计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2017.20160011
54(5):998-1011,2017

知识嵌入的贝叶斯 MA 型模糊系统

顾晓清[1,2]　　王士同[1]
[1](江南大学数字媒体学院　江苏无锡　214122)
[2](常州大学信息科学与工程学院　江苏常州　213164)
(czxqgu@163.com)

Knowledge Embedded Bayesian MA Fuzzy System

Gu Xiaoqing[1,2] and Wang Shitong[1]
[1](*School of Digital Media, Jiangnan University, Wuxi, Jiangsu* 214122)
[2](*School of Information Science and Engineering, Changzhou University, Changzhou, Jiangsu* 213164)

Abstract The most distinctive characteristic of fuzzy system is its high interpretability. But the fuzzy rules obtained by classical cluster based fuzzy systems commonly need to cover all features of input space and often overlap each other. Specially, when facing the high-dimension problem, the fuzzy rules often become more sophisticated because of too much features involved in antecedent parameters. In order to overcome these shortcomings, based on the Bayesian inference framework, knowledge embedded Bayesian Mamdan-Assilan type fuzzy system (KE-B-MA) is proposed by focusing on the Mamdan-Assilan (MA) type fuzzy system. First, the DC (don't care) approach is incorporated into the selection of fuzzy membership centers and features of input space. Second, in order to enhance the classification performance of obtained fuzzy systems, KE-B-MA learns both antecedent and consequent parameter of fuzzy rules simultaneously by a Markov chain Monte Carlo (MCMC) method, and the obtained parameters can be guaranteed to be global optimal solutions. The experimental results on a synthetic dataset and several UCI machine datasets show that the classification accuracy of KE-B-MA is comparable to several classical fuzzy systems with distinctive ability of providing explicit knowledge in the form of interpretable fuzzy rules. Rather than being rivals, fuzziness in KE-B-MA and probability can be well incorporated.

Key words classification; Bayesian inference; Mamdan-Assilan type fuzzy system; knowledge embedded; Markov chain Monte Carlo (MCMC) method

摘　要 模糊系统的独特优势在于其高度的可解释性,然而传统的基于聚类的模糊系统往往需要使用输入空间的全部特征且常出现模糊集交叉的现象,系统的可解释性不高;此外,此类模糊系统对高维数据处理时还会因使用大量的特征而使规则过于复杂.针对此问题,探讨了一种知识嵌入的贝叶斯 MA 型模糊系统(knowledge embedded Bayesian Mamdan-Assilan type fuzzy system, KE-B-MA).首先,KE-B-MA 使用 DC(don't care)方法进行知识嵌入的模糊集划分,对模糊隶属度函数中心和输入空间特征

收稿日期:2016-01-05;**修回日期**:2016-10-10
基金项目:国家自然科学基金项目(61572236,61502058,61572085);江苏省自然科学基金项目(BK20160187);中央高校基本科研业务费专项资金项目(JUSRP51614A);江苏省高校自然科学基金项目(15KJB520002)
This work was supported by the National Natural Science Foundation of China (61572236, 61502058, 61572085), the Natural Science Foundation of Jiangsu Province of China (BK20160187), the Fundamental Research Funds for the Central Universities (JUSRP51614A), and the Natural Science Foundation of Jiangsu Higher Education Institutions (15KJB520002).

的选择进行有效指导,其获得的规则可对应于不同的特征空间.其次,KE-B-MA 基于贝叶斯推理使用马尔可夫蒙特卡洛(Markov chain Monte Carlo, MCMC)方法对模糊规则的前后件参数同时学习,所得结果为全局最优解.实验结果表明:与一些经典模糊系统相比,KE-B-MA 具有令人满意的分类性能且具有更强的可解释性和清晰性.

关键词 分类;贝叶斯推理;Mamdan-Assilan 型模糊系统;知识嵌入;马尔可夫链蒙特卡洛方法

中图法分类号 TP18; TP391.4

目前,基于规则的模糊系统已成功地应用于系统建模、模式识别和图像处理等领域[1-5].模糊系统与其他人工智能方法的主要区别在于:1)模糊系统具备不确定和模糊信息的处理能力,能够将自然语言直接转译成与人类推理机制相似的模糊规则;2)模糊系统不仅具有强大的学习能力,其构建的模糊规则还具有高度的可解释性,而这一特性是其他人工智能方法所不具备的.

广泛使用的提取模糊规则的方法主要可分为 3 类:

1) 通过对输入空间的固定划分或基于网格的空间划分方法来获取模糊规则[6-8],其优点是空间的模糊划分清晰度高,但在对高维数据进行训练时,会因采用大量的特征而使其规则变得过于复杂.假设使用均匀 5 网格的方法对输入空间的每一维进行模糊划分,对于特征数是 n 的数据集而言,最终提取的模糊规则数为 5^n.为此,进化算法和自我评价学习算法[9-11]常用于模糊规则的结构简化.

2) 使用支持向量机(support vector machine, SVM)技术[12-13],将每一个支持向量作为一条模糊规则.SVM 技术遵循结构风险最小化原则,能够最大限度地减少模型的泛化误差,然而支持向量的数量通常是很大的,尤其是对于复杂的分类任务,因此这类方法设计得到的模糊规则也易出现规模较大且易冗余的情况.为此,研究者提出了一系列的删减准则用于模糊规则的约减[14].

3) 使用聚类算法,模糊规则的个数等于聚类的个数.文献[15-16]中,模糊规则的前件参数则使用自适应模糊聚类算法获得,模糊规则的后件参数则使用 SVM 的方法获得;文献[17]使用密度聚类对模糊系统的结构进行初始化;文献[18-20]使用经典的模糊 C 均值聚类(fuzzy c-means, FCM)算法来获得输入空间的模糊划分,并使用 Ho-Kashyap 算法得到模糊规则的后件参数;文献[21]利用 FCM 聚类算法确定模糊规则的前件参数后,基于极大极小概率学习理论产生了一种有分类可靠性保证的模糊分类器.虽然近几年基于聚类的模糊系统得到了一定的发展,但这类方法普遍存在 2 点不足:

1) 模糊空间的划分通过聚类算法获得,通常情况下不能保证所得的聚类中心,即模糊集中心具有可解释性,得到的模糊集往往存在过度的交叉或重叠现象,模糊划分的清晰性自然得不到保证.

2) 模糊规则前件对应于输入空间的所有特征,在处理高维数据时,每条规则所对应模糊集的数量大大增加,模糊规则前件则变得愈加复杂,规则的可解释性也必然大大下降.

针对上述不足,本文提出了全新的知识嵌入的贝叶斯 MA 型模糊系统(knowledge embedded Bayesian Mamdan-Assilan type fuzzy system, KE-B-MA).鉴于 MA(Mamdan-Assilan)型模糊系统的简单性和受到广泛应用等特点,本文将其作为研究对象.首先,通过 DC 方法(don't care approach)实现对模糊隶属中心和输入空间特征的选择,其获得的模糊划分物理意义明显且有效降低所构建系统的复杂性.其次,引入贝叶斯概率框架并使用马尔可夫链蒙特卡洛(Markov chain Monte Carlo, MCMC)方法构造一条马尔可夫链,同时对模糊规则的前件和后件参数进行学习,保证获得参数具有全局最优值.因而,本文所提方法能够在保证模糊规则的可解释性的同时兼顾模糊系统分类准确度.

1 相关工作

1.1 MA 型模糊系统基本概念

经典模糊系统可分为 3 类[22-23]:Mamdan 型模型、Takagi-Sugeno-Kang 型模型和 Generalized Fuzzy 型模型.其中,Mamdan 型模糊系统由于其输出的简洁性,使得构造的模糊规则后件参数具有较高的可解释性受到较多关注.MA 型模糊系统的第 k 个模糊规则的表示为[20]

Rule k: if x_1 is A_1^k and x_2 is A_2^k and … and x_d is A_d^k, then

$$f_k(\boldsymbol{x})=v_k, \tag{1}$$

其中,$k=1,2,\cdots,K$,每一条规则都有对应的输入向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_d)^{\mathrm{T}}$,并把输入空间的模糊子集 $A^k\subset\mathbb{R}^d$ 投影到输出空间的模糊集 $f_k(\boldsymbol{x})$. 式(1)中 d 是样本的特征数,K 是模糊规则数,and 是模糊合取操作,A_i^k 为第 i 维输入在第 k 条模糊规则中对应的模糊子集,令 $\mu_k(\boldsymbol{x})$ 为第 k 条规则对应的模糊隶属度函数,若采用高斯型隶属函数,则其可以表示为

$$\mu_k(\boldsymbol{x})=\prod_{i=1}^{d}\exp\left(\frac{-(x_i-c_{ki})^2}{2\delta_{ki}}\right),\tag{2}$$

其中,参数 c_{ki} 和 δ_{ki} 分别是隶属函数的中心和方差. 在经过一系列的操作及去模糊化处理之后,可得 MA 型模糊系统的实值输出:

$$\tilde{y}=\sum_{k=1}^{K}\frac{\mu_k(\boldsymbol{x})}{\sum_{k'=1}^{K}\mu_k'(\boldsymbol{x})}f_k(\boldsymbol{x})=\sum_{k=1}^{K}\tilde{\mu}_k(\boldsymbol{x})v_k,\tag{3}$$

其中,$\tilde{\mu}_k(\boldsymbol{x})$ 为 A_i^k 对应的归一化后的模糊隶属度函数. 以二元分类为研究对象,模糊系统的决策函数可以表示为

$$F(\boldsymbol{x})=\begin{cases}\mathrm{sgn}(\tilde{y})\geqslant 0,\ \boldsymbol{x}\in\text{ 正类.}\\ \mathrm{sgn}(\tilde{y})<0,\ \boldsymbol{x}\in\text{ 负类.}\end{cases}\tag{4}$$

1.2 基于模糊聚类和 SVM 技术的 MA 型模糊系统

设给定样本 $\boldsymbol{X}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N)$ 和对应的标签 $\boldsymbol{Y}=(y_1,y_2,\cdots,y_N)$,其中 $y_i\in\{-1,+1\}$,$1\leqslant i\leqslant N$. 传统基于模糊聚类的模糊系统参数学习方法如图 1 所示,模糊系统的规则前件参数可以直接通过模糊聚类算法的学习获得. 由于模糊 C 均值聚类 FCM[24] 所获得的空间划分具有模糊性,且具有实现简单和有效的优点,其被广泛应用于模糊系统规则前件参数的学习中. 此时,式(2)中的隶属度函数中心 $\boldsymbol{c}_k=(c_{k1},c_{k2},\cdots,c_{kd})^{\mathrm{T}}$ 为聚类中心,方差 $\boldsymbol{\delta}_k=(\delta_{k1},\delta_{k2},\cdots,\delta_{kd})^{\mathrm{T}}$ 可计算得到:

$$\delta_{ki}=h\sum_{j=1}^{N}u_{jk}(x_{ji}-c_{ki})^2\Big/\sum_{j=1}^{N}u_{jk},\tag{5}$$

其中,u_{jk} 为输入向量 $\boldsymbol{x}_j=(x_{j1},x_{j2},\cdots,x_{jd})^{\mathrm{T}}$ 隶属于第 k 类的隶属度,尺度参数 h 为一正常数.

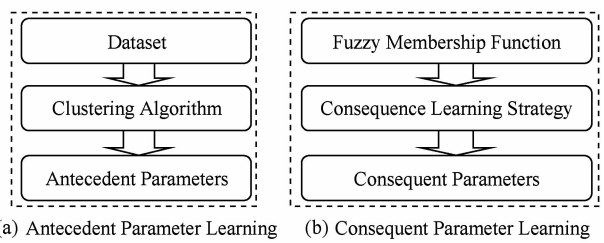


Fig. 1 The cluster based TSK fuzzy classifier's parameter learning mechanism

图 1 基于聚类算法的模糊系统参数学习的示意图

正如引言所述,支持向量机(SVM)算法由于同时考虑模型的结构风险最小化和经验风险最小化,通过其获得的后件参数具有较强的泛化性能. 对于每一个输入样本 $\boldsymbol{x}$,令 $\boldsymbol{\phi}=(\tilde{\mu}_1(\boldsymbol{x}),\tilde{\mu}_2(\boldsymbol{x}),\cdots,\tilde{\mu}_K(\boldsymbol{x}))^{\mathrm{T}}$,此时,模糊系统的后件参数可转化为以下最优化问题:

$$\begin{aligned}&\min_{\boldsymbol{V},\boldsymbol{\xi}}\frac{1}{2}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}+C'\sum_{i=1}^{N}\xi_i,\\&\text{s.t. }\ y_i(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{\phi}_i+b)\geqslant 1-\xi_i,\\&\xi_i\geqslant 0,i=1,2,\cdots,N.\end{aligned}\tag{6}$$

其中,后件参数 $\boldsymbol{V}=(v_1,v_2,\cdots,v_K)^{\mathrm{T}}$,$C'$ 是正则化参数.

通过最优化理论,式(6)可转化为对偶问题:

$$\begin{aligned}&\min_{\boldsymbol{\alpha}}\sum_{n=1}^{N}\alpha_n-\frac{1}{2}\sum_{n=1}^{N}\sum_{i=1}^{N}\alpha_n\alpha_i y_n y_i\phi_n\phi_i,\\&\text{s.t. }\sum_{n=1}^{N}\alpha_n y_n=0,0\leqslant\alpha_n\leqslant C'.\end{aligned}\tag{7}$$

由此可得后件参数的全局最优解:

$$\begin{aligned}&v_k=\sum_{i=1}^{N}y_i\tilde{\alpha}_i\tilde{\mu}_k(\boldsymbol{x}_i)=\sum_{i\in SV}y_i\tilde{\alpha}_i\tilde{\mu}_k(\boldsymbol{x}_i),\\&k=1,2,\cdots,K.\end{aligned}\tag{8}$$

最终,对于任意给定的样本 $\boldsymbol{x}$,MA 型模糊系统的实值输出为

$$\begin{aligned}\tilde{y}&=\boldsymbol{V}^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{x})+b=\\&\sum_{k=1}^{K}\Big(\sum_{i=1}^{N}y_i\tilde{\alpha}_i\tilde{\mu}_k(\boldsymbol{x}_i)\Big)\tilde{\mu}_k(\boldsymbol{x})+b,\end{aligned}\tag{9}$$

其中,偏移量 b 可通过求解式(7)获得.

值得注意的是,偏移量 b 可以看作是除 K 条模糊规则之外的第 $K+1$ 条规则,其形式为

$$\begin{aligned}&\text{Rule } K+1\text{: if } x_1 \text{ is } A_1^{K+1} \text{ and } x_2 \text{ is } A_2^{K+1}\\&\text{and }\cdots\text{ and } x_d \text{ is } A_d^{K+1}\text{, then}\\&f_{K+1}(\boldsymbol{x})=b,\end{aligned}\tag{10}$$

其中,$\tilde{\mu}_{K+1}(\boldsymbol{x})\equiv 1$,$\tilde{\alpha}_{K=1}=0$.

2 知识嵌入的贝叶斯 MA 型模糊系统

针对第 1 节分析的传统基于聚类的模糊系统的缺陷,本文探讨了一种知识嵌入的贝叶斯 MA 型模糊系统,即 KE-B-MA 的构建方法. KE-B-MA 模糊系统的工作原理和参数学习示意图如图 2 所示. 1)KE-B-MA 使用 DC 方法给出系统对应的模糊隶属度函数中心和每条规则使用的输入空间特征;

2)构造用于模糊规则前件和后件参数同时学习的概率模型框架;3)使用 MCMC 方法来估计模型的最大后验概率,得到模型参数的全局最优解. 对比图 1 和图 2 可知,传统基于聚类的模糊系统前件和后件参数的学习是相对独立的,一旦模糊规则的前件参数通过聚类算法确定后,在随后的模糊规则后件参数学习过程中不能更改. 然而聚类方法的本质是按照某种相似程度的度量来完成输入空间的划分,其无法考量模糊系统中输入空间和输出空间之间的联系. KE-B-MA 从贝叶斯推理的角度建立输入空间和输出空间之间的联系,保证了在复杂问题中所得模糊系统的分类准确度.

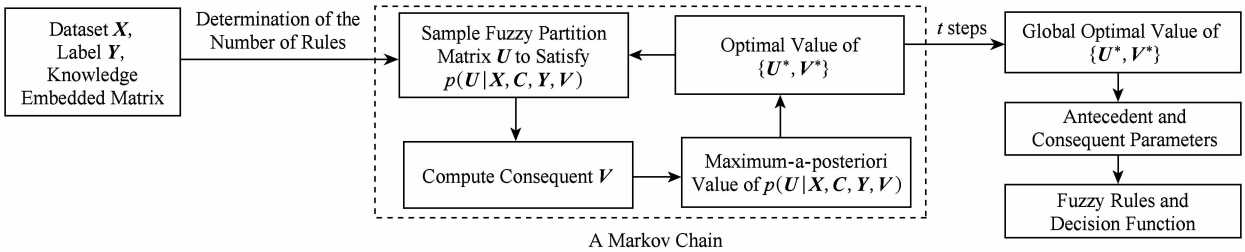


Fig. 2 System diagram of the KE-B-MA and parameter learning mechanism

图 2 KE-B-MA 工作原理和参数学习示意图

2.1 KE-B-MA 中模糊规则构建

日常生活和生产活动中通常积累有大量的专家经验,把这些有用的专家经验嵌入到模糊系统中能起到提高模糊规则可解释性的作用. 如在糖尿病患者的诊断系统中,胰岛素释放曲线常作为一个特征进行考察. 胰岛素释放曲线一般可分为:胰岛素分泌正常型、胰岛素分泌不足型、胰岛素分泌增多型和胰岛素释放障碍型 4 种. 在这一特征进行模糊划分时,如果按照 4 种胰岛素释放类型的平均数值进行预先设定可得 4 个具有可解释性的模糊隶属中心.

根据这一思路,KE-B-MA 模糊系统使用知识嵌入的 DC 方法[25-26]对每条规则使用的输入空间特征和模糊隶属度函数中心进行选择. DC 方法在模糊规则每一维上定义 5 个高斯型隶属度函数并赋予了其各自的语意标签:very low,low,medium,high,very high,分别用数字 1～5 表示. 隶属度函数中心相应地设定为(0,0.25,0.5,0.75,1);而 DC 则用来表示特征在规则中不被选择. 以下给出 KE-B-MA 模糊系统对应的模糊规则示例:表 1 是使用 DC 方法在样本维度为 5 的数据集上设置的知识嵌入矩阵. 其中第 1 条规则选择的输入特征(和对应的隶属度函数中心)分别是第 1 维(0),第 2 维(0.25)和第 5 维(0.75),而第 3 维和第 4 维不选择;第 2 条规则上选择的输入特征(和对应的隶属度函数中心)分别是第 1(0.25)、第 3(0.5)、第 4(0)和第 5 维(1)、第 2 维不选择. DC 方法中的知识嵌入矩阵在实际应用中往往按照专家经验人为设定.

Table 1 A Knowledge Embedded Matrix in DC Approach

表 1 DC 方法中知识嵌入矩阵示例

Rules	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5
Rule 1	1	2	DC	DC	4
Rule 2	2	DC	3	1	5
⋮	⋮	⋮	⋮	⋮	⋮

由于篇幅所限,以下给出第 1 条模糊规则的 IF 语句部分的格式:

if x_1 is very low and x_2 is low and x_3 is "don't care" and x_4 is "don't care" and x_5 is high,

进一步,该模糊规则的 if 语句部分可以简洁地写为

if x_1 is very low and x_2 is low and x_5 is high,

显然,与经典 MA 型模糊规则相比,本文所用模糊规则具有 2 个特点:1)if 语句部分保证模糊划分的清晰性;2)允许每条模糊规则的输入空间不同,即使用非连续的隶属度函数,这样既缩短了规则前件对应于输入向量的特征数,又可使每条规则从不同的视角进行推理,增强了模糊系统的可解释性.

2.2 KE-B-MA 模糊系统构建

当使用知识嵌入的 DC 方法预设了模糊隶属度函数的中心 $\boldsymbol{C}=(\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_K)^{\mathrm{T}}$ 和各规则的输入空间特征,KE-B-MA 模糊系统中待求的前件参数是模糊隶属度函数的方差 $\boldsymbol{\delta}=(\boldsymbol{\delta}_1,\boldsymbol{\delta}_2,\cdots,\boldsymbol{\delta}_K)^{\mathrm{T}}$. 值得注意的是,每一条规则的方差 $\boldsymbol{\delta}_k$ 与中心 $\boldsymbol{c}_k$ 在同一输入空间下. 此外,由式(2)(5)(7)可知,$\boldsymbol{\delta}_k$ 的值与

模糊聚类的隶属度 u_{nk}($1\leqslant n\leqslant N, 1\leqslant k\leqslant K$)有关,$u_{nk}$表示样本 $\boldsymbol{x}_n$ 与第 k 个聚类中心 $\boldsymbol{c}_k$ 的划分关系. 因此,求解 $\boldsymbol{\delta}$ 的问题就转化为求解模糊划分矩阵 $\boldsymbol{U}$ 的问题,加上待求的模糊规则后件参数 $\boldsymbol{V}=(v_1, v_2, \cdots, v_K)^{\mathrm{T}}$,KE-B-MA 模糊系统中需同时学习的模糊规则前件和后件参数为 $\boldsymbol{U}$ 和 $\boldsymbol{V}$.

根据贝叶斯理论,KE-B-MA 模糊系统中数据和各参数的联合似然估计可以写为

$$p(\boldsymbol{X},\boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})= p(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})p(\boldsymbol{U}|\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})p(\boldsymbol{C},\boldsymbol{V},\boldsymbol{Y}), \quad (11)$$

先验分布 $p(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$ 与样本标签 $\boldsymbol{Y}$ 和后件 $\boldsymbol{V}$ 无关,因此,先验分布 $p(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$ 等价于 $p(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{C})$. 依据文献[27]中样本分布的假设,在给定划分矩阵和聚类中心的情况下,每个样本 $\boldsymbol{x}_n$ 的分布可以看成 K 个高斯先验的乘积. 其中,K 个高斯先验的中心分别与 K 个固定的聚类中心对应,协方差为由模糊隶属度 u_{nk}^m 构成的单位阵,即:

$$p(\boldsymbol{x}_n \mid \boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}) = \prod_{k=1}^{K} N(\boldsymbol{x}_n \mid \boldsymbol{c}_k, \boldsymbol{\Lambda}), \quad (12)$$

其中,$\boldsymbol{\Lambda}^{-1}=u_{nk}^m\boldsymbol{I}$,$\boldsymbol{I}$ 是一个 $n\times n$ 的单位阵;m 是模糊指数,其作用等同于 FCM 聚类中的模糊指数. 式(12)每个样本点均具有独立的先验分布,对于全体样本 $\boldsymbol{X}$ 而言,其先验分布可以表示为所有样本的高斯先验的乘积,即:

$$p(\boldsymbol{X} \mid \boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}) = \prod_{n=1}^{N} p(\boldsymbol{x}_n \mid \boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}) = \prod_{n=1}^{N}\prod_{k=1}^{K} N(\boldsymbol{x}_n \mid \boldsymbol{c}_k, \boldsymbol{\Lambda}) = \left(\prod_{n=1}^{N}\prod_{k=1}^{K} u_{nk}^{md/2}\right) \exp\left(-\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K} u_{nk}^m \left\|\boldsymbol{x}_n-\boldsymbol{c}_k\right\|^2\right). \quad (13)$$

对式(13)取负对数,可得:

$$-\lg(p(\boldsymbol{X} \mid \boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})) = \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K} u_{nk}^m \left\|\boldsymbol{x}_n-\boldsymbol{c}_k\right\|^2 - \lg\left(\prod_{n=1}^{N}\prod_{k=1}^{K} u_{nk}^{md/2}\right), \quad (14)$$

式(14)等号右边第 1 项与 FCM 聚类的目标函数相同,用来保证在聚类中心固定的情况下得到 $\boldsymbol{X}$ 合理的模糊划分. 假设模糊划分矩阵 $\boldsymbol{U}$ 的先验分布 $p(\boldsymbol{U}|\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$ 只与聚类中心 $\boldsymbol{C}$ 有关,即 $p(\boldsymbol{U}|\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$ 等价于 $p(\boldsymbol{U}|\boldsymbol{C})$,同时不同样本所属的模糊隶属度之间相互独立,则先验分布 $p(\boldsymbol{U}|\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$ 可以定义为

$$p(\boldsymbol{U} \mid \boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}) = \prod_{n=1}^{N} p(\boldsymbol{u}_n \mid \boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}) = \prod_{n=1}^{N}\left(\prod_{k=1}^{K} u_{nk}^{-md/2}\right) Dirichlet(\boldsymbol{u}_n \mid \boldsymbol{\alpha}). \quad (15)$$

式(15)由 2 项因子 $F_1=\prod_{n=1}^{N}\prod_{k=1}^{K} u_{nk}^{-md/2}$ 和 $F_2=\prod_{n=1}^{N} Dirichlet(\boldsymbol{u}_n \mid \boldsymbol{\alpha})$ 构成,F_1 因子的作用是消去 $p(\boldsymbol{X} \mid \boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$ 中的 $\prod_{n=1}^{N}\prod_{k=1}^{K} u_{nk}^{md/2}$ 项. F_2 因子则被定义为 N 个狄利克雷分布(Dirichlet)的乘积,其中,Dirichlet 分布的具体形式如下:

$$Dirichlet(\boldsymbol{u}_n \mid \boldsymbol{\alpha}) = \prod_{n=1}^{N} \frac{\Gamma\left(\sum_{k=1}^{K}\alpha_k\right)}{\prod_{k=1}^{K}\Gamma(\alpha_k)} \prod_{k=1}^{K} u_{nk}^{\alpha_k-1}, \quad (16)$$

其中,$\boldsymbol{\alpha}=(\alpha_1, \alpha_2, \cdots, \alpha_K)^{\mathrm{T}}$ 是 Dirichlet 参数. Dirichlet 分布能保证所得的 u_{nk} 均为正值且 $\sum_{k=1}^{K} u_{nk}=1$.

本文所提 KE-B-MA 假设先验分布 $p(\boldsymbol{C},\boldsymbol{V},\boldsymbol{Y})$ 服从指数分布,形式为

$$p(\boldsymbol{C},\boldsymbol{V},\boldsymbol{Y}) = \prod_{n=1}^{N} \exp(-(1-f(\boldsymbol{C},y_n,\boldsymbol{V}))), \quad (17)$$

其中,函数 $f(\boldsymbol{C},y_n,\boldsymbol{V})$ 的定义为

$$f(\boldsymbol{C},y_n,\boldsymbol{V})=\begin{cases}1, & y_n\tilde{y}_n\geqslant 1.\\ y_n\tilde{y}_n, & y_n\tilde{y}_n<1.\end{cases} \quad (18)$$

其中,$\tilde{y}_n$ 是样本 $\boldsymbol{x}_n$ 对应的模糊系统的实值输出.

先验分布 $p(\boldsymbol{C},\boldsymbol{V},\boldsymbol{Y})$ 的值与 KE-B-MA 模糊系统的输出直接相关. 如果样本标签 y_n 和实际输出值 $\tilde{y}_n$ 的乘积$\geqslant 1$,即 $y_n\tilde{y}_n\geqslant 1$,那么表明该样本位于分类超平面正确的一边,且在分类面的外侧,则设置该样本对应的 $f(\boldsymbol{C},y_n,\boldsymbol{V})=1$. 相反,如果 $y_n\tilde{y}_n<1$,表明该样本点位于分类超平面的内侧或被错分,则设置该样本对应的 $f(\boldsymbol{C},y_n,\boldsymbol{V})$ 值为样本标签 y_n 和实际输出值 $\tilde{y}_n$ 的乘积. 当 $p(\boldsymbol{C},\boldsymbol{V},\boldsymbol{Y})$ 的值达到最大,即分类准确度达到最高时,KE-B-MA 模糊系统将得到其模型的最大后验(maximum-a-posteriori, MAP)值.

KE-B-MA 假设先验分布 $p(\boldsymbol{Y})$ 为常数. 将式(13)(15)(17)相乘,得到 KE-B-MA 的数据和参数的联合分布,其形式为

$$p(\boldsymbol{X},\boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})= p(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})p(\boldsymbol{U}|\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})p(\boldsymbol{C},\boldsymbol{V}|\boldsymbol{Y})p(\boldsymbol{Y})\propto \exp\left(-\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K} u_{nk}^m \left\|\boldsymbol{x}_n-\boldsymbol{c}_k\right\|^2\right)\times\left(\prod_{n=1}^{N}\prod_{k=1}^{K} u_{nk}^{\alpha_k-1}\right)\times \exp\left(-\frac{1}{2}\sum_{n=1}^{N}(1-f(\boldsymbol{C},y_n,\boldsymbol{V}))\right). \quad (19)$$

通过对式(19)取对数,可得 KE-B-MA 目标函数:

$$J(\boldsymbol{X},\boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})=-\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K}u_{nk}^{m}\|\boldsymbol{x}_n-\boldsymbol{c}_k\|^2+(\alpha_k-1)\sum_{n=1}^{N}\sum_{k=1}^{K}u_{nk}-\frac{1}{2}\sum_{n=1}^{N}(1-f(\boldsymbol{C},y_n,\boldsymbol{V})).\quad(20)$$

从式(20)可以看出,KE-B-MA 将模糊规则前件和后件参数的学习融入到一个概率模型中,当模型的联合分布达到 MAP 值时,参数$\{\boldsymbol{U},\boldsymbol{V}\}$可同时得到最优解.

2.3 模糊规则前件和后件参数的学习策略

为了得到参数$\{\boldsymbol{U},\boldsymbol{V}\}$的全局最优解,如图 2 所示,本文使用蒙特卡洛(MCMC)方法[28]来求解式(19)的 MAP 值. 具体地,本文通过 MH(Metropolis-Hastings)采样算法构造一条满足一定后验分布且平稳分布的 Markov 链. 该 Markov 链的第 t 次迭代的步骤如下:

1) 产生采样值 $\tilde{\boldsymbol{u}}_n$

由 Dirichlet 先验分布 Dirichlet($\boldsymbol{\alpha}$)产生一个新的采样值 $\tilde{\boldsymbol{u}}_n$,$n=1,2,\cdots,N$,即:

$$\tilde{\boldsymbol{u}}_n\sim Dirichlet(\boldsymbol{\alpha}),\quad(21)$$

其中,参数 $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_K)$决定了 Dirichlet 分布的形状,当 α_k 值越接近于 0,产生的 u_{nk} 值趋于二值化(0 或 1);当 α_k 值在 1 附近时,产生的 u_{nk} 值具有模糊性;随着 α_k 值的继续增大,产生的 u_{nk} 值趋向于均匀分布. 实验中,由于数据分布的未知性,$\boldsymbol{\alpha}$ 参数设为 $\boldsymbol{\alpha}=\mathbf{1}_K$.

2) 计算 $\tilde{\boldsymbol{u}}_n$ 对应的接受概率 a_u

根据贝叶斯概率基本原理,$p(\boldsymbol{U}\mid\boldsymbol{X},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})p(\boldsymbol{X},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})=p(\boldsymbol{X},\boldsymbol{U},\boldsymbol{Y},\boldsymbol{C},\boldsymbol{V})=p(\boldsymbol{X},\boldsymbol{U}\mid\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})p(\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$. 显然,当给定后件参数 $\boldsymbol{V}$ 时,条件概率 $p(\boldsymbol{U}\mid\boldsymbol{X},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$正比于 $p(\boldsymbol{X},\boldsymbol{U}\mid\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$. 同时,样本的模糊隶属度相互无关. 因此,采样值 $\tilde{\boldsymbol{u}}_n$ 对应的接受概率 a_u 可以定义为

$$a_u=\min\left(1,\frac{p(\boldsymbol{x}_n,\tilde{\boldsymbol{u}}_n\mid\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})}{p(\boldsymbol{x}_n,\boldsymbol{u}_n^{(t)}\mid\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})}\right),\quad(22)$$

其中,

$$p(\boldsymbol{x}_n,\boldsymbol{u}_n\mid\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})=p(\boldsymbol{x}_n\mid\boldsymbol{u}_n,\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})p(\boldsymbol{u}_n\mid\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})=\prod_{k=1}^{K}N(\boldsymbol{x}_n\mid\boldsymbol{c}_k,\boldsymbol{\Lambda})u_{nk}^{-md/2}Dirichlet(\boldsymbol{u}_n\mid\boldsymbol{\alpha})\propto\prod_{k=1}^{K}\exp\left(-\frac{1}{2}u_{nk}^{m}\|\boldsymbol{x}_n-\boldsymbol{c}_k\|^2\right)u_{nk}^{\alpha_k-1}.\quad(23)$$

值得注意的是,采样值 $\tilde{\boldsymbol{u}}_n$ 的分布和当前样本 $\boldsymbol{x}_n$ 无关,且 $\tilde{\boldsymbol{u}}_n/\boldsymbol{u}_n$ 在不同时刻状态独立,因此接受概率 a_u 不需要进行 Hasting 校验.

3) 以概率 a_u 接受 $\tilde{\boldsymbol{u}}_n$ 为下一个 $\boldsymbol{u}_n^{(t+1)}$

概率 a_u 的计算方法可写为

$$\boldsymbol{u}_n^{(t+1)}=\begin{cases}\tilde{\boldsymbol{u}}_n, & Unif(0,1)\leqslant a_u,\\ \boldsymbol{u}_n^{(t)}, & \text{otherwise},\end{cases}\quad(24)$$

其中,函数 $Unif(0,1)$产生一个在[0,1]范围内均匀分布的随机数.

4) 更新 $\boldsymbol{u}_n$ 的 MAP 值 $\boldsymbol{u}_n^*$

将采样值 $\tilde{\boldsymbol{u}}_n$ 和当前 $\boldsymbol{u}_n^*$ 分别代入式(23),进行 $p(\boldsymbol{x}_n,\tilde{\boldsymbol{u}}_n\mid\boldsymbol{C}^*,\boldsymbol{Y},\boldsymbol{V}^*)$与 $p(\boldsymbol{x}_n,\boldsymbol{u}_n^*\mid\boldsymbol{C}^*,\boldsymbol{Y},\boldsymbol{V}^*)$的比较,如果前者大,将当前 $\tilde{\boldsymbol{u}}_n$ 替换为 $\boldsymbol{u}_n^*$,否则保留当前 $\boldsymbol{u}_n^*$,即:

$$\boldsymbol{u}_n^*=\begin{cases}\tilde{\boldsymbol{u}}_n, & p(\boldsymbol{x}_n,\tilde{\boldsymbol{u}}_n\mid\boldsymbol{C}^*,\boldsymbol{Y},\boldsymbol{V}^*)>\\ & \quad p(\boldsymbol{x}_n,\boldsymbol{u}_n^*\mid\boldsymbol{C}^*,\boldsymbol{Y},\boldsymbol{V}^*),\\ \boldsymbol{u}_n^*, & \text{otherwise}.\end{cases}\quad(25)$$

5) 计算第 t 次迭代后件 $\boldsymbol{V}^{(t+1)}$ 和当前后件 $\boldsymbol{V}^*$

模糊划分矩阵 $\boldsymbol{U}$ 和聚类中心 $\boldsymbol{C}$ 的确定为构建模糊子集对应的模糊隶属度函数提供了依据,第 k 条规则的模糊子集 A_i^k 的模糊隶属度函数可由式(2)计算得到. KE-B-MA 分别使用$\{\boldsymbol{U}^{(t+1)},\boldsymbol{C}\}$和$\{\boldsymbol{U}^*,\boldsymbol{C}\}$计算样本 $\boldsymbol{X}$ 对应的模糊隶属度函数,将样本 $\boldsymbol{X}$ 分别特征映射为$\{\phi_1^{(t+1)},\phi_2^{(t+1)},\cdots,\phi_N^{(t+1)}\}$和$\{\phi_1^*,\phi_2^*,\cdots,\phi_N^*\}$的形式,通过求解标准二次规划式(7),可以得到第 t 次迭代后件参数 $\boldsymbol{V}^{(t+1)}$ 和当前后件参数 $\boldsymbol{V}^*$.

6) 更新 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 的 MAP 值$\{\boldsymbol{U}^*,\boldsymbol{V}^*\}$

在 Markov 链第 t 次迭代的最后,进行 $p(\boldsymbol{X},\boldsymbol{U}^{(t+1)},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}^{(t+1)})$与 $p(\boldsymbol{X},\boldsymbol{U}^*,\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}^*)$值的比较,值大者成为当前新的 MAP 值$\{\boldsymbol{U}^*,\boldsymbol{V}^*\}$,即:

$$\{\boldsymbol{U}^*,\boldsymbol{V}^*\}=\begin{cases}\{\boldsymbol{U}^{(t+1)},\boldsymbol{V}^{(t+1)}\}, & p(\boldsymbol{X},\boldsymbol{U}^{(t+1)},\\ & \quad \boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}^{(t+1)})>p(\boldsymbol{X},\boldsymbol{U}^*,\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}^*),\\ \{\boldsymbol{U}^*,\boldsymbol{V}^*\}, & \text{otherwise}.\end{cases}\quad(26)$$

通过重复上述步骤,当迭代次数足够大时,可以保证得到的参数$\{\boldsymbol{U}^*,\boldsymbol{V}^*\}$是全局收敛.

2.4 KE-B-MA 算法描述和算法分析

根据 2.2 节利用 MCMC 理论构造 Markov 链求解$\{\boldsymbol{U}^*,\boldsymbol{V}^*\}$的叙述,本节给出 KE-B-MA 模糊系统学习算法.

算法 1. KE-B-MA 算法.

输入:训练样本 $\boldsymbol{X}$ 和样本标签 $\boldsymbol{Y}$、知识嵌入矩阵、模糊指数 m、规则数 K、正则化参数 C'、隶属函数尺度参数 h、Markov 链最大迭代次数 $t_{\max}$;

输出:KE-B-MA 模糊系统的模糊规则和决策函数.

初始化:根据知识嵌入矩阵为每条模糊规则的输入向量选取对应的维数;设置隶属度 $\boldsymbol{u}_n^{(1)} \sim Dirichlet(\boldsymbol{\alpha})$, $n=1,2,\cdots,N$;通过式(7)计算后件 $\boldsymbol{V}^{(1)}$;设置$\{\boldsymbol{U}^*,\boldsymbol{C}^*,\boldsymbol{V}^*\}=\{\boldsymbol{U}^{(1)},\boldsymbol{C}^{(1)},\boldsymbol{V}^{(1)}\}$.

步骤 1. 设置 $iter=1$;

步骤 2. 设置 $iterationU=1$;

步骤 3. 对于 $\boldsymbol{x}_n$,通过式(21)采样隶属度 $\bar{\boldsymbol{u}}_n$;

步骤 4. 根据式(22)计算接受概率 $a_{\boldsymbol{u}}$;

步骤 5. 由式(24)使用 $a_{\boldsymbol{u}}$ 计算$(t+1)$次迭代的 $\boldsymbol{u}_n^{(t+1)}$;

步骤 6. $iterationU=iterationU+1$;

步骤 7. 如果 $iterationU \leqslant N$,转到步骤 3;

步骤 8. 根据式(7)(8)计算后件参数 $\boldsymbol{V}^{(t+1)}$ 和其对应的实际输出值 $\tilde{\boldsymbol{Y}}^{(t+1)}$,并计算后件参数 $\boldsymbol{V}^*$ 和其对应的实际输出值 $\tilde{\boldsymbol{Y}}^*$;

步骤 9. 根据式(19)分别计算 $p(\boldsymbol{X},\boldsymbol{U}^{(t+1)},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}^{(t+1)})$和 $p(\boldsymbol{X},\boldsymbol{U}^*,\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}^*)$;

如果 $p(\boldsymbol{X},\boldsymbol{U}^{(t+1)},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}^{(t+1)})>p(\boldsymbol{X},\boldsymbol{U}^*,\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}^*)$,那么 $\boldsymbol{U}^*=\boldsymbol{U}^{(t+1)},\boldsymbol{V}^*=\boldsymbol{V}^{(t+1)}$;

步骤 10. $iter=iter+1$;如果 $iter<t_{\max}$,转到步骤 2;

步骤 11. 使用式(5)计算规则前件参数 $\boldsymbol{\delta}$,并得到模糊规则;

步骤 12. 由式(9)得到系统的决策函数.

通过对算法 1 的分析可知,算法 1 的时间复杂度主要有 2 部分构成:条件概率 $p(\boldsymbol{x}_n,\boldsymbol{u}_n|\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$(式(15))和后件参数 $\boldsymbol{V}$(式(7))的计算,时间复杂度分别是 $O(Kd_k)$和 $O(N^3)$,其中 d_k 为第 k 条规则的输入向量的特征数.因此,算法 1 的时间复杂度小于等于$O(t_{\max}(NKd+KN^3))$.可见,算法 1 运行的时间效率与样本的容量和结构有关,大部分情况下数据集的维数 d 与样本容量 N 之间存在 $d \ll N^2$ 的关系,因此,算法 1 执行时间复杂度近似为 $O(t_{\max}KN^3)$,即等于求解后件参数对应的二次规划所需的时间.为了在一定程度上提高算法的执行效率,可以采用许多成熟的快速算法来提高二次规划的求解速度,例如 SMO(sequential minimal optimization)[29]和 parallel mixture[30]等,其执行效率在样本容量为 N 时分别能够达到$O(N^2)$和$O(N)$.本文采用 SMO 算法来求解后件参数.

MH 采样算法的全局收敛性在文献[28]中得到证明,根据随机游走的策略,无论其参数的初值如何设置,算法最终一定收敛于平稳分布.而本文提出的 KE-B-MA 模糊系统在采用 MH 采样算法的同时,引入了计算模糊规则后件参数的 SVM 算法,其 QP 解法也可保证获得结果的全局收敛性.因此,算法 1 得到的 KE-B-MA 的前件和后件参数是全局最优解.

3 实验研究

3.1 实验设置

为了验证本文方法的有效性,本节将分别通过人工合成数据集以及 UCI 标准数据库数据对 KE-B-MA 算法进行分析与验证.有关实验数据的详细描述分别于 3.2 节和 3.3 节给出.此外,与 KE-B-MA 进行比较的算法包括 4 种典型的基于聚类的模糊系统:FCM-IRLS[20],FCM-SVM-FS[16](模糊系统前件参数学习使用 FCM 聚类)和 L2-TSK-FS[31],以及经典分类算法 SVM.在本文实验部分,各调节参数的设置采取 5 重交叉验证法来选取最优值,参数的详细设置如表 2 所示.所有算法在 Matlab2010b 环境下实现,SVM 算法由 LIBSVM 软件实现,采用 L1-SVM 的形式.

Table 2 Definition and Settings of the Related Parameters for All Algorithms

表 2 实验中各算法参数的定义和设置

Algorithms	Parameters Description
FCM-IRLS, FCM-SVM-FS, L2-TSK-FS, KE-B-MA	The number of rules $K \in \{2,3,\cdots,20\}$, the fuzzy index $m \in \{1.1,1.5,2,2.5,3\}$, the scale parameter $h \in \{0.2^2,0.4^2,\cdots,2^2\}$, the regularization parameter $C' \in \{10^{-3},10^{-2},10^3\}$. $t_{\max}$ in KE-B-MA being $t_{\max}=300$.
SVM	The regularization parameter $C' \in \{10^{-3},10^{-2},10^3\}$, Gaussian kernel width parameter $\sigma \in \{10^{-2},10^{-1},\cdots,10^2\}$.

为对各算法的性能进行全面比较,实验采用 4 个评价指标:1)分类精度和方差;2)模糊系统的规则数和 SVM 的支持向量数;3)各模糊系统规则中包含的特征数;4)各算法所对应模型的变量数.SVM 包含的变量数为 $L+L(s+1)+1$,其中 L 表示支持向量数,s 表示样本的维数;L2-TSK-FS 包含的变量数为 $2Kd+K(d+1)$,K 表示模糊规则数,d 表示模糊规则对应输入空间的维数;FCM-SVM-FS 和 KE-B-MA 包含的变量数为 $2Kd+K+1$;FCM-IRLS 包含的变量数为 $2Kd+K$.

3.2 人工数据集实验

人工数据集由 500 个 10 维在[0,1]之间均匀分布的随机数 $\boldsymbol{x}$ 构成.每个样本对应的类标签定义为 $y=\mathrm{sgn}(x_1+x_2+\cdots+x_{10})$.

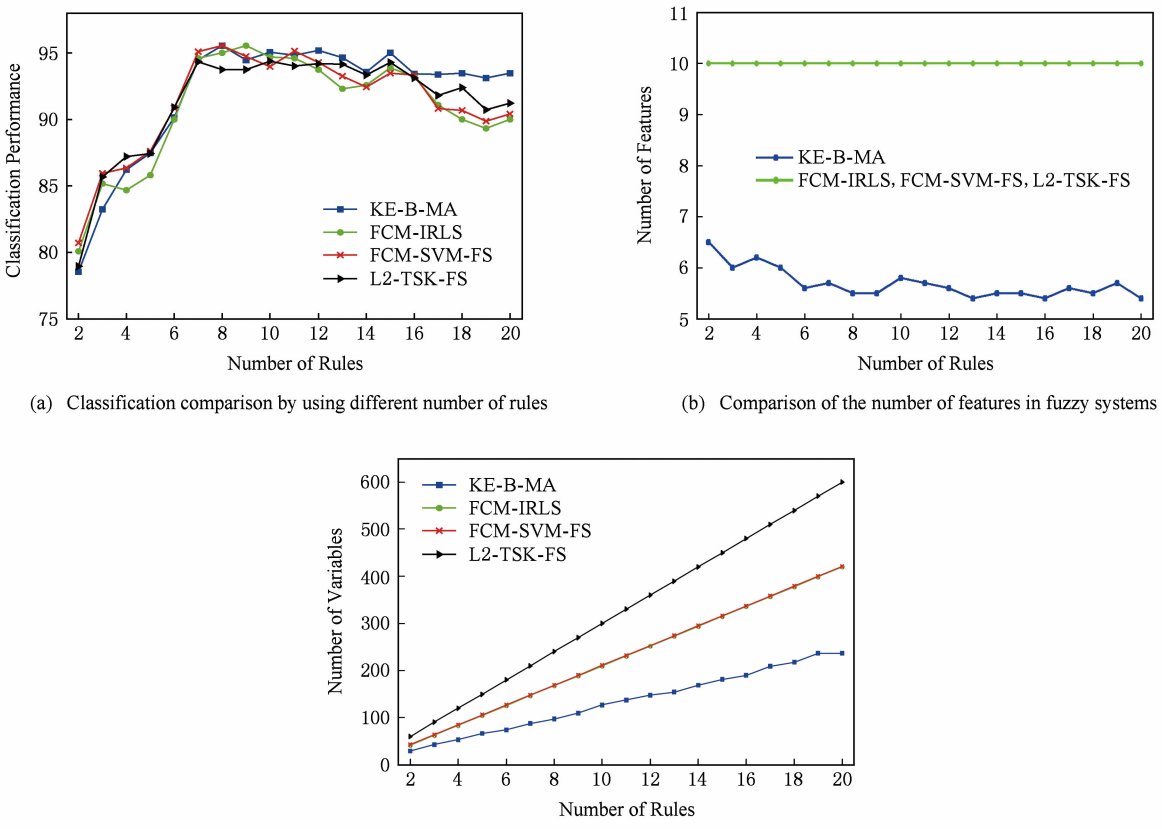
为了验证 KE-B-MA 中基于贝叶斯理论的前件后件参数同时学习策略的有效性,本节实验中 5 重交叉验证的每一折中知识嵌入矩阵通过随机选择样本特征和随机选择语义标签的方式获得. KE-B-MA 中知识嵌入矩阵中各元素的选择原则是:模糊集应覆盖输入空间所有的特征. 实验比较了 KE-B-MA 与其他 4 种对比算法的分类平均准确率和标准差、规则数、规则包含的特征数和模型包含的变量数 4 个指标,结果如图 3 所示(最佳结果用黑体标出). 其中,SVs 表示 SVM 获得的支持向量数.

Table 3 Accuracy Performance Comparison of Several Algorithms on the Artificial Dataset

表 3 各种算法在人工数据集上的性能比较

Algorithms	Classification Performance (Standard Deviation)	Model Structure	Features in Fuzzy Systems	Variables in Fuzzy Systems
SVM	0.9534(0.0155)	50.2 SVs	10	603.4
FCM-IRLS	0.9554(0.0078)	8 rules	10	168
FCM-SVM-FS	**0.9556**(0.0171)	8 rules	10	169
L2-TSK-FS	0.9438(0.0155)	10 rules	10	310
KE-B-MA	0.9555(0.0169)	**7.8** rules	**5.8**	**99.28**

SVs: Support vectors



(a) Classification comparison by using different number of rules

(b) Comparison of the number of features in fuzzy systems

(c) Comparison of the number of variables in fuzzy rules

Fig. 3 Performance comparison between KE-B-MA, FCM-SVM-FS, FCM-IRIS and L2-TSK-FS on the artificial dataset

图 3 KE-B-MA, FCM-SVM-FS, FCM-IRIS, L2-TSK-FS 在人工集上的性能比较

由表 3 和图 3 可以看出:

1) 本文方法与实验 4 种对比算法相比,在人工数据集上取得了相当的分类性能,KE-B-MA 的分类准确度仅比最佳分类效果低 0.01%.

2) FCM-IRLS, FCM-SVM-FS, L2-TSK-FS 模糊系统采用样本的全部 10 个特征来构造模糊规则,而 KE-B-MA 构造的模糊规则通过知识嵌入矩阵只使用了 5.8 个特征,说明 KE-B-MA 虽然没有使用

样本的全部特征,但通过规则前件和后件共同学习的策略能够充分考虑输入空间和输出空间之间的内在联系,保证了模糊系统的分类精确度.这一点是FCM-IRLS,FCM-SVM-FS,L2-TSK-FS不具备的.

3) 表3的最后一行比较了实验中5种算法各自模型中的变量数,显然KE-B-MA模糊系统具有绝对的优势,其模型中的变量数远小于另外4种算法.

4) 对于模糊系统而言,模糊规则的可解释性不仅与模糊规则数有关,同时也与规则在输入空间的划分清晰度有关.KE-B-MA中通过使用知识嵌入矩阵设定了语义清晰的模糊隶属度函数中心,保证了输入空间的划分清晰度,这一优点也是实验中其他3种模糊系统所不具备的.

为了对KE-B-MA所得模糊规则这一特性有更直观的了解,下面给出了实验中设定的知识嵌入矩阵和最佳分类精度时得到的模糊规则,分别如表4和图4所示.

Table 4 The Knowledge Embedded Matrix Corresponding to Eight Fuzzy Rules in One Running Time on the Artificial Dataset

表4 某次运行时人工集上产生的8条模糊规则对应的知识嵌入矩阵

Rules	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
Rule 1	5	2	DC	3	DC	1	3	DC	5	DC
Rule 2	DC	1	DC	DC	1	2	1	DC	2	2
Rule 3	3	3	3	1	3	DC	DC	1	1	DC
Rule 4	DC	DC	DC	4	4	5	4	DC	3	5
Rule 5	2	5	2	DC	DC	DC	5	DC	4	DC
Rule 6	DC	4	DC	4	5	DC	5	3	3	4
Rule 7	DC	DC	5	DC	2	4	2	2	DC	3
Rule 8	1	DC	1	2	DC	3	DC	5	DC	1

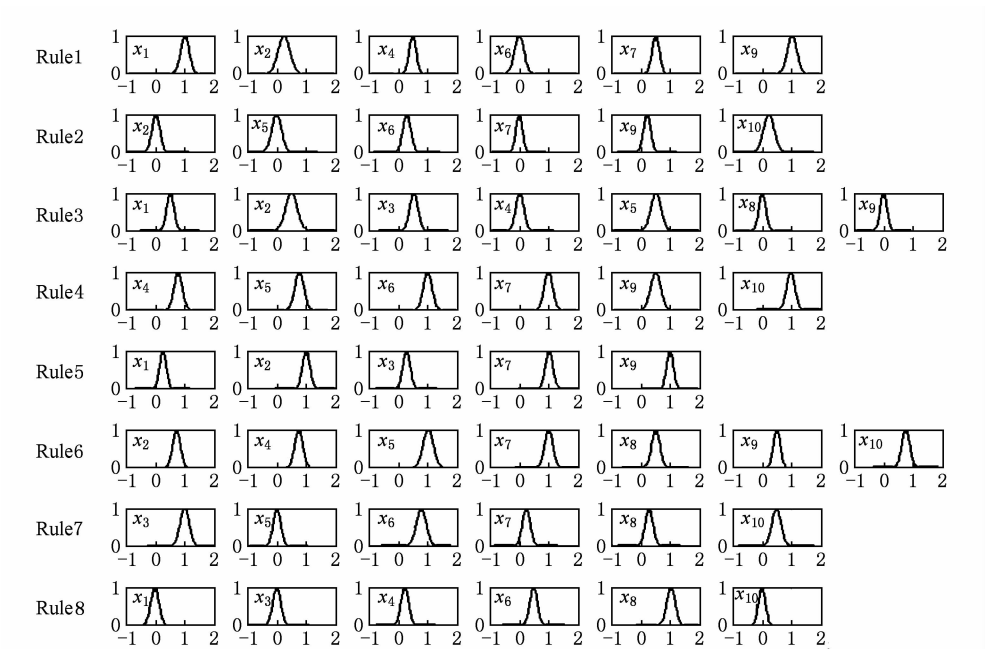


Fig. 4 Fuzzy membership functions obtained by KE-B-MA by using the knowledge embedded matrix in table 4

图4 KE-B-MA模糊系统根据表4知识嵌入矩阵得到的模糊隶属度函数

图4所示规则对应的后件参数为 $\boldsymbol{V}=(0.9731, 0.8931, 1.2387, 1.7321, -1.2971, -1.8662, -07582, -0.9265)^{\mathrm{T}}$.由于篇幅所限,下面给出所得第1条模糊规则的形式:

R_1: if x_1 is very high and x_2 is low and x_4 is medium and x_6 is very low and x_7 is medium and x_9 is very high, then $v_1=0.9731$.

3.3 UCI数据集实验

本节通过5个真实的UCI机器学习库[32]中的数据集来验证KE-B-MA的性能:Pima Indians'

Diabetis(Diabetis)数据集、Breast Cancer(Breast)数据集、Liver 数据集、Image Segmentation(Image)数据集和 Ringnorm 数据集. Diabetis 数据集共有 768 个数据样本,每个样本有 8 个特征,数据集分为糖尿病类和正常类,样本数分别为 500 和 268. Breast 共有 288 个数据样本,其中属于乳腺癌类样本 85 个,属于良性乳腺癌类样本 201 个,每个样本有 9 个特征. Liver 数据集共有 345 个数据样本,包含肺部异常类样本 145 个和正常类样本 200 个,每个样本有 6 个特征. Image 数据集为图像分割测试数据集,共有 2 310 个数据样本,含 18 个特征,其中正类 1 320 个样本,负类 990 个样本. Ringnorm 数据集包含 7 400 个样本,其特征数为 20 维,正类和负类样本分别为 3 364 和 3 736 个. 实验中 5 个 UCI 集的数据均归一化至[0,1]范围内. 本节实验中知识矩阵的建立借助于 F-score 方法[33],优先选择 F-score 值较大的特征,并确保模糊集覆盖输入空间所有的特征. 需要说明的是,实际应用中,知识嵌入矩阵常通过事先积累的专家经验获得.

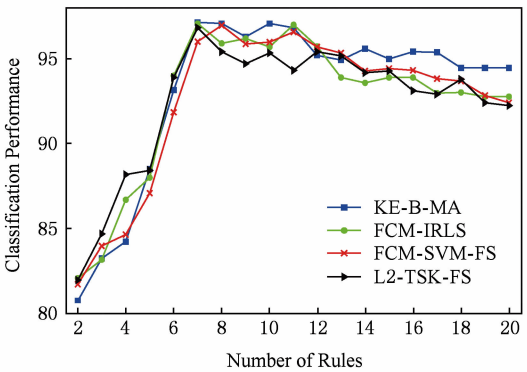
表 5 给出了 5 种分类算法在 UCI 数据集上的实验结果. 与人工数据集实验类似,实验比较了 KE-B-MA 与 3 种对比算法在规则数{2,3,…,20}范围内分类精确度、规则包含的特征数和模型包含的变量数这 3 个指标,结果如图 5 所示.

Table 5 Performance Comparisons on UCI Datasets with Different Algorithms

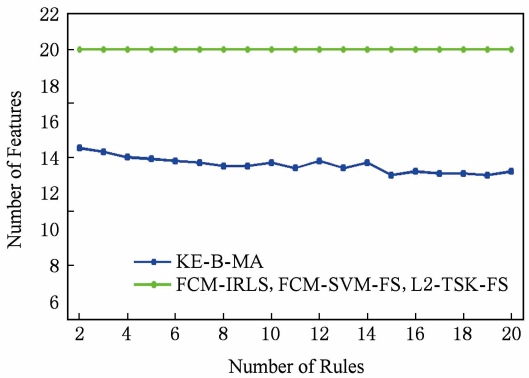
表 5 各种算法在 UCI 数据集上的性能比较

Datasets	Algorithms	Classification Performance (Standard Deviation)	Model Structure	Features in Fuzzy Systems	Variables in Fuzzy Systems
Diabetis	SVM	0.7698(0.0155)	57.0 SVs		571.0
	FCM-IRLS	0.7758(0.0178)	**10.0 rules**	8.0	170.0
	FCM-SVM-FS	0.7763(0.0178)	**10.0 rules**	8.0	171.0
	L2-TSK-FS	0.7766(0.0157)	12.0 rules	8.0	300.0
	KE-B-MA	**0.7775**(0.0151)	10.7 rules	**5.1**	**120.8**
Breast	SVM	**0.7402**(0.0332)	102.0 SVs		1123.0
	FCM-IRLS	0.7306(0.0250)	9.0 rules	9.0	171.0
	FCM-SVM-FS	0.7308(0.0215)	**8.0 rules**	9.0	153.0
	L2-TSK-FS	0.7300(0.0323)	10.0 rules	9.0	280.0
	KE-B-MA	0.7319(0.0332)	8.2 rules	**5.7**	**102.7**
Liver	SVM	0.7689(0.0241)	246.8 SVs		2222.2
	FCM-IRLS	0.8000(0.0185)	10.0 rules	9.0	190.0
	FCM-SVM-FS	0.8026(0.0197)	**9.0** rules	9.0	172.0
	L2-TSK-FS	**0.8100**(0.0082)	10.0 rules	9.0	280.0
	KE-B-MA	0.8076(0.0245)	10.3 rules	**5.7**	**128.7**
Image	SVM	**0.9197**(0.0111)	206.7 SVs		4135.0
	FCM-IRLS	0.9181(0.0182)	**13.0 rules**	18.0	481.0
	FCM-SVM-FS	0.9160(0.0155)	**13.0 rules**	18.0	482.0
	L2-TSK-FS	0.9113(0.0135)	14.0 rules	18.0	770.0
	KE-B-MA	0.9190(0.0141)	13.6 rules	**10.4**	**297.5**
Ringnorm	SVM	0.9403(0.0069)	1665.0 SVs		36631.0
	FCM-IRLS	0.9708(0.0078)	**7.0 rules**	20.0	287.0
	FCM-SVM-FS	0.9697(0.0086)	8.0 rules	20.0	329.0
	L2-TSK-FS	0.9683(0.0062)	**7.0 rules**	20.0	427.0
	KE-B-MA	**0.9712**(0.0074)	7.1 rules	**13.8**	**204.1**

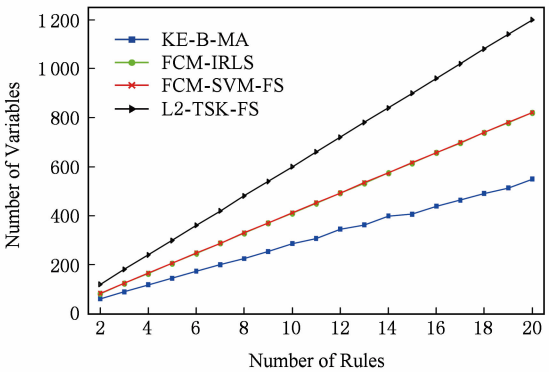
SVs: Support vector



(a) Classification comparison by using different number of rules



(b) Comparison of the number of features in fuzzy systems



(c) Comparison of the number of variables in fuzzy rules

Fig. 5 Performances comparison between KE-B-MA, FCM-SVM-FS FCM-IRIS and L2-TSK-FS on Ringnorm dataset

图 5 KE-B-MA, FCM-SVM-FS, FCM-IRIS, L2-TSK-FS 在 Ringnorm 集上的性能比较

从这些实验结果,我们可以看出:

1) UCI 数据集上获得的实验结果与人工数据集的实验结果是一致的. 由于使用了知识嵌入矩阵实现了模糊隶属度函数中心和规则特征的选择,可以保证得到的模糊规则具有高度的可解释性,同时使用基于贝叶斯理论的前件和后件同时学习的策略,保证了所得模型参数的全局最优解,并建立起输入空间和输出空间的内在联系,保证了分类器的分类精度,且结果偏差不大,较为稳定.

2) 从图 5 可以看出,当设定的模糊规则数较小时($K<4$),KE-B-MA 的性能略逊于其他 3 种模糊系统,当设定的模糊规则数逐渐增大时,KE-B-MA 的性能快速提升,5 个 UCI 数据集上除了 Liver 数据集上分类准确度略差于 FCM-IRLS,另 4 个 UCI 数据集上的分类准确度在 4 种模糊系统中是最优的. 究其原因在于:KE-B-MA 构造的每条模糊规则的输入特征不同,在规则数过小时不足以映射完整的输入空间,但是一旦模糊规则数增加到合适值,KE-B-MA 能获得较之经典模糊系统更好或可比的分类性能. 同时,虽然 FCM-SVM-FS, FCM-IRIS, L2-TSK-FS 模糊系统在规则数过小时分类性能略优于 KE-B-MA,但此时的分类精度也低于其最高值,从获得优良分类效果的角度看,此时的规则数不会被最终选择. 而 FCM-SVM-FS, FCM-IRIS, L2-TSK-FS 在 UCI 数据集上获得最佳分类性能时模型的变量数和规则的复杂程度远大于 KE-B-MA 的变量数和规则的复杂度,也再次说明了 KE-B-MA 得到的模型规则具有高度的可解释性.

3.4 参数敏感性实验

本文所提 KE-B-MA 实现需要协调的实验参数有:K, m, h, C',其中 K 为模糊规则数,C'为结构化风险正则化参数,h 为隶属函数尺度参数,m 为模糊指数. 模糊规则数 K 的取值与样本的分布有关,在基于聚类的模糊系统中,选择合适的模糊规则数常用 2 种方法:

1) 借助于 Xie-Beni 指数[34]、Mountain potential 指数[35]等来确定,但这种方法本质上只是从聚类的角度对输入空间进行划分,没有考虑模糊系统输入空间和输出空间之间的联系,应用在模糊系统中往往效果不佳;

2) 通过手动设定的方式来确定模糊规则数,也是本文所用方法. KE-B-MA 为了保证所得规则具有高度的可解释性使用知识嵌入矩阵对模糊规则的特征和隶属度函数中心进行选择,在这种情况下,通过交叉验证策略对模糊规则数 K 进行寻优是合适的.

本节重点评价 KE-B-MA 中 m,h,C' 这 3 个参数对分类精度的影响,实验使用人工集(Artificial dataset),Liver 和 Ringnorm 数据集作为实验数据,模糊规则数分别固定为 8,10,7 条,另外采用固定其他参数寻优的方法,图 6~8 分别显示了上述 3 个参数所提方法的性能影响曲线.

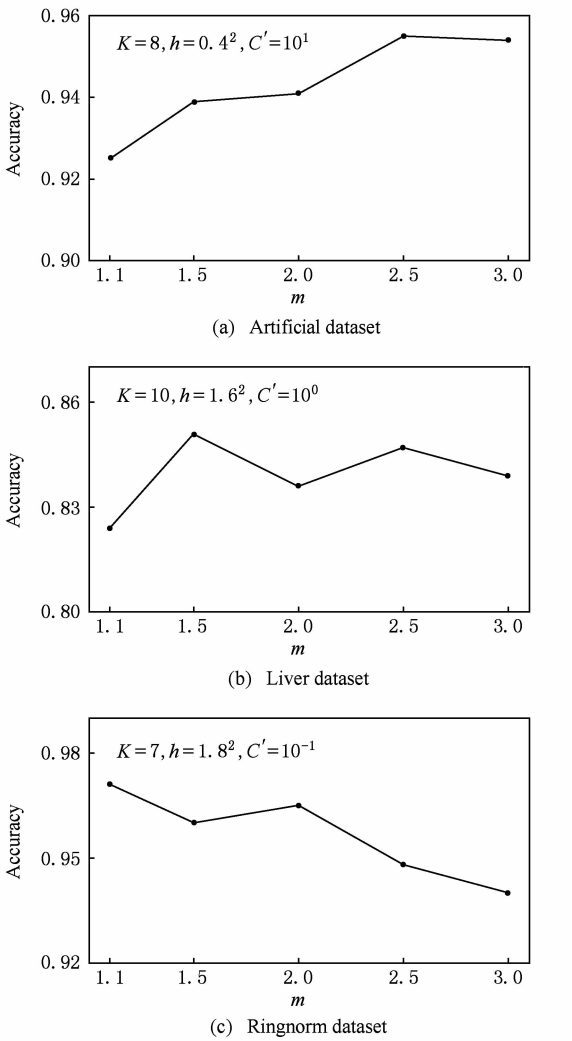


Fig. 6 Parameter m sensitivity analysis on three datasets

图 6 参数 m 的敏感性实验

由此可得 3 个结论:

1) KE-B-MA 中模糊指数 m 的作用等同于 FCM 聚类中的作用,m 在 FCM 中必须设置为大于 1,否则得不到模糊隶属度的解析解,但在 KE-B-MA 中 m 在理论上可以取任何值,甚至为负值. 根据文献[36]对模糊指数的物理意义解释,本文设定 m 的调节区间是 $m\in\{1.1,1.5,2,2.5,3\}$. 由图 6 可以出,KE-B-MA 的分类精度对 m 值较敏感. 因此,通过交叉验证策略对参数 m 进行寻优是合理的.

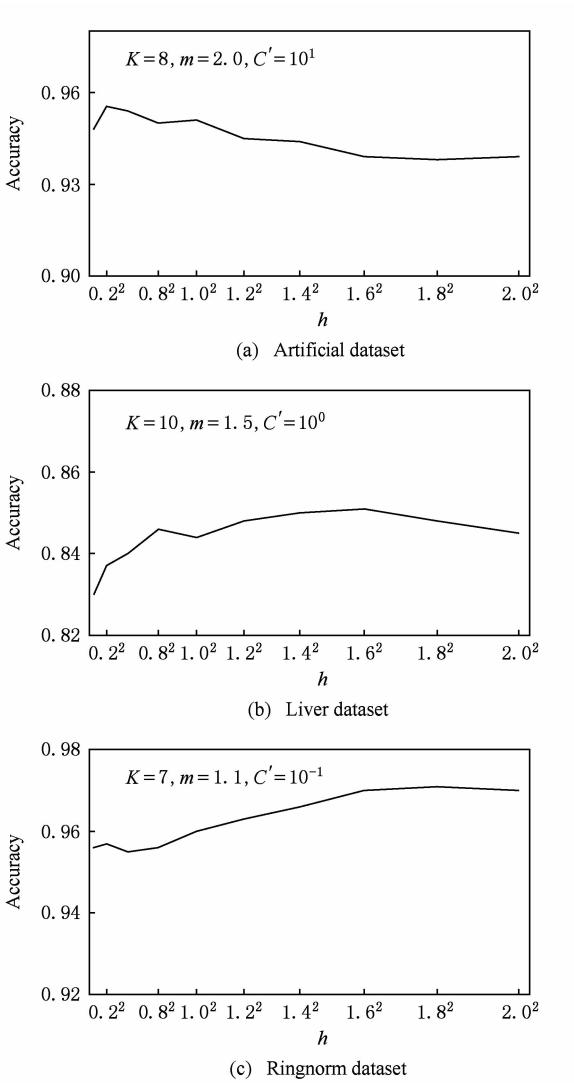


Fig. 7 Parameter h sensitivity analysis on three datasets

图 7 参数 h 的敏感性实验

2) 尺度参数 h 对 KE-B-MA 的分类性能起着相对温和的影响. 从图 7 可以看出,在 h 的设定区间内,KE-B-MA 在 3 个实验数据集上分类效果的最高值与最低值的变化幅度不超过 4%. 同时,正如我们熟知的,尺度参数 h 的设置对模糊规则的清晰性和可解释性起到了重要的影响,因此,对该参数的获取也应使用交叉验证的方法.

3) KE-B-MA 中模糊规则后件的,学习基于典型的 SVM 模型,所有对正则化参数 C' 具有较强的敏感性,C' 在一定范围内的不同取值明显影响所提

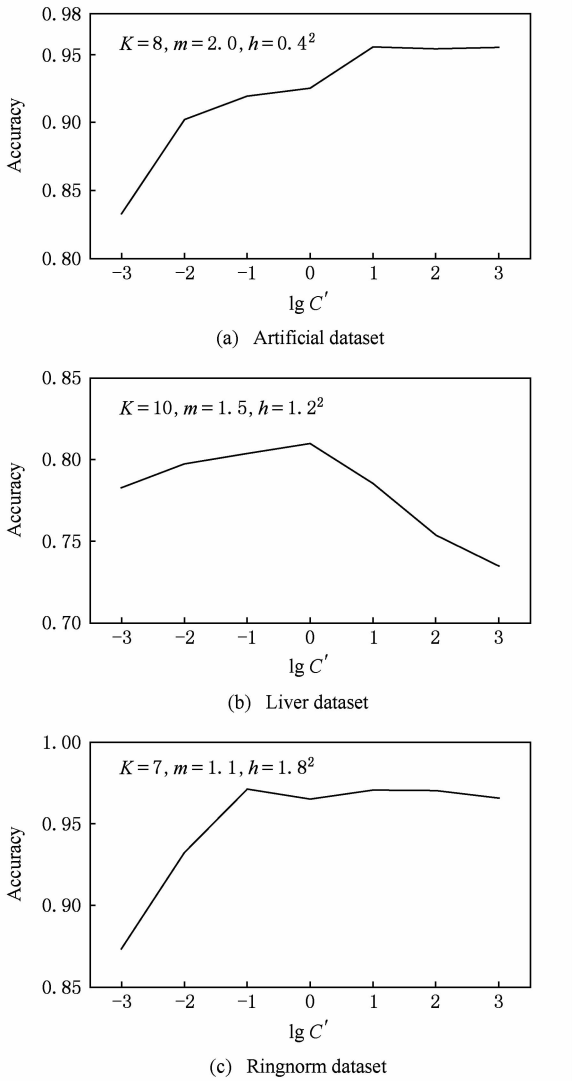


Fig. 8 Parameter C' sensitivity analysis on three datasets

图 8 参数 C'的敏感性实验

方法的泛化性能,且 C'的取值与数据集的分布有关,这也进一步说明了对 C'寻优的重要性.

4 总 结

本文使用贝叶斯概率模型提出了一种新的 MA 型模糊系统,即 KE-B-MA 模糊系统. KE-B-MA 运用 DC 方法实现对模糊规则的特征和隶属度函数中心进行选择,保证了模糊划分的清晰性和有效降低了所建系统的复杂性,解决了传统基于聚类的模糊系统可解释性不高,且需要数据的全部特征来构造模糊规则易造成规则繁杂的问题. KE-B-MA 运用 MH 采样构建了一个 Markov 链实现模糊规则前件和后件参数的同时学习,保证了所得结果的全局最优解. 这些特性是传统的基于聚类的模糊系统所不具备的. 通过人工数据和真实数据集的仿真实验,结果亦表明了本文算法的分类性能较之传统方法相当,但获得的模糊规则更具可解释性. 应当指出,目前本文算法仍存在一些不足之处,例如对 KE-B-MA 模糊系统能否有效解决大样本等问题没有进行深入探讨,当数据容量极大时,从 MH 采样和二次规划求解角度而言,KE-B-MA 仍面临进一步提高实用性的挑战,这将作为我们下阶段的研究重点.

参 考 文 献

[1] Huo Weigang, Shao Xiuli. A fuzzy associative classification method based on multi-objective evolutionary algorithm [J]. Journal of Computer Research and Development, 2011, 48(4): 567-575 (in Chinese)
(霍纬纲, 邵秀丽. 一种基于多目标进化算法的模糊关联分类方法[J]. 计算机研究与发展, 2011, 48(4): 567-575)

[2] Sanz J, Fernández A, Bustince H, et al. IVTURS: A linguistic fuzzy rule-based classification system based on a new interval-valued fuzzy reasoning method with tuning and rule selection [J]. IEEE Trans on Fuzzy Systems, 2013, 21(3): 399-411

[3] Juang C F, Hsiao C M, Hsu C H. Hierarchical cluster-based multispecies particle-swarm optimization for fuzzy-system optimization [J]. IEEE Trans on Fuzzy Systems, 2010, 18(1): 14-26

[4] Alcalá R, Alcala-Fdez J, Casillas J, et al. Local identification of prototypes for genetic learning of accurate TSK fuzzy rule-based systems [J]. International Journal of Intelligent Systems, 2007, 22(9): 909-941

[5] Lughofer E, Buchtala O, Reliable all-pairs evolving fuzzy classifiers [J]. IEEE Trans on Fuzzy Systems, 2013, 21(4): 625-641

[6] Fazzolari M, Alcalá R, Nojima Y, et al. A review of the application of multiobjective evolutionary fuzzy systems: Current status and further directions [J]. IEEE Trans on Fuzzy Systems, 2013, 21(1): 45-65

[7] Ishibuchi H, Yamamoto T. Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining [J]. Fuzzy Sets and Systems, 2004, 141(1): 59-88

[8] Derrac J, Verbiest N, García S, et al. On the use of evolutionary feature selection for improving fuzzy rough set based prototype selection [J]. Soft Computing, 2013, 17(2): 223-238

[9] Alcala-Fdez J, Alcalá R, Herrera F. A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning [J]. IEEE Trans on Fuzzy Systems, 2011, 19(5): 857-872

[10] Chen Yicheng, Pal N R, Chung I. An integrated mechanism for feature selection and fuzzy rule extraction for classification [J]. IEEE Trans on Fuzzy System, 2012, 20(4): 683-698

[11] Chen Yixin, Wang J Z. Support vector learning for fuzzy rule-based classification systems [J]. IEEE Trans on Fuzzy System, 2003, 11 (6): 716-728

[12] Chiang J H, Hao P Y. Support vector learning mechanism for fuzzy rule-based modeling: A new approach [J]. IEEE Trans on Fuzzy Systems, 2004, 12 (1): 1-12

[13] Juang C F, Chen G C. A TS fuzzy system learned through a support vector machine in principal component space for real-time object detection [J]. IEEE Trans on Industrial Electronics, 2012, 59(8): 3309-3320

[14] Lin C T, Yeh C M, Liang S F, et al. Support-vector-based fuzzy neural network for pattern classification [J]. IEEE Trans on Fuzzy Systems, 2006, 14 (1): 31-41

[15] Juang C F, Chiu S H, Shiu S J. Fuzzy system learned through fuzzy clustering and support vector machine for human skin color segmentation [J]. IEEE Trans on System, Man and Cybernetics-Part A: Systems and Humans, 2007, 37(6): 1077-1087

[16] Cheng W, Juang C. An incremental support vector machine-trained TS-type fuzzy system for online classification problems [J]. Fuzzy Sets and Systems, 2011(163): 24-44

[17] Pan Weimin, He Jun. Neuro-fuzzy system modeling with density-based clustering [J]. Journal of Computer Research and Development, 2010, 47(11): 1986-1992 (in Chinese) (潘维民, 何骏. 基于密度聚类的神经模糊系统建模算法 [J]. 计算机研究与发展, 2010, 47(11): 1986-1992)

[18] Leski J M. An ε-margin nonlinear classifier based on fuzzy if-then rules [J]. IEEE Trans on System, Man and Cybernetics, Part B: Cybernetics, 2004, 34(1): 68-76

[19] Leski J M. TSK-fuzzy modeling based on ε-insensitive learning [J]. IEEE Trans on Fuzzy System, 2005, 13(2): 181-193

[20] Leski J M. Fuzzy ($c+p$)—means clustering and its application to a fuzzy rule-based classifier: Towards good generalization and good interpretability [J]. IEEE Trans on Fuzzy System, 2015, 23(4): 802-812

[21] Deng Zhaohong, Cao Longbing, Jiang Yizhang, et al. Minimax probability TSK fuzzy system classifier: A more transparent and highly interpretable classification model [J]. IEEE Trans on Fuzzy System, 2015, 23(4): 813-826

[22] Deng Zhaohong, Jiang Yizhang, Chung F L, et al. Knowledge-leverage based fuzzy system and its modeling [J]. IEEE Trans on Fuzzy System, 2013, 21(4): 597-609

[23] Azeem M F, Hanmandlu M, Ahmad N. Generalization of adaptive neural-fuzzy inference systems [J]. IEEE Trans on Neural Networks, 2000, 11(6): 1332-1346

[24] Bezdek J C. Pattern Recognition with Fuzzy Objective Function Algorithms [M]. Amsterdam, Netherlands: Kluwer Academic Publishers, 1981

[25] Dehzangi O, Zolghadri M, Taheri S, et al. On equivalence of FIS and ELM for interpretable rule-based knowledge representation [J]. IEEE Trans on Neural Networks and Learning Systems, 2015, 26(7): 1417-1430

[26] Lan Y, Soh Y C, Huang G B. Two-stage extreme learning machine for regression [J]. Neurocomputing, 2010, 73(16): 3028-3038

[27] Glenn T C, Zare A, Gader P D. Bayesian fuzzy clustering [J]. IEEE Trans on Fuzzy System, 2015, 23(5): 1545-1561

[28] Robert C, Casella G. Monte Carlo Statistical Methods [M]. Berlin: Springer, 2005

[29] Chang C C, Lin C J. LIBSVM: A library for support vector machines [EB/OL]. 2001 [2015-09-28]. http://www.csie.ntu.edu.tw/~cjlin/libsvm

[30] Collobert R, Bengio S, Bengio Y. A parallel mixture of SVMs for very large scale problems [J]. Neural Computation, 2002, 14 (5): 1105-1114

[31] Deng Zhaohong, Choi K S, Chung F L, et al. Scalable TSK fuzzy modeling for very large datasets using minimal-enclosing-ball approximation [J]. IEEE Trans on Fuzzy Systems, 2011, 19(2): 210-226

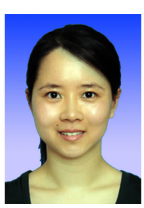
[32] Bache K, Lichman M. UCI database [EB/OL]. 2013 [2015-09-28]. http://www.ics.uci.edu/%20mlearn/ MLRepository.html

[33] Chen Y W, Lin C J. Feature Extraction: Foundations and Applications [M]. Berlin: Springer, 2007

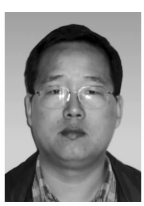
[34] Xie Xiaoliang, Beni G. A validity measure for fuzzy clustering [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1991, 13(4): 841-846

[35] Mezquida E T, Rubio A, Sánchez-Palomares O. Evaluation of the potential index model to predict habitat suitability of forest species: The potential distribution of mountain pine (pinus uncinata) in the Iberian peninsula [J]. European Journal of Forest Research, 2010, 129(1): 133-140

[36] Pal N R, Bezdek J C. On cluster validity for the fuzzy c-means model [J]. IEEE Trans on Fuzzy Systems, 1995, 3 (3): 370-379



Gu Xiaoqing, born in 1981. PhD candidate and lecturer. Her main research interests include pattern recognition and machine learning.



Wang Shitong, born in 1964. Professor and PhD supervisor. His main research interests include artificial intelligence, neuro-fuzzy systems, pattern recognition, and image processing.