文章编号: 1001-0920(2019)06-1319-06 DOI: 10.13195/j.kzyjc.2017.1589

基于D-vine Copula理论的贝叶斯分类器设计

王 蓓[1†], 孙玉东[1], 金 晶[1], 张 涛[2], 王行愚[1]

(1. 华东理工大学 化工过程先进控制和优化技术教育部重点实验室，上海 200237；2. 清华大学 自动化系，北京 100084)

摘 要: 高斯判别分析、朴素贝叶斯等传统贝叶斯分类方法在构建变量的联合概率分布时,往往会对变量间的相关性进行简化处理,从而使得贝叶斯决策理论中类条件概率密度的估计与实际数据之间存在一定的偏差. 对此,结合Copula函数研究特征变量之间的相关性优化问题,设计基于D-vine Copula理论的贝叶斯分类器,主要目的是为了提高类条件概率密度估计的准确性. 将变量的联合概率分布分解为一系列二元Copula函数与边缘概率密度函数的乘积,采用核函数方法对边缘概率密度进行估计,通过极大似然估计对二元Copula函数的参数分别进行优化,进而得到类条件概率密度函数的形式. 将基于D-vine Copula理论的贝叶斯分类器应用到生物电信号的分类问题上,并对分类效果进行分析和验证. 结果表明,所提出的方法在各项分类指标上均具备良好的性能.

关键词: 贝叶斯决策；相关性分析；类条件概率密度估计；D-vine Copula；模式识别；生物电信号

中图分类号: TP273 **文献标志码:** A

Bayesian classifier based on D-vine Copula theory

WANG Bei[1†], *SUN Yu-dong*[1], *JIN Jing*[1], *ZHANG Tao*[2], *WANG Xing-yu*[1]

(1. Key Laboratory of Advanced Control and Optimization for Chemical Processes，Ministry of Education，East China University of Science and Technology，Shanghai 200237，China；2. Department of Automation，Tsinghua University，Beijing 100084，China)

Abstract: In the traditional Bayesian classifiers such as the Gaussian discriminant analysis method and the Naive Bayesian method, the correlation between variables are commonly simplified when constructing the joint probability distribution of variables. Accordingly, the estimation of the class conditional probability density would have differences with the actual data. In this study, a Bayesian classifier based on the D-vine Copula theory is developed by investigating on the correlation between variables. The main objective is to improve the accuracy of the class conditional probability density estimation. The joint probability distribution of variables is decomposed into a series of pair Copula functions and marginal probability density functions. The kernel function method is adopted to estimate the marginal probability density. The parameters of pair Copula functions are optimized by the maximum likelihood estimation. The developed method is analyzed and validated on the classification of neurophysiological signals. The obtained results show that it has better performance on several classification indexes.

Keywords: Bayesian decision；correlation analysis；class conditional probability density estimation；D-vine Copula；pattern recognition；neurophysiological signal

0 引 言

模式识别方法在文本处理、图像处理、统计学习、数据挖掘等方面发挥着重要作用[1-3]. 在诸多流行的模式识别分类器中,贝叶斯分类器是其中之一,其基本思想可以认为是从先验信息中推断后验信息的过程. 对于贝叶斯分类器而言,常用的方法有高斯判别分析、朴素贝叶斯分类器[4]等. 高斯判别分析中,假设待分类的每一类对象都服从多元高斯分布,这种假设较为普遍,主要是因为该假设可以近似地模拟实际应用中多种数据的分布,从而简化复杂分布问题的分析[5]. 然而,这种假设与数据的真实分布还是有一定差距的. 首先,多元高斯分布的协方差矩阵仅能描述特征之间的线性相关性[6];其次,多元高斯分布的边缘分布为一元高斯分布,而实际应用中,特征是否服从高斯分布还有待商榷. 朴素贝叶斯分类器对待分类对象的特征作了条件独立性假设,该假设略去了特

收稿日期: 2017-11-23；**修回日期:** 2018-07-05.

基金项目: 国家自然科学基金项目(61773164)；上海市自然科学基金项目(16ZR1407500).

责任编委: 陈虹.

†通讯作者. E-mail: beiwang@ecust.edu.cn.

征之间的相关性,从而忽略了实际数据中互相关联的成分[7]. 这两种模型对特征之间的相关性尚没有提供相对准确的构建方式,如何分析、提取和利用多维样本数据的信息来对复杂的相关性结构进行建模,从而对贝叶斯分类器中的类条件概率密度实现准确估计是值得研究的问题.

Copula理论是Sklar在1959年提出的,主要应用于金融风险管理领域[8]. 在控制科学相关的研究中,Copula理论对算法性能的提升也得到了关注. 王丽芳等[9]将多元Copula函数应用于分布估计算法的研究中,通过仿真实验表明,引入Copula理论的cEDA算法能够更快地收敛于最优解. 韩敏等[10]将Copula理论应用于互信息估计,通过二维高斯数据的仿真实验表明,基于Copula熵的互信息估计算法在计算复杂度和精度方面,相比于核方法、k近邻方法和直方图法,具备更高的性能. 许民利等[11]将Copula函数与CVaR相结合,构建了随机需求与随机价格之间的决策模型,给出该模型的具体求解方法,并证明了该模型的解的唯一性.

Vine Copula是在Copula理论的基础上发展起来的. 传统的多元Copula函数[12]如椭圆Copula簇、阿基米德Copula簇等在处理高维变量时,变量间的相关性优化问题比较复杂,且计算量较大. 早期,Joe[13]借助Vine结构将多元Copula函数分解成一系列二元Copula函数的乘积,该模型通过一系列二元Copula函数可以构建变量间复杂的相关性,并且减少了计算的复杂程度,使得Vine Copula受到关注. 在后来的研究中,Aas等[14]检验了Vine Copula在计算复杂度和拟合能力上比传统的多元阿基米德Copula模型具有更好的性能. Czado[15]通过三维D-vine Copula的模型,使用不同种类的二元Copula函数描述变量间的不对称性,并展示了Vine Copula模型的灵活性. 近年来,Vine Copula函数作为随机变量相关性建模工具被广泛应用在资产收益波动[16]、化工故障诊断[17]、能源管理[18-19]等领域,并取得了较为显著的效果.

为了提高贝叶斯分类器中的类条件概率密度估计的准确性,同时考虑特征之间存在的复杂相关性,本文设计并提出基于D-vine Copula理论的贝叶斯分类器. 首先,通过D-vine Copula理论将变量的联合概率分布分解成一系列二元Copula函数与边缘概率密度函数乘积的形式;然后,依据特征样本之间相互关联的特性,选取合适的二元Copula函数,并采用核函数方法对边缘概率密度函数进行估计,从而构建贝叶斯分类器中的类条件概率密度函数的形式;最后,将该分类器应用于实际生物电信号的分类问题,对模型进行分析和验证. 通过与高斯判别分析、朴素贝叶斯分类器以及支持向量机模型的比较,本文所提出的分类器在相关的分类指标上具备良好的性能,能够为类条件概率密度的估计提供一种新的实现途径.

1 D-vine Copula贝叶斯分类器

1.1 贝叶斯决策理论

对于未知样本$\boldsymbol{x}=\{x_1,x_2,\cdots,x_n\}$,$n$是样本的特征个数,由贝叶斯公式可知

$$P(C_k|\boldsymbol{x})=\frac{p(\boldsymbol{x}|C_k)P(C_k)}{p(\boldsymbol{x})}. \tag{1}$$

其中:C_k为类别标签;$k=1,2,\cdots,K$,K是类别数量;$P(C_k|\boldsymbol{x})$为该类别对应的后验概率,根据最小化分类错误率准则,将样本$\boldsymbol{x}$归为$P(C_k|\boldsymbol{x})$最大的一类;$P(C_k)$是样本所属类别的先验概率,通常可由训练样本中估算出来[20];$p(\boldsymbol{x}|C_k)$是相应的类条件概率密度函数. 类条件概率密度的估计是本文研究的重点,将结合D-vine Copula模型来构建贝叶斯分类器中的类条件概率密度函数.

1.2 Copula函数

由Sklar定理可知,一个n维的联合概率分布可以分解为n个一元的边缘分布函数与一个n维Copula函数的乘积. 假设F是n维随机变量$\boldsymbol{x}=\{x_1,x_2,\cdots,x_n\}$的联合概率分布,边缘分布函数为$F_1(x_1),F_2(x_2),\cdots,F_n(x_n)$,则存在一个Copula函数$C$使得

$$F(x_1,x_2,\cdots,x_n)=C(F_1(x_1),F_2(x_2),\cdots,F_n(x_n)). \tag{2}$$

如果各变量的边缘分布函数是连续的,则Copula函数是唯一的. 当Copula函数可微时,可以得到随机变量$\boldsymbol{x}$的联合概率分布

$$f(x_1,x_2,\cdots,x_n)=f(x_1)\times f(x_2)\times\cdots\times f(x_n)\times c(F_1(x_1),F_2(x_2),\cdots,F_n(x_n);\theta), \tag{3}$$

$f(x_i)$是随机变量$\boldsymbol{x}$的边缘概率密度函数. Copula函数的密度函数定义为

$$C(F_1,F_2,\cdots,F_n;\theta)=\frac{\partial^n C}{\partial F_1\partial F_2\cdots\partial F_n}. \tag{4}$$

由式(3)和(4)可知,当Copula函数的类型和参数确定后,结合边缘概率密度函数,联合概率分布便可以确定.

1.3 二元Copula函数

对于随机变量$\boldsymbol{x}=\{x_1,x_2,\cdots,x_n\}$,假设其联合概率分布函数为$f(x_1,x_2,\cdots,x_n)$,则该联合概率

分布函数可以被分解为

$$f(x_1,x_2,\cdots,x_n)=$$
$$f(x_n)\times\prod_{t=1}^{n-1}f(x_t|x_{t+1},x_{t+2},\cdots,x_n). \tag{5}$$

对式(5)中的条件概率密度函数再次分解[21],可得

$$f(x_i|\boldsymbol{v})=$$
$$c_{x,v_j|\boldsymbol{v}_{-j}}(F(x_i|\boldsymbol{v}_{-j}),F(v_j|\boldsymbol{v}_{-j}))\times f(x_i|\boldsymbol{v}_{-j}). \tag{6}$$

其中: x_i 是随机变量 $\boldsymbol{x}$ 中的任意一个成分; $\boldsymbol{v}$ 表示随机变量 $\boldsymbol{x}$ 去掉 x_i 后的 $n-1$ 维向量; $\boldsymbol{v}_{-j}$ 表示 $\boldsymbol{v}$ 中去掉 v_j 后的向量, v_j 为 $\boldsymbol{v}$ 中任意一个成分; $c_{x,v_j|\boldsymbol{v}_{-j}}$ 为相应条件下的二元Copula条件概率密度函数.

1.4　D-vine Copula模型

Copula建模的本质是用样本来拟合式(4)的过程,并通过优化准则来估计相应Copula函数的参数. 对于二元样本而言,其优化过程较易实现. 然而,随着维数的增大,会出现"维数灾难",即一个 m 元Copula函数的待优化参数个数将远远大于 m. Vine Copula是解决上述问题的有效途径[22].

在构建多维变量的联合概率分布函数时,有多种二元Copula结构可以用来描述变量之间的相关性,其中C-vine Copula和D-vine Copula是最典型的两种[23]. 考虑到D-vine Copula结构具有更多的灵活性,本文采用D-vine Copula结构对随机变量 $\boldsymbol{x}=\{x_1,x_2,\cdots,x_n\}$ 的复杂相关性进行建模. 该模型由树 $T_j(j=1,2,\cdots,n-1)$ 构成,每棵树由节点和边组成,第1颗树的节点表示随机变量 $\boldsymbol{x}$ 的各个特征,连接节点的边表示各个特征之间的相互关联关系,即以相邻节点的边缘分布函数为自变量构成的二元Copula函数. 除第1颗树外,后续的每一棵树的节点都来自其前一棵树的边,总共有 $n(n-1)/2$ 个二元Copula函数,由此可知,D-vine Copula结构的联合概率分布函数可以分解为

$$f(x_1,x_2,\cdots,x_n)=$$
$$\prod_{k=1}^{n}f(x_k)\times$$
$$\Big\{\prod_{i=1}^{n-1}\prod_{j=1}^{n-i}c_{i,j+i|j+1}(F(x_j|x_{j+1},\cdots,x_{j+i-1}),$$
$$F(x_{j+i}|x_{j+1},\cdots,x_{j+i-1});\theta_{j,j+i|j+1:j+i-1})\Big\}. \tag{7}$$

其中: $F(x_j|x_{j+1},\cdots,x_{j+i-1})$ 和 $F(x_{j+i}|x_{j+1},\cdots,x_{j+i-1})$ 为条件分布函数, $c_{j,j+i|j+1:j+i-1}$ 为二元Copula条件概率密度函数, $\theta_{j,j+i|(j+1):(j+i-1)}$ 表示相应二元Copula函数的参数.

与传统Copula函数的分解式(3)相比,高维变量的相关性优化问题被转化为一系列二元Copula函数的参数估计问题,其求解过程避免了式(4)中的多次求导带来的复杂计算问题,条件分布函数的表达形式为

$$F(x|\boldsymbol{v})=\frac{\partial c_{x,v_j|\boldsymbol{v}_{-j}}(F(x|\boldsymbol{v}_{-j}),F(v_j|\boldsymbol{v}_{-j}))}{\partial F(v_j|\boldsymbol{v}_{-j})}. \tag{8}$$

当 $\boldsymbol{v}$ 是标量时

$$F(x|v)=\frac{\partial c_{x,v}(F(x),F(v))}{\partial F(v)}. \tag{9}$$

通过式(8)可以对(7)中的条件分布函数进行求解.

由式(7)可知,联合概率分布函数被分解成一系列二元Copula函数与各边缘概率密度函数的乘积形式. 本文采用核函数方法对边缘概率密度函数进行估计. 核函数方法是用于概率密度估计的一种非参数方法, $x_1,x_2,\cdots,x_n$ 为独立同分布的 n 个样本点,设其概率密度函数为 f,则测试样本 x 的概率密度估计值为

$$\hat{f}_n(x)=\frac{1}{n}\sum_{i=1}^{n}K_n(x-x_i)=\frac{1}{nh}\sum_{i=1}^{n}K\Big(\frac{x-x_i}{h}\Big). \tag{10}$$

其中: $K(\cdot)$ 为高斯核函数; $h>0$,称作带宽.

本文采用极大似然估计法对式(7)中二元Copula函数的参数进行估计,有

$$\hat{\theta^o}_{j,j+i|(j+1):(j+i-1)}=$$
$$\arg\max\sum_{k=1}^{M}\log[c^o_{j,j+i|(j+1):(j+i-1)}(F^k(x_j|x_{j+1},\cdots,$$
$$x_{j+i-1}),F^k(x_{j+i}|x_{j+1},\cdots,x_{j+i-1});$$
$$\theta^o_{j,j+i|(j+1):(j+i-1)})]. \tag{11}$$

其中: M 表示D-vine Copula模型中二元Copula函数的总数; F^k 表示第 k 个分布函数,由式(8)求出; c^o 表示选择的第 o 个二元Copula函数,由赤池信息准则(AIC)进行选择.

通过上述求取方式,可以得到贝叶斯决策中的类条件概率密度函数的形式,也就是待识别的类别状态下变量 $\boldsymbol{x}$ 的联合概率分布函数.

2　基于生物电信号的觉醒度状态识别

生物电信号中蕴含了与人体状态变化息息相关的特征信息. 从清醒状态到困倦再进入睡眠状态,人的觉醒度呈现逐渐下降的过程,随之脑电等生物电信号的特征也会发生相应的变化. 觉醒度反映了人的注意力集中程度和对外界的反应能力,尤其是对需要保持高度注意力的职业,如飞行员和驾驶员等,觉醒度的下降往往会引发事故.

生物电信号的采集,如脑电信号是通过在头皮上安放电极来获取的,对应于相应大脑功能区大量神经元细胞的综合电活动. 由上亿个神经元细胞构成的错综复杂的结构,使得不同位置记录下来的脑电等生物电信号的特征之间存在着较为复杂且不容忽视的相关性. 因此,本文选用短时睡眠过程中同步记录下来的多维电生理信号作为分析对象,采用D-vine Copula贝叶斯分类方法对清醒(Awake)和困倦(Drowsiness)两种不同的觉醒度状态进行识别,用来分析和验证本文方法的有效性.

2.1 数据采集

本文选用8名受试者的白天短时睡眠数据. 实验数据来源于日本佐贺大学先进控制实验室(数据采集遵照国际通用的道德声明,并获得受试者的知情同意书). 受试者处于安静的环境下,采集其午后30 min的睡眠数据. 按照多导睡眠图(PSG, Ploysomnograph)的检测方式记录受试者的脑电、眼电和肌电信号.

脑电信号分别采集左右半脑中部区域的C_3和C_4位置,以及枕部区域的O_1和O_2位置,以异侧的耳垂作为参考电位,采样频率为100 Hz;眼电信号的采集位置在眼睑附近,即LOC和ROC通道,均以左侧耳垂为参考电位,采样频率为100 Hz;肌电信号的采集位置在下颚部位,记为Chin-EMG,采样频率为200 Hz. 本文主要提取脑电信号和眼电信号的特征来对清醒和困倦状态进行分类识别.

2.2 特征提取

将短时睡眠过程中记录的数据划分为每5 s的连续数据段,并进行快速傅里叶变换提取脑电信号和眼电信号的频域特征,依据频域特征对受试者的觉醒度状态进行识别.

特征计算公式如表1所示. 提取的特征包括C_3和C_4通道θ波的能量与C_3和C_4通道脑电信号全频段能量比值的最大值x_1,O_1和O_2通道α波与O_1和O_2通道脑电信号全频段能量比值的最大值x_2,以及水平和垂直眼电活动的能量x_3和x_4.

表1 脑电信号和眼电信号中提取的特征参数

参数	计算公式
x_1	$\max\left\{\frac{E_\theta(C_3)}{E_T(C_3)},\frac{E_\theta(C_4)}{E_T(C_4)}\right\}$
x_2	$\max\left\{\frac{E_\alpha(O_1)}{E_T(O_1)},\frac{E_\alpha(O_2)}{E_T(O_2)}\right\}$
x_3	$E_{\rm LOC}(\rm LOC)$
x_4	$E_{\rm ROC}(\rm ROC)$

表1中:E表示频域中的相应频率段的能量和,其下标θ和α对应于脑电信号的不同频率段成分,分别是2～7 Hz和8～13 Hz;T表示脑电信号全频段成分0.5～25 Hz;LOC和ROC对应的是眼电信号,计算2～10 Hz的能量和.

2.3 模式识别

2.3.1 D-vine Copula模型

首先,对数据集进行归一化处理,使样本特征映射到[0, 1]之间,并求得两种状态下特征的Kendall秩相关系数. 其中清醒状态下的秩相关系数为

$$\tau_{c1}=\begin{bmatrix}1 & -0.05 & -0.05 & -0.04\\ -0.05 & 1 & 0.41 & 0.43\\ -0.05 & 0.41 & 1 & 0.81\\ -0.04 & 0.42 & 0.81 & 1\end{bmatrix},\tag{12}$$

秩相关系数表示了特征之间的相关性. 从式(12)中可以发现秩相关系数有正有负,说明特征之间存在一定的依赖关系,其依赖关系的强弱程度不同. 困倦状态下的秩相关系数也具有类似的特点.

然后,通过核函数方法求出每个特征的边缘概率密度值,进而根据根节点选择准则和AIC准则确定D-vine Copula模型结构. 由根节点选择准则计算得到每行Kendall秩相关系数之和,由每行的Kendall秩相关系数和可以获知清醒状态下第1棵树的节点从左往右依次为特征x_4、x_3、x_2、x_1,困倦状态下第1棵树的节点从左往右依次为特征x_3、x_4、x_2、x_1.

表2 基于AIC准则选取的二元Copula函数类型

	二元Copula函数	Copula类型
清醒状态	C_{43}	Clayton
	C_{32}	Frank
	C_{21}	Independence Copula
	$C_{42\|3}$	Survival BB8
	$C_{13\|2}$	Independence Copula
	$C_{14\|32}$	Independence Copula
困倦状态	C_{34}	Gumbel
	C_{42}	Frank
	C_{21}	Frank
	$C_{23\|4}$	Survival Clayton
	$C_{14\|2}$	Frank
	$C_{14\|32}$	Independence Copula

表2分别给出了清醒状态和困倦状态下二元Copula函数的选取结果. 对于两种不同状态,从常用的Copula函数[12]中基于AIC准则分别得到了6个二元Copula函数类型,并由此可以确定清醒状态和困倦状态下的D-vine Copula模型结构.

最后,采用极大似然估计对二元Copula函数的参数进行估计. 将二元Copula函数与边缘概率密度函数值相结合,构建类条件概率密度函数的形式.

2.3.2 相关性分析

在D-vine Copula模型中确定了特征变量顺序,两两特征之间的相关性分布情况如图1所示. 各特征在D-vine Copula结构中是第1颗树的各个节点,相应的二元Copula函数如表2中的前3行所示.

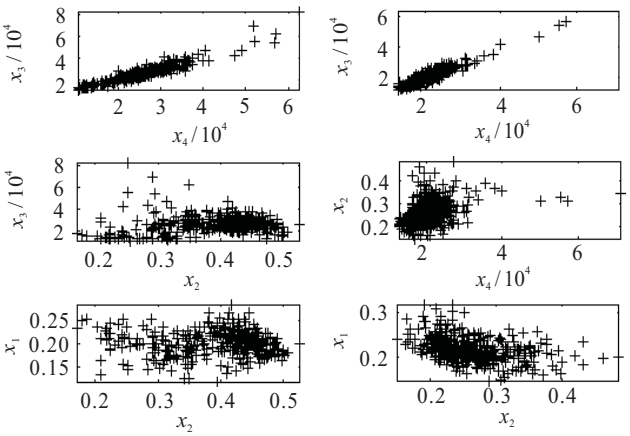

图 1　样本数据散点图

由图1可以看出,在清醒状态下,变量x_4与x_3之间有很强的下尾相关性(即一个变量减小的前提下,另一个变量也随着减小的概率增大,上尾相关性则相反),右上角分布呈分散趋势,上尾相关性不明显,因此,D-vine Copula模型选取了刻画数据的下尾相关性较好的Clayton Copula函数. 清醒状态下变量x_3与x_2并没有强烈的尾部相关性,除左上角部分有一些离群点外,分布总体呈左右对称趋势,因此,在AIC准则下选取了构建对称分布能力较好的Frank Copula函数. 变量x_2和x_1分布较为离散,并没有很强的相关性,因此,选取了独立的Copula函数. 困倦状态下,变量x_4和x_3的分布与清醒状态下较为相似,但D-vine Copula模型选取了具有较好刻画上、下尾相关性能力的Gumbel Copula函数来构建二者之间相关性. 由图1中不难推测出,当样本数量增加时,困倦状态下x_3与x_4的上尾相关性特征将更为明显,而清醒状态下x_3与x_4的分布在上尾部分将更为分散. 变量x_4与x_2之间,x_2与x_1之间的相关性并不是很强烈,但总体分布呈对称形状,因此,D-vine Copula模型选取了Frank Copula函数.

由此表明,D-vine Copula模型可以从众多不同的二元Copula函数中选取最为贴切的Copula函数来拟合特征之间存在的线性或非线性相关性.

2.4 结果比较

将基于D-vine Copula理论的贝叶斯分类算法与高斯判别分析、朴素贝叶斯分类器和带径向基核的支持向量机进行比较,以70%的数据作为训练集,30%作为测试集.

首先,通过ROC曲线分析和比较4种方法的分类性能. ROC曲线是反映真阳率和假阳率连续变化的综合指标,通过将连续变量设定出多个不同的临界值,从而计算出一系列的真阳率和假阳率,再以真阳率为纵坐标,假阳率为横坐标绘制成曲线,曲线越靠近左上角,其分类准确率越高. 图2给出了4种方法的ROC曲线,由于D-vine Copula贝叶斯方法的ROC曲线相比于另外3条ROC曲线处在最左上角的位置,具备较好的分类性能.

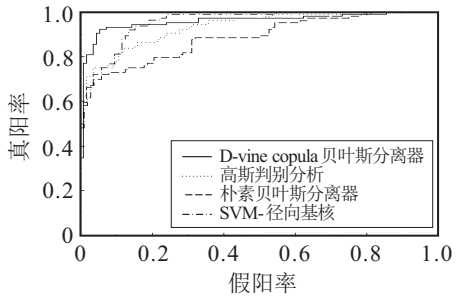

图 2　不同分类方法的ROC曲线比较

然后,通过计算灵敏度和特异度来分析和比较不同分类器的分类性能. 灵敏度指标和特异度指标分别对应于待识别的两种状态的分类准确率. 表3给出了4种方法的灵敏度和特异度. 从表3中可以观察到,D-vine Copula贝叶斯分类器的灵敏度为93.3%,特异度为91.8%,均高于其他3个分类器的灵敏度和特异度值.

表3　不同分类方法的灵敏度和特性度比较　%

分类方法	灵敏度	特异度
D-vine Copula贝叶斯分类器	93.3	91.8
高斯判别分析	83.8	84.7
朴素贝叶斯分类器	86.6	77.5
SVM-径向基核	89.5	86.5

3 结　论

本文结合D-vine Copula函数,将变量的联合概率分布分解成一系列二元Copula函数与边缘概率密度函数的乘积,通过对边缘概率密度函数与二元Copula函数分别进行优化估计,将特征之间的复杂相关性构建在类条件概率密度函数中. 与其他分类模型相比,D-vine Copula贝叶斯分类器在生物电信号分类问题中得到了较好的分类结果.

D-vine Copula模型的优势在于Copula函数作为连接特征的边缘分布函数,可以从众多不同的二元Copula函数中选取最合适的函数来拟合特征之间存在的线性或非线性相关性,且连接形式不受特征边缘分布的限制. 相对于传统贝叶斯分类算法,该方法在

变量的相关性构建上更具灵活性,能够为贝叶斯分类器中类条件概率密度的估计提供一种新的实现途径,从而提高贝叶斯分类器在处理特征之间具有复杂相关性时的分类性能.

参考文献(References)

[1] Shi C Z, Gao S, Liu M T, et al. Stroke detector and structure based models for character recognition: A comparative study[J]. IEEE Trans on Image Processing, 2015, 24(12): 4952-4964.

[2] Duchenne O, Joulin A, Ponce J. A graph-matching kernel for object categorization[C]. Int Conf on Computer Vision. Spain: IEEE, 2011: 1792-1799.

[3] O'Mara-Eves A, Thomas J, Mcnaught J, et al. Using text mining for study identification in systematic reviews: A systematic review of current approaches[J]. Systematic Reviews, 2015, 4(1): 4-5.

[4] Duda R O, Hart P E, Stork D G. Pattern classification[M]. Beijing: China Machine Press, 2004: 4-23.

[5] Sathe S. A novel Bayesian classifier using Copula functions[J]. Computer Science, 2006, 11(2): 23-45.

[6] Qian D, Wang B, Zhang T, et al. Classification algorithm based on Copula theory and Bayesian decision theory[J]. CAAI Trans on Intelligent Systems, 2016, 11(1): 78-83.

[7] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012: 48-53.
(Li H. Statistical learning method[M]. Beijing: Tsinghua University Press, 2012: 48-53.)

[8] Salvatierra I D L, Patton A J. Dynamic Copula models and high frequency data[J]. J of Empirical Finance, 2015, 30: 120-135.

[9] 王丽芳, 曾建潮, 洪毅. 利用Copula函数估计概率模型并采样的分布估计算法[J]. 控制与决策, 2011, 26(9): 1333-1337.
(Wang L F, Zeng J C, Hong Y. Estimation of distribution algorithm modeling and sampling by means of Copula[J]. Control and Decision, 2011, 26(9): 1333-1337.)

[10] 韩敏, 刘晓欣. 基于Copula熵的互信息估计方法[J]. 控制理论与应用, 2013, 30(7): 86-90.
(Han M, Liu X X. Mutual information estimation based on Copula entropy[J]. Control Theory & Application, 2013, 30(7): 86-90.)

[11] 许民利, 李展. 需求依赖于价格情境下基于Copula-CVaR的报童决策[J]. 控制与决策, 2014, 29(6): 1083-1090.
(Xu M L, Li Z. Newsvendor decision based on Copula-CVaR with price-dependent demand[J]. Control and Decision, 2014, 29(6): 1083-1090.)

[12] Nelsen R B. An introduction to Copulas[M]. New York: Spinger, 2006: 17-28, 141-146.

[13] Joe H. Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters[J]. Lecture Notes-Monograph Series, 1996, 28: 120-141.

[14] Aas K, Berg D. Models for construction of multivariate dependence — A comparison study[J]. European J of Finance, 2009, 15(7/8): 639-659.

[15] Czado C. Pair-Copula constructions of multivariate Copulas[M]. Berlin: Springer, 2010: 93-109.

[16] Diβmann J, Brechmann E C, Czado C, et al. Selecting and estimating regular vine Copulae and application to financial returns[J]. Computational Statistics & Data Analysis, 2013, 59(1): 52-69.

[17] Ahooyi T M, Soroush M, Arbogast J E, et al. Maximum-likelihood maximum-entropy constrained probability density function estimation for prediction of rare events[J]. Aiche J, 2014, 60(3): 1013-1026.

[18] Becker R. Generation of time-coupled wind power infeed scenarios using pair-coupled construction[J]. IEEE Trans on Sustainable Energy, 2017, 9(3): 1298-1306.

[19] Sun M, Konstantelos I, Strbac G. C-vine Copula mixture model for clustering of residential electrical load pattern data[J]. IEEE Trans on Power Systems, 2017, 32(3): 2382-2393.

[20] Bishop C M. Pattern recognition and machine learning[M]. New York: Springer, 2006: 122-127.

[21] Smith M, Min A, Almeida C, et al. Modeling longitudinal data using a Pair-Copula decomposition of serial dependence[J]. J of the American Statistical Association, 2010, 105(492): 1467-1479.

[22] Mazo G, Girard S, Forbes F. A class of multivariate Copulas based on products of bivariate Copulas[J]. J of Multivariate Analysis, 2015, 140: 363-376.

[23] Brechmann E C, Schepsmeier U, Grün B, et al. Modeling dependence with C- and D-vine Copulas: The R package CDVine[J]. J of Statistical Software, 2013, 52(3): 1-27.

作者简介

王蓓 (1976−), 女, 副研究员, 博士, 从事智能信号处理、模式识别等研究, E-mail: beiwang@ecust.edu.cn;

孙玉东 (1992−), 男, 硕士生, 从事模式识别与智能系统的研究, E-mail: y30160657@mail.ecust.edu.cn;

金晶 (1981−), 男, 教授, 博士生导师, 从事智能信号处理、模式识别等研究, E-mail: jinjing@ecust.edu.cn;

张涛 (1969−), 男, 教授, 博士生导师, 从事非线性控制理论与应用、机器人智能控制等研究, E-mail: taozhang@tsinghua.edu.cn;

王行愚 (1944−), 男, 教授, 博士生导师, 从事智能控制、模式识别和控制理论等研究, E-mail: xywang@ecust.edu.cn.

(责任编辑: 李君玲)