异步动态贝叶斯网络分类器研究

王双成[1), 2)] 张 立[1)] 郑 飞[1)]

[1)](上海立信会计金融学院信息管理学院 上海 201620)

[2)](上海立信会计金融学院数据科学交叉研究院 上海 201209)

摘 要 时间序列数据普遍存在，对其进行分类预测有着广泛的需求. 虽然有一些时间序列数据分类方面的研究，但主要是面向时序同步分类（类与属性同步变化），还需要进行更有实际意义的异步分类（类与属性不同步变化）方面的探索. 本文结合时间序列的离散化、变量的时序转换、变量的错位变换、依据变量顺序和打分搜索的分类器结构学习和类变量的丢失数据修复等，建立异步动态贝叶斯网络分类器，这种分类器能够有效利用多变量时间序列数据中所蕴含的时滞、非时滞和混合分类信息，以及属性为类提供的传递依赖信息、直接导出依赖信息和间接导出依赖信息进行分类计算，来提高分类器的可靠性. 分别使用 UCI、金融和宏观经济时间序列数据进行实验，结果显示所建立的异步动态贝叶斯网络分类器具有良好的分类准确性.

关键词 时间序列；动态贝叶斯网络；分类器；同步分类；异步分类

中图法分类号 TP181 **DOI 号** 10.11897/SP.J.1016.2020.01737

Asynchronous dynamic Bayesian network classifiers

WANG Shuang-Cheng[1), 2)] ZHANG Li[1)] ZHENG Fei[1)]

[1)](*School of information management, Shanghai Lixin University of Accounting and Finance, Shanghai* 201620)

[2)](*Institute of data science and interdisciplinary studies, Shanghai Lixin University of Accounting and Finance, Shanghai* 201209)

Abstract Time series is one of the main forms of data in the real world. It widely exists in various large databases, such as macroeconomic, finance, industry, management, internet and so on. A large number of time series record all kinds of important information of the system at different time points (or time slices). There is abundant and valuable knowledge about causality, classification rules and regression functions in these information. They are often the important basis for diagnosing the operation of the system and formulating corresponding strategies. Classification is a computer simulation of human concept learning (also known as concept learning). It is one of the core technologies in machine learning and data mining. Many famous classifiers have been developed, such as decision tree, neural network, support vector machine, nearest neighbor classifier and so on. They have their own characteristics and are widely used in many fields, but these classifiers are mainly for non time series data. Bayesian network is a probabilistic graphical model to describe the dependence and restriction relationship between random variables. It has the characteristics of multi-function, effectiveness and openness, and is a powerful tool to deal with uncertainty. Classical Bayesian networks are mainly used for causal knowledge representation and uncertainty reasoning. The Bayesian network for classification is generally called Bayesian network classifier. There are many researches on Bayesian network classifiers, but these classifiers are all for non time series data and can not be directly used in the classification calculation of time series data. Dynamic Bayesian network is a timing extension of Bayesian network. It is mainly used to solve the uncertainty of

收稿日期：2019-09-16；在线发布日期：2020-02-07. 本课题得到国家社会科学基金（18BTJ020）资助. 王双成（通信作者），博士，教授，中国计算机学会（CCF）高级会员，主要研究领域为人工智能、机器学习、数据挖掘与应用. E-mail: wangsc@lixin.edu.cn. 张 立，博士，讲师，主要研究领域为机器学习与信息系统. 郑 飞，博士，副教授，主要研究领域为信息安全和数据挖掘.

time series. The dynamic Bayesian network for time series data classification is generally called dynamic Bayesian network classifier. The research on dynamic Bayesian network classifier is relatively less, and it is mainly synchronous classification (synchronous change of class and attributes). It's also needed to explore more practical asynchronous classification (asynchronous change of class and attributes). In this paper, we combine the discretization of time series, the timing conversion of variables, the dislocation transformation of variables, classifier structure learning based on variable order and search & scoring method, and missing data restoration for class variable to build an asynchronous dynamic Bayesian network classifier. This classifier can effectively utilize the time delay, no time delay and mixed classification information contained in multivariate time series data, as well as the transitive dependency, direct induction dependency and indirect induction dependency information to do classification calculation to improve the reliability. We use UCI, financial (stock, exchange rate, futures and fund) and macroeconomic time series data respectively, and carry out experiments and analysis from four aspects: the comparison of classification accuracy, the impact of classification order on classification accuracy, the impact of time delay information on classification accuracy and the impact of discretization methods on classification accuracy. The experimental results reveal the constraints of various factors on the classification accuracy of the asynchronous dynamic Bayesian network classifier, and verify its effectiveness, practicability and reliability for the classification calculation of multivariable time series of different orders.

Keywords time series; dynamic Bayesian network; classifier; synchronous classification; asynchronous classification

1 引 言

时间序列是现实世界中数据的主要表现形式之一，它广泛存在于宏观经济、金融、工业、管理和互联网等各种大型的数据库中. 大量时间序列真实地记录了系统在不同时间点（或时间片）的各种重要信息，其中蕴含着因果关系、分类规则和回归函数等方面丰富而有价值的知识，这些知识往往是诊断系统运行状况，以及制定相应策略的重要依据. 分类是人类概念学习的计算机模拟（也称为概念学习），是机器学习和数据挖掘等领域的核心技术之一. 现已发展了许多著名的分类器，如决策树、神经网络、支持向量机和最近邻分类器等，它们各有特点，在许多领域都得到了广泛的应用，但这些分类器主要针对的是非时间序列数据. 近些年，循环神经网络（Recurrent Neural Network）被广泛用于多变量时间序列数据的分类计算，它以时间序列数据为输入，在序列的演进方向进行递归，一个时间点（或时间片）的分类结果作为下一个时间点（或时间片）分类的影响因素，可使时滞信息得到有效的利用，后来又演化出长短期记忆网络（Long Short-Term Memory networks，LSTM）和门循环单元网络（Gated Recurrent Unit networks, GRU，LSTM 的变体）等，这些多变量时间序列分类器均收到了良好的效果，但在迭代分类计算过程中，历史（或时滞）信息可能产生叠加，对历史（或时滞）信息的使用没有选择性，而且无法进行异步分类计算.

贝叶斯网络（Bayesian network）[1]是描述随机变量（简称为变量）之间依赖与制约关系的概率图模型，它具有多功能性、有效性和开放性等特征，是处理不确定性问题的有力工具. 经典的贝叶斯网络主要用于因果知识表示和不确定性推理，用于分类的贝叶斯网络一般称为贝叶斯网络分类器. 对贝叶斯网络分类器已有许多研究，如 Flach（2004）[2]、Stephens（2017）[3]和 Xu（2019）[4]等对朴素贝叶斯分类器的研究，Zheng（2012）[5]、Flores（2014）[6]和 Cai（2018）[7] 等对半朴素贝叶斯分类器的研究，Friedman（1997）[8]、Jing（2008）[9]和 Namrata（2019）[10]等对树增强朴素贝叶斯分类器和变体的研究，王双成（2013, 2016）[11,12]和 Yang（2019）[13]等对贝叶斯网络分类器的研究. 这些分类器均是针对非时间序列数据，不能直接用于时间序列数据的分类计算.

1998 年 Friedman[14]在隐马尔科夫模型（Hidden Markov model）和卡尔曼滤波模型（Kalman filtering model）的基础上给出受平稳性与马尔科夫性两个假

设约束的动态贝叶斯网络（dynamic Bayesian network）定义和基于打分-搜索的学习方法，2002 年 Murphy[15]对动态贝叶斯网络进行了系统的理论分析和应用展望，从此揭开了动态贝叶斯网络研究进程. 早期的动态贝叶斯网络主要关注的是隐马尔科夫模型、卡尔曼滤波模型和两个模型的变体，以及它们在语音识别、视频分析和信息滤波等方面的应用研究. 近些年更关注于将动态贝叶斯网络用于动态识别、预警、诊断和评估等方面的研究（隐马尔科夫模型应用的扩展），如 Yang（2010）[16]使用动态贝叶斯网络识别驾驶员的疲劳程度，Dabrowski（2016）[17]运用动态贝叶斯网络对系统性银行危机进行预警，Tanjin（2019）[18]依据动态贝叶斯网络进行故障诊断与路径分析，Heng（2019）[19]基于动态贝叶斯网络做正交异性钢桥面疲劳可靠性的系统评估等. 建立这些动态贝叶斯网络主要依靠专家知识，它们更适合于动态分析和推理计算（其结构中的有向边更突出因果含义，而不是强调信息的传递渠道和方式），直接将其用于分类计算效果往往并不理想. 在动态贝叶斯网络分类器的研究方面，Palacios-Alonso（2010）[20]使用遗传算法来优化动态朴素贝叶斯分类器，并将其用于手势识别；Alkhateeb（2011）[21]将具有复杂结构的动态贝叶斯网络分类器和隐马尔科夫模型用于手写阿拉伯语的单词识别；王双成（2011）[22]建立基于高斯函数估计属性密度的动态贝叶斯网络分类器，并用于经济周期波动转折点预测；Kafai （2012）[23]基于专业知识构建动态贝叶斯网络分类器结构，并将其用于视频中的车辆分类；Premebida （2017）[24] 采用动态贝叶斯混合模型（一种动态贝叶斯网络的变体）对移动机器人进行语义位置分类；王双成（2017）[25,26]分别基于高斯函数和高斯核函数估计属性密度建立动态朴素和动态完全贝叶斯网络分类器，并将它们用于宏观经济指标的增减性预测；Rishu （2019）[27]利用动态贝叶斯网络对基于智能手机的驾驶员行为进行上下文感知分类. 这些动态贝叶斯网络分类器均取得了较好的分类效果，但它们都是同步分类器，而且在确定分类器结构方面，主要依据专家知识、采用平凡结构（朴素或完全结构）和使用整体打分-搜索方法，因此，不利于实现训练与泛化之间的均衡，以及更有效的分类信息提取等，但这些分类器方面的成果为本文的异步动态贝叶斯网络分类器研究奠定了基础. 关于异步分类计算，王双成（2017）[28]给出了时间序列数据的异步回归计算模型，为异步动态贝叶斯网络分类器研究提供了可借鉴的方法. 也有一些其它的关于异步时间序列数据方面的研究，但主要集中在动态贝叶斯网络学习和推理，一般通过隐藏变量对不同步的时间序列数据（或变量）进行整合，并基于 EM（Expectation-Maximization）算法或 MCMC（Markov Chain Monte Carlo）方法进行迭代学习和推理计算，以两种特殊的动态贝叶斯网络（隐马尔科夫模型和卡尔曼滤波模型）以及它们的一些变体最具代表性，这些研究所侧重的是变量之间的近似推理计算，而不是属性对类的有效信息传递，因此并不适合于异步多变量时间序列分类预测，但其中一些思想提供了有价值的参考.

本文的主要贡献如下：

（1）在变量时序转换（构建转换数据集[28]）和变量错位变换（建立属性和类之间的错位对应关系）的基础上，对于给定的变量顺序（类排在首位），我们基于贪婪打分-搜索发现类的非时滞子结点（也实现了非时滞属性选择），再通过类和非时滞属性的时滞父结点学习（同样实现了类的时滞变量和非时滞属性的时滞变量选择），得到类的近似马尔科夫毯（Markov blanket），在理论上，马尔科夫毯中的属性子集是最优属性子集，在概率模式存在完全图时还是最小属性子集.

（2）我们将变量的时序转换和错位变换、类的马尔科夫毯学习、由错位变换而导致类丢失数据的修复和动态分类计算等相结合，给出了具有离散属性的异步动态贝叶斯网络分类器（Asynchronous Dynamic Bayesian Network Classifiers，ADBNC）. 时序转换实现了时滞与非时滞信息的统一，错位变换为异步分类计算奠定了基础；ADBNC 的马尔科夫毯结构，使三种分类信息（传递依赖信息、直接导出依赖信息和间接导出依赖信息）均得到充分的利用；类的缺失数据修复则可避免由错位变换而导致的分类信息丢失.

（3）分别使用 UCI、金融（股票、汇率、期货和基金）和宏观经济时间序列数据，从分类准确性比较、分类阶数对分类准确性的影响、时滞信息对分类准确性的影响和离散化方法对分类准确性的影响四个方面进行实验与分析，以揭示各种因素对 ADBNC 分类准确性的制约，并验证 ADBNC 对多变量时间序列不同阶分类计算的有效性、实用性和可靠性.

文章分为四个部分，第一部分是对贝叶斯网络和动态贝叶斯网络（包括分类器）的发展进行回顾与分析；第二部分给出时间序列数据的预处理、具有离散属性的 ADBNC 学习与分类方法和算法；第

三部分是使用 UCI、金融和宏观经济时间序列数据进行的实验与分析；第四部分是结论和进一步的工作. 再有，文中将概率模式中的变量和对应的图形模式中的结点有时不加区分.

2 ADBNC

分别用 $X_1[t], X_2[t], ..., X_n[t], C[t]$ 表示时间序列属性变量（简称为属性）和类变量（简称为类），其中 t 取离散的时间点且 $1 \leqslant t \leqslant T$，$x_1[t], x_2[t], ..., x_n[t], c[t]$ 是它们的具体取值，$D[n,T]=\{(x_1[t], x_2[t], ..., x_n[t], c[t]) | 1 \leqslant t \leqslant T\}$ 是具有 T 个记录的时间序列分类数据集，$D[n,T]$ 中的记录之间不满足独立同分布的假设，而是具有时序依赖. 时间序列数据的一般分类预测可以描述为：首先基于给定的时间序列数据集 $D[n,T]$ 建立分类器，然后使用分类器对 $C[T+\varphi]$ 进行分类预测，其中 $\varphi \geqslant 0$. 我们将 φ 称为分类器的阶数，$\varphi=0$ 为同步分类，$\varphi>0$ 为异步分类. 下面从时间序列预处理、ADBNC 的结构学习、由错位变换而形成的类丢失数据的修复、ADBNC 的分类计算和 ADBNC 中的属性为类提供的信息分析五个方面研究 ADBNC.

2.1 时间序列预处理

时间序列预处理是建立 ADBNC 之前的数据准备，包括 $D[n, T]$中时间序列的规范化、时间序列的离散化和数据集的错位变换三个部分，其中时间序列的规范化已有许多成熟的方法，可根据需要进行选择.

2.1.1 时间序列的离散化

在时间序列数据的分类问题中，如果作为类的时间序列是由连续值构成，则必须根据需求进行离散化. 属性时间序列可以离散化，也可以不离散化，本文研究将作为属性和类的时间序列均离散化的 ADBNC. 属性和类可以采用一致或不一致的离散化方法，离散化方法可大致分成时序离散化（与时间有关）和非时序离散化（与时间无关）两种，具体情况如图 1 所示.

2.1.2 时间序列数据集的错位变换

对给定的分类器阶数 $\varphi(\varphi>0)$，$D[n,T]$ 的错位变换（或变量的错位变换）是时间序列 $\{x_1[t], x_2[t], ..., x_n[t]\}$ 和 $\{c[t]\}$ 错 φ 位重新建立对应关系，形成新的时间序列数据集的过程. $D[n,T]$ 经过错位变换后会形成一些丢失数据（没有观测到的数据），用 $D_{(Full)}[n,T]$ 表示将 $D[n,T]$ 错位变换后具有类丢失数据的时间序列数据集 $\{(x_1[t], x_2[t], ..., x_n[t], c[t+\phi]) | 1 \leqslant t \leqslant T\}$，$D_{(Part)}[n,T]$ 表示在 $D_{(Full)}[n,T]$ 中删除丢失数据所在行后得到的时间序列数据集 $\{(x_1[t], x_2[t], ..., x_n[t], c[t+\varphi]) | 1 \leqslant t \leqslant T-\varphi\}$.

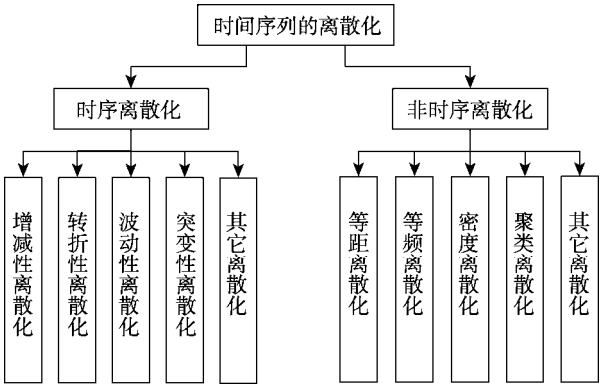

图 1 时间序列的离散化

2.2 ADBNC的结构学习

根据贝叶斯网络理论[1]，$C[t+\varphi]$ 的马尔科夫毯是最优属性子集，也就是给定马尔科夫毯中的变量时 $C[t+\varphi]$ 与其它变量条件独立，因此，通过建立 $C[t+\varphi]$ 的马尔科夫毯，能够显著减少不相关和冗余的属性. $C[t+\varphi]$ 的马尔科夫毯中包括三种结点，它们是 $C[t+\varphi]$ 的父结点、子结点和子结点的父结点. 我们将 ADBNC 的结构学习划分为三个部分（或阶段），分别是 $C[t+\varphi]$ 在 $\{X_1[t], X_2[t], ..., X_n[t]\}$ 中的子结点（称为非时滞子结点）学习、$C[t+\varphi]$ 在 $\{C[1+\varphi], C[2+\varphi], ..., C[t+\varphi-1]\}$ 中的父结点（称为时滞父结点）学习和 $C[t+\varphi]$ 的子结点的父结点（称为非时滞子结点的时滞父结点）学习. 我们结合变量顺序和贪婪打分-搜索方法进行 ADBNC 的结构学习（或类的马尔科夫毯学习）. 主要的贝叶斯网络结构打分标准有 MDL（Minimum Description Length）、BD（Bayesian Dirichlet）和 K2，鉴于 BD 和 K2 会使学习得到的贝叶斯网络结构倾向于复杂化，不适合于分类器的结构学习，而 MDL 能够实现拟合数据与网络复杂程度之间的均衡，因此更适合于分类器的结构学习（实现训练与泛化之间的均衡）. 我们结合 MDL 和贪婪搜索方法来学习 ADBNC 的结构.

2.2.1 $C[t+\varphi]$ 的非时滞子结点学习

发现 $C[t+\varphi]$ 的非时滞子结点就是确定 $C[t+\varphi]$ 是否为 $X_i[t]$ 父结点的过程. 以 $C[t+\varphi], X_1[t], X_2[t], ..., X_n[t]$ 为变量的顺序，采用贪婪打分-搜索方法，通过发现 $X_i[t]$ 的父结点集来确定 $C[t+\varphi]$ 在 $X_1[t], X_2[t], ..., X_n[t]$ 中的子结点集. 使用 PNS 表示 $X_i[t]$ 的父结点集，CNS 表示 $C[t+\varphi]$ 的子结点集，$CPNS$ 表示 $X_i[t]$ 的候选父结点集，$MDL(X_i[t] | SET, D_{(Part)}[n,T])$ 表示基于数据集

$D_{(Part)}[n,T]$，$X_i[t]$具有父结点集 SET 的 MDL 打分，初始化 $CNS=\Phi$，$C[t+\varphi]$的非时滞子结点学习如算法 1 所示.

算法 1.　$C[t+\varphi]$的非时滞子结点学习

输入: 时间序列数据集 $D_{(Part)}[n,T]$

输出: $C[t+\varphi]$的非时滞子结点集 CNS

FOR i=1 TO n

　$PNS=\Phi$, $CPNS=\{C[t+\phi],X_1[t],...,X_{i-1}[t]\}$

　FOR h=1 TO Δ // 限定 $X_i[t]$父结点的数量为 Δ

　　发现具有最小 MDL 打分的结点 $X_{h_0}[t]$

　　IF $MDL(X_i[t]\,|\,PNS\cup\{X_{h_0}[t]\},D_{(Part)}[n,T])$ $<MDL(X_i[t]\,|\,PNS,D_{(Part)}[n,T])$ THEN

　　$PNS=PNS+\{X_{h_0}[t]\}$

　　$CPNS=CPNS-\{X_{h_0}[t]\}$

　　ELSE

　　Exit FOR

　　END IF

　END FOR

　IF $C[t+\varphi]\in PNS$ THEN

　$CNS=CNS+X_i[t]$

　END IF

END FOR

在 $C[t+\varphi]$的子结点学习算法中，主要的运算是 MDL 打分，需要进行不超过 $\frac{\Delta}{2}(n^2-\Delta n)$ 次的 MDL 打分，Δ 是一个与 n 无关的量，因此，相对于 MDL 打分运算，$C[t+\varphi]$的非时滞子结点学习算法的时间复杂度是 $O(n^2)$. 使用 $\{X_{w_1}[t],X_{w_2}[t],...,X_{w_m}[t]\}$表示 $C[t+\varphi]$在 $\{X_1[t],X_2[t],...,X_n[t]\}$中的非时滞子结点集，其中 $1\leqslant m\leqslant n$.

2.2.2　$X_{w_k}[t]$和 $C[t+\varphi]$的时滞父结点学习

对给定的时滞阈值 Γ，发现 $X_{w_k}[t]$的时滞父结点就是确定 $X_{w_k}[t]$在 $\{X_{w_k}[t-1],X_{w_k}[t-2],...,X_{w_k}[t-\Gamma]\}$中父结点集的过程. 我们仍然使用 PNS 表示 $X_{w_k}[t]$的时滞父结点集，$CPNS$ 表示 $X_{w_k}[t]$的候选时滞父结点集，$MDL(X_{w_k}[t]\,|\,PNS,D_{(Part)}[n,T])$ 表示 $X_{w_k}[t]$具有时滞父结点集 PNS 的 MDL 打分，$\|PNS\|$表示集合 PNS 中元素的个数，$\|PNS\|\leqslant\Gamma$，初始化转换数据集[28] $TSET=\{x_{w_k}[t-\Gamma],...,x_{w_k}[t-2],x_{w_k}[t-1],\ x_{w_k}[t]\}$，其中 $\Gamma+1\leqslant t\leqslant T-\varphi$，$PNS=\{C[t+\varphi]\}$，$CPNS=\{X_{w_k}[t-\Gamma],...,X_{w_k}[t-2],X_{w_k}[t-1]\}$. $X_{w_k}[t]$的时滞父结点学习如算法 2 所示.

算法 2.　$X_{w_k}[t]$的时滞父结点学习

输入: 时间序列数据集 $\{x_{w_k}[t]\}\subseteq D_{(Part)}[n,T]$

输出: $X_{w_k}[t]$的时滞父结点集

采用文献[28]的方法建立转换数据集

FOR h=1 TO Δ

　发现具有最小 MDL 打分的结点 $X_{w_k}[t-v_h^{w_k}]$

　IF $MDL(X_{w_k}[t]\,|\,PNS\cup\{X_{w_k}[t-v_h^{w_k}]\},D_{(Part)}[n,T])$ $<MDL(X_{w_k}[t]\,|\,PNS,D_{(Part)}[n,T])$ THEN

　$PNS=PNS+\{X_{w_k}[t-v_h^{w_k}]\}$

　$CPNS=CPNS-\{X_{w_k}[t-v_h^{w_k}]\}$

　ELSE

　Exit FOR

　END IF

END FOR

在 $X_i[t]$的时滞父结点学习算法中，主要的运算也是 MDL 打分，需要不超过 $\frac{\Delta}{2}(2\Gamma-\Delta+1)$ 次的 MDL 打分，其中 Δ 是与 Γ 无关的量，因此，相对于 MDL 打分运算，$X_{w_k}[t]$的时滞父结点学习算法的时间复杂度是 $O(\Gamma)$，所有变量（包括属性和类）的时滞父结点学习算法的时间复杂度是 $O(n\Gamma)$. 我们可以采用类似的方法发现 $C[t+\varphi]$的时滞父结点集. 分别使用 $\{X_{w_k}[t-v_1^{w_k}],X_{w_k}[t-v_2^{w_k}],...,X_{w_k}[t-v_{q_{w_k}}^{w_k}]\}$ 和 $\{C[t+\phi-v_1^c],C[t+\phi-v_2^c],...,C[t+\phi-v_{q_c}^c]\}$ 表示 $X_{w_k}[t]$和 $C[t+\varphi]$的时滞父结点集，可得到 ADBNC 的局部结构，如图 2 所示.

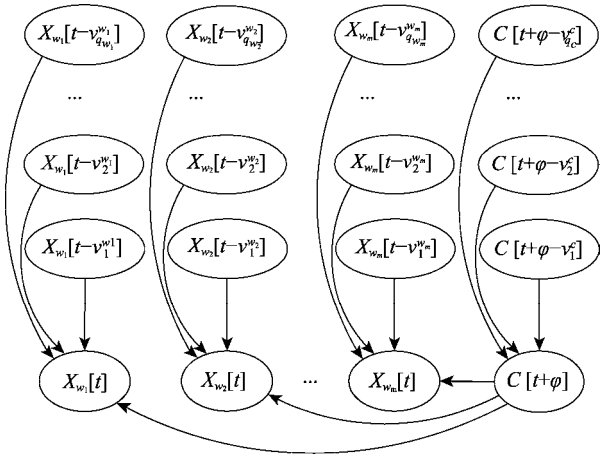

图 2　ADBNC 的时间点（或时间片）局部结构

在图 2 中，$X_{w_1}[t],X_{w_2}[t],...,X_{w_m}[t]$为 $C[t+\varphi]$提供非时滞分类信息，$\{C[t+\varphi-v_1^c],C[t+\varphi-v_2^c],...,C[t+\varphi-v_{q_c}^c]\}$ 为 $C[t+\varphi]$ 提供时滞分类信息，$\{X_{w_k}[t-v_1^{w_k}],X_{w_k}[t-v_2^{w_k}],...,X_{w_k}[t-v_{q_{w_k}}^{w_k}]\}$ 为 $C[t+\varphi]$提供混合分类信息，其中 $1\leqslant k\leqslant m$.

依据概率公式和图 2 中的条件独立性关系，可以得到：

$$\begin{aligned}&p(c[t+\varphi]\mid c[1+\varphi],...,c[t+\varphi-1],x_1[1],...,x_1[t],...,x_n[1],...,x_n[t])\\&=\frac{p(c[t+\varphi],c[1+\varphi],...,c[t+\varphi-1],x_1[1],...,x_1[t],...,x_n[1],...,x_n[t]))}{p(c[1+\varphi],...,c[t+\varphi-1],x_1[1],...,x_1[t],...,x_n[1],...,x_n[t])}\\&=\alpha p(c[t+\varphi],c[1+\varphi],...,c[t+\varphi-1],x_1[1],...,x_1[t],...,x_n[1],...,x_n[t])\\&=\alpha p(c[t+\varphi]\mid\pi(c[t+\varphi]))\prod_{k=1}^{m}p(x_{w_k}[t]\mid\pi(x_{w_k}[t]),c[t+\varphi])\end{aligned}\tag{1}$$

其中 $\pi(c[t+\varphi])$ 是 $C[t+\varphi]$ 的时滞父结点集 $\Pi(C[t+\varphi])$ 的配置，$\pi(x_{w_k}[t])$ 是 $X_{w_k}[t]$ 的时滞父结点集 $\Pi(X_{w_k}[t])$ 的配置.

ADBNC 可以表示为

$$\underset{c[t+\varphi](\pi(c[t+\varphi]),x_{w_1}[t],\pi(x_{w_1}[t]),...,x_{w_m}[t],\pi(x_{w_m}[t]))}{\arg\max}\left\{p(c[t+\varphi]\mid\pi(c[t+\varphi]))\prod_{k=1}^{m}p(x_{w_k}[t]\mid\pi(x_{w_k}[t]),c[t+\varphi])\right\}\tag{2}$$

2.3 类的丢失数据修复

在完整错位变换数据集中，$C[T+1],C[T+2],...,C[T+\varphi-1]$ 的位置可以看作是丢失数据（这些变量的值未知）. 当然我们可以使用 $D_{(Part)}[n,T]$ 中的数据进行分类器学习，但数据集 $D_{(Part)}[n,T]$ 与数据集 $D_{(Full)}[n,T]$ 之间差异部分的数据中所蕴含的信息将得不到利用，从而造成信息丢失，φ 越大，丢失的信息越多. 本文采用 Gibbs 抽样的方法来修复丢失的类数据，丢失数据的修复过程是一个迭代，按照 $C[T+1],C[T+2],...,C[T+\varphi-1]$ 的顺序依次修复每一个丢失数据，修复完所有的丢失数据实现一次迭代，当满足终止条件（可以采用相邻两次迭代一致性判断或给定迭代次数作为终止条件）时结束迭代.

假设 $C[t]$的值域是 $\{c^1,c^2,...,c^H\}$，用 $\hat{c}[t+i]$表示 $c[t+i]$ 的修正值，随机初始化 $c[T+1],c[T+2],...,c[T+\varphi-1]$，然后进行迭代，$c[T+i]$（$1\leqslant i\leqslant\phi-1$）按照下面的方法进行修正.

对产生的随机数 λ，变量 $C[T+i]$的修正值为

$$\hat{c}[t+i]=\begin{cases}c^1, & 0<\lambda\leqslant\psi(1)\\ \cdots & \\ c^h, & \sum_{j=1}^{h-1}\psi(j)<\lambda\leqslant\sum_{j=1}^{h}\psi(j)\\ \cdots & \\ c^H, & \lambda>\sum_{j=1}^{H-1}\psi(j)\end{cases}\tag{3}$$

其中 $\psi(h)$ 是满条件分布. 我们仍然使用 $D_{(Full)}[n,T]$ 表示类丢失数据修复后的时间序列数据集.

2.4 ADBNC的分类计算

假设基于 $D_{(Part)}[n,T]$ 已经建立了 ADBNC 的结构，由错位变换而形成的丢失数据也得到修复，ADBNC 的分类计算如算法 3 所示.

算法 3. ADBNC 的分类计算

输入: 时间序列数据集 $D_{(Full)}[n,T]$

输出: $c[T+\varphi]$

估计概率 $p(c[T+\varphi]\mid\pi(c[T+\varphi]),D_{(Full)}[n,T])$

FOR k=1 TO m

估计概率 $p(x_{w_k}[T]\mid\pi(x_{w_k}[T]),c[T+\varphi],D_{(Full)}[n,T])$

END FOR

合成计算得到 $c[T+\varphi]$

在 ADBNC 的分类计算中，运算的主要部分是概率估计，共需要 $m+1$ 次的概率估计，其中 $m\leqslant n$ 因此，关于概率估计，算法的时间复杂度是 $O(m)$.

2.5 属性为类提供的信息分析

根据贝叶斯网络分类器中属性为类提供的信息构成理论[11,12]，在贝叶斯网络分类器中，属性可为类提供三种信息，分别是传递依赖信息（最主要的信息）、直接导出依赖信息（重要信息）和间接导出依赖信息（辅助和补充信息），对于动态贝叶斯网络分类器也是如此，在 ADBNC 中属性为类提供信息的情况如图 3 所示.

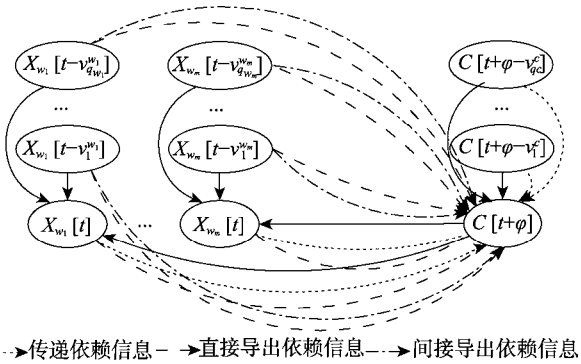

图 3 属性为类提供的信息构成

在图 3 中，属性可为类提供所有的三种信息，其中 $C[t-\varphi_k^c](1\leqslant k\leqslant q_c)$ 为 $C[t+\varphi]$ 提供传递依赖信息（时滞信息），$X_{w_h}[t](1\leqslant h\leqslant m)$ 为 $C[t+\varphi]$ 提供传递依赖信息和直接导出依赖信息（非时滞信息），而 $X_{w_h}[t-\varphi_k^{w_h}](1\leqslant h\leqslant m,1\leqslant k\leqslant q_{w_h})$ 能够为 $C[t+\varphi]$ 提供直接导出依赖信息和间接导出依赖信息（混合信息），因此更有助于提高分类器的分类准确性. 根据属性为类提供的信息分析，我们也可以

将为 $C[t+\varphi]$ 提供分类信息的变量分成两种，一种是类的时滞变量和非时滞属性，这种变量为类提供传递依赖信息，对分类至关重要，另一种是非时滞属性的时滞变量，它们为类提供导出依赖信息，起到次要和补充的作用.

3　实验与分析

在 UCI 和 Wind 数据库中选择用于实验的时间序列数据（16 个 UCI，18 个金融和 6 个宏观经济多变量时间序列数据集），采用滑动平均的方法修复缺失的数据，对较大的数据集进行截取，通过差分的方法对单调时间序列进行平稳化处理，依据文献[26]中的时序递进分类准确性标准计算分类器的分类准确率（或分类错误率）. 在 UCI 时间序列数据集中，选择一个时间序列变量作为类，其它的时间序列变量作为属性；在金融时间序列（股票、期货、汇率和基金）数据集中，同样也是选择一个时间序列变量（一只股票、一种期货、一个基金和一种汇率）作为类，其它的时间序列变量作为属性；在一个含有 12 个指标的宏观经济数据集中，分别选择其中的 6 个指标依次作为类，当一个指标作为类时，其它的指标作为属性. 在 ADBNC 的结构学习和分类过程中，对作为属性和类的指标均采用随时间的增减性变化来进行离散化（也可以根据需要采用其它的时序离散化方法），分别从分类准确性比较、分类器阶数对分类准确性的影响、时滞信息对分类准确性的影响和离散化方法对分类准确性的影响四个方面进行实验与分析，其中 $\Gamma=20$，$\Delta=4$，$M=10$（Γ、Δ 和 M 是根据实验测试的经验值）. 用于分类实验的时间序列数据集的情况如表 1 所示，其中 T、T_0 和 n 分别表示时间序列数据集中的记录数量，测试阈值和非时滞属性数量.

表 1　用于实验的多变量时间序列数据集情况

序号	数据集	$T(T_0)$	n	序号	数据集	$T(T_0)$	n
1	Adult	528（113）	6	21	股票_飞机制造_开盘价	417（113）	13
2	AllUsers	629（113）	8	22	股票_玻璃行业_开盘价	405（113）	15
3	Ann	496（113）	5	23	股票_公路桥梁_最低价	399（113）	16
4	Arabic_Digit	399（113）	12	24	股票_传媒娱乐_最低价	382（113）	14
5	Cmc	385（113）	9	25	股票_电力行业_收盘价	366（113）	46
6	EEG_Eye_State	399（113）	13	26	股票_服装鞋类_收盘价	324（113）	23
7	Eighthr1	432（113）	22	27	股票_纺织行业_最高价	348（113）	25
8	Eighthr2	455（113）	19	28	期货_上海_收盘价	243（113）	12
9	Eighthr3	369（113）	18	29	期货_上海_涨跌	243（113）	10
10	Hill_Valley	399（113）	20	30	期货_上海_成交量	243（113）	9
11	Reaction_Network_Undirected	371（113）	15	31	汇率_英镑_瑞士_加元	561（113）	29
12	Relation_Network_Directed	418（113）	13	32	汇率_香港	561（113）	9
13	Synthetic_Control1	507（113）	19	33	基金_债券_累积净值	423（113）	19
14	Synthetic_Control2	493（113）	19	34	基金_债券_日增长率	423（113）	19
15	Synthetic_Control3	374（113）	19	35	全国居民消费价格总指数	278（107）	11
16	White_wine_quality	533（113）	10	36	各项贷款合计同比增长率	278（107）	11
17	股票_sw 银行_开盘价	374（113）	13	37	出口商品总额同比增长率	278（107）	11
18	股票_sw 银行_收盘价	374（113）	13	38	M2 同比增长率	278（107）	11
19	股票_sw 银行_最低价	374（113）	13	39	全国商品零售价格总指数	278（107）	11
20	股票_sw 银行_最高价	374（113）	13	40	固定资产投资额同比增长率	278（107）	11

3.1　分类准确性比较

选择十个分类器与 ADBNC 进行分类准确性比较，其中分类器 SVM 和 XGboost 还需要建立转换数据集[28]（增加时滞信息），分类器的具体情况如下：

- DNBC：离散属性动态朴素贝叶斯分类器[20].
- GDNB：采用高斯函数估计属性密度的动态朴素贝叶斯分类器[22].
- GDFB：基于高斯函数估计属性密度的动态完全贝叶斯分类器[22].
- KDNB：使用高斯核函数估计属性密度的动态朴素贝叶斯分类器[25].
- KDFB：基于高斯核函数估计属性密度的动态完全贝叶斯分类器[26].

- DBN：依据 Friedman 的整体打分-搜索方法建立动态贝叶斯网络而得到的分类器[24].
- RNN：循环神经网络（Recurrent Neural Network）.
- LSTM：长短期记忆网络（Long Short-Term Memory）.
- SVM: 支持向量机（Support Vector Machine）.
- XGboost: 结合决策树的结构特点,使用函数空间的梯度下降法，实现优化损失函数的集成模型.
- ADBNC: 异步动态贝叶斯网络分类器.

RNN 的参数配置：1 个隐藏层，units=32，激活函数为 tanh（默认），损失函数 loss='mean_squared_error'，优化算法采用 optimizer=rmsprop，metrics=['accuracy']，epochs=100，batch_size=32；LSTM 的参数配置：1 个隐藏层，units=32，激活函数为'relu'，损失函数 loss='mean_squared_error'，优化算法采用 optimizer='adam'，metrics=['accuracy']，epochs=100，batch_size=z15；SVM 的参数配置：Cost=200，Gamma=0.01；XGboost 的参数配置：subsample=0.6，max_depth=2，eta=0.2，reg_lambda=0.001. 我们使用 Wilcoxon Signed-Ranks Test 和 Friedman Test with post-hoc Bonferroni test[29]进行两个分类器分类错误率（分类错误率=1-分类准确率）之间差异的置信打分，其中※表示 ADBNC 和用于比较的分类器相对于给定的检验方法差别显著,十个分类器与 ADBNC 的分类错误率实验结果如表 2 所示.

表 2 11 个分类器在 40 个时间序列数据集上的分类错误率实验结果（$\varphi=1$）

数据集	DNBC	GDNB	GDFB	KDNB	KDFB	DBN	RNN	LSTM	SVM	XGboost	ADBNC
Adult	0.3805	0.3628	0.3894	0.4336	0.3982	0.4159	0.3540	0.4159	0.3717	0.2523	0.3451
AllUsers	0.3363	0.3540	0.3806	0.4336	0.3009	0.3806	0.2478	0.3009	0.2920	0.2430	0.3009
Ann	0.1858	0.1770	0.1681	0.2301	0.3805	0.1770	0.1770	0.1858	0.1150	0.1869	0.1327
Arabic_Digit	0.3271	0.3805	0.3540	0.3982	0.3540	0.3982	0.3451	0.3717	0.3274	0.2991	0.2832
Cmc	0.3982	0.3894	0.3628	0.4159	0.2920	0.3982	0.3186	0.3717	0.3363	0.3458	0.3540
EEG_Eye_State	0.2212	0.4867	0.2743	0.4248	0.2478	0.3628	0.2832	0.2566	0.3628	0.2897	0.1947
Eighthr1	0.1681	0.2832	0.1416	0.3805	0.3097	0.1495	0.0885	0.0708	0.1327	0.1495	0.1062
Eighthr2	0.2743	0.2832	0.2566	0.3097	0.3097	0.1869	0.2212	0.2035	0.2655	0.2056	0.2212
Eighthr3	0.1593	0.4956	0.4513	0.4248	0.2124	0.1947	0.2212	0.1947	0.2566	0.1869	0.1593
Hill_Valley	0.2301	0.2301	0.1593	0.1239	0.1062	0.1416	0.2124	0.0088	0.1416	0.0467	0.0000
Reaction_Network_Undirected	0.0708	0.1593	0.1416	0.1327	0.2755	0.1327	0.0531	0.0088	0.0885	0.0841	0.0088
Relation_Network_Directed	0.2389	0.3009	0.2389	0.2832	0.2035	0.1593	0.1150	0.0973	0.1239	0.0935	0.1593
Synthetic_Control1	0.3186	0.3540	0.4248	0.3894	0.3363	0.3645	0.3451	0.3805	0.3097	0.3645	0.2920
Synthetic_Control2	0.2212	0.4867	0.3717	0.4602	0.3097	0.2212	0.2478	0.3097	0.2743	0.3178	0.2035
Synthetic_Control3	0.3451	0.3363	0.3363	0.3363	0.2478	0.2655	0.2832	0.2655	0.2389	0.3178	0.1770
White_wine_quality	0.2920	0.2920	0.2301	0.2655	0.2920	0.2150	0.2212	0.1593	0.2212	0.2150	0.1593
股票_sw 银行_开盘价	0.2743	0.4956	0.3894	0.1770	0.1947	0.3894	0.2920	0.2920	0.2565	0.3364	0.2301
股票_sw 银行_收盘价	0.2565	0.4513	0.3274	0.3894	0.3186	0.3551	0.2478	0.2389	0.4159	0.3551	0.2124
股票_sw 银行_最低价	0.2920	0.4867	0.3805	0.4956	0.4071	0.4867	0.2655	0.2920	0.3717	0.2430	0.2389
股票_sw 银行_最高价	0.2920	0.4867	0.3717	0.4779	0.3186	0.3717	0.3805	0.2832	0.4425	0.2991	0.2566
股票_飞机制造_开盘价	0.2301	0.4779	0.3717	0.3982	0.2212	0.2301	0.3451	0.1947	0.3274	0.3458	0.1504
股票_玻璃行业_开盘价	0.2655	0.3451	0.4425	0.4513	0.3628	0.3451	0.3717	0.3540	0.4336	0.4579	0.2743
股票_公路桥梁_最低价	0.3186	0.4779	0.4336	0.4248	0.3805	0.3274	0.4159	0.3274	0.3982	0.4486	0.2832
股票_传媒娱乐_最低价	0.2301	0.4071	0.3982	0.4071	0.2478	0.2301	0.2566	0.2301	0.3274	0.2991	0.1858
股票_电力行业_收盘价	0.3628	0.4602	0.4513	0.4867	0.3982	0.3628	0.4248	0.3628	0.4425	0.4206	0.3274
股票_服装鞋类_收盘价	0.2832	0.4779	0.4602	0.4425	0.3540	0.2832	0.3363	0.3628	0.4513	0.3925	0.2566
股票_纺织行业_最高价	0.3717	0.4956	0.4425	0.4602	0.4159	0.3894	0.4336	0.3894	0.4779	0.3832	0.3363
期货_上海_收盘价	0.3451	0.3363	0.3451	0.3186	0.2655	0.2124	0.2832	0.2124	0.3274	0.2991	0.2035
期货_上海_涨跌	0.2743	0.2478	0.2035	0.3805	0.3009	0.3186	0.2743	0.3186	0.2743	0.2710	0.2389
期货_上海_成交量	0.3274	0.4779	0.4512	0.4602	0.4512	0.2920	0.4071	0.2655	0.4425	0.4953	0.2920
汇率_英镑_瑞士_加元	0.2743	0.4956	0.3628	0.4336	0.2655	0.3451	0.1770	0.1858	0.3451	0.4112	0.2212

（续　表）

数据集	DNBC	GDNB	GDFB	KDNB	KDFB	DBN	RNN	LSTM	SVM	XGboost	ADBNC
汇率_香港	0.3274	0.3982	0.4425	0.4956	0.3451	0.3982	0.4248	0.3540	0.3363	0.3832	0.2832
基金_债券_累积净值	0.1504	0.3009	0.2478	0.4310	0.2743	0.4310	0.4513	0.2212	0.4867	0.4425	0.1150
基金_债券_日增长率	0.2301	0.3540	0.3805	0.3894	0.2212	0.2301	0.2832	0.3540	0.3540	0.2897	0.1681
全国居民消费价格总指数	0.1682	0.1869	0.2243	0.1869	0.2243	0.1308	0.1121	0.2243	0.2056	0.1308	0.1215
各项贷款合计同比增长率	0.1869	0.2336	0.2150	0.2150	0.2336	0.1869	0.2056	0.2336	0.2336	0.1776	0.1589
出口商品总额同比增长率	0.2710	0.2523	0.3084	0.2430	0.2804	0.2710	0.1869	0.3084	0.1869	0.2336	0.2056
M2 同比增长率	0.2336	0.1869	0.1776	0.2430	0.3084	0.2356	0.1682	0.2336	0.1589	0.1776	0.1963
全国商品零售价格总指数	0.1869	0.2243	0.2243	0.2056	0.2991	0.1308	0.1963	0.1121	0.1121	0.1215	0.1308
固定资产投资额同比增长率	0.3178	0.2617	0.2897	0.2617	0.2617	0.3178	0.3364	0.3551	0.2523	0.1963	0.2804
平均	0.2659	0.3598	0.3256	0.3579	0.2982	0.2859	0.2753	0.2577	0.298	0.2802	0.2116
Wilcoxon SR test	※−5.48	※−5.43	※−5.38	※−5.44	※−5.23	※−3.54	※−4.66	※−4.45	※−4.79	※−4.18	ADBNC
Friedman/Bonferroni test	※−3.07	※−6.52	※−5.61	※−6.88	※−4.34	※−4.13	※−3.06	※−3.01	※−3.73	※−3.35	ADBNC

在表 2 中，使用 Wilcoxon Signed-Ranks Test 和 Friedman Test with post-hoc Bonferroni test 的检验结果显示，ADBNC 与十个分类器在分类错误率方面差异显著. 再考察总体平均值，ADBNC 优于其它十个分类器的程度依次是 7.40%、23.15%、16.90%、22.78%、12.34%、10.40%、8.79%、6.21%、12.31% 和 9.53%，可见 ADBNC 具有优势的程度也非常明显. ADBNC 与其它分类器，在 40 个数据集的分类错误率散点图如图 4 所示. 图中每一个点的坐标是用于比较的两个分类器的分类错误率，在 45 度线上方和下方的点分别表示 ADBNC 的分类错误率小于和大于用于比较的分类器.

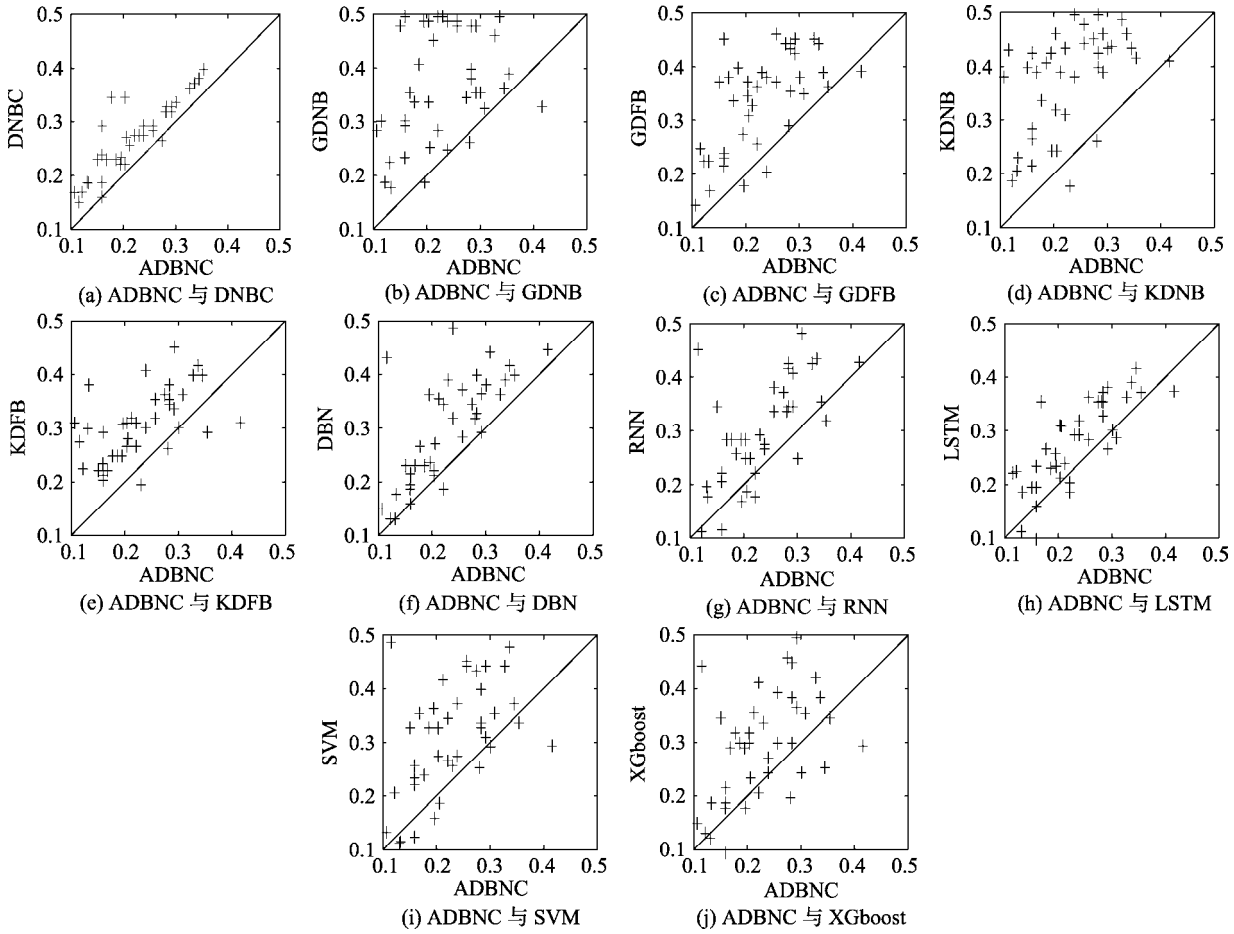

图 4　分类错误率散点图

从图 4 中的十个散点图来看，每一个散点图中的绝大部分点都在 45 度线的上方，因此，ADBNC 的分类错误率明显优于其它分类器. 综合分类器之间的分类准确性差异的显著性检验、分类准确性平均值比较和分类错误率散点图三方面的结果，显示了 ADBNC 相对于其它十个分类器在分类准确性方面具有明显的优势. ADBNC 与动态朴素贝叶斯分类器和动态完全贝叶斯分类器：动态朴素贝叶斯分类器（DNBC、GDNB 和 KDNB）虽然具有高效率，但条件独立性假设可能导致分类器与数据的欠拟合，而动态完全贝叶斯分类器（GDFB 和 KDFB）不考虑变量之间的条件独立性，易于导致分类器与数据的过拟合，ADBNC 的结构是类的属性马尔科夫毯，可以避免两种分类器所存在的问题. ADBNC 与循环神经网络（RNN 和 LSTM）：在 RNN 和 LSTM 的迭代中，时滞信息可能产生叠加，对时滞与非时滞信息的使用也没有选择性，而且只能采用分类结果来填充由错位变换而导致的丢失数据，这样易于产生丢失数据填充值的极端化，ADBNC 以理论上最优的马尔科夫毯来选择时滞与非时滞信息，并结合马尔科夫毯和 Gibbs 来迭代修复丢失的数据，因此在分类器的泛化能力方面具有优势. ADBNC 与非时间序列数据分类器（SVM 和 XGboost）：非时间序列数据分类器无法直接利用时滞信息，需要先建立时间序列数据集的转换数据集，转换数据集中的时滞信息量很难界定，而时滞信息的量又对分类准确性有较大的影响. ADBNC 与 DBN 的比较：采用 Friedman 的整体打分-搜索方法建立的 DBN，所突出的是变量之间的因果联系和推理计算，而不是更有效的分类信息传递和提取，因此，不利于实现训练与泛化之间的均衡.

同样使用表 1 中的数据集，通过发现类的非时滞子结点来选择属性，然后进行 RNN 的分类计算，分类错误率发生了一些变化（有增有减），总体平均值为 0.2635（有所下降），Wilcoxon Signed-Ranks Test 和 Friedman Test with post-hoc Bonferroni test 的结果为–4.02 和–3.07，与 ADBNC 之间的差异仍然显著. RNN 与 ADBNC 基于完全不同的机制提取分类信息进行分类计算，RNN 采用 BP（Back Propagation）神经网络的基本计算模式，并通过将前一个时间点的分类输出作为下一个时间点的分类输入（增加一个输入单元）来进行迭代分类计算，能够有效利用属性和类的时滞信息，但所采用的时间点拟合（不需要平稳性假设）方式，也存在过度拟合的风险. ADBNC 使用时间段拟合（需要局部平稳性假设）的方式，因此，往往具有更好的泛化性能.

3.2 分类阶数对分类准确性的影响

选择表 1 中的 Ann、EEG_Eye_State、Eighthr2、Relation_Network_Directed、Synthetic_Control2、股票_sw 银行_开盘价、股票_飞机制造_开盘价、期货_上海_收盘价、汇率_香港、基金_债券_日增长率、各项贷款合计同比增长率、出口商品总额同比增长率、M2 同比增长率、全国商品零售价格总指数和固定资产投资额同比增长率 15 个时间序列数据集，依次将它们编号为 1,2,...,15，分别取 $\varphi=0,1,2,...,50$ 计算分类准确率，从无时滞信息和有时滞信息两种情况进行实验与分析，实验结果如表 3 和表 4 所示，其中表的横向表示数据集，纵向表示分类阶数，后三行分别是 $\varphi=1,2,...,50$ 的分类准确率平均值、最大值和最小值，括号中的数字是最大值和最小值出现的次数，表中的加粗数字表示最大和最小的分类准确率.

表 3 无时滞信息的不同阶分类实验结果

阶数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0.8407	0.7965	0.8053	0.8850	0.7965	0.7522	0.8407	0.8142	0.7168	0.6549	0.8318	0.7383	0.8037	0.9159	0.6822
1	**0.6903**	0.5575	0.6637	0.6106	0.6549	0.5398	0.6106	0.5752	0.6106	**0.8850**	**0.6822**	**0.6822**	**0.6262**	**0.6355**	**0.6355**
2	0.5310	**0.6637**	**0.5044**	0.5398	0.5664	0.5310	0.6372	0.6283	0.5929	0.6637	0.5421	0.5327	0.5794	0.5981	0.5794
3	0.5752	0.5398	0.6372	0.5664	0.6106	0.5398	0.5221	0.5929	0.5664	0.8584	0.5421	0.5607	0.5888	**0.5140**	0.5514
4	**0.4867**	0.6106	0.5752	0.5664	0.6372	0.5752	0.5929	0.6195	0.6106	0.6283	0.5421	0.5327	0.5327	0.6168	0.5701
5	0.5929	0.6018	0.6372	0.5929	0.6195	0.5752	0.5664	0.6018	0.5487	0.5929	0.6262	0.5701	0.6168	0.6262	0.6168
6	0.5841	0.5929	0.6372	0.5575	0.5664	0.6018	0.5398	0.6018	0.5752	0.5133	0.6068	0.6168	0.6068	0.5327	0.5327
7	0.5752	0.5575	0.6195	0.5575	0.5929	0.6106	**0.6814**	0.6106	0.5929	0.5752	0.5421	0.5421	0.5234	0.6168	**0.4860**
8	0.5133	0.5929	0.5929	0.5841	0.5841	0.5841	0.5752	0.5752	0.5398	0.6018	0.5421	0.5421	0.5607	0.6168	0.5327
9	0.5841	0.5310	0.6195	0.5664	0.6283	0.6195	0.5575	0.6018	**0.4956**	0.5310	0.6075	**0.5047**	0.6075	0.6262	0.5701
10	0.5841	0.6372	**0.6991**	0.5398	**0.6903**	0.5841	0.5929	0.5575	0.5221	0.5664	0.5140	**0.5047**	0.5234	0.5421	0.5888
11	0.5752	0.5221	0.6106	0.5752	0.6195	0.5664	0.6372	0.5398	0.6106	0.5841	0.5794	0.5327	0.5514	0.6168	0.6168
12	0.5310	0.5487	0.5752	0.5664	0.6283	0.6018	0.5487	0.6018	**0.6195**	0.5575	0.5514	0.6262	0.5514	0.5514	0.5421

（续　表）

阶数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
13	0.5044	0.5310	0.5929	0.5221	0.5841	0.5310	0.5133	0.6106	0.5841	0.5841	0.5514	0.6168	0.5327	0.5234	0.5701
14	0.5487	0.6372	0.6018	0.5310	0.6106	0.5752	0.5310	**0.6726**	0.5487	0.6106	0.5701	0.5701	0.6075	0.5981	0.5701
15	0.5398	0.5221	**0.5044**	0.5133	0.5487	0.5575	0.5221	0.5752	0.5575	0.5310	0.5607	0.5981	0.5888	0.5794	0.5981
16	0.4867	0.5841	0.5841	0.5664	0.5221	0.6283	0.6018	0.5487	0.5044	0.5398	0.6355	0.6542	0.5888	0.5981	0.6075
17	0.5664	**0.6637**	0.6460	0.6018	0.5664	0.5929	0.5044	0.5929	0.5664	0.5752	0.5981	0.5794	**0.5047**	0.5421	0.5607
18	0.6106	0.5841	0.5752	0.5929	0.5310	0.5841	0.5752	**0.6726**	0.5664	0.5575	0.6449	0.6168	0.6168	0.5327	0.5421
19	0.5310	0.6018	0.6195	0.5310	0.6195	0.5752	0.5752	0.5398	0.5664	0.5575	0.5701	0.6355	0.6168	0.5327	0.5514
20	0.5752	0.5575	0.6283	**0.4956**	0.6018	0.5841	0.5752	0.5575	0.5487	0.5575	0.5327	0.5514	0.5514	0.5327	0.5514
21	0.5575	0.5841	0.6018	0.5487	0.6106	0.5752	0.5044	0.5929	0.5044	0.6018	0.5701	0.6075	0.5327	0.5514	0.5327
22	0.5841	0.5575	0.5929	0.5310	0.6106	0.6018	0.5221	0.6195	0.5487	0.5841	0.6355	0.5327	0.5421	0.5794	0.5327
23	0.5929	0.5664	0.5487	0.5133	0.6195	0.6018	0.6283	0.5487	0.5575	0.5310	0.5140	0.5140	0.5327	0.5981	0.5701
24	0.6018	0.6460	0.5929	0.5487	0.6018	0.5487	0.5221	0.6549	0.5487	0.5664	0.5888	0.6075	0.5794	0.6075	0.5607
25	0.5575	0.6283	0.5752	0.5929	0.5752	0.5133	0.6106	0.6460	0.5929	0.5487	0.6168	0.5888	0.6168	0.5514	0.5327
26	0.5310	0.5664	0.5929	0.5752	0.5221	0.6018	0.5575	0.6195	0.5398	0.5487	0.5327	0.5701	0.5140	0.5421	0.5794
27	**0.4867**	0.6106	0.5929	0.5752	**0.5133**	0.6460	0.5664	0.6460	0.5575	**0.4867**	0.5607	**0.5047**	0.5327	0.5514	0.5047
28	0.5221	**0.5044**	0.5398	0.5841	0.5841	**0.6549**	0.5487	0.5310	0.5752	0.5398	0.5607	0.6168	0.6075	0.5607	0.5607
29	0.6106	0.6460	0.5133	0.5398	0.5487	0.5487	0.5841	0.6106	0.5221	0.6195	0.5981	0.5888	0.5981	0.6168	0.5607
30	0.5398	0.5664	0.5752	0.5487	0.5221	0.5310	0.5133	0.5664	0.5487	0.5487	0.6168	0.5327	0.5327	0.5981	0.6168
31	0.5487	0.5929	0.6283	0.5133	0.5575	0.5752	0.5398	0.6372	0.5310	0.5664	0.5514	0.5794	0.5421	0.5327	0.5794
32	0.5929	0.5841	0.5841	0.6195	0.5310	0.5841	0.5664	0.5929	0.5841	0.6283	0.6168	0.5421	0.5888	0.5421	0.5794
33	0.6018	0.5752	0.6018	0.5487	0.5133	**0.4956**	**0.5133**	0.5575	0.5398	0.6106	0.5421	0.5607	0.5514	0.5327	0.5327
34	0.5929	0.5664	0.6283	0.5133	0.6460	0.6372	0.6549	0.5929	0.5221	0.5575	0.5234	0.5888	0.5981	0.6168	0.5607
35	0.6195	0.5044	0.6283	0.5221	0.5575	0.5664	0.5841	0.5664	0.5575	0.5221	0.5888	0.5888	0.5794	0.5421	0.5421
36	0.6283	0.5310	0.6283	0.5664	0.5841	0.5310	0.5398	0.6283	0.5664	0.5752	0.5327	0.5421	0.5888	0.5421	0.6355
37	0.5221	0.5133	0.6372	0.5398	0.5929	0.6106	0.4867	0.6018	0.5487	0.5398	0.5327	0.6449	0.5981	0.5514	0.6168
38	0.5221	0.5487	0.6372	**0.6372**	0.5929	0.5133	0.6018	**0.5221**	0.5664	0.5752	0.5794	0.5140	0.5794	0.5514	0.6262
39	0.5487	0.5487	0.5929	0.5575	0.5221	0.5044	0.6195	0.5929	0.6106	0.5221	0.5981	0.5140	0.6075	0.5514	0.5981
40	0.5664	0.5310	0.5841	0.6018	0.5841	0.5929	0.5929	**0.6726**	0.5398	0.6372	0.5981	0.5607	0.5607	0.5327	0.5327
41	0.6195	0.5575	0.5841	0.5575	0.5398	0.6018	0.5841	0.5664	0.5044	0.5575	0.5234	0.5794	0.5981	0.5140	0.5888
42	0.6372	**0.5044**	0.6195	0.6106	0.6106	0.5221	0.5487	0.5841	0.5664	0.5841	0.5421	0.5607	0.5794	0.5981	0.5421
43	0.5044	0.6018	0.6372	0.6018	0.5929	0.6018	0.5398	0.5575	0.5575	0.5752	0.5514	0.5234	0.5981	0.6262	0.6168
44	0.6106	0.5133	0.6106	0.5133	0.5575	0.6106	0.6106	0.5664	0.5487	0.5310	0.5514	0.5421	0.5327	0.5607	0.5514
45	0.5752	0.6549	0.5575	0.5487	0.5929	0.5398	0.5752	0.5929	0.5221	0.6018	**0.5047**	0.5514	0.5327	0.5981	0.5981
46	0.5398	0.5221	0.6372	0.5664	0.6106	0.6018	0.5664	0.5841	0.5929	0.6195	0.6075	0.6168	0.5514	0.5794	0.5607
47	0.5575	0.6106	0.5752	0.5310	0.5929	0.5575	0.5398	0.6018	**0.4956**	0.6283	0.5421	0.5794	0.5607	0.5327	0.5234
48	0.5664	0.5664	0.5664	0.5929	0.5752	0.5841	0.6106	0.5841	0.5575	0.5841	0.5140	0.5888	0.5327	0.6262	0.5981
49	0.6018	0.5929	0.5841	**0.4956**	0.6283	0.6106	0.6195	0.5929	0.5752	0.5841	0.5607	0.5794	0.5981	0.5234	0.5888
50	0.6106	0.5841	0.6106	0.5664	0.6018	0.5929	0.6283	**0.5221**	0.6018	0.5575	0.5514	0.5421	0.5140	0.5327	0.6168
平均	0.5667	0.5743	0.5996	0.5588	0.5855	0.5763	0.5708	0.5926	0.5582	0.5841	0.5690	0.5707	0.5691	0.5695	0.5703
最大	0.6903（1）	0.6637（2）	0.6991（1）	0.6372（1）	0.6903（1）	0.6549（1）	0.6814（1）	0.6726（3）	0.6195（1）	0.8850（1）	0.6822（1）	0.6822（1）	0.6262（1）	0.6355（1）	0.6355（1）
最小	0.4867（2）	0.5044（3）	0.5044（2）	0.4956（2）	0.5133（2）	0.4956（1）	0.4867（1）	0.5221（2）	0.4956（2）	0.4867（1）	0.5047（3）	0.5047（3）	0.5047（1）	0.5140（2）	0.4860（1）

表 4 有时滞信息的不同阶分类实验结果

阶数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0.8584	0.8053	0.7965	0.8319	0.8053	0.7522	0.8496	0.7965	0.7257	0.8850	0.8318	0.7850	0.8037	0.8505	0.7103
1	0.8673	0.8053	**0.7788**	0.8407	0.7965	0.7699	0.8496	0.7965	0.7168	**0.8938**	0.8411	**0.7944**	0.8037	**0.8692**	0.7196
2	0.8584	0.8053	**0.7788**	0.8496	0.8053	0.7699	0.8407	0.8142	0.7168	0.8850	0.8411	0.7664	0.7944	0.8505	0.7196
3	0.8673	0.7965	0.8053	**0.8319**	0.8053	**0.7611**	0.8584	0.8142	0.7168	0.8850	0.8505	**0.7944**	0.8037	0.8505	0.7290
4	0.8761	0.8053	0.7965	0.8584	0.7965	0.7788	0.8496	0.8053	0.7257	0.8761	0.8411	0.7850	0.8037	0.8411	0.7009
5	0.8673	0.8053	0.8053	0.8584	0.8053	0.7788	0.8496	0.8053	0.7257	0.8673	0.8411	0.7757	0.7944	0.8505	0.7196
6	0.8673	0.8053	0.7965	0.8584	**0.8142**	0.7788	0.8584	0.8053	0.7168	0.8761	0.8411	0.7757	0.8037	0.8411	0.7290
7	0.8761	0.8053	0.8053	0.8496	**0.8142**	0.7876	**0.8673**	0.8053	0.7257	0.8761	0.8318	0.7850	0.8037	0.8505	0.7103
8	0.8584	0.8053	0.8142	0.8673	**0.8142**	0.7788	0.8584	0.8053	0.7168	0.8761	0.8411	0.7664	0.7850	0.8505	0.7009
9	0.8761	0.8053	0.8142	0.8673	0.7965	0.7965	**0.8673**	0.8230	0.7257	0.8761	0.8318	0.7664	0.7944	0.8318	0.7103
10	0.8938	0.8230	0.8053	0.8584	**0.8142**	0.7876	**0.8673**	0.8142	**0.7345**	0.8850	0.8505	**0.7944**	0.7944	0.8411	0.7196
11	0.8761	0.8142	0.8053	0.8496	**0.8142**	0.7965	**0.8673**	0.8053	0.7257	**0.8938**	0.8598	0.7664	0.7757	0.8318	0.7103
12	0.8850	0.8053	0.8053	0.8584	0.7965	0.8142	**0.8673**	0.7965	0.7080	0.8761	0.8505	0.7757	0.8037	0.8318	0.7103
13	0.8761	0.8230	0.8142	0.8584	0.7876	0.8053	**0.8673**	0.8053	0.7080	0.8850	0.8411	0.7664	0.8131	0.8411	0.7290
14	0.8850	0.8230	0.8142	0.8584	0.7876	0.8230	0.8584	0.8142	0.7168	0.8850	0.8224	0.7477	0.8037	0.8411	0.7196
15	0.8850	0.8407	0.7876	0.8673	0.7876	0.8142	0.8496	0.8142	0.7080	0.8850	0.8318	0.7757	0.8131	0.8411	0.7196
16	0.8761	0.8230	0.8053	0.8673	0.7788	0.8230	0.8496	0.8142	0.7168	0.8850	0.8411	0.7570	0.7944	0.8411	0.7383
17	0.8761	0.8230	0.8230	0.8761	0.7876	0.8319	0.8584	0.8142	0.6991	0.8761	0.8224	0.7757	0.8037	**0.8692**	0.7103
18	0.8761	0.8407	0.8053	0.8673	0.7699	0.8230	0.8584	0.8053	0.7080	0.8850	0.8318	0.7477	0.7944	0.8411	0.7383
19	0.8761	0.8230	0.8142	**0.8938**	0.7699	0.8142	0.8407	0.7965	0.6903	0.8850	0.8318	0.7477	0.8131	0.8411	0.7196
20	0.8761	0.8142	0.8230	0.8761	0.7699	0.8319	0.8584	0.7876	0.7168	0.8850	0.8318	0.7664	**0.8224**	0.8411	0.7383
21	0.8850	0.8319	0.8230	0.8673	0.7699	0.8319	0.8496	0.7965	0.6903	0.8850	0.8224	0.7664	**0.8224**	0.8505	**0.7664**
22	0.8761	0.8407	0.8053	0.8761	0.7699	0.8407	0.8496	**0.7699**	0.6991	0.8850	0.8131	0.7383	0.8037	0.8505	0.7196
23	0.8850	0.8230	0.8230	0.8673	0.7611	0.8407	0.8584	0.7876	0.7080	0.8761	0.8224	0.7664	0.8131	0.8318	0.7196
24	0.8761	0.8230	0.8230	0.8673	0.7699	0.8319	0.8584	0.7788	0.6991	0.8850	0.8131	0.7477	**0.8224**	0.8131	0.7290
25	0.8761	0.8230	0.8319	0.8761	0.7699	0.8496	**0.8673**	0.7788	0.6991	0.8761	0.8131	0.7664	0.8037	0.8318	0.7290
26	0.8761	0.8230	0.8142	**0.8938**	0.7788	0.8407	0.8584	0.7788	0.7257	0.8850	0.8131	0.7570	0.8037	0.8318	0.7103
27	0.8673	0.8230	0.8319	0.8761	0.7522	0.8496	0.8496	0.7699	0.7257	0.8850	0.8037	0.7477	0.7850	0.8505	0.7477
28	0.8496	0.8142	**0.8407**	0.8850	0.7611	0.8496	0.8319	0.7788	0.7080	0.8850	0.8224	**0.7290**	0.8131	0.8505	**0.6916**
29	0.8673	0.8230	0.8230	0.8761	0.7611	0.8407	0.8407	0.7876	0.6991	0.8761	0.8318	0.7477	0.7944	0.8505	0.7196
30	0.8496	0.8230	0.8142	0.8850	0.7522	0.8407	0.8496	0.8053	0.6814	0.8761	0.8318	0.7477	0.8037	0.8598	0.7009
31	0.8584	0.8230	0.8230	0.8761	0.7522	0.8407	0.8496	0.7788	0.6903	0.8850	0.8224	0.7570	0.8037	0.8318	0.7196
32	0.8673	0.8319	0.8230	0.8673	0.7699	0.8407	0.8407	0.7876	0.7168	0.8850	0.8224	0.7570	0.7664	0.8505	0.6916
33	0.8496	0.8407	0.8142	0.8761	0.7611	0.8407	0.8496	0.8053	0.6991	0.8850	0.8318	0.7383	0.7664	0.8692	0.7196
34	0.8584	0.8584	0.8053	0.8761	0.7522	0.8584	0.8584	0.7876	0.6814	0.8761	0.8411	0.7570	0.7944	0.8598	0.7290
35	0.8496	0.8584	0.7965	0.8850	0.7611	0.8407	0.8407	0.7965	0.6726	0.8761	0.8224	0.7477	0.7944	0.8411	0.7290
36	0.8319	0.8584	0.7876	0.8673	0.7522	0.8319	0.8496	0.7965	0.6814	0.8761	0.8318	0.7383	0.7757	0.8411	0.7103
37	**0.8230**	0.8584	0.8142	**0.8938**	0.7434	0.8407	0.8496	0.7965	0.6726	0.8850	0.8224	0.7570	**0.7664**	0.8224	0.7103
38	0.8407	0.8584	0.8142	0.8850	0.7434	0.8319	0.8407	0.8053	0.6726	0.8850	0.8505	0.7664	0.7757	0.8224	0.7196
39	0.8407	**0.8673**	0.7965	0.8761	0.7434	0.8319	0.8496	0.8142	0.6814	0.8850	0.8411	0.7757	0.7850	0.8224	0.7196
40	0.8407	0.8496	0.7965	0.8850	0.7345	0.8319	0.8407	0.7965	0.6903	0.8850	0.8411	0.7477	0.7850	0.8224	0.7477
41	0.8407	0.8407	0.7876	0.8761	0.7434	0.8319	0.8407	0.8142	**0.6637**	0.8761	0.8505	0.7477	0.7944	0.8411	0.7383
42	0.8496	0.8407	**0.7788**	0.8761	0.7434	0.8319	0.8496	0.8230	0.6726	0.8761	0.8692	0.7664	0.7944	0.8318	0.7290
43	0.8319	0.8584	**0.7788**	0.8850	0.7611	0.8407	0.8496	**0.8407**	**0.6637**	0.8761	0.8692	0.7477	0.8037	0.8224	0.7103
44	0.8496	0.8584	0.8230	0.8850	0.7257	0.8319	0.8319	0.7876	0.6903	0.8761	0.8505	0.7383	0.7944	0.8224	0.7196
45	0.8319	0.8496	0.7876	**0.8938**	0.7345	0.8319	0.8319	0.8142	**0.6637**	0.8673	0.8598	0.7570	0.8037	0.8131	0.7290

（续　表）

阶数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
46	0.8319	**0.8673**	0.7876	**0.8938**	0.7345	0.8496	0.8496	0.8053	0.6903	0.8673	0.8598	0.7383	0.8037	0.8131	0.7196
47	0.8496	0.8584	0.8053	**0.8938**	0.7257	0.8319	0.8407	0.8319	0.6726	0.8673	0.8505	0.7477	0.7757	**0.7944**	0.7196
48	0.8584	0.8230	0.8053	**0.8938**	0.7168	0.8407	0.8407	0.7965	0.6726	0.8673	0.8692	0.7757	0.7944	0.8037	0.7196
49	0.8584	0.8496	0.8053	0.8850	0.7168	0.8319	0.8407	0.7965	0.6726	0.8673	0.8692	0.7383	0.7944	0.8131	0.7196
50	0.8673	0.8496	0.8053	**0.8938**	**0.7080**	0.8319	**0.8319**	0.7965	0.6726	**0.8584**	**0.8785**	0.7570	**0.7664**	0.8131	0.7196
平均	0.8625	0.8299	0.8071	0.8717	0.7696	0.8209	0.8510	0.8009	0.7000	0.8800	0.8374	0.7606	0.7972	0.8381	0.7207
最大	0.8938（1）	0.8673（2）	0.8407（1）	0.8938（8）	0.8142（5）	0.8584（1）	0.8673（7）	0.8407（1）	0.7345（1）	0.8938（2）	0.8785（1）	0.7944（3）	0.8224（3）	0.8692（3）	0.7664（1）
最小	0.8230（1）	0.7965（1）	0.7788（4）	0.8319（1）	0.7080（1）	0.7522（1）	0.8319（1）	0.7699（2）	0.6637（3）	0.8584（1）	0.8037（5）	0.7290（1）	0.7664（4）	0.7944（1）	0.6916（2）

从无时滞信息（表 3）和有时滞信息（表 4）两个方面，通过简单的统计运算，我们可以得到表 5. 在表 5 中，无时滞信息或有时滞信息的前两列表示 φ（$\varphi>0$）阶异步分类（在表 3 或表 4 中编号为 φ 的行）在 15 个数据集中具有最大（在 50 阶中的最大）和最小（在 50 阶中的最小）分类准确率的数量（最好和最坏情况的数量分布），后三列表示 φ 阶异步分类的分类准确率大于、等于和小于同步分类器（$\varphi=0$），在 15 个数据集中出现的频率（在表 3 或表 4 编号为 φ 的记录中，大于、等于和小于同步分类在 15 个数据集中所占的比例）.

对于无时滞信息的情况，宏观经济数据最大分类准确率集中在一阶（$\varphi=1$）分类器（宏观经济数据具有很强的马尔科夫性），最小分类准确率没有反

表 5　不同阶分类器的分类结果分布

阶数	无时滞信息					有时滞信息				
	最大	最小	大于	等于	小于	最大	最小	大于	等于	小于
1	7	0	0.0667	0.0000	0.9333	3	1	0.5333	0.2667	0.2000
2	1	1	0.0667	0.0000	0.9333	0	1	0.3333	0.3333	0.3334
3	0	1	0.0667	0.0000	0.9333	1	3	0.5333	0.3333	0.1334
4	0	1	0.0000	0.0000	1.0000	0	0	0.3333	0.4000	0.2667
5	0	0	0.0000	0.0000	1.0000	0	0	0.4667	0.3333	0.2000
6	0	0	0.0000	0.0000	1.0000	1	0	0.5333	0.2000	0.2667
7	1	1	0.0000	0.0000	1.0000	2	0	0.4667	0.4667	0.0666
8	0	0	0.0000	0.0000	1.0000	1	0	0.4667	0.2000	0.3333
9	0	2	0.0000	0.0000	1.0000	1	0	0.4000	0.2667	0.3333
10	2	1	0.0000	0.0000	1.0000	5	0	0.8000	0.0667	0.1333
11	0	0	0.0000	0.0000	1.0000	3	0	0.6667	0.1333	0.2000
12	1	0	0.0000	0.0000	1.0000	1	0	0.4000	0.2667	0.3333
13	0	0	0.0000	0.0000	1.0000	1	0	0.6667	0.0667	0.2666
14	1	0	0.0000	0.0000	1.0000	0	0	0.5333	0.1333	0.3334
15	0	1	0.0000	0.0000	1.0000	0	0	0.4667	0.2000	0.3333
16	0	0	0.0000	0.0000	1.0000	0	0	0.5333	0.1333	0.3334
17	1	1	0.0000	0.0000	1.0000	1	0	0.5333	0.1333	0.3334
18	1	0	0.0000	0.0000	1.0000	0	0	0.5333	0.1333	0.3334
19	0	0	0.0000	0.0000	1.0000	1	0	0.4667	0.2000	0.3333
20	0	1	0.0000	0.0000	1.0000	1	0	0.5333	0.1333	0.3334
21	0	0	0.0000	0.0000	1.0000	2	0	0.4667	0.2667	0.2666
22	0	0	0.0000	0.0000	1.0000	0	1	0.4000	0.2667	0.3333
23	0	0	0.0000	0.0000	1.0000	0	0	0.6000	0.0000	0.4000
24	0	0	0.0000	0.0000	1.0000	1	0	0.5333	0.0667	0.4000

（续 表）

阶数	无时滞信息					有时滞信息				
	最大	最小	大于	等于	小于	最大	最小	大于	等于	小于
25	0	0	0.0000	0.0000	1.0000	1	0	0.4667	0.0667	0.4666
26	0	0	0.0000	0.0000	1.0000	1	0	0.4000	0.2000	0.4000
27	0	4	0.0000	0.0000	1.0000	0	2	0.4000	0.2667	0.3333
28	1	1	0.0000	0.0000	1.0000	1	2	0.3333	0.1333	0.5334
29	0	0	0.0000	0.0000	1.0000	0	0	0.4000	0.1333	0.4667
30	0	0	0.0000	0.0000	1.0000	0	0	0.4000	0.2000	0.4000
31	0	0	0.0000	0.0000	1.0000	0	0	0.3333	0.2667	0.4000
32	0	0	0.0000	0.0000	1.0000	0	2	0.3333	0.1333	0.5334
33	0	2	0.0000	0.0000	1.0000	1	1	0.4667	0.2000	0.3333
34	0	0	0.0000	0.0000	1.0000	1	0	0.5333	0.0667	0.4000
35	0	1	0.0000	0.0000	1.0000	0	0	0.3333	0.1333	0.5334
36	0	0	0.0000	0.0000	1.0000	0	0	0.2000	0.3333	0.4667
37	0	1	0.0000	0.0000	1.0000	1	2	0.2667	0.2667	0.4666
38	1	1	0.0000	0.0000	1.0000	0	0	0.4667	0.0667	0.4666
39	0	0	0.0000	0.0000	1.0000	1	0	0.4000	0.2000	0.4000
40	1	0	0.0000	0.0000	1.0000	0	0	0.3333	0.2000	0.4667
41	0	0	0.0000	0.0000	1.0000	0	1	0.4000	0.0000	0.6000
42	0	1	0.0000	0.0000	1.0000	0	1	0.4000	0.0667	0.5333
43	0	0	0.0000	0.0000	1.0000	1	2	0.3333	0.2000	0.4667
44	0	0	0.0000	0.0000	1.0000	0	0	0.4000	0.0000	0.6000
45	0	1	0.0000	0.0000	1.0000	1	1	0.4000	0.0667	0.5333
46	0	0	0.0000	0.0000	1.0000	2	0	0.4000	0.1333	0.4667
47	0	1	0.0667	0.0000	0.9333	1	1	0.4667	0.0000	0.5333
48	0	0	0.0000	0.0000	1.0000	1	0	0.4000	0.1333	0.4667
49	0	1	0.0000	0.0000	1.0000	0	0	0.4000	0.1333	0.4667
50	0	1	0.0000	0.0000	1.0000	2	4	0.4667	0.0667	0.4666
平均	0.3600	0.5000	0.0053	0.0000	0.9947	0.8000	0.5000	0.4467	0.1733	0.3800

应出明显的规律，其它数据集也没有明显的倾向性.$\varphi>1$和$\varphi=0$的比较，只存在一个数据集“基金_债券_日增长率”,有 6%异步分类的分类准确率大于同步分类，其它的数据集，异步分类的分类准确率均明显低于同步分类；异步分类与同步分类的平均值、最大值和最小值之间差值的平均值依次是–0.2170、–0.1116 和–0.2903，取到最大值与最小值的数量分别是 18 和 28，大于、等于和小于情况的分布也非常极端；在 15 个数据集中，一次异步分类不小于同步分类的可能性大约是 0.0053. 也就是当不考虑时滞信息时，总体来看，异步分类的效果远不如同步分类（阶数对分类准确率有较大的影响），因此不适宜进行异步分类计算. 对于有时滞信息的情况，具有最大和最小分类准确率的情况没有明显的倾向性，取到最大值与最小值的数量分别是 40（与无时滞信息情况相比，增加了 122%）和 29（与无时滞信息情况相当）；异步分类与同步分类的平均值、最大值和最小值之间差值的平均值依次是 0.0104、0.0419 和–0.0267，我们能够发现，有时滞信息取到最大值的可能性比较大，这更适合于对分类器进行优化；对于异步分类器，大约以 0.62 的可能性在一次分类中的分类准确性不小于同步分类，因此，在增加时滞信息的情况下，可以进行异步分类计算.

3.3 时滞信息对分类准确性的影响

分别对表 3 和表 4 中的数据按数据集与分类阶数进行平均，得到无时滞信息和有时滞信息的分类准确率平均值，如图 5 所示，其中图 5（a）的横轴表示分类阶数，图 5（b）的横轴表示数据集，纵轴都表示平均分类准确率.

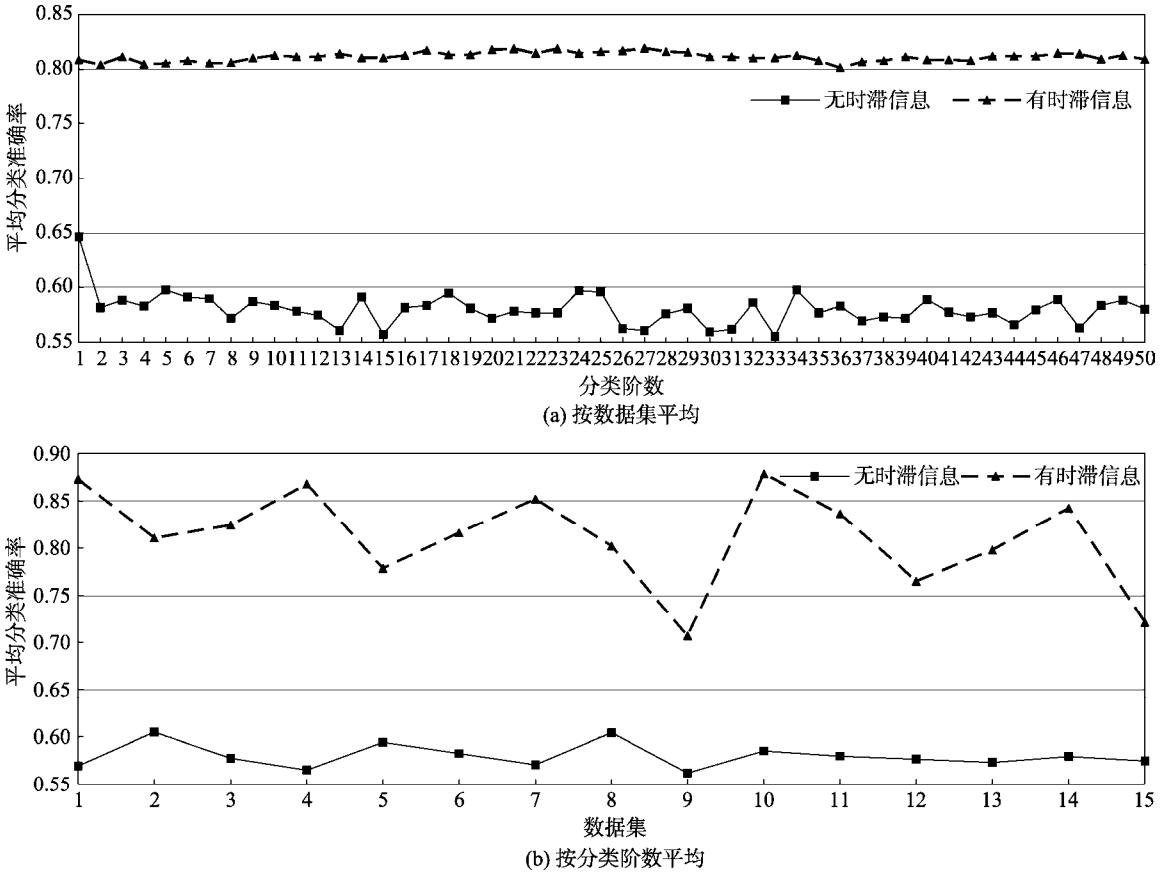

图 5　时滞信息对分类准确性的影响

从图 5 中我们能够看出，增加时滞信息能够显著提高分类器的分类准确性. 对于按数据集的平均，有时滞信息与无时滞信息的分类准确率差异的最大值、最小值和平均值分别是 0.2587、0.1620 和 0.2316；关于按分类阶数的平均，有时滞信息与无时滞信息的分类准确率差异的最大值、最小值和平均值分别是 0.3036、0.1450 和 0.2316. 对表 3 和表 4 中 15 个数据集的同步分类准确率和异步平均分类准确率，再按数据集平均，结果分别是 0.7916 和 0.5744（无时滞信息），以及 0.8058 和 0.8098（有时滞信息）. 我们发现无时滞信息的同步分类准确率远大于异步平均分类准确率，而有时滞信息的同步分类准确率却小于异步平均分类准确率，具有时滞信息的同步和异步平均分类准确率也均大于无时滞信息的情况，这验证了时滞信息也是一种重要的分类信息. 但时滞信息（或时滞变量）不是越多越好，过多的时滞信息（或时滞变量）会产生大量的冗余，反而会导致分类准确性的下降，也会降低效率. 机器学习研究的一个核心问题是实现训练与泛化之间的均衡，基于数据建立分类器是一种归纳学习，大量的冗余信息会对分类器学习产生误导，也会导致分类器与数据的过度拟合，从而会降低分类器的可靠性.

时间序列分类数据集（数据集的记录之间不满足独立同分布的假设，而是具有时序依赖）与非时间序列分类数据集（数据集的记录之间满足独立同分布的假设）的最大不同就是含有时滞信息，我们已经通过大量的实验验证了时滞信息也是分类的重要信息. 时滞信息对各种分类器的分类结果均有影响，但影响程度会有差异.

3.4　离散化方法对分类准确性的影响

在表 3 的 15 个数据集中选择 5 个数据集，它们是 EEG_Eye_State、股票_sw 银行_开盘价、股票_飞机制造_开盘价、期货_上海_收盘价和 M2 同比增长率，类统一按增减性进行离散化，属性则分别使用非时序离散化（等频、等距和分位数）和时序离散化（增减、转折、增增和减减）方法进行离散化，对 5 个数据集的分类准确率进行平均，7 种离散化方法对 50 阶具有时滞信息的平均分类准确率情况如图 6 所示，其中横轴表示分类阶数，纵轴表示平均分类准确率.

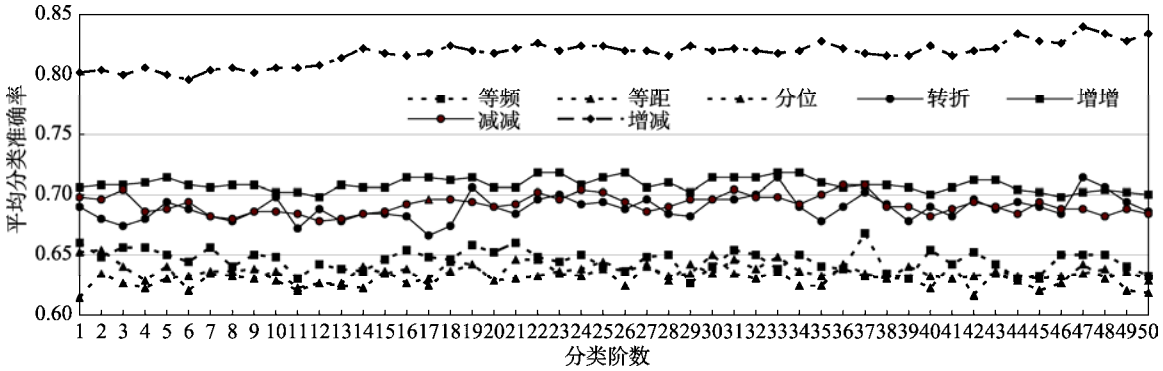

图 6 离散化方法对分类准确性的影响

在图 6 中我们能够发现，七种离散化方法的效果明显地形成三种情况. 第一种是非时序离散化（包括等频、等距和分位数），这种情况的分类效果最差，因为丢失了时滞信息；第二种是非一致时序离散化（包括转折、增增和减减），部分时滞信息会得到有效的利用，因此具有较好的分类效果；第三种是一致时序离散化（属性和类采用同样的离散化方法），时滞信息会得到充分的利用，分类效果最好，对其它的时间序列数据集和离散化方法也得到了类似的结果. 我们可以得出结论：如果需要将时间序列数据离散化后进行分类计算，那么采用一致离散化方法离散化属性和类是一种比较好的选择.

4 结论和进一步的工作

我们结合时间序列的离散化，变量的时序转换，变量的错位变换，类的子结点、父结点和子结点的父结点学习等，给出了可用于多变量时间序列分类预测的 ADBNC，使用 UCI、金融和宏观经济时间序列数据进行实验的结果显示，ADBNC 具有良好的分类准确性.

在 ADBNC 的学习与分类过程中，变量的时序转换使属性与类的时滞信息（或历史信息）得到有效的利用，从而有利于提高分类器的分类准确性；通过变量的错位变换（或数据的错位变换）构建新的多变量时间序列数据集，可实现异步分类预测；结合类的非时滞子结点、时滞父结点和非时滞子结点的时滞父结点学习，我们能够得到类的近似马尔科夫毯，而在理论上，马尔科夫毯分类器是最优分类器；通过对由属性和类的错位变换而形成丢失数据的修复，可避免时间序列数据集中的信息丢失，使 ADBNC 的分类计算更加可靠.

进一步的工作是通过对高阶分类问题的深入研究，揭示多变量时间序列所蕴含的关联与制约机制，并将其向异步动态贝叶斯网络回归模型推广.

参 考 文 献

[1] Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. San Mateo, USA: Morgan Kaufmann, 1988, 383-408

[2] Flach P A, Lachiche N. Naive Bayesian classification of structured data. Machine Learning, 2004, 57(3): 233-269

[3] Stephens C R, Huerta H F, Linares A R. When is the naive Bayes approximation not so naive? Machine Learning, 2017, 92(1): 1-45

[4] Xu W Q, Jiang L X, Yu L J. An attribute value frequency-based instance weighting filter for naive Bayes. Journal of Experimental & Theoretical Artificial Intelligence, 2019, 31(2): 225-236

[5] Zheng F, Webb G I, Pramuditha S. Subsumption resolution: an efficient and effective technique for semi-naïve Bayesian learning. Machine Learning, 2012, 87(1): 93-125

[6] Flores M J, Gámez J A, Martínez A M. Domains of competence of the semi-naive Bayesian network classifiers. Information Sciences, 2014, 260(1): 120-148

[7] Cai Q, Liu H, Zhou S. An adaptive-scale active contour model for inhomogeneous image segmentation and bias field estimation. Pattern Recognition, 2018, 82(10): 79-93

[8] Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. Machine Learning, 1997, 29(2-3): 131-161

[9] Jing Y S, Pavlović V, Rehg J M. Boosted Bayesian network classifiers. Machine Learning, 2008, 73(2): 155-184

[10] Namrata S, Pradeep S, Sabu M T et al. A novel bagged naive Bayes-decision tree approach for multi-class classification problems. Journal of Intelligent & Fuzzy Systems, 2019, 36(3): 2261-2271

[11] Wang S C, Xu G L, Du R J. Restricted Bayesian classification networks. Science China Information Sciences, 2013, 56(5): 2122-2137

[12] Wang Shuang-Cheng, Gao Rui, Du Rei- Jie. Restricted Bayesian network classifier based on Gaussian Copula. Chinese Journal of Computers, 2016, 39(8): 1612-1625 (in Chinese)
(王双成, 高瑞, 杜瑞杰. 基于高斯 Copula 的约束贝叶斯网络分类器研究. 计算机学报, 2016, 39(8): 1612-1625)

[13] Yang Y L, Ding M X. Decision function with probability feature weighting based on Bayesian network for multi-label classification. Neural Computing & Applications, 2019, 31(9):

4819-4828
[14] Friedman N, Murphy K, Russell S. Learning the structure of dynamic probabilistic networks//Proceedings of the 14th International Conference on Uncertainty in Artificial Intelligence, Madison, USA, 1998, 139-147
[15] Murphy K. Dynamic Bayesian networks: Representation, inference and learning [Ph.D. Thesis], UC Berkeley, Computer Science Division, USA, 2002
[16] Yang G S, Lin Y Z, Bhattacharya P. A driver fatigue recognition model based on information fusion and dynamic Bayesian network. Information Sciences, 2010, 180(10): 1942-1954
[17] Dabrowski J J, Beyers C, de Villiers J P. Systemic banking crisis early warning systems using dynamic Bayesian networks. Expert systems with applications, 2016, 62(11): 225-242
[18] Tanjin A M, Faisal K, Syed I. Fault detection and pathway analysis using a dynamic Bayesian network. Chemical Engineering Science, 2019, 195(2): 777-790
[19] Heng J L, Zheng K F, Kaewunruen S et al. Dynamic Bayesian network-based system-level evaluation on fatigue reliability of orthotropic steel decks. Engineering Failure Analysis, 2019, 105(11): 1212-1228
[20] Palacios-Alonso M A, Brizuela C A, Sucar L E. Evolutionary learning of dynamic naive Bayesian classifiers. Journal of Automated Reasoning, 2010, 45(1): 21-37
[21] Alkhateeb J H, Pauplin O, Ren J et al. Performance of hidden Markov model and dynamic Bayesian network classifiers on handwritten Arabic word recognition. Knowledge-Based Systems, 2011, 24(5): 680-688
[22] Wang Shuang-Cheng, Pei Zhen, Bi Yu-Jiang. Dynamic Bayesian network classifier model for predicting the cyclical turning points of economic fluctuation. Journal of Industrial Engineering and Engineering Management, 2011, 25(2): 173-177 (in Chinese)
(王双成, 裴瑱, 毕玉江. 经济周期转折点预测的动态贝叶斯网络分类器模型. 管理工程学报, 2011, 25(2): 173-177)
[23] Kafai M, Bhanu B. Dynamic Bayesian networks for vehicle classification in video. IEEE Transactions on Industrial Informatics, 2012, 8(1): 100-109
[24] Premebida C, Faria D R, Nunes U. Dynamic Bayesian network for semantic place classification in mobile robotics. Autonomous Robots, 2017, 41(5): 1161-1172
[25] WANG Shuang-Cheng, GAO Rui, DU Rui-Jie. Learning and optimization of dynamic naive Bayesian classifiers for small time series. Control and Decision, 2017, 32(1): 163-166 (in Chinese)
(王双成, 高瑞, 杜瑞杰. 小时间序列动态朴素贝叶斯分类器学习与优化. 控制与决策, 2017, 32(1): 163-166)
[26] WANG Shuang-Cheng, ZHENG Fei, GAO Rui. Dynamic full Bayesian ensemble classifiers for small time series. Science China: Information Science, 2017, 47(11): 1445-1463 (in Chinese)
(王双成, 郑飞, 高瑞. 小时间序列动态完全 Bayesian 集成分类器研究. 中国科学: 信息科学, 2017, 47(11): 1445-1463)
[27] Rishu C, Rama K C, Seema V et al. Smartphone based context-aware driver behavior classification using dynamic Bayesian network. Journal of Intelligent & Fuzzy Systems, 2019, 36(5): 4399-4412
[28] Wang Shuang-Cheng, Gao Rui, Du Rei-Jie. With super parent node Bayesian network ensemble regression model for time series. Chinese Journal of Computers, 2017, 40(12): 2748-2761 (in Chinese)
(王双成, 高瑞, 杜瑞杰. 具有超父结点时间序列贝叶斯网络集成回归模型. 计算机学报, 2017, 40(12): 2748-2761)
[29] Demsar J. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research, 2006, 7(1): 1-30

WANG Shuang-Cheng, Ph. D., professor. His main research interests include artificial intelligence, machine learning, data mining and their application.

ZHANG Li, Ph. D., lecturer. His main research interests include machine learning and information system.

ZHENG Fei, Ph. D., associate professor. His main research interests include information safety and data mining.

Background

We have made a thorough study of Bayesian network classifier for non-time series data. They include Bayesian network classifiers with discrete attributes, Markov network classifiers, Markov blanket classifiers, as well as Bayesian network classifiers for continuous attributes based on estimating attribute density by Gaussian function, Gaussian kernel function and Gaussian Copula. They all show good classification accuracy in solving the corresponding classification problems, but these classifiers can not be directly used in the classification of time series data. We have also studied dynamic Bayesian network classifiers with discrete and continuous attributes. They include dynamic naive Bayesian classifiers with discrete attributes, dynamic chain extended Bayesian classifiers and dynamic tree extended Bayesian classifiers as well as the dynamic Bayesian network classifier with continuous attributes based on Gaussian function, Gaussian kernel function and Gaussian copula to estimate attribute density. These dynamic Bayesian network classifiers are synchronous classifiers (synchronous changes of class and attributes). The same is true for recurrent neural networks, long short-term memory networks and gated recurrent unit

networks that can be used for time series data classification. Asynchronous classification is ubiquitous, but there is still a lack of in-depth exploration for this kind of classification problems. In this paper, we combine time series preprocessing, the timing conversion of variables, the dislocation transformation of variables, classifier structure learning based on variable order and search & scoring method, missing data restoration for class variable, the criteria for timing progressive classification accuracy to build an asynchronous dynamic Bayesian network classifier. This classifier can effectively utilize the time-delay, non-time-delay and mixed classification information contained in multivariate time series data, as well as the transitive dependency, direct induction dependency and indirect induction dependency information to do classification calculation. It can implement the classification prediction for short-, medium- and long-term in actual demand. Our research content is an important part of the National Natural Science Foundation (No. 61672065) and the National Social Science Foundation (No. 18BTJ020). Our further work is to reveal the relationship and restriction mechanism of multivariate time series by exploring the problem of high-order classification, and to extend it to the asynchronous dynamic Bayesian network regression model.