

DOI: 10.7641/CTA.2015.40541

基于权重阈值寻优的案例推理分类器特征约简

赵辉^{1,2}, 严爱军^{1,2,3†}, 王普^{1,2}

(1. 北京工业大学 电子信息与控制工程学院, 北京 100124; 2. 数字社区教育部工程研究中心, 北京 100124;
3. 计算智能与智能系统北京市重点实验室, 北京 100124)

摘要: 为提高案例推理(case-based reasoning, CBR)分类器的分类准确率并降低时间复杂度, 本文提出了一种基于权重阈值寻优的特征约简策略. 首先通过基于数据驱动的方法对特征权重进行分配, 得到每个特征的权重结果; 其次, 设计特征权重重要度阈值的适应度函数, 并利用遗传算法对该重要度阈值进行优化搜索, 最后根据得到的优化阈值与特征的权重分配情况, 删除权重小于该阈值的特征从而完成特征的约简过程. 通过对比实验, 本文所提策略能够有效提高CBR分类器的分类准确率并降低时间复杂度, 表明了权重阈值寻优约简策略的可行性与优越性. 验证了本文方法不仅可以降低CBR分类器的时间复杂度, 而且能够提高CBR的决策与学习能力.

关键词: 案例检索; 特征权重; 阈值寻优; 特征约简

中图分类号: TP18 文献标识码: A

Feature reduction method based on threshold optimization for case-based reasoning classifier

ZHAO Hui^{1,2}, YAN Ai-jun^{1,2,3†}, WANG Pu^{1,2}

(1. College of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China;
2. Engineering Research Center of Digital Community, Ministry of Education, Beijing 100124, China;
3. Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing 100124, China)

Abstract: To improve the performance of case-based reasoning (CBR) classifier, we propose a feature reduction method based on threshold optimization for CBR classifier. First a data-driven method is adopted to conduct the feature weight distribution. Then, a weight threshold is introduced, where a genetic algorithm is utilized to obtain an appropriate threshold result, together with the feature weight and the threshold, the features of which the weights are lower than the threshold are deleted to accomplish the feature reduction process. The experimental results indicate that the weight distribution method and the threshold optimization method can improve the performance of CBR classifier, which confirms that the proposed reduction method is able to achieve a higher classification accuracy, decrease the time complexity, and improve the learning ability of CBR classifier.

Key words: case retrieval; feature weight; threshold optimization; feature reduction

1 引言(Introduction)

案例推理(case-based reasoning, CBR)是20世纪80年代人工智能领域中兴起的一种新的推理技术^[1]. 其问题求解过程包括检索、重用、修正和存储4个环节^[2], 即4R循环. 目前, 案例推理在故障诊断、智能决策、医疗诊断等方面得到了成功应用^[3-6]. 在模式分类领域中, CBR分类器的分类性能包括准确性^[7]、可靠性^[8]、学习能力^[9]等方面得到了广泛的关注. 由于CBR分类器对噪声和不精确的数据十分敏感, 而在实际应用中, CBR系统中可能存在大量的案例特征, 这

些特征数据往往存在着大量的信息冗余与重叠, 这些冗余的案例特征对系统带来的干扰不仅会造成算法时间复杂度的增加, 还会导致案例推理准确率的下降. 因此, 对于冗余特征的约简成为提高CBR分类器的分类准确率与减小运行时间的一个主要研究方向^[10].

特征约简是将与结果不相关或作用较小的冗余特征进行删除. 目前, 常用的特征约简算法主要有两种方式: 一种是在特征集中选择出冗余度最小的特征子集, 如主成分分析法^[11]、粗糙集^[12]等. 上述方法能够得到特征约简子集, 但还存在一定的局限性. 如主成

收稿日期: 2014-06-11; 录用日期: 2014-12-23.

†通信作者. E-mail: yanaijun@bjut.edu.cn; Tel.: +86 10-67396376.

国家自然科学基金项目(61374143), 北京市自然科学基金项目(4152010), 城市轨道交通北京实验室课题资助.

Supported by National Natural Science Foundation of China (61374143), Beijing Natural Science Foundation of China (4152010) and Beijing Laboratory for Urban Mass Transit.

分分析中提取出的主元是各个特征的组合,其具体含义难以说明;粗糙集需对数据进行离散化,可能会导致有用信息被约简.而另一种约简方式则是剔除重要度较小的特征^[13],即根据某种标准对每个案例特征分配相应的权重,然后设定一个特征重要度的阈值,将权重小于该阈值的特征剔除从而完成约简.这种方式的优点在于其约简效率较高,因而应用较为广泛.但其不足之处在于合理的重要度阈值难以得到.目前该重要度阈值主要由人工设定,如果设定了不合理的重要度阈值,就有可能导致约简掉有用的特征信息,进而影响整个系统的分类准确率和效率,不利于系统准确率的提高与时间复杂度的降低.因此,如何得到合理的特征重要度阈值来进行特征约简需要进一步研究.

针对上述问题,为提高CBR分类器的准确率与求解速度,本文提出了一种基于权重阈值寻优的特征约简策略.首先利用基于数据驱动的方法对案例的特征权重进行分配,然后利用遗传算法(genetic algorithm, GA)对案例空间进行搜索给出一个最优的重要度阈值,并将此阈值与特征权重结合,得到特征约简的子集.本文结构安排如下,第2节介绍传统的CBR分类器及存在的问题,第3节详细描述了权重阈值寻优的特征约简过程,第4节进行仿真实验和结果分析,最后,对本文所提方法与未来的研究方向进行了总结与展望.

2 CBR分类器与问题分析(CBR classifier and problem analysis)

2.1 CBR分类器(CBR classifier)

根据Aamodt & Plaza提出的4R认知模型^[2],如图1所示,典型CBR分类器的工作原理如下.

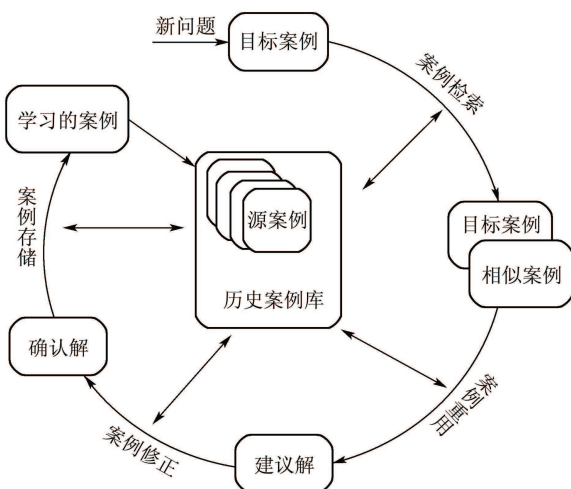


图1 CBR分类器原理图

Fig. 1 Schematic diagram of CBR

具体求解过程如下:

设历史案例库中的源案例可表示为如下的二元组形式:

$$H: \langle X_i; Y_i \rangle, i = 1, 2, \dots, m, \quad (1)$$

其中: m 表示源案例总数, X_i 和 Y_i 分别是第 i 个源案例的问题描述和分类结论, X_i 可用特征向量表示:

$$X_i = (x_{i1}, \dots, x_{ij}, \dots, x_{in}), \quad (2)$$

其中: n 为案例描述的特征个数, x_{ij} 为第 i 个案例描述中的第 j 个特征的值,并经过归一化处理.

记目标案例为 $T = (t_1, t_2, \dots, t_n)$,对应的分类结论为 Y_T .采用基于欧氏距离的KNN(K-nearest neighbor)检索策略^[14],计算目标案例 T 与每一个源案例 X_i 的相似度,记为

$$\begin{cases} s(T, X_i) = 1 - \sqrt{\sum_{j=1}^n \omega_j (t_j - x_{ij})^2}, \\ i = 1, 2, \dots, m, \end{cases} \quad (3)$$

其中 t_j 为目标案例中第 j 个特征的描述值,则 $s(T, X_i) \in [0, 1]$ 表示目标案例 T 与历史案例 X_i 之间的相似度; $\omega_j (j = 1, 2, 3, \dots, n)$ 表示第 j 个特征的权重,满足

$$\sum_{j=1}^n \omega_j = 1, 0 < \omega_j < 1. \quad (4)$$

在得到案例间的相似度后,将其按照降序进行排列:

$$\begin{cases} S = [s_1 \ s_2 \ \dots \ s_m], \\ s_1 > s_2 > \dots > s_m. \end{cases} \quad (5)$$

选取前 K 个相似度对应的源案例为 H_1, H_2, \dots, H_K ,根据重用原理,可采用最大相似度重用方式($K = 1$)或多数重用方式($K > 1$ 且为奇数)得到分类结果^[15].

$$\begin{cases} Y_T = Y(H_1), K = 1, \\ Y_T = \text{majority}(Y(H_i)), K > 1. \end{cases} \quad (6)$$

最后将相应的目标案例及类别存储进历史案例库中,完成CBR的知识存储和经验学习.

2.2 问题分析(Problem analysis)

上述过程即为传统的CBR分类器模型,当案例特征数据中存在着大量的信息冗余与重叠时,这些冗余的案例特征对系统带来的干扰不仅会造成算法时间复杂度的增加,还会导致分类准确率的下降.因此需要对冗余的特征进行约简.本文引入特征的重要度阈值来指导约简过程,即通过设定一个重要度的阈值,将权重小于该阈值的特征剔除从而完成约简.此约简策略存在着重要度阈值的合理选取问题.

由式(3)可知特征权重的结果不仅会影响分类器的分类准确率,还直接影响着后续特征约简的合理性;其次,重要度阈值的确定直接关系到特征约简的结果进而影响CBR分类器的分类性能.目前此阈值主要由

用户或专家根据自身经验来确定的, 主观性过强, 不利于系统准确率提高与时间复杂度的降低. 若选取不合适的重要度阈值, 就有可能导致约简掉有用的特征, 进而影响整个系统的分类准确率和效率.

3 特征约简(Feature reduction)

本节首先介绍权重的分配方法, 再通过遗传算法的选择、交叉、变异等算子对约简特征的权重阈值进行寻优. 在此基础上, 将权重小于此阈值的特征删除以完成约简过程.

3.1 权重分配(Weight distribution)

采用基于数据驱动的方法对特征权重进行优化分配, 设定权重指标如下:

$$C_{\omega} = \sum_{j=1}^n \log_2(1 + \rho_j^2 \omega_j), \quad (7)$$

其中 C_{ω} 值越大说明该特征提供的信息量越大, 重要度越高, 相应的特征权重也越大. 因此, 若要得到最优的权重分配方案需要使该指标达到最大值. ρ_j^2 则表示第 j 个特征的重要度指标, 即案例库中第 j 个特征的数据波动性.

为了使权重指标 C_{ω} 达到最大值, 在式(4)的约束条件下, 构造拉格朗日函数如下:

$$L(\omega, \lambda) = \sum_{j=1}^n \log_2(1 + \rho_j^2 \omega_j) + \lambda(1 - \sum_{j=1}^n \omega_j), \quad (8)$$

其中 λ 是拉格朗日算子, 各特征分配到的权重可利用该函数对权重 ω_j 求偏导后得到

$$\omega_j = \max(0, (\varepsilon - 1/\rho_j^2)), \quad (9)$$

其中 ε 为门限值, 其值可通过约束条件和式(9)求得. 详细求解过程见文献[13].

3.2 权重阈值寻优(Weight threshold optimization)

利用特征权重完成约简的策略需要解决另一个重要问题, 即重要度阈值的确定问题. 目前此阈值主要由用户或专家根据自身经验来确定的, 主观性过强, 有可能导致约简掉有用的特征而不利于系统准确率的提高与时间复杂度的降低. 因此, 本文采用遗传算法(GA)对特征的重要度阈值进行寻优, 以期搜索到最为合理的重要度阈值从而在保证系统分类准确率的同时降低时间复杂度.

遗传算法是一种求解问题的高度并行的全局随机搜索方法^[16-17]. 它遵循“优胜劣汰”的自然选择策略, 在搜索过程中自动获取和积累有关搜索空间的知识, 并自适应地控制搜索过程以求得最优解. 利用遗传算法实现特征重要度阈值的寻优, 需要实现以下操作:

1) 编码. 本文选择最常用的二进制编码, 采用了5位2进制编码;

2) 种群初始化. 随机选择 P_s 个染色体组成一个群体, 群体内个体的数量就是群体规模, 每个初始个体

就表示这问题的初始解;

3) 适应度函数. 在GA寻优过程中, 一个好的适应度函数能够指导寻优的方向, 因此, 为了保证搜索到的重要度阈值能够有效地提高整体系统的准确率, 定义如下的适应度函数:

$$\text{Fitness} = N'/N, \quad (10)$$

其中: N' 表示在当前阈值下进行特征约简后, 能够正确分类的案例个数, N 表示用于训练的案例总数. 该函数可以保证找到分类准确率高且所含特征数目小的重要度阈值, 即最优解;

4) 选择. 本文按照轮盘赌法选择合适的个体, 设种群大小为 P_s , 其中每个个体的适应度为 f_i , 则个体 i 被选择的概率为

$$P_i = f_i / \sum_{i=1}^{P_s} f_i. \quad (11)$$

5) 交叉. 本文采用单点交叉: 在个体基因串中随机设定交叉点, 当实行染色体交叉时, 两个个体交叉点前后的基因以概率 P_c 互换, 并生成两个新个体;

6) 变异. 本文采用基本变异算子, 对群体中染色体随机挑选一个或几个基因位置以概率 P_m 进行变异, 进行取反操作.

本文在得到各个特征权重后, 采用GA进行重要度阈值寻优过程的伪代码如下所示:

```

BEGIN
Initialize threshold population  $P_o(t)$ ;
population size  $P_s$ ;
mutation probability  $P_m$ ;
crossover probability  $P_c$ ;
iteration number Iter, the feature weight; // 初始化,
给定种群数量, 交叉变异概率, 迭代次数与得到的权重
While ( $t < \text{Iter}$ )
{
Decode the threshold,
Reduct the attribute, // 此阈值下进行特征约简
Evaluate Fitness; // 适应度计算
Select operation,
Crossover operation,
Mutation operation; // 选择, 交叉, 变异操作
 $P_o(t+1) = P_o(t)$ ;
 $t++$ ;
}
Return the best fitness;
Decode the threshold; // 得到最优的重要度阈值
END

```

通过上述步骤, 就可得到优化的重要度阈值结果 S_{λ} , 利用此结果就结合式(9)确定的特征权重将权重小

于该阈值的特征删除掉,从而完成特征约简过程.

4 实验结果(Experimental results)

为了考察本文提出的基于权重阈值寻优特征约简方法的有效性,选取UCI数据库中的10组分类数据集进行实验^[18],具体信息如表1所示.

表1 实验数据集基本信息

Table 1 Basic information of the datasets

序号	数据集	案例个数/个	特征个数/个	类别/类
1	glass	214	9	6
2	seeds	210	7	3
3	fertility	100	9	2
4	Parkinsons	195	23	2
5	blood transfusion	748	4	2
6	statlog (vehicle)	846	18	4
7	ionosphere	351	33	2
8	pima	768	8	2
9	image segment	2310	18	7
10	bench	208	60	2

为了测试本文方法的效果,采用5折交叉验证方法,并与一些典型分类方法进行分类准确率的对比,设计实验方案如下:

实验1 为验证本文约简策略中特征权重分配的

合理性与优越性,首先将本文采用的数据驱动方法分配权重的CBR分类器(记为WF&CBR),并与传统均权CBR(记为MCCR)、熵权法分配权重的CBR(记为EN&CBR)、遗传算法分配权重的CBR(记为GA&CBR)以及常用的分类算法(选取J48决策树算法与RBF神经网络)作对比;

实验2 为验证权重阈值寻优策略的可行性,在权重的基础上利用GA进行重要度阈值的优化搜索,并利用此阈值指导特征约简,并将约简后特征子集的分类准确率记为WF&CBR&RD.

实验中用到的一些参数值设定为:实验中 $K = 7$;RBFNN中采用Gaussian函数作为径向基函数,实验2遗传算法(GA)中设定初始种群规模 $P_s = 30$,二进制编码位数为5位,变异概率 $P_m = 0.06$,交叉概率 $P_c = 0.3$,迭代次数 $Iter = 30$.

4.1 约简前的结果(Results without reduction)

根据实验1,利用式(7)–(9)对数据集中每个特征进行权重的分配,根据每个特征的重要度指标得出优化后的权重;并利用此权重进行CBR的分类得到WF&CBR的分类准确率,并与传统MCCR, GA&CBR, EN&CBR, J48, RBFNN 进行准确率以及运行时间的对比,对比结果如表2所示.

表2 6种方法的分类准确率与运行时间对比

Table 2 Comparison results between 6 methods

数据	MCCR		EN&CBR		GA&CBR		WF&CBR		J48		RBFNN	
	准确率/%	时间/s	准确率/%	时间/s	准确率/%	时间/s	准确率/%	时间/s	准确率/%	时间/s	准确率/%	时间/s
glass	67.26	0.047	65.90	0.047	69.62	4.914	71.04	0.078	65.42	0.050	68.69	0.764
seeds	92.86	0.016	92.38	0.026	92.38	4.243	93.81	0.016	89.05	0.010	90.95	0.718
fertility	81.00	0.016	82.00	0.008	81.00	1.123	84.00	0.016	84.00	0.010	87.00	0.359
Parkinsons	91.79	0.031	90.26	0.040	88.21	6.911	92.82	0.031	89.74	0.020	82.05	0.655
blood transfusion	61.12	0.234	61.25	0.221	61.65	43.560	63.52	0.265	77.27	0.080	78.61	1.123
statlog (vehicle)	68.92	0.561	66.19	0.555	71.52	113.650	72.34	0.593	73.29	0.110	65.37	1.529
ionosphere	85.75	0.140	85.75	0.152	90.01	30.330	89.16	0.156	89.74	0.050	90.88	0.764
pima	69.92	0.296	66.28	0.294	67.46	59.610	69.77	0.296	71.22	0.080	75.26	1.061
image segment	97.27	3.900	96.36	4.035	97.23	861.620	97.58	4.134	96.45	0.280	87.32	4.711
bench	77.83	0.078	79.29	0.096	78.87	17.000	78.77	0.094	71.15	0.080	74.04	0.640
mean	79.37	0.532	78.57	0.547	79.80	114.296	81.28	0.568	80.73	0.077	80.02	1.232

由表2可看出,10个数据集中WF&CBR的分类准确率大多要比MCCR, EN&CBR与GA&CBR的准确率要高,表明了本文利用数据驱动方法进行特征权重分配能够得到更为精确的权重结果,提高了CBR的分类准确率,为后续阈值寻优约简提供了合理的权重结果.与J48, RBFNN方法相比,WF&CBR在多数的数据集中分类准确率要更优越些,将表2中10个数据集的结果计算出平均分类准确率,

并作图如图2所示.由图2中可看出WF&CBR的平均分类准确率最高,体现出了此方法在分类准确率上的优越性,进一步验证了本文采用的特征权重分配方法能够充分挖掘出特征中潜在的重要度信息,得到更加合理的权重分配结果从而提高了CBR分类器的准确率.此外,从运行时间上考虑,按升序排列为: J48, MCCR, EN&CBR, WF&CBR, RBFNN, GA&CBR. 可看出WF&CBR在运行时间上并不占

优势. 为解决此问题, 本文在后续进行了重要度阈值的寻优, 通过将冗余的特征删除掉, 来减少WF&CBR的时间复杂度.

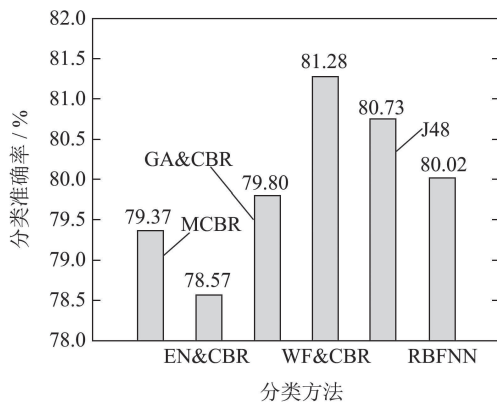


图2 平均分类准确率

Fig. 2 Average classification accuracy

表3 阈值寻优约简后的分类准确率与时间对比

Table 3 Comparison results of classification accuracy and the cost time with GA for reduction

数据	WF&CBR/%	WF&CBR&RD/%	约简的特征/个	约简率/%	约简前运行时间/s	约简后运行时间/s
glass	71.04	72.91	5	55.56	0.078	0.016
seeds	93.81	93.81	0	0	0.016	0.016
fertility	84.00	88.00	6	66.67	0.016	0.010
Parkinsons	92.82	93.33	14	60.87	0.031	0.010
blood transfusion	63.52	69.66	2	50.00	0.265	0.203
statlog (vehicle)	72.34	72.69	7	38.89	0.593	0.406
ionosphere	89.16	90.30	11	33.33	0.156	0.109
pima	69.77	69.77	4	50.00	0.296	0.218
image segment	97.58	97.58	4	22.22	4.134	3.416
bench	78.77	78.77	34	56.67	0.094	0.031
mean	81.28	82.68	8.7	43.42	0.568	0.443

表3与表2中其他分类方法相比, WF & CBR & RD的平均分类准确率达到82.68%, 比MCBR, EN & CBR, GA & CBR, WF & CBR, J48, RBFNN分别高出约3.31%, 4.11%, 2.88%, 1.40%, 1.95%, 2.66%. 表明本文所采用的阈值寻优方法能够有效地约简掉冗余的特征, 从而进一步提高了CBR的分类准确率; 从系统的运行时间看, 特征约简后的分类运行时间得到了一定程度的减低, 平均来看约简后的运行时间比约简前的降低了约21.91%.

此外, 为考察GA算法中, 种群数量 P_s 与迭代次数Iter对分类结果及收敛速度的影响, 分别选取了种群数量为10, 15, 20, 30时, 10个数据集分类准确率的变化情况, 如图3所示. 由图3可看出, 在设定的迭代次数Iter内, 10个数据集的平均分类准确率均达到了收敛, 随着种群数量的增加, 平均分类准确率达到收敛的迭代次数Iter有减少的趋势, 且在

4.2 阈值寻优约简的结果 (Results of feature reduction)

根据表2得到的结果可得知, 利用特征权重分配方法能够显著提高CBR的分类准确率, 与其他权重分配方法的对比表明了该权重结果的优越性. 为后续特征的约简提供了合理的前提条件.

本节在得到优化的特征权重后, 引入了重要度阈值, 采用了GA对重要度阈值进行寻优, 并根据此阈值完成对案例特征集的约简以减小算法的时间复杂度. 并利用约简后的数据集进行CBR的分类实验, 其分类准确率结果与运行时间如表3所示. 从表3可以看出, 10个数据集都约简掉了若干特征, 不同数据集其约简率也有所变化. 从分类准确率方面看, 10个数据集在特征约简后的分类准确率不仅没有下降反而均有所提高.

$P_s = 30$ 时, 收敛速度最快.

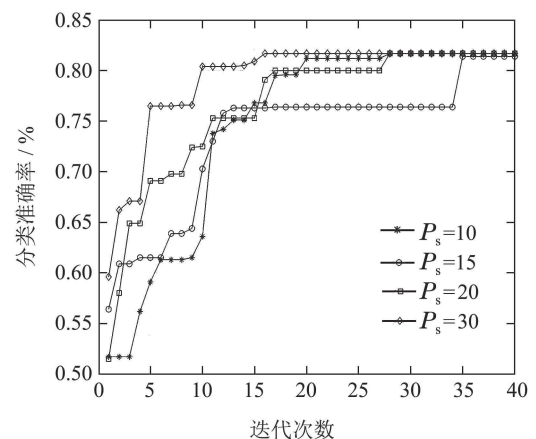


图3 不同种群数量与迭代次数情况下的平均分类准确率

Fig. 3 Average classification accuracy under different population and iteration

4.3 显著性检验(Significance test)

为了验证本文所提约简方法是否能够显著提高分类性能,可进行 T 检验,通过检验的概率 P 值与显著水平的大小,来判断两个样本平均数与各自所代表的总体的差异是否显著:当 P 大于显著水平 α 时,则表示二者无差异或差异不明显; P 小于显著水平 α 时,则表示有差异或差异明显, P 值越小差异越显著.表4则是WF&CBR&RD, EN&CBR, GA&CBR, J48, RBFNN与传统MBCR间的分类性能的成对 T 检验结果(显著水平为0.05).

表4 分类性能的显著性检验($\alpha = 0.05$)

Table 4 Significance test for the classification performance

方法	GA&CBR	EN&CBR	WF&CBR&RD	J48	RBFNN
P	0.291	0.071	0.0034	0.255	0.406
结论	不显著	不显著	显著	不显著	不显著

由表4看出,在显著水平 α 为0.05时,WF&CBR&RD检验得到的概率 P 值小于0.05,而GA&CBR, EN&CBR, J48, RBFNN检验得到的 P 值均大于0.05,说明与其他4种方法相比,本文所提基于权重阈值寻优的特征约简策略能够显著提高CBR的分类性能.

综上所述,本文提出的权重阈值寻优特征约简策略包含特征的权重分配与重要度阈值的寻优,由上述实验表明通过数据驱动方法对特征权重进行分配能够有效利用特征中潜在的重要度信息,从而得到了更为精确的权重分配结果,与传统均权CBR, EN & CBR, GA & CBR和常用的分类方法J48以及RBFNN的对比表明,WF&CBR能够提高分类准确率,验证了该方法的可行性与权重结果的优越性;在得到合理的权重后,本文采用了GA对特征重要度阈值进行了寻优,并利用此阈值进行了特征的约简,实验结果表明约简后的特征不仅降低了运行的时间还进一步提高了CBR的分类准确率,表明了本文所提约简策略能够提高CBR系统的性能与效率.

5 结论(Conclusions)

为提高CBR的分类准确率并降低时间复杂度,本文提出了一种基于权重阈值寻优的特征约简策略.首先通过数据驱动方法计算出特征权重,其次引入重要度阈值,并采用GA对数据集的重要度阈值进行优化搜索,根据得到的阈值指导特征约简.对比实验表明,基于权重阈值寻优的特征约简方法能够进一步提高CBR的分类准确率,并降低了系统的运行时间,提高了CBR的学习能力.由于特征权

重分配的合理程度直接影响阈值寻优以及特征约简的结果,虽然实验表明,本文采用的数据驱动的方法具有一定的合理性,但这种分配机制是一种静态的分配机制,缺乏迭代学习能力.而CBR是一个增量学习过程,随着CBR分类器的运行,案例库规模会不断增加,此时静态的权重分配机制不足以得出合理的权重结果进而影响到特征约简的结果.因此,下一步的研究主要包括对权重动态学习机制的研究,以保证CBR分类器的运行性能.

参考文献(References):

- [1] SCHANK R. *Dynamic Memory* [M]. New York: Cambridge University Press, 1982.
- [2] AAMODT A, PLAZA E. Case-based reasoning: foundational issues, methodological variations, and system approaches [J]. *AI Communications*, 1994, 7(1): 39 – 59.
- [3] FENTON B, MCGINNITY M, MAGUIRE L. Fault diagnosis of electronic system using artificial intelligence [J]. *IEEE Instrumentation & Measurement Magazine*, 2006, 5(3): 16 – 20.
- [4] KAVITHA D M K, VENKATESH P. An intelligent decision support system for e-purchasing using CBR and CF [J]. *International Journal of Agent-Oriented Software Engineering*, 2009, 3(3): 212 – 229.
- [5] LOPEZ B, POUS C, GAY P, et al. A framework for case-based medical diagnosis development and experimentation [J]. *Artificial Intelligence in Medicine*, 2011, 51(2): 81 – 91.
- [6] GU D X, LIANG C Y, BICHINDARUTZ I, et al. A case-based knowledge system for safety evaluation decision making of thermal power plants [J]. *Knowledge-Based Systems*, 2012, 26(1): 185 – 195.
- [7] SAMPAIO L N, TEDESCO P, MONTEIRO J, et al. A knowledge and collaboration-based CBR process to improve network performance-related support activities [J]. *Expert Systems with Applications*, 2014, 41(11): 5466 – 5482.
- [8] XU X, WANG K, MA W, et al. Improving the reliability of case-based reasoning systems [J]. *International Journal of Computational Intelligence Systems*, 2010, 3(3): 256 – 265.
- [9] GUO Y, HU J, PENG Y H. A CBR system for injection mould design based on ontology: a case study [J]. *Computer-Aided Design*, 2012, 44(6): 496 – 508.
- [10] SALAMO M, LOPEZ S M. Rough set based approaches to feature selection for Case-Based Reasoning classifiers [J]. *Pattern Recognition Letters*, 2011, 32(2): 280 – 292.
- [11] 范雪莉, 冯海泓, 原猛. 基于互信息的主成分分析特征选择算法 [J]. *控制与决策*, 2013, 28(6): 915 – 919. (FAN Xueli, FENG Haihong, YUAN Meng. PCA based on mutual information for feature selection [J]. *Control and Decision*, 2013, 28(6): 915 – 919.)
- [12] 常春光, 汪定伟, 胡馄元, 等. 基于粗糙集的案例属性约简技术 [J]. *控制理论与应用*, 2006, 23(6): 867 – 872. (CHANG Chunguang, WANG Dingwei, HU Kunyuan, et al. Rough-set based reduction technique for case attributes [J]. *Control Theory & Applications*, 2006, 23(6): 867 – 872.)
- [13] 赵辉, 严爱军, 王善. 提高案例推理分类器的可靠性研究 [J]. *自动化学报*, 2014, 40(9): 2029 – 2036. (ZHAO Hui, YAN Aijun, WANG Pu. On improving reliability of case-based reasoning classifier [J]. *Acta Automatica Sinica*, 2014, 40(9): 2029 – 2036.)

- [14] 魏书宁, 王耀南, 印峰, 等. 基于 k -最近邻分类增强学习的除冰机器人抓线控制 [J]. 控制理论与应用, 2012, 29(4): 470 – 476.
(WEI Shuning, WANG Yaonan, YIN Feng, et al. Line-grasping control of de-icing robot based on k -nearest neighbor reinforcement learning [J]. *Control Theory & Applications*, 2012, 29(4): 470 – 476.)
- [15] CARLOS A C, MARCOS A L, CELSON P L, et al. Assessment of a product range model concept to support design reuse using rule based systems and case based reasoning [J]. *Advanced Engineering Informatics*, 2012, 26(2): 292 – 305.
- [16] KIM M, LEE S, WOO S, et al. Approximate cost estimating model for river facility construction based on case-based reasoning with genetic algorithms [J]. *KSCE Journal of Civil Engineering*, 2012, 16(3): 283 – 292.
- [17] 金敏, 鲁华祥. 一种遗传算法与粒子群优化的多子群分层混合算法 [J]. 控制理论与应用, 2013, 30(10): 1231 – 1238.
(JIN Min, LU Huaxiang. A multi-subgroup hierarchical hybrid of genetic algorithm and particle swarm optimization [J]. *Control Theory & Applications*, 2013, 30(10): 1231 – 1238.)
- [18] FRANK A, ASUNCION A. *UCI machine learning repository* [D]. Irvine, CA: University of California. <http://archive.ics.uci.edu/ml>.

作者简介:

赵 辉 (1988–), 男, 博士研究生, 主要研究方向为人工智能、智能控制等, E-mail: zhaohui2012@emails.bjut.edu.cn;

严爱军 (1970–), 男, 副教授, 主要研究方向为人工智能、过程建模与优化控制等, E-mail: yanaijun@bjut.edu.cn;

王 普 (1962–), 男, 教授, 主要研究方向为信息处理、智能控制等, E-mail: wangpu@bjut.edu.cn.