

## 案例推理属性权重的分配模型比较研究

严爱军<sup>1</sup> 钱丽敏<sup>1</sup> 王普<sup>1</sup>

**摘要** 案例推理系统中各属性权重的赋值决定了案例之间的相似度大小, 进而对推理结果的正确与否产生显著影响. 以属性加权  $K$ -最近邻相似案例检索为基础, 讨论了使用注水原理分配属性权重的机理, 并通过建立权重分配的合理性指标, 构造拉格朗日函数对权重进行优化求解, 得到了收敛的注水分配算法. 通过五折交叉的模式分类实验, 分别对属性权重的平均分配法、注水分配算法和遗传算法分配法进行了比较研究, 案例推理分类结果证明, 在引入注水分配算法后, 其分类性能得到有效改善.

**关键词** 案例推理, 属性权重, 注水原理, 模式分类

**引用格式** 严爱军, 钱丽敏, 王普. 案例推理属性权重的分配模型比较研究. 自动化学报, 2014, 40(9): 1896–1902

**DOI** 10.3724/SP.J.1004.2014.01896

### A Comparative Study of Attribute Weights Assignment for Case-based Reasoning

YAN Ai-Jun<sup>1</sup> QIAN Li-Min<sup>1</sup> WANG Pu<sup>1</sup>

**Abstract** The attribute weights assignment in case-based reasoning (CBR) system may determine the similarities between cases, and thus it has a significant impact on the correctness of reasoning. To improve the reasoning performance, the water-filling theory is introduced to the attribute weights assignment in this paper. Reasonable indicators of weight distribution are established, an associated Lagrange function is constructed and the weight optimization solution can be achieved. Thereby a convergent water-filling assignment (WFA) algorithm is obtained which can be used in the weighted  $K$ -nearest neighbor rule to retrieve similar cases. Classification experiments for comparison between the mean assignment method, WFA method and genetic algorithms for the attribute weights using the 5-fold cross-validation method are conducted. The results show that the classification performance of CBR can be further increased after the attribute weights are assigned by WFA.

**Key words** Case-based reasoning (CBR), attribute weights, water-filling theory (WFT), pattern classification

**Citation** Yan Ai-Jun, Qian Li-Min, Wang Pu. A comparative study of attribute weights assignment for case-based reasoning. *Acta Automatica Sinica*, 2014, 40(9): 1896–1902

案例推理 (Case-based reasoning, CBR) 是人工智能领域一种较新的推理技术和机器学习方法, 是一种利用过去解决类似问题的经验进行推理求解新问题的方法. 1982 年, Schank<sup>[1]</sup> 提出了动态记忆理论, 描述了 CBR 的雏形. 1983 年, Kolodner<sup>[2]</sup> 开发完成了第一个问答式的案例推理系统. 1994 年, Aamodt 等<sup>[3]</sup> 给出了经典的案例推理认知模型, 即案例检索、案例重用、案例校正和案例存储. 从此, 基于案例的机器学习和推理研究在如何提高其求解性能方面得到了广泛关注<sup>[4–5]</sup>, 特别是 CBR 在优化控制、模式识别等领域取得了卓有成效的工作<sup>[6–9]</sup>, 显示出 CBR 的研究价值及应用潜力.

在 CBR 的认知模型中, 案例检索是问题求解的关键, 从国内外关于案例检索的研究现状看, 目前应用较广泛的是基于  $K$ -最近邻 ( $K$ -nearest neighbor, KNN)<sup>[10]</sup> 的检索策略. 依据属性权重及其特征值计算出待求解的目标案例与案例库中的源案例之间的相似度, 然后选择一个或一些相似度高的源案例解答作为案例重用的基础<sup>[11]</sup>. 在计算相似度时, 权重的分配情况会对计算结果和问题求解的质量产生显著影响. 一般起主要作用的属性, 分配较大的权重; 反之, 则赋予较小的权重, 而 KNN 一般使用均权法, 会影响计算结果的可靠性, 因而权重的分配成为一个重要的研究方向<sup>[12]</sup>. 目前, 确定权重的方法主要有主观分析法和客观分析法. 主观分析法常用的定权方法有专家咨询法、调查统计法、无差异折衷法、相关分析法<sup>[13]</sup>、层次分析法<sup>[14]</sup> 等. 其中, 层次分析法应用较为广泛. 专家咨询法、调查统计法和无差异折衷法一般是获取专家的先验知识, 然后采用相应方法确定属性的权重值. 相关分析法是一种基于数学统计的方法, 相对于前 3 种方法虽然有一定的

收稿日期 2013-05-29 录用日期 2014-02-26  
Manuscript received May 29, 2013; accepted February 26, 2014  
国家自然科学基金项目 (61374143)  
Supported by National Natural Science Foundation of China (61374143)  
本文责任编辑 刘德荣  
Recommended by Associate Editor LIU De-Rong  
1. 北京工业大学电子信息与控制工程学院 北京 100124  
1. College of Electronic Information & Control Engineering, Beijing University of Technology, Beijing 100124

进步, 但主观性仍然较强. 所以, 从总体来看, 这几种确定权重的方法过分依赖于主观判断和经验, 会给相似案例检索的质量带来不确定性影响. 于是, 对权重进行优化分配的客观分析法相继被提出, 比如遗传算法 (Genetic algorithms, GA)<sup>[15]</sup>、神经网络法<sup>[16-17]</sup>、模拟退火算法<sup>[18]</sup>等, 这些方法在权重的优化方面进行了有益探索, 但由于其自身固有一些缺陷 (如神经网络的结构很难有一个标准, 遗传算法容易陷入局部极小等), 因此还需进一步研究权重优化分配的学习机制, 以提高问题求解的质量.

本文针对案例推理属性权重的分配问题. 首先, 根据注水原理建立权重分配的合理性指标; 然后, 通过构造拉格朗日函数进行优化计算, 获得一种权重的注水分配 (Water-filling assignment, WFA) 算法; 最后, 采用五折交叉验证法对 CBR、KNN、支持向量机 (Support vector machine, SVM) 的学习性能进行了比较, 同时测试了权重的平均分配法 (Mean assignment method, MA)、遗传算法、注水分配法的效果. 本文的结构是: 第 1 节介绍 CBR 模型的实现算法, 第 2 节讨论属性权重的分配方法和时间复杂度, 第 3 节分析算法的收敛性, 第 4 节通过五折交叉验证法对各种方法的性能进行考察, 最后给出结论.

## 1 案例推理模型及算法

根据经典的 CBR 认知模型<sup>[3]</sup>, CBR 用于问题求解时的一个循环过程是: 新问题出现时形成目标案例, 首先通过案例检索将目标案例与源案例进行类比, 检索出相似案例; 接着重用相似案例的解决方案 (称为建议解), 如果建议解不足以解决新问题, 则对此建议解进行修正, 以获得确认解; 最后将新问题与对应的解决方案存储至案例库中, 用来解决未来的问题. 具体算法如下:

采用属性特征值描述法<sup>[19]</sup>, 案例库中第  $k$  条源案例记录  $C_k$  可表示为如下的二元组形式:

$$C_k : < X_k; Y_k >, k = 1, 2, \dots, p \quad (1)$$

其中,  $p$  是源案例的总数,  $X_k$  和  $Y_k$  分别是第  $k$  条源案例记录中的问题描述和决策属性.

将待求解的目标案例记为  $X = (x_1, \dots, x_n)$ , 决策结果记为  $Y$ . 按照 KNN 规则,  $X$  与  $X_k$  的相似度计算式为<sup>[20]</sup>:

$$s_k = 1 - \sqrt{\sum_{i=1}^n \omega_i (x_i - x_{i,k})^2}, k = 1, 2, \dots, p \quad (2)$$

其中,  $\omega_i, i = 1, 2, \dots, n$  是第  $i$  个属性的权重,  $x_{i,k}$  为第  $k$  条记录中第  $i$  个属性的特征值. 权重的约束条件为:

$$\sum_{i=1}^n \omega_i = 1, \omega_i \geq 0 \quad (3)$$

计算出相似度后, 根据最大相似度重用原则, 目标案例  $X$  的重用结果  $Y$  可用下式描述:

$$Y = \arg \max_{Y_k} (s_1, \dots, s_p) \quad (4)$$

若重用后能够正确解决问题, 则将二元组  $< X, Y >$  存储于案例库中; 否则, 可采用专家修正方法对结果  $Y$  进行不断调整, 从而完成基于实践检验的案例存储<sup>[21]</sup>, 此时式 (1) 转化为如下形式:

$$C_k : < X_k; Y_k >, k = 1, 2, \dots, p + 1 \quad (5)$$

通过每一次问题的求解, 案例库的记录均会增加, 案例记录的总数由  $p$  变为  $p + 1$  个, 表现出增量学习的特点, 使得 CBR 模型表现出一定的学习能力<sup>[22]</sup>.

由式 (2) 可知,  $X$  与  $X_k$  的相似度大小受两方面因素的影响: 1) 受  $X$  与  $X_k$  中各特征值的距离影响. 距离越近, 说明  $X$  与  $X_k$  越相似. 这一点为定值计算, 无需调整; 2) 受各属性的权重分配情况的影响. 权重的不同组合对相似度的结果会产生影响, 进而影响到问题求解的性能.

在满足式 (3) 的条件下, MA 方法将各属性的重要程度视为均等, 即  $\omega_1 = \dots = \omega_n = 1/n$ , 事实上, 不论是模式分类问题还是回归预测问题, 各属性对求解结果的影响程度不尽相同, 如果平均分配, 得到的相似度顺序可能会变得不正确, 导致求解结果的不可信, 因而通过 MA 法获取权重的求解过程虽然简单, 但结果并不客观. 以 GA 为典型代表的迭代寻优方法<sup>[15]</sup> 从案例本身提供的数据入手, 将问题求解的质量作为适应度函数去寻找合适的权重, 提高了权重分配的客观性, 但 GA 算法易陷入局部最优, 且时间复杂度会大幅增加. 比如, 对样本量是  $p$  的案例库来说, 检索的时间复杂度为  $O(p)$ , 假设 GA 法的迭代次数为  $N$ , 则寻优权重的时间复杂度为  $O(p \times N)$ , 这会显著降低问题求解的效率. 因此, 研究一种既能合理分配权重, 又不显著增加时间复杂度的分配方法, 仍是一个亟待解决的问题<sup>[12]</sup>.

## 2 属性权重分配方法

借鉴通信领域中的注水原理 (Water-filling theory, WFT)<sup>[23]</sup>, 将信道功率的优化分配机制应用于属性权重的分配中, 得到一种基于 WFT 的分配方

法(即 WFA). 在 WFT 中, 认为子信道具有不同的衰减特性, 信噪比高的子信道得到的功率多, 反之功率少, 从而保证信号的传输效率. 其分配机理可用下式描述:

$$C = \sum_{i=1}^n \log_2 \left( 1 + \frac{\alpha_i^2 P_i}{\sigma_i^2} \right) \quad (6)$$

其中,  $C$  为信道容量,  $\alpha_i$  为第  $i$  个子信道的增益,  $P_i$  表示第  $i$  个子信道分配到的功率,  $\sigma_i^2$  为第  $i$  个特征子信道的噪声方差.

根据式 (2) 相似度计算的分析, 各属性具有不同的重要程度, 重要的属性赋予较大的权重值, 反之取较小值, 从而保证权重分配的合理性. 可见, 基于 WFT 的信道功率分配机制与基于属性权重的分配机制非常相似. 由此, 根据 WFT 得到属性权重分配的描述式为:

$$R = \sum_{i=1}^n \log_2 (1 + \alpha_i^2 \omega_i) \quad (7)$$

式中,  $R$  为权重分配的合理性指标 (对应于 WFT 中信道功率分配的容量最大化指标).  $\alpha_i$  表示第  $i$  个属性的重要程度 (对应于 WFT 中信噪比的高低), 即案例库中第  $i$  个属性与输出的相关程度, 可通过计算其相关系数来确定, 即:

$$\alpha_i = \frac{\sum_{k=1}^p (x_{i,k} - \bar{x}_i)(y_k - \bar{y})}{\sqrt{\sum_{k=1}^p (x_{i,k} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^p (y_k - \bar{y})^2}} \quad (8)$$

其中,  $\bar{x}_i$  表示第  $i$  个属性特征值的均值,  $\bar{y}$  表示案例库中所有决策的均值.

为了计算属性权重, 需按 WFT 的求解思路构造拉格朗日函数<sup>[23]</sup>. 将式 (7) 和约束条件 (3) 相结合, 得到拉格朗日函数如下:

$$L(\omega, \lambda) = \sum_{i=1}^n \log_2 (1 + \alpha_i^2 \omega_i) + \lambda \left( 1 - \sum_{i=1}^n \omega_i \right) \quad (9)$$

其中,  $\lambda$  是拉格朗日算子.  $L(\omega, \lambda)$  的意义是在满足式 (3) 的约束条件下, 权重的赋值是合理的. 对式 (9) 中的  $\omega_i$  求偏导后得:

$$\frac{\partial L}{\partial \omega_i} = \frac{1}{\ln 2} \frac{\alpha_i^2}{(1 + \alpha_i^2 \omega_i)} - \lambda = 0, i = 1, 2, \dots, n \quad (10)$$

从中可解得各属性所分配到的权重, 考虑到式 (3) 中  $\omega_i \geq 0$  的约束条件, 权重计算公式为:

$$\omega_i = \left[ \varepsilon - \frac{1}{\alpha_i^2} \right]^+ \quad (11)$$

其中,  $[\cdot]^+$  表示 0 和  $Z$  的最大值;  $\varepsilon$  为门限值, 根据式 (3) 中  $\sum_{i=1}^n \omega_i = 1$  的约束条件, 可得:

$$\varepsilon = \frac{1}{n} \left( 1 + \sum_{i=1}^n \frac{1}{\alpha_i} \right) \quad (12)$$

式 (11) 表明, 权重的分配受属性的重要程度影响, 属性对应的  $\alpha_i^2$  越大, 则分配到的权重较大. 由此, 得到了满足式 (3) 和式 (7) 的权重分配合理化的算法.

综上所述, 基于 WFA 的权重计算步骤是: 首先通过式 (8) 计算每个属性与决策的相关系数, 然后根据式 (12) 得到门限值, 最后由式 (11) 得到各属性的权重.

相对于 GA 迭代寻优权重的时间复杂度来说, 假设 GA 法的迭代次数为  $N$ , 则寻优权重的时间复杂度为  $O(p \times N)$ , 而 WFA 法只与样本中属性个数  $n$  有关, 表示为  $O(n)$ . 当  $p > n$  时 (满足此条件的情况较为常见), 显然  $O(p \times N) > O(n)$ , 可见 WFA 法在求解的时间复杂度方面比 GA 法效率高, 具有显著优势.

### 3 收敛性分析

案例推理模型的收敛性是防止计算结果出现较大差异, 是保证算法在各种条件下可靠计算与正确解决问题的前提.

式 (2) 中计算相似度采用的方法是最近邻规则; 式 (7) 的优化权重指标函数和式 (3) 的权重约束引入了拉格朗日乘子法对权重进行优化计算, 从而得到了式 (8)、式 (11) 和式 (12). 如果这两处的计算不收敛, 将会导致推理结果的差异性变大, 性能将变得不可靠. 为此, 对式 (1)~(2) 的计算过程作如下约定: 式 (1) 有  $p$  对数据,  $X_k$  是取值于一个  $p$  维的度量空间  $\Omega$ . 对目标案例  $X$  来说, 从式 (2) 的计算结果中选择出具有最大相似度所对应的案例描述  $X_k$ , 即为  $X$  的最近邻点. 则最近邻算法的收敛条件定义为: 如果  $X_k \rightarrow X$  的概率为 1, 则算法是收敛的. 这一结论已经由 Cover 等<sup>[10]</sup> 给出证明, 下面给出简要的介绍.

设  $S_X(r)$  是以  $X$  为中心  $r$  为半径的球体  $\{\tilde{X} \in \Omega : d(X, \tilde{X}) \leq r\}$ ,  $d$  是定义在  $\Omega$  上的度量. 首先, 假设点  $X \in \Omega$ , 则每一个球体  $S_X(r)$  ( $r > 0$ )

均有非 0 的概率度量. 那么, 对于任意的  $\delta > 0$ , 有:

$$P \left\{ \min_{k=1,2,\dots,p} d(X_k, X) \geq \delta \right\} = 1 - P(S_X(\delta))^p \rightarrow 0 \quad (13)$$

因此,  $d(X_k, X)$  在  $k$  处是单调递减的,  $X$  的最近邻点以概率 1 收敛于  $X$ .

此外, 根据式 (8)、式 (11) 和式 (12) 对权重进行计算时, 需要注意的是由式 (11) 得到的权重是否有界, 如果有界, 则证明权重的计算过程是收敛的, 由此给出以下定理:

**定理 1.** 由式 (11) 得到的权重  $\omega_i$  ( $i = 1, 2, \dots, n$ ), 如果对每一个权重均存在一个正实数  $B$ , 满足  $0 \leq \omega < B$ , 则权重是一致有界的.

**证明:** 式 (8) 中的相关系数满足:  $0 < |\alpha| \leq 1$ , 可以得到  $1 \leq 1/\alpha^2 < B_\alpha$  ( $B_\alpha > 0$ ), 则式 (12) 中的门限值满足  $1/n \leq \varepsilon < B_\varepsilon$  ( $B_\varepsilon > 0$ ), 通过式 (11) 可知  $|\varepsilon - 1/\alpha^2| < B$ , 从而  $0 \leq \omega < B$  得到满足, 因此权重的计算过程是收敛的.  $\square$

## 4 实验设计及结果

为了测试 CBR 方法及权重分配方法的效果, 本文设计了模式分类实验, 将 CBR 同典型的分类器 KNN 和 SVM 进行对比, 同时对 MA、GA、WFA 等 3 种属性权重分配方法与 KNN 和 CBR 结合形成的分类器进行对比实验.

### 4.1 实验数据

分类数据集从 UCI 机器学习资源库中获得<sup>[24]</sup>, 本文实验采用的数据集信息如表 1 所示, 其中, 红葡萄酒质由葡萄牙米尼奥大学提供, 共有 6 种质量分数; 葡萄酒来源于意大利的一个医药食品分析和技术研究所, 共有 3 种不同种类的酒; 计划和放松是关于 2 种心理阶段的分类问题; 图像分割来源于美国马萨诸塞大学的视觉小组, 共 7 个类别, 分别是砖面、天空、树叶、水泥、窗户、道路和草坪; 胎心监护由葡萄牙波尔图大学医学院提供, 用于对胎儿状态进行分类, 分别是正常、病态和疑似病态 3 种类别; 脊柱来源于巴西塞阿拉联邦大学, 共有 3 种结果, 分别是正常、脊柱前移和椎间盘突出; 电离层由美国约翰霍普金斯大学应用物理实验室提供, 可对电离层中的雷达回波质量进行分类, 有好和坏 2 种状态; 钢板缺陷来自意大利的一个通讯科学研究中心, 有 7 种缺陷, 分别是裂纹、Z 刮伤、K 刮伤、鳞皮、结疤、凸起和其他.

表 1 数据集信息

Table 1 Data sets for experiments

编号	数据集名称	样本数	属性数	类别数
1	红葡萄酒质	1599	11	6
2	葡萄酒	178	13	3
3	计划和放松	182	12	2
4	图像分割	2310	18	7
5	胎心监护	2126	22	3
6	脊柱	310	6	3
7	电离层	351	33	2
8	钢板缺陷	1941	27	7

### 4.2 实验设计

为了综合检验各种分类方法的效果, 采用五折交叉验证法<sup>[19]</sup>, 实验用计算机环境为 Intel Core 2 Quad Q9550, 2.83 GHz, 4 GB 内存, 算法软件采用 Matlab 7.0 实现. 实验中涉及到的基本分类方法是 KNN、SVM 和 CBR, 其中的 KNN 和 CBR 分类的属性权重分配方法有 3 种, 分别是: 对 KNN 方法, 权重平均分配法记为 MA-KNN, 遗传算法记为 GA-KNN, 注水分配法记为 WFA-KNN; 对 CBR 方法, 按上述顺序分别记为 MA-CBR、GA-CBR、WFA-CBR. 主要步骤如下:

**步骤 1.** 对数据集进行归一化处理. 每个属性的特征值都被映射到  $[0, 1]$  区间.

**步骤 2.** 设定实验参数. 在 GA 算法中, 初始种群取 20, 交叉概率取 0.4, 变异概率取 0.05, 染色体长度用 3 位二进制进行编码, 迭代次数设定为 10 次, 寻优权重时对数据采用留一法进行训练; 在 SVM 算法中, 惩罚因子取 10, 核函数取高斯径向基函数, 宽度取 0.3, 对于多分类数据, 采用一对多法训练分类器; KNN 规则中的  $K = 1$ .

**步骤 3.** 将每一个数据集分为 5 份, 其中的 4 份作为训练样本, 余下的 1 份作为测试样本, 进行五折交叉实验, 并记录不同分类器的准确率和时间 (时间为每一折实验用时, 包含训练权重、分类器和测试分类结果的时间).

### 4.3 结果分析

根据上述实验步骤, 可得到不同分类器针对每一个数据集的实验结果, 表 2 为其中一个数据集的结果. 将所有数据集的实验结果进行平均统计, 得到 KNN (包括 MA、GA、WFA)、SVM、CBR (包括 MA、GA、WFA) 方法的平均准确率分别是 75.69%, 73.35% 和 79.83%, 平均时间分别是 78.81 秒、1369 秒和 87.75 秒. 可以看出, 在这三种方法中, CBR 的分类准确率最高, 效率居中.

根据每个分类器对 8 个数据集的实验结果, 求出各自的分类准确率和时间的均值, 见图 1 和图 2.

从图 1 可以看出, 3 种方法分配权重后, 分类准确率从高到低的顺序是 WFA、GA 和 MA, 说明 WFA 比 GA 和 MA 分配的权重更为合理, 能提高分类准确率.

表 2 分类准确率及时间对比

Table 2 Comparison results of classification accuracy and cost time

数据集名称	分类器	准确率 (%)	时间 (s)
图像分割	MA-KNN	97.10	0.68
	GA-KNN	97.40	579
	WFA-KNN	97.23	0.82
	SVM	96.41	6866
	MA-CBR	97.27	0.79
	GA-CBR	97.66	578
	WFA-CBR	97.32	0.79

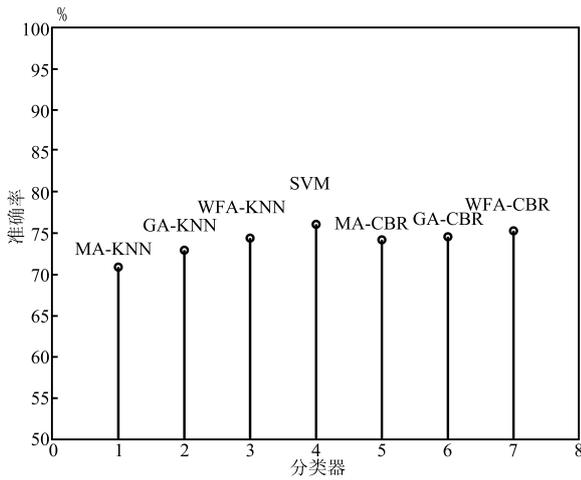


图 1 不同分类器的平均准确率对比

Fig. 1 Comparison among average accuracy of the different classifiers

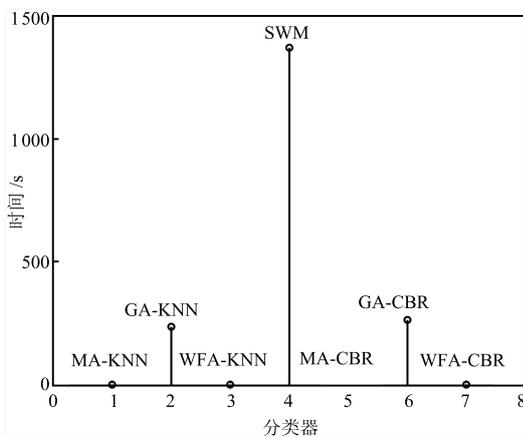


图 2 不同分类器的平均时间对比

Fig. 2 Comparison among average cost time of the different classifiers

从图 2 可以看出, 时间从快到慢的顺序是 MA-

KNN、WFA-KNN、MA-CBR、WFA-CBR、GA-KNN、GA-CBR、SVM. 其中, SVM 在处理多分类问题或大样本量 (对应数据集编号为 1、4、5、8) 时会大幅增加时间成本, GA 在寻优权重时也会增加时间复杂度, 使得分类效率降低, 而 WFA 和 MA 方法的差别并不明显. 综合图 1 和图 2 可以看出, WFA 分配权重的分类器性能较优.

另外, 从表 1 中选出具有较多样本的 4 个数据集 (对应数据集编号为 1、4、5、8), 计算不同分类器准确率的平均值, 如图 3 所示. 从图 3 可以看出, GA-CBR (相对于 MA-CBR、WFA-CBR) 和 GA-KNN (相对于 MA-KNN、WFA-KNN) 均有较好的结果. 说明具有大样本量时, GA 通过迭代寻优, 能搜索到合理的属性权重, 在提高分类准确率方面, 比其他方法具有明显优势.

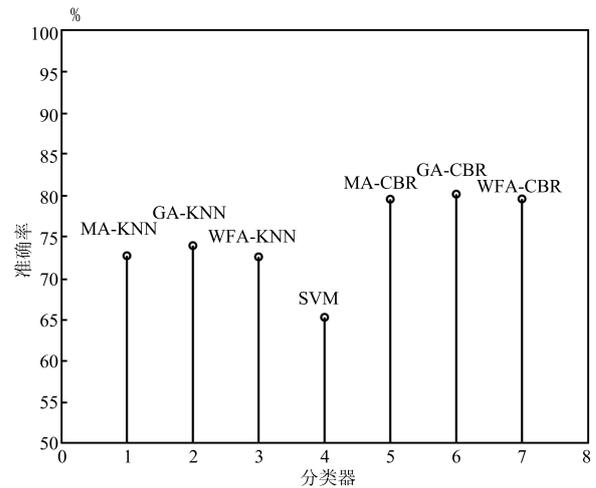


图 3 大样本量时不同分类器的平均准确率对比

Fig. 3 Comparison among average accuracy of the different classifiers under large sample

在表 1 的数据集中有 2 个二分类问题 (对应数据集编号为 3 和 7), 分别计算不同分类器准确率的平均值, 如图 4 所示. 从图 4 可以看出, SVM 比其他分类方法具有较高的准确率; 从花费的时间成本上看, 表 2 说明了 SVM 比 WFA-CBR 大幅增加了时间复杂度, 综合性能与 WFA-CBR 仍有差距.

综上所述, 采用 WFA、GA 和 MA 对属性权重进行分配时, 在处理多分类问题或大样本量时, WFA 比 GA 的分类效率更高. 此外, 在处理大样本量时, GA 的应用在提高分类准确率方面具有明显的优势, 说明通过 GA 迭代寻优能搜索到更为合理的属性权重; 对二分类问题来说, SVM 比其他分类方法具有较高的准确率; 从花费的时间成本上看, SVM 比 WFA-CBR 大幅增加了时间复杂度; 从统计意义上看, CBR 相比 SVM 和 KNN, 分类准确率

最高, 所用时间居中, 综合性能较好. 另外, 对属性权重进行分配时, 综合考虑分类效率和分类准确率 2 个性能指标, WFA 方法表现出较好的综合性能.

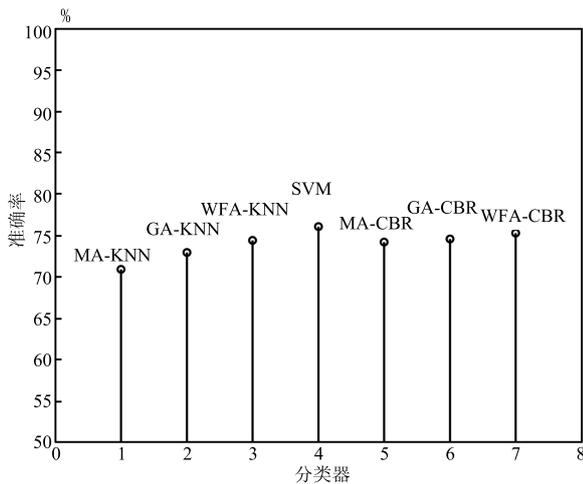


图 4 二分类问题不同分类器的平均准确率对比

Fig. 4 Comparison among average accuracy of the different classifiers in binary classification tasks

## 5 结论

本文讨论了 CBR 推理系统中属性权重的分配方法, 给出了基于注水原理的权重分配算法, 并分析了其收敛条件. 通过模式分类数据集的实验结果分析, 证明注水原理应用于 CBR 属性权重赋值的有效性, 表明注水分配算法能够充分利用案例属性的潜在知识, 提高分类的准确率. 而 GA 算法寻优权重虽然也可以提高分类准确率, 但相比 WFA 分配权重会大幅增加时间复杂度. 在后续的研究中, 为了增强 CBR 在问题求解方面的学习性能, 研究重点是基于注水原理的权重自适应迭代学习算法, 以寻求更为合理的权重分配. 同时, 应考虑一种衡量时间复杂度和求解质量方面的评价函数, 以指导如何提高问题求解时的综合性能.

## References

- Schank R C. *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*. New York: Cambridge University Press, 1982.
- Kolodner J L. Maintaining organization in a dynamic long-term memory. *Cognitive Science*, 1983, **7**(4): 243–280
- Aamodt A, Plaza E. Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Communications*, 1994, **7**(1): 39–59
- Liu Y H, Yang C S, Yang Y B, Lin F H, Du X M, Ito T. Case learning for CBR-based collision avoidance systems. *Applied Intelligence*, 2012, **36**(2): 308–319
- Xing G S, Ding J L, Chai T Y, Afshar P, Wang H. Hybrid intelligent parameter estimation based on grey case-based reasoning for laminar cooling process. *Engineering Applications of Artificial Intelligence*, 2012, **25**(2): 418–429
- Chai Tian-You. Operational optimization and feedback control for complex industrial processes. *Acta Automatica Sinica*, 2013, **39**(11): 1744–1757 (柴天佑. 复杂工业过程运行优化与反馈控制. *自动化学报*, 2013, **39**(11): 1744–1757)
- Tadrat J, Boonjing V, Pattaraintakorn P. A new similarity measure in formal concept analysis for case-based reasoning. *Expert Systems with Applications*, 2012, **39**(1): 967–972
- Wang H C, Huang T H. An enhanced case-based reasoning model for supporting inference missing attribute and its feature weight. *Journal of Internet Technology*, 2012, **13**(1): 45–56
- Carmona M A, Barbancho J, Larios D F, León C. Applying case based reasoning for prioritizing areas of business management. *Expert Systems with Applications*, 2013, **40**(9): 3450–3458
- Cover T M, Hart P E. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 1967, **13**(1): 21–27
- Lin S W, Chen S C. Parameter tuning, feature selection and weight assignment of features for case-based reasoning by artificial immune system. *Applied Soft Computing Journal*, 2011, **11**(8): 5042–5052
- Kim K J, Kim K. Preliminary cost estimation model using case-based reasoning and genetic algorithms. *Journal of Computing in Civil Engineering*, 2010, **24**(6): 499–505
- Kristi R, Qiang Y. Redundancy detection in semi-structured case bases. *IEEE Transactions on Knowledge and Data Engineering*, 2001, **13**(3): 513–518
- Park C S, Han I. A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction. *Expert Systems with Applications*, 2002, **23**(3): 255–264
- Ahn H, Kim K, Man I. Global optimization of feature weights and the number of neighbors that combine in a case-based reasoning system. *Expert Systems*, 2006, **23**(5): 290–301
- Pian Jin-Xiang, Chai Tian-You, Li Jie-Jia. Application of case-based reasoning and iterative learning in laminar cooling process control. *Acta Automatica Sinica*, 2012, **38**(12): 2032–2037 (片锦香, 柴天佑, 李界家. 案例推理及迭代学习在层流冷却控制中的应用. *自动化学报*, 2012, **38**(12): 2032–2037)
- Pian Jin-Xiang, Chai Tian-You, Li Jie-Jia. Rule and data driven strip coiling temperature model in laminar cooling process. *Acta Automatica Sinica*, 2012, **38**(11): 1861–1869 (片锦香, 柴天佑, 李界家. 规则与数据驱动层流冷却过程带钢卷取温度模型. *自动化学报*, 2012, **38**(11): 1861–1869)
- Mishra N, Petrovic S, Sundar S. A self-adaptive case-based reasoning system for dose planning in prostate cancer radiotherapy. *Medical Physics*, 2011, **38**(12): 6528–6538
- Bergmann R, Kolodner J, Plaza E. Representation in case-based reasoning. *Knowledge Engineering Review*, 2005, **20**(3): 209–213
- Liu C H, Chen H C. A novel CBR system for numeric prediction. *Information Sciences*, 2012, **185**(1): 178–190
- Kaedi M, Ghasem-Aghae N. Biasing Bayesian optimization algorithm using case based reasoning. *Knowledge-Based Systems*, 2011, **24**(8): 1245–1253

- 22 Lopez de Mantaras R, Plaza E. Case-based reasoning: an overview. *AI Communications*, 1997, **10**(1): 21–29
- 23 Luo B, Cui Q M, Wang H, Tao X F. Optimal joint water-filling for coordinated transmission over frequency-selective fading channels. *IEEE Communications Letters*, 2011, **15**(2): 190–192
- 24 Frank A, Asuncion A. UCI machine learning repository [Online], available: <http://archive.ics.uci.edu/ml>, March 1, 2014



**严爱军** 北京工业大学电子信息与控制工程学院副教授. 2006 年获东北大学博士学位. 主要研究方向为人工智能, 过程建模与优化控制. 本文通信作者.

E-mail: yanaijun@bjut.edu.cn

(**YAN Ai-Jun** Associate professor in the College of Electronic Information & Control Engineering, Beijing University of Technology. He received his Ph.D. degree from Northeastern University in 2006. His research interest covers artificial intelligence, process modeling and optimization control. Corresponding author of this paper.)

University of Technology. He received his Ph.D. degree from Northeastern University in 2006. His research interest covers artificial intelligence, process modeling and optimization control. Corresponding author of this paper.)



**钱丽敏** 北京工业大学硕士研究生. 主要研究方向为案例推理及其应用.

E-mail: qlm0408@126.com

(**QIAN Li-Min** Master student at Beijing University of Technology. Her research interest covers case-based reasoning and its applications.)



**王 普** 北京工业大学电子信息与控制工程学院教授. 1988 年获中国矿业大学博士学位. 主要研究方向为信息处理, 智能控制. E-mail: wangpu@bjut.edu.cn

(**WANG Pu** Professor in the College of Electronic Information & Control Engineering, Beijing University of Technology. He received his Ph.D. degree from China University of Mining Technology in 1988. His research interest covers information process and intelligent control.)

China University of Mining Technology in 1988. His research interest covers information process and intelligent control.)