计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200700036

基于直觉模糊集的集成学习算法

戴宗明　胡　凯　谢　捷　郭　亚
江南大学轻工业先进过程控制教育部重点实验室　江苏 无锡 214122
江南大学物联网工程学院　江苏 无锡 214122
(dzm_1995@163.com)

摘　要　为提高传统机器学习算法的分类精度和泛化能力,提出一种基于直觉模糊集的集成学习算法。根据传统分类器分类精度构建直觉模糊偏好关系矩阵,确定分类器权重,结合多属性群决策方法确定样本分类结果。在UCI中的7个数据集上进行测试,与目前流行的传统分类算法以及集成学习分类算法SVM,LR,NB,Boosting,Bagging相比,提出的算法分类平均精度分别提升了1.91%,3.89%,7.80%,3.66%,4.72%。该算法提高了传统分类方法的分类精度和泛化能力。

关键词:直觉模糊集;集成学习;分类;多属性群决策
中图法分类号　TP301.6

Ensemble Learning Algorithm Based on Intuitionistic Fuzzy Sets

DAI Zong-ming,HU Kai,XIE Jie and GUO Ya
Key Laboratory of Advanced Process Control for Light Industry,Ministry of Education,Jiangnan University,Wuxi,Jiangsu 214122,China
School of Internet of Things,Jiangnan University,Wuxi,Jiangsu 214122,China

Abstract　In order to improve the classification accuracy and generalization ability of traditional machine learning algorithms,this paper proposes an ensemble learning algorithm based on intuitionistic fuzzy sets (IFS-EL). The algorithm constructs an intuitionistic fuzzy preference relation (IFPR) matrix according to the classification accuracy of the traditional classifier. The matrix is used to determine the weights of the classifiers and the multi-criteria group decision making (MCGDM) is used to determine the sample classification result. The experimental data uses 7 classification data sets in UCI,and the training set and test set are divided into 7:3. The classification results are compared with the current popular traditional classification algorithms and ensemble learning classification algorithms,SVM,LR,NB,Boosting,Bagging,the average accuracy of the algorithm in this paper is improved by 1.91%,3.89%,7.80%,3.66%,4.72%. The experimental results show that the IFS-EL can improve the classification accuracy and generalization ability.

Keywords　Intuitionistic fuzzy sets,Ensemble learning,Classification,Multi-criteria group decision making

1　引言

分类是机器学习和模式识别等研究领域中最基本的任务之一,目前相关的分类算法已经十分成熟,例如决策树、朴素贝叶斯(Naïve Bayes,NB)、KNN、支持向量机(Support Vector Machine,SVM)、逻辑回归(Logistic Regression,LR)和神经网络[1]。但是,这些单一的模型存在泛化能力不足的问题,因此从上世纪末开始,集成学习(EnsembleLearning)算法成为机器学习领域的热门研究点,有权威学者认为集成学习是机器学习研究四大方向之首[2]。经过多年的实验证明,集成学习算法得到的模型要明显优于单个学习算法的模型,且泛化能力明显提高。1990年,Schapire证明了弱学习算法可以变成强学习算法,为集成学习奠定了理论基础,经典集成学习算法Boosting也是这之后产生[3]。1996年,Breiman提出了一种和Boosting相当的集成学习算法Bagging,该算法对不稳定学习器能够取得更好的结果[4]。这两种集成学习算法是最经典的,并且在分类任务中取得了良好的性能[5-6]。

经典模糊集理论是Zadeh于1965年提出的[7],经过长时间的发展,相继出现了区间模糊集、二型模糊集和直觉模糊集(Intuitionistic Fuzzy Set,IFS)等[8]。模糊集理论在处理不确定性时有巨大的优势,所以也有学者将模糊集理论应用在多属性群决策(Multi-Criteria Group Decision Making,MCGDM)中[9-11]。对于分类任务来说,分类器的结果也只是样本属于某一类的概率(0～1),而不是一个精确整数值(0或1),这与模糊理论中的隶属度是十分相似的,模糊集理论能够解决这种模糊性和不确定性。因此,对于多样本的分类过程也

基金项目:国家自然科学基金项目(71904064);测绘遥感信息工程国家重点实验室重点开放基金项目(18I04);江苏省自然科学基金(BK20190580);中央高校自主科研基金青年项目(JUSRP11922)
This work was supported by the National Natural Science Foundation of China(71904064),Open Research Fund of State Laboratory of Information Engineering in Surveying,the Mapping and Remote Sensing,Wuhan University(18I04),Natural Science Foundation of Jiangsu Province (BK20190580) and 111 Project and the Fundamental Research Funds for the Central Universities(JUSRP11922).
通信作者:郭亚(guoy@jiangnan.edu.cn)

相当于模糊决策中的多属性群决策问题。决策技术中最重要的层次分析法(Analytic Hierarchy Process，AHP)[12]是由 Saaty 提出的。该方法通过将复杂的问题分解为目标、标准、子标准和替代方案的多层层次结构，从而获得方案的排名。在经典的层次分析法模型中，不同条件下成对比较的相对幅度由数字表示。然而，在某些现实情况下，人们发现由于一些客观或主观原因，例如知识限制、个人兴趣、个人偏好、事物的复杂性和模糊性等，他们难以在比较判断中分配清晰的评估值。因此，为了提高层次分析法的能力，模糊集理论和直觉模糊集理论等已与经典层次分析法相结合，出现了一系列扩展方法[13-14]。

本文针对现存分类方法对分类结果的不确定性，提出一种基于直觉模糊集的集成学习算法(Ensemble learning algorithm based on intuitionistic fuzzy sets，IFS-EL)。不同于现存模糊决策系统中决策人判断的主观性，该算法所用的“决策人”为机器学习分类算法，更具有客观性。本算法旨在使用模糊决策方法提高传统机器学习算法的分类精度和泛化能力。

2 分类算法

2.1 单一分类器

本实验中拟采用 3 种常用的机器学习分类算法，分别为 SVM，NB 和 LR。

SVM 算法是以统计学习理论为基础的一种数据挖掘方法，能够非常成功地处理机器学习中的分类任务[15]。SVM 的核心思想在高维空间是寻找一个满足分类要求的最佳分类超平面，使得该平面能够将更多的正样本和负样本分开。对于线性不可分的情况，SVM 使用核函数将输入数据转换为一个更高维的向量空间中，找到最优超平面[16]。理论上，SVM 算法能够实现对线性数据的最优分类。

NB 算法是一种非规则分类，并且假定在给定类标记时属性值之间是相互独立的[17]。因此在给定类标记情况下，联合概率是每个属性概率的乘积。分类原则是通过使用贝叶斯公式计算每个属性的先验概率，即对象属于每个类的概率，然后计算后验概率并比较概率值的大小以确定类别。

LR 模型用于二分类因变量回归分析时是一种非线性的统计方法[18]。该模型不需要关于变量分布的假设条件，也不需要假设变量之间存在多元正态分布。在拟合 LR 模型参数估计时不采用最小二乘法，而是采用最大似然估计方法，最终的分类结果为事件发生的概率。

2.2 集成学习分类器

本实验还采用两种经典的集成学习算法，分别为 Boosting 和 Bagging。

Boosting 的基本思想是采用组合学习的方法，将预测精度不高的弱学习器集合成精度高的强学习器[19]。现在常使用的 Boosting 算法为 Adaboost 算法[20]。Boosting 算法的训练过程是对所有弱学习器赋予一个相同的权重，然后使用这些弱学习器对样本进行训练，对那些分类错误的样本赋予更大的权重再次重新训练。重复这一过程直到得出一个满意的结果。

Bagging 的基本思想是给定一弱学习器和训练集，每次训练是从训练集中随机抽取 n 个样本构成新的训练集。某些样本可能在训练集中出现多次，也有可能一次不出现。最终的分类结果是采用投票法将多次训练结果结合起来。Bagging 通过这种方法增加了学习器的差异度，从而提高了泛化能力[21-22]。

3 直觉模糊集理论和多属性群决策

对于 Zadeh 提出的经典模糊集理论，设 $X=\{x_1,x_2,\cdots,x_n\}$ 为非空集合，X 上的经典模糊集 A 定义为：

$$A=\{\langle x,\mu_A(x)\rangle \mid x\in X\} \tag{1}$$

其中，$\mu_A(x):X\to[0,1]$ 为隶属度函数，表明 x 属于 A 的程度。

为了更好地表达不确定信息，Atanassov 提出了直觉模糊集的概念[23]。设非空集合 $X=\{x_1,x_2,\cdots,x_n\}$，直觉模糊集 A 定义为：

$$A=\{\langle x,\mu_A(x),v_A(x)\rangle \mid x\in X\} \tag{2}$$

其中，$\mu_A(x):X\to[0,1]$ 和 $v_A(x):X\to[0,1]$ 分别表示 x 属于 A 的隶属度和非隶属度，二者满足 $0\leqslant\mu_A(x)+v_A(x)\leqslant 1$。

对于一个多属性群决策问题[24-25]，有 $A=\{A_1,A_2,\cdots,A_m\}$ 表示备选方案集，$C=\{C_1,C_2,\cdots,C_n\}$ 表示决策指标，$W=\{w_1,w_1,\cdots,w_n\}^{\mathrm{T}}$ 表示指标的权重，$E=\{e_1,e_2,\cdots,e_l\}$ 表示多个决策人的权重。在处理多属性决策问题的时候，通常采用二者的优劣来表达偏好关系[26]。相较于使用数值大小来评估单个方案，人们更容易比较两个方案的优劣[27]。直觉模糊集理论很容易通过比较两者的偏好关系构建直觉模糊偏好关系(Intuitionistic Fuzzy Preference Relation，IFPR)矩阵。

对于 $C=\{C_1,C_2,\cdots,C_n\}$，IFPR 矩阵如式(3)所示：

$$R=\begin{array}{c} \\ C_1\\ C_2\\ \vdots\\ C_n\end{array}\begin{array}{c}\begin{array}{cccc} C_1 & C_2 & \cdots & C_n\end{array}\\ \begin{pmatrix}(\mu_{11},\nu_{11}) & (\mu_{12},\nu_{12}) & \cdots & (\mu_{1n},\nu_{1n})\\ (\mu_{21},\nu_{21}) & (\mu_{22},\nu_{22}) & \cdots & (\mu_{2n},\nu_{2n})\\ \vdots & \vdots & \ddots & \vdots\\ (\mu_{n1},\nu_{n1}) & (\mu_{n2},\nu_{n2}) & \cdots & (\mu_{mn},\nu_{mn})\end{pmatrix}\end{array} \tag{3}$$

其中，$r_{ij}=(\mu_{ij},\nu_{ij})$ 为直觉模糊数(Intuitionistic fuzzy numbers，IFNs)，μ_{ij} 表示 C_i 相对偏好于 C_j 的程度，ν_{ij} 表示 C_j 相对偏好于 C_i 的程度。$\mu_{ij},\nu_{ij}\in[0,1]$，$\mu_{ij}=\nu_{ji}$，$\mu_{ij}+\nu_{ij}\leqslant 1$，$\mu_{ii}=\nu_{ii}=0.5$。比较直觉模糊数大小，可得到决策指标 C 中的最优指标 C_{best} 和最差指标 C_{worst}。

对 IFPR 矩阵中的任意两个直觉模糊数 $r_{ik}=(\mu_{ik},\nu_{ik})$ 和 $r_{tl}=(\mu_{tl},\nu_{tl})$，有如下运算法则[28]：

$$r_{ik}\oplus r_{tl}=(\mu_{ik}+\mu_{tl}-\mu_{ik}\cdot\mu_{tl},\nu_{ik}\cdot\nu_{tl}) \tag{4}$$

$$r_{ik}\otimes r_{tl}=(\mu_{ik}\cdot\mu_{tl},\nu_{ik}+\nu_{tl}-\nu_{ik}\cdot\nu_{tl}) \tag{5}$$

$$\lambda r_{ik}=(1-(1-\mu_{ik})^{\lambda},\nu_{ik}^{\lambda}),\lambda>0 \tag{6}$$

$$r_{ik}^{\lambda}=(\mu_{ik}^{\lambda},1-(1-v_{ik})^{\lambda}),\lambda>0 \tag{7}$$

在获得 IFPR 矩阵之后，需要判断一致性。在多属性群体决策问题中，一致性是一个非常重要的主题，缺乏一致性可能会导致产生误导性的结果。所以，获得合理的解决方案之前，需要先检查 IFPR 是否一致[29]。对于 IFPR 矩阵 $\boldsymbol{R}=(r_{ij})_{n\times n}$，其中 $r_{ij}=(\mu_{ij},\nu_{ij})$，有：

$$\begin{cases}\dfrac{1}{2}(1+\log_9\mu_{best,i})\times\dfrac{1}{2}(1+\log_9\mu_{ij})=\\ \quad\dfrac{1}{2}(1+\log_9\mu_{best,j})\\ \dfrac{1}{2}(1+\log_9\mu_{ij})\times\dfrac{1}{2}(1+\log_9\mu_{j,worst})=\\ \quad\dfrac{1}{2}(1+\log_9\mu_{i,worst})\end{cases} \tag{8}$$

$$\begin{cases}\frac{1}{2}(1+\log_9 \nu_{best,i})\times\frac{1}{2}(1+\log_9 \nu_{ij})=\\ \quad\frac{1}{2}(1+\log_9 \nu_{best,j})\\ \frac{1}{2}(1+\log_9 \nu_{ij})\times\frac{1}{2}(1+\log_9 \nu_{j,worst})=\\ \quad\frac{1}{2}(1+\log_9 \nu_{i,worst})\end{cases} \tag{9}$$

其中，$\mu_{best,worst}\in[0.5,1]$，$\nu_{best,worst}\in[0,0.5]$。

若式(8)和式(9)成立，IFPR 矩阵满足严格一致性。但是，当等式冲突时，一致性程度会降低。对于 μ_{ij} 和 ν_{ij}，最大程度上存在：

$$\begin{cases}\mu_{best,j}=\mu_{j,worst}=\mu_{best,worst}\\ \nu_{best,j}=\nu_{j,worst}=\nu_{best,worst}\end{cases} \tag{10}$$

引入变量 δ 和 ε，则式(8)和式(9)变为：

$$\begin{cases}(\frac{1}{2}(1+\log_9 \mu_{best,worst})-\varepsilon)\times(\frac{1}{2}(1+\log_9 \mu_{best,worst})-\varepsilon)=\frac{1}{2}(1+\log_9 \mu_{best,worst})+\varepsilon\\ (\frac{1}{2}(1+\log_9 \nu_{best,worst})-\delta)\times(\frac{1}{2}(1+\log_9 \nu_{best,worst})-\delta)=\frac{1}{2}(1+\log_9 \nu_{best,worst})+\delta\end{cases} \tag{11}$$

根据 $\mu_{best,worst}$ 和 $\nu_{best,worst}$ 的取值范围能够求出 δ 和 ε，可视为一致性指标 1(CI_1)和一致性指标 2(CI_2)。对于确定一致性程度，我们引入下面两个数学模型：

$$\begin{aligned}&\min\ \xi\\ &\text{s.t.}\ |\phi_{best}/\phi_j-9^{(2\times\mu_{best,j}^{-1})}|\leqslant\xi\\ &\quad|\phi_j/\phi_{worst}-9^{(2\times\mu_{j,worst}^{-1})}|\leqslant\xi\\ &\quad\sum_{j=1}^{n}\phi_j=1\\ &\quad\phi_{best}\geqslant\cdots\geqslant\phi_j\geqslant\cdots\geqslant\phi_{worst}\\ &\quad\phi_j\geqslant0,\xi\geqslant0\end{aligned} \tag{12}$$

$$\begin{aligned}&\min\ \zeta\\ &\text{s.t.}\ |\varphi_{best}/\varphi_j-9^{(2\times\nu_{best,j}^{-1})}|\leqslant\zeta\\ &\quad|\varphi_j/\varphi_{worst}-9^{(2\times\nu_{j,worst}^{-1})}|\leqslant\zeta\\ &\quad\sum_{j=1}^{n}\varphi_j=1\\ &\quad\varphi_{best}\leqslant\cdots\leqslant\varphi_j\leqslant\cdots\leqslant\varphi_{worst}\\ &\quad\varphi_j\geqslant0,\zeta\geqslant0\end{aligned} \tag{13}$$

根据这两个数学模型，算得 ξ^*，$(\phi_1^*,\phi_2^*,\cdots,\phi_n^*)^{\mathrm{T}}$，$\zeta^*$，$(\varphi_1^*,\varphi_2^*,\cdots,\varphi_n^*)^{\mathrm{T}}$ 和决策指标 $C=\{C_1,C_2,\cdots,C_n\}$ 的权重 $W^*=(w_1^*,w_2^*,\cdots,w_n^*)^{\mathrm{T}}=((\phi_1^*,\varphi_1^*),(\phi_2^*,\varphi_2^*),\cdots,(\phi_n^*,\varphi_n^*))^{\mathrm{T}}$。结合 ξ^*，ζ^*，CI_1 和 CI_2，由式(14)计算一致性比率 CR：

$$CR=\max\left\{\frac{\xi^*}{CI_1},\frac{\zeta^*}{CI_2}\right\} \tag{14}$$

CR 是检查权重 W^* 是否可靠的度量。CR 越小，说明一致性程度越好，IFPR 矩阵构建合理。当 $CR=1$ 时，说明一致性最差，IFPR 矩阵构造不合理，需要重新调整。

一致性检验通过之后，构建决策矩阵 $\boldsymbol{D}=(d_{ij})_{m\times n}$，其中 $d_{ij}=(\mu_{ij},\nu_{ij})$ 表示备选方案集 A_i 和决策指标 C_j 之间的直觉模糊关系。备选方案评估值 $U(A_i)$ 如式(15)所示，根据 $U(A_i)$ 获得所有备选方案集的排序。

$$U(A_i)=\bigoplus_{j=1}^{n}(w_j\otimes d_{ij})=\bigoplus_{j=1}^{n}((\phi_j,\varphi_j)\otimes(\mu_{ij},\nu_{ij})) \tag{15}$$

4 基于直觉模糊集的集成学习算法

本文提出的基于直觉模糊集的集成学习算法主要包括 5 个步骤，具体流程如图 1 所示。

(1)将原始数据集分为训练集和测试集，对数据预处理。

(2)依次采用 SVM，LR 和 NB 分类器进行训练和预测，并获得分类精度。

(3)由分类器精度对分类器进行排序，并构建分类器模糊偏好关系矩阵 IFPR。

(4)引入一致性指标，建立数学模型检查 IFPR 是否一致，若不一致，返回步骤(3)调整 IFPR；若一致，结合分类器性能排序，推导每个分类器权重。

(5)根据每个样本对每个分类器的隶属度，结合多属性群决策方法对所有测试样本进行排序，确定分类结果。

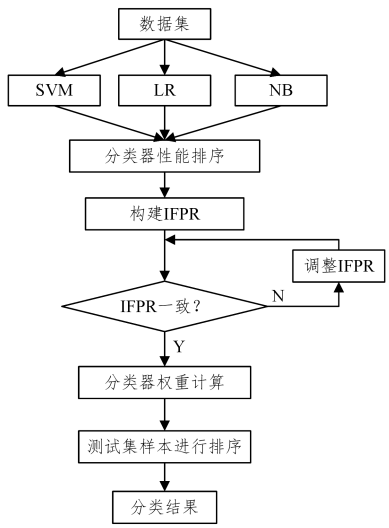


图 1 基于直觉模糊集的集成学习算法流程

Fig. 1 Process of IFS-EL

基于直觉模糊集的集成学习算法最重要的是构建 IFPR 矩阵，并由该矩阵计算分类器权重。

通过对分类器{SVM，LR，NB}分类精度的成对比较来确定 IFPR 矩阵 $\boldsymbol{R}$：

$$\boldsymbol{R}=\begin{array}{c}\\ \text{SVM}\\ \text{LR}\\ \text{NB}\end{array}\begin{array}{c}\begin{array}{ccc}\text{SVM}&\text{LR}&\text{NB}\end{array}\\ \begin{pmatrix}(\mu_{11},\nu_{11})&(\mu_{12},\nu_{12})&(\mu_{13},\nu_{13})\\ (\mu_{21},\nu_{21})&(\mu_{22},\nu_{22})&(\mu_{23},\nu_{23})\\ (\mu_{31},\nu_{31})&(\mu_{32},\nu_{32})&(\mu_{33},\nu_{33})\end{pmatrix}\end{array} \tag{16}$$

其中，$r_{12}=(\mu_{12},\nu_{12})$ 表示分类器 SVM 优于 LR 的程度。μ_{ij}，$\nu_{ij}\in[0,1]$，$\mu_{ij}=\nu_{ji}$，$\mu_{ij}+\nu_{ij}=1$，$\mu_{ii}=\nu_{ii}=0.5$。针对 $r_{ij}=(\mu_{ij},\nu_{ij})$，本文采用的策略如下：

$$\begin{cases}\mu_{ij}^*=\mu_{ij}+\nu_{ij}\times10\%\\ \nu_{ij}^*=\nu_{ij}-\mu_{ij}\times10\%\end{cases},\mu_{ij}>\nu_{ij} \tag{17}$$

$$\begin{cases}\mu_{ij}^*=\mu_{ij}\\ \nu_{ij}^*=\nu_{ij}\end{cases},\mu_{ij}=\nu_{ij} \tag{18}$$

$$\begin{cases}\mu_{ij}^*=\mu_{ij}-\nu_{ij}\times10\%\\ \nu_{ij}^*=\nu_{ij}+\mu_{ij}\times10\%\end{cases},\mu_{ij}<\nu_{ij} \tag{19}$$

根据上述公式，IFPR 矩阵 R 转变为 R^*：

$$\boldsymbol{R}^*=\begin{array}{c}\\ \text{SVM}\\ \text{LR}\\ \text{NB}\end{array}\begin{array}{c}\begin{array}{ccc}\text{SVM}&\text{LR}&\text{NB}\end{array}\\ \begin{pmatrix}(\mu_{11}^*,\nu_{11}^*)&(\mu_{12}^*,\nu_{12}^*)&(\mu_{13}^*,\nu_{13}^*)\\ (\mu_{21}^*,\nu_{21}^*)&(\mu_{22}^*,\nu_{22}^*)&(\mu_{23}^*,\nu_{23}^*)\\ (\mu_{31}^*,\nu_{31}^*)&(\mu_{32}^*,\nu_{32}^*)&(\mu_{33}^*,\nu_{33}^*)\end{pmatrix}\end{array} \tag{20}$$

很容易得到 $\mu_{ij}^*,\nu_{ij}^*\in[0,1]$,$\mu_{ij}^*=\nu_{ji}^*$,$\mu_{ij}+\nu_{ij}\leqslant 1$ 以及 $\mu_{ii}^*=\nu_{ii}^*=0.5$。

5 算法评估

5.1 数据集

为验证本文所提出的基于直觉模糊集的集成学习算法的性能,使用 UCI 机器学习数据库[30]中的 7 个分类数据集。由于多分类问题可以看作多个二分类问题,因此本实验针对二分类问题。这 7 个数据集信息如表 1 所列。样本数为 155～699,属性数为 9～60。

表 1 实验所用数据集

Table 1 Experimental data set

数据集	样本数	属性数	类别数
Breast	699	9	2
Credit	690	14	2
Heart	270	13	2
Hepatitis	155	19	2
Ionosphere	351	34	2
Sonar	208	60	2
Voting	435	16	2

5.2 评价指标

本实验采用分类精度(Accuracy)来评估算法的性能。对于二分类类别 C,有混淆矩阵如表 2 所列。

表 2 类别 C 的混淆矩阵

Table 2 Confusion matrix of class C

真实值	预测值	
	属于类别 C	不属于类别 C
属于类别 C	TP	FN
不属于类别 C	FP	TN

表 2 中,TP 为被正确分类为属于类别 C 的数量;FN 为被错误分类为不属于类别 C 的数量;FP 为被错误分类为属于类别 C 的数量;TN 为被正确分类为不属于类别 C 的数量。

根据分类的 4 种情况,分类的精度定义如下:

$$Accuracy=\frac{TP+TN}{TP+FN+FP+TN} \tag{21}$$

5.3 实验结果

本实验以传统分类算法 SVM,LR,NB 以及目前两个流行的集成学习算法 Boosting 和 Bagging 作为基准,对本文提出的算法进行评估。其中 Boosting 和 Bagging 采用默认的决策树算法。针对每个数据集,选择其中 70%作为训练集,剩余 30%作为测试集。表 3 列出了基准算法(SVM,LR,NB,Boosting 和 Bagging)以及本文所提出的基于直觉模糊集的集成学习算法在全部数据集上的精度比较。

下面针对 Breast 数据集介绍本文提出的集成学习算法的分类过程。

根据式(16),构建 SVM,LR 和 NB 之间的 IFPR 矩阵 $\boldsymbol{R}$,如式(22)所示,其中 $\{SVM,LR,NB\}=\{C_1,C_2,C_3\}$。$\mu_{ij}=\frac{Accuracy_{C_i}}{Accuracy_{C_i}+Accuracy_{C_j}}$,$\nu_{ij}=\frac{Accuracy_{C_j}}{Accuracy_{C_i}+Accuracy_{C_j}}$。

$$\boldsymbol{R}=\begin{pmatrix}(0.5,0.5) & (0.5025,0.4975) & (0.5037,0.4963)\\ (0.4975,0.5025) & (0.5,0.5) & (0.5013,0.4987)\\ (0.4963,0.5037) & (0.4987,0.5013) & (0.5,0.5)\end{pmatrix} \tag{22}$$

根据式(17)、式(18)和式(19),矩阵 $\boldsymbol{R}$ 转为 $\boldsymbol{R}^*$,如式(23)所示。

$$\boldsymbol{R}^*=\begin{pmatrix}(0.5,0.5) & (0.5522,0.4473) & (0.5534,0.4459)\\ (0.4473,0.5522) & (0.5,0.5) & (0.5511,0.4486)\\ (0.4459,0.5534) & (0.4486,0.5511) & (0.5,0.5)\end{pmatrix} \tag{23}$$

由矩阵 $\boldsymbol{R}^*$,可得 $C_{best}=C_1$,$C_{worst}=C_3$。由数学模型(12)得 $(\phi_1^*,\phi_2^*,\phi_3^*)^{\mathrm{T}}=(0.3854,0.3302,0.2844)^{\mathrm{T}}$,$\xi^*=0.0910$。由数学模型(13)可得到 $(\varphi_1^*,\varphi_2^*,\varphi_3^*)^{\mathrm{T}}=(0.2821,0.3312,0.3867)^{\mathrm{T}}$,$\zeta^*=0.0587$。分类器权重为:

$$\begin{aligned}\boldsymbol{W}^*&=(w_1^*,w_2^*,w_3^*)^{\mathrm{T}}\\&=((0.3845,0.2851),(0.3302,0.3312),(0.2844,0.3867))^{\mathrm{T}}\end{aligned}$$

根据 C_{best} 和 C_{worst},得 $\mu_{best,worst}=\mu_{1,3}=0.5534$,$\nu_{best,worst}=\nu_{1,3}=0.4459$,由式(11)得出一致性指标 CI_1 为 1.8124,一致性指标 CI_2 为 1.7556。由式(14)得出一致性比例 $CR=\max\left\{\frac{\xi^*}{CI_1},\frac{\zeta^*}{CI_2}\right\}=\max\left\{\frac{0.0910}{1.8124},\frac{0.0587}{1.7556}\right\}=0.0334$,一致性比例远低于 1,说明 IFPR 矩阵合理。最后,根据分类器对测试集样本的分类隶属度,构建决策矩阵 $\boldsymbol{D}=(d_{ij})_{m\times n}$,由式(15)得出测试集的排序,最终得到分类结果。表 4 列出了基于直觉模糊集的集成学习算法在各个数据集中的权重和一致性比例。表 3 中的一致性比例说明了本实验构建的 IFPR 具有合理性。

表 3 不同算法在各数据集上的精度比较

Table 3 Accuracy comparison of different algorithms on each dataset

(单位:%)

数据集	SVM	LR	NB	Boosting	Bagging	IFS-EL
Breast	96.19	95.24	94.76	94.76	95.24	**96.67**
Credit	83.57	85.20	84.06	82.61	84.01	**86.47**
Heart	82.72	82.72	85.19	77.78	79.01	**86.42**
Hepatitis	87.23	89.36	65.96	87.23	82.98	**91.49**
Ionosphere	91.51	85.85	89.62	**94.34**	90.57	92.45
Sonar	**90.48**	79.37	74.60	84.13	79.37	**90.48**
Voting	97.71	97.71	93.89	96.18	**98.47**	97.71

表 4 基于直觉模糊集的集成学习算法中各分类器权重和一致性比例

Table 4 Weights and consistency ratio of classifiers in IFS-EL

数据集	分类器权重	一致性比例
Breast	$((0.3845,0.2581),(0.3302,0.3312),(0.2844,0.3867))^{\mathrm{T}}$	0.0334
Credit	$((0.2841,0.3871),(0.3858,0.2817),(0.3301,0.3312))^{\mathrm{T}}$	0.0334
Heart	$((0.3047,0.3604),(0.3047,0.3604),(0.3907,0.2792))^{\mathrm{T}}$	<0.01
Hepatitis	$((0.3546,0.2908),(0.4138,0.2401),(0.2316,0.4692))^{\mathrm{T}}$	0.0277
Ionosphere	$((0.3917,0.2741),(0.2757,0.3992),(0.3326,0.3267))^{\mathrm{T}}$	0.0323
Sonar	$((0.4199,0.2459),(0.3191,0.3341),(0.2610,0.4200))^{\mathrm{T}}$	0.0298
Voting	$((0.3608,0.3024),(0.3608,0.3024),(0.2784,0.3952))^{\mathrm{T}}$	<0.01

如表 3 所列,本文提出的 IFS-EL 算法在数据集 Breast,Credit,Heart 和 Hepatitis 上均获得最高精度,分别为 96.67%,86.47%,86.42%,91.49%;对于 Sonar 数据集,IFS-

EL 精度和 SVM 分类器精度同为最优，为90.48%；对于 Ionosphere数据集，IFS-EL 精度(92.45%)仅低于 Boosting 集成算法(94.34%)，但高于 Bagging 集成算法(90.57%)；对于 Voting 数据集，IFS-EL 精度(97.71%)低于 Bagging 集成算法(98.47%)，但高于 Boosting 集成算法(96.18%)。我们从 UCI 数据集进行分析，二分类类别比例如表 5 所列。

表 5 不同数据集类别比例

Table 5 Proportion of different dataset categories

数据集	正负类比例
Breast	0.66:0.34
Credit	0.56:0.44
Heart	0.56:0.44
Hepatitis	0.55:0.45
Ionosphere	0.64:0.36
Sonar	0.53:0.47
Voting	0.86:0.14

从表 5 可知，当数据不平衡时(例如 Breast，Ionosphere 和 Voting)，经典集成学习算法 Boosting 和 Bagging 能克服这种困难，取得很好的分类结果，并且在数据集 Ionosphere 和 Voting 中，分类精度高于 IFS-EL 算法。而当数据平衡时，IFS-EL 算法精度大幅度提高。图 2 给出不同算法在所有数据集上分类精度的平均值。

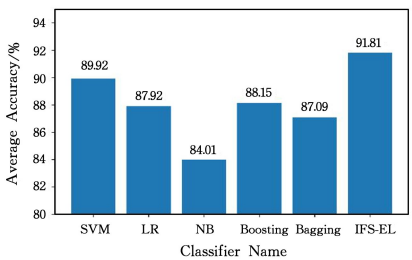


图 2 不同算法在各数据集上的平均精度比较

Fig. 2 Comparison of average accuracy of different algorithms on each dataset

由图 2 所示，本文提出的基于直觉模糊集的集成学习算法的平均精度为 91.81%，比 3 个传统分类算法 SVM，LR，NB 平均精度分别提高 1.91%，3.89%，7.80%，比集成学习算法中的 Boosting 高出 3.66%，比 Bagging 算法高出4.72%。结果表明，本文提出的集成学习算法是有效的。

结束语 为提高传统机器学习算法的分类精度和泛化能力，本文提出一种基于直觉模糊集的集成学习算法。该算法根据分类器的分类精度，结合直觉模糊集理论构建 IFPR 矩阵，利用该矩阵确定分类器权重，最后使用多属性群决策方法对样本进行分类。此外，对该算法和传统分类算法 SVM，LR，NB 以及集成学习算法 Boosting 和 Bagging 在 UCI 数据集上进行性能比较。实验表明，基于直觉模糊集的集成学习算法分类精度优于传统分类算法和集成学习算法，并且泛化能力有了一定的提高。

参 考 文 献

[1] HE Q, LI N, LUO W J, et al. A Survey of Machine Learning Algorithms for Big Data[J]. Pattern Recognition and Artificial Intelligence, 2014, 27(4): 327-336.

[2] DIETTERICH T G. Machine-learning research-Four current directions[J]. Ai Magazine, 1997, 18(4): 97-136.

[3] SCHAPIRE R E. The strength of weak learnability [J]. Proceedings of the Second Annual Workshop on Computational Learning Theory, 1989, 5(2): 197-227.

[4] BREIMAN L. Bagging predictors[J]. Machine Learning, 1996, 24(2): 123-140.

[5] BAUER E, KOHAVI R. An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants[J]. Machine Learning, 1999, 36(1): 105-139.

[6] JIANG Y, CHEN N, MING L T, et al. Bagging-based Probabilistic Neural Network Ensemble Classification Algorithm[J]. Computer Science, 2013, 40(5): 242-246.

[7] ZADEH L A. Fuzzy Sets[J]. Information & Control, 1965, 8(3): 338-353.

[8] XU J, SUN G, ZHAO J. Novel Correlation Coefficients Between Hesitant Fuzzy Sets and Their Applications in Multi-attribute Decision Making[J]. Acta Electronica Sinica, 2018, 46(6): 1327-1335.

[9] CHEN Y B, LIU P D. Multi-attribute decision-making approach based on intuitionistic trapezoidal fuzzy generalized heronian OWA operator[J]. Journal of Intelligent & Fuzzy Systems, 2014, 27(3): 1381-1392.

[10] XU Z S, YAGER R R. Dynamic intuitionistic fuzzy multi-attribute decision making[J]. International Journal of Approximate Reasoning, 2008, 48(1): 246-262.

[11] ZHANG X M, XU Z S. A new method for ranking intuitionistic fuzzy values and its application in multi-attribute decision making[J]. Fuzzy Optimization and Decision Making, 2012, 11(2): 135-146.

[12] SAATY T L. An exposition of the AHP in reply to the paper "remarks on the analytic hierarchy process"[J]. Management Science, 1990, 36(3): 259-268.

[13] LAI H F. Applying fuzzy AHP to evaluate the sustainability of knowledge-based virtual communities in healthcare industry [C]// Proceedings of the International Conference on Service Systems and Service Management. 2010.

[14] LIAO H, XU Z. Consistency of the fused intuitionistic fuzzy preference relation in group intuitionistic fuzzy analytic hierarchy process [J]. Applied Soft Computing, 2015, 35: 812-826.

[15] DING S F, QI B J, TAN H Y. An Overview on Theory and Algorithm of Support Vector Machines[J]. Journal of University of Electronic Science and Technology of China, 2011, 40(1): 1-10.

[16] SUN A, LIM E P, LIU Y. On strategies for imbalanced text classification using SVM: A comparative study[J]. Decision Support Systems, 2010, 48(1): 191-201.

[17] JIANG L X. Research on Naïve Bayes Classifiers and Its Improved Algorithms[D]. China University of Geosciences, 2009.

[18] JIANG G H, ZHANG F R, CHEN J W, et al. Analysis of the driving forces of change of rural residential areas in Beijing mountainousareas based on Logistic regression model [J]. Transactions of the Chinese Society of Agricultural Engineering, 2007, 5: 81-87.

[19] KHAN N M, KSANTINI R, AHMAD I S, et al. Covariance-guided One-Class Support Vector Machine[J]. Pattern Recognition, 2014, 47(6): 2165-2177.

(下转第 280 页)

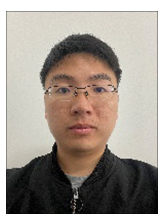
tection from Captured Images[C]// 2019 International Conference on Opto-Electronics and Applied Optics (Optronix). Kolkata, India, 2019: 1-4.
[7] NASRABADI N M. DeepTarget: An Automatic Target Recognition Using Deep Convolutional Neural Networks [J]. IEEE Transactions on Aerospace and Electronic Systems, 2019, 55(6): 2687-2697.
[8] MOU X, CHEN X, GUAN J, et al. Marine Target Detection Based on Improved Faster R-CNN for Navigation Radar PPI Images [C]// 2019 International Conference on Control, Automation and Information Sciences (ICCAIS). Chengdu, China, 2019: 1-5.
[9] HE H, WANG S, YANG D, et al. SAR target recognition and unsupervised detection based on convolutional neural network [C]// 2017 Chinese Automation Congress (CAC). Jinan, 2017: 435-438.
[10] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
[11] CHEN L B, CHEN Y Z, WANG X C, et al. Underwater image super-resolution reconstruction method based on deep learning [J]. Application Research of Computers, 2019, 39(9): 2738-2743.
[12] DU D, QI Y, YU H, et al. The unmanned aerial vehicle benchmark: Object detection and tracking [C]// Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 370-386.
[13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer Vision. Springer, Cham, 2016.
[14] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger [C]// IEEE. 2017: 6517-6525.
[15] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[J]. arXiv: 1804.02767, 2018.

NIU Kang-li, born in 2000, postgraduate. His main research interests include artificial intelligence and image processing.

CHEN Yu-zhang, born in 1984, Ph.D, associate professor. His main research interests include laser and LED in water, night vision or underwater scattering medium radiation transmission theory and computer simulation, image acquisition and restoration and reconstruction algorithms, image processing algorithms embedded including the research of Android development.

(上接第 274 页)

[20] FREUND Y, SCHAPIRE R E. A decision-theoretic generalization of on-line learning and an application to boosting[J] Journal of Computer and System Sciences, 1997, 55(1): 119-139.
[21] WANG L L. Research on Ensemble Learming Algorithm[D]. Guangxi University, 2006.
[22] SHEN X H, ZHOU Z H, WU J X, et al. Survey of Boosting and Bagging[J]. Computer Engineering and Applications, 2000, 12: 31-32, 40.
[23] ATANASSOV K T. On Intuitionistic Fuzzy Sets Theory [M]. Springer Berlin Heidelberg, 2012.
[24] XU X S. Approaches to Multiple Attribute Decision Making with Intuitionistic Fuzzy Preference Information [J]. Systems Engineering-Theory & Practice, 2007, 11: 62-71.
[25] XU Z S, SUN Z D. A Method Based on Satisfactory Degree of Alternative for Uncertainly Multi-Attribution Decision-Making [J]. Systems Engineering, 2001, 3: 76-79.
[26] XU Z. A survey of preference relations[J]. International Journal of General Systems, 2007, 36(2): 179-203.
[27] BUSTINCE H, PAGOLA M, MESIAR R, et al. Grouping, Overlap, and Generalized Bientropic Functions for Fuzzy Modeling of Pairwise Comparisons [J]. IEEE Transactions on Fuzzy Systems, 2012, 20(3): 405-415.
[28] XU Z. Intuitionistic preference relations and their application in group decision making[J]. Information Sciences, 2007, 177(11): 2363-2379.
[29] MOU Q, XU Z S, LIAO H C. A graph based group decision making approach with intuitionistic fuzzy preference relations [J]. Computers & Industrial Engineering, 2017, 110: 138-150.
[30] DUA D, GRAFF C. UCI Machine Learning Repository[D]. School of Information and Computer Sciences, 2017.



DAI Zong-ming, born in 1995, postgraduate. His main research interests include text classification and fuzzy decision.

GUO Ya, born in 1977, Ph.D, professor, Ph.D supervisor. His main research interests include system modeling and control, and deep learning.