方法生成的集成系统规模降低 17%左右。

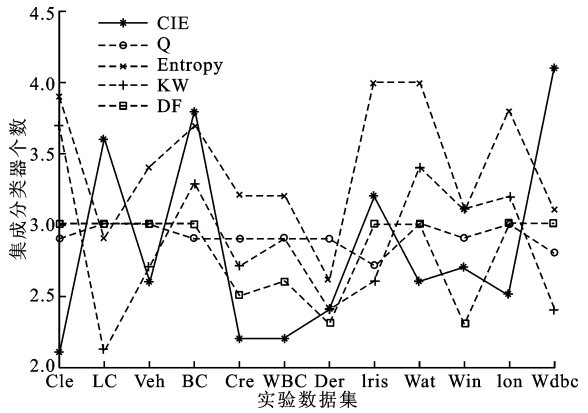

图 6 以决策树为基分类器时几种差异性度量方法集成的分类器个数比较

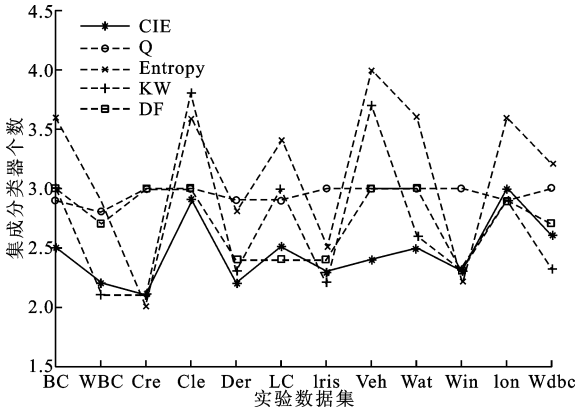

图 7 以 SVM 为基分类器时几种差异性度量方法集成的分类器数比较

通过上述实验，基于 CIE 差异性评价的集成算法具有在选择较少基分类器的基础上，保持或提高分类器系统性能的能力。互补信息熵差异性度量方法在度量多分类器系统差异性方面是有效的，在分类器集成过程中的应用也是可行的。

4 结 论

为了满足直接度量分类器差异性的多样性需求，提高分类数据处理的能力，本文提出了一种互补信息熵差异性度量方法，并利用分类器重要性评价选择基分类器进行集成。该方法能够直接处理分类器的输出结果，不受 0/1 模式限制；此外，通过对分类器系统信息量的直接度量，省略了对分类结果正确性的判别，适用于半标记和未标记数据的处理。实验结果验证了本文方法在分类器集成应用方面的有效性和可行性。

需要指出的是，CIE 集成方法在分类器选择过程中仅采用了差异性指标，虽然有效降低了系统的集成规模，但未考虑与集成精度的平衡问题，对系统的泛化能力可能会有一定影响。目前在如何实现分类器系统差异性和集成精度的有效平衡以及对系统的影响方面尚缺乏理论依据，下一步工作将在集成系统的优化方面进行研究和探索。

参考文献：

[1] KUNCHEVA L I, SKURICHINA M, DUIN R P W. An experimental study on diversity for bagging and boosting with linear classifiers [J]. Information Fu-

sion，2002，3(4)：245-258.

[2] BROWN G，KUNCHEVA L I. “Good” and “bad” diversity in majority vote ensembles [C]// Proceedings of International Conference on Multiple Classifier Systems. Berlin，Germany：Springer，2010：124-133.

[3] 张宏达，王晓丹，韩钧，等. 分类器集成差异性研究 [J]. 系统工程与电子技术，2009，31(12)：307-3012.
ZHANG Hongda，WANG Xiaodan，HAN Jun，et al. Survey of diversity researches on classifier ensembles [J]. Systems Engineering and Electronics，2009，31(12)：3007-3012.

[4] NASCIMENTO D，COELHO A，CANUTO A. Integrating complementary techniques for promoting diversity in classifier ensembles：a systematic study [J]. Neurocomputing，2014，138：347-357.

[5] KUNCHEVA L I，WHITAKER C J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy [J]. Machine Learning，2003，51：181-207.

[6] WINDEATT T. Diversity measures for multiple classifier system analysis and design [J]. Information Fusion，2005，6(1)：21-36.

[7] HAGHIGHI M S，VAHEDIAN A，YAZDI H S. Creating and measuring diversity in multiple classifier systems using support vector data description [J]. Applied Soft Computing，2011，11(8)：4931-4942.

[8] KRAWCZYK B，WOZNIAK M. Diversity measures for one-class classifier ensembles [J]. Neurocomputing，2004，126：36-44.

[9] YIN X C，HUANG K Z，HAO H W，et al. A novel classifier ensemble method with sparsity and diversity [J]. Neurocomputing，2014，134：214-221.

[10] BI Y X. The impact of diversity on the accuracy of evidential classifier ensembles [J]. International Journal of Approximate Reasoning，2012，53(4)：584-607.

[11] AKSELA M，LAAKSONEN J. Using diversity of errors for selecting members of a committee classifier [J]. Pattern Recognition，2006，39(4)：608-623.

[12] RASHEED S，STASHUK D W，KAMEL M S. Diversity-based combination of non-parametric classifiers for EMG signal decomposition [J]. Pattern Anal Applic，2008，11(3/4)：385-408.

[13] 杨春，殷绪成，郝红卫，等. 基于差异性的分类器集成有效性分析及优化集成 [J]. 自动化学报，2014，40(4)：660-674.
YANG Chun，YIN Xucheng，HAO Hongwei，et al. Classifier ensemble with diversity：effectiveness analysis and ensemble optimization [J]. Acta Automatica Sinica，2014，40(4)：660-674.

[14] 杨长盛，陶亮，曹振田，等. 基于成对差异性度量的选择性集成方法 [J]. 模式识别与人工智能，2010，23(4)：565-571.
YANG Changsheng，TAO Liang，CAO Zhentian，et al. Pairwise diversity measures based selective ensemble method [J]. PR&AI，2010，23(4)：565-571.

[15] 谷雨. 分类器集成中的多样性度量 [J]. 云南民族大学学报：自然科学版，2012，21(1)：59-65.
GU Yu. Measure diversity classifier ensemble [J]. Journal of Yunnan National University：Natural Science，2012，21(1)：59-65.

[16] LIU W Y，WU Z H，PAN G. An entropy-based diversity measure for classifier combining and its application to face classifier ensemble thinning [C]// Proceedings of International Conference on Sinobiometrics. Berlin，Germany：Springer，2004：118-124.

[17] LIANG J，CHIN K，DANG C. A new method for measuring uncertainty and fuzziness in rough set theory [J]. International Journal of General Systems，2002，31(4)：331-342.

[18] YU D，HU Q，WU C. Uncertainty measures for fuzzy relations and their applications [J]. Applied Soft Computing，2007，7(3)：1135-1143.

[19] ZHAO J，ZHANG Z，HAN C，et al. Complement information entropy for uncertainty measure in fuzzy rough set and its application [J]. Soft Computing，2015，19(7)：1997-2010.

[20] BLAKE C L. MERZ C L. UCI repository of machine learning databases [EB/OL]. (2007-10-12) [2015-05-08]. http：// www. ics. uci. edu/～ mlearn/MLRepository. html.

[本刊相关文献链接]

兰景宏，刘胜利，吴双，等. 用于木马流量检测的集成分类模型. 2015，49(8)：84-89. [doi：10. 7652/xjtuxb201508014]

喻明让，张英杰，陈琨，等. 考虑调整时间的作业车间调度与预防性维修集成方法. 2015，49(6)：16-21. [doi：10. 7652/xjtuxb201506003]

杨宏晖，王芸，孙进才，等. 融合样本选择与特征选择的 AdaBoost 支持向量机集成算法. 2014，48(12)：63-68. [doi：10. 7652/xjtuxb201412010]

王羡慧，覃征，张选平，等. 采用仿射传播的聚类集成算法. 2011，45(8)：1-6. [doi：10. 7652/xjtuxb201108001]

马超，陈西宏，徐宇亮，等. 广义邻域粗集下的集成特征选择及其选择性集成算法. 2011，45(6)：34-39. [doi：10. 7652/xjtuxb201106006]

（编辑　刘杨）