



基于不确定性的多元时间序列分类算法研究

张旭 张亮 金博 张红哲

Uncertainty-based Multivariate Time Series Classification

ZHANG Xu, ZHANG Liang, JIN Bo, ZHANG Hong-Zhe

在线阅读 View online: <https://doi.org/10.16383/j.aas.c210302>

您可能感兴趣的其他文章

多元时间序列因果关系分析研究综述

Survey on Causality Analysis of Multivariate Time Series

自动化学报. 2021, 47(1): 64–78 <https://doi.org/10.16383/j.aas.c180189>

基于小样本学习的图像分类技术综述

Survey on Image Classification Technology Based on Small Sample Learning

自动化学报. 2021, 47(2): 297–315 <https://doi.org/10.16383/j.aas.c190720>

基于扩散方法的分布式随机变分推断算法

Distributed Stochastic Variational Inference Based on Diffusion Method

自动化学报. 2021, 47(1): 92–99 <https://doi.org/10.16383/j.aas.c200445>

基于异常序列剔除的多变量时间序列结构化预测

Structural Prediction of Multivariate Time Series Through Outlier Elimination

自动化学报. 2018, 44(4): 619–634 <https://doi.org/10.16383/j.aas.2017.c160707>

基于改进差分进化和回声状态网络的时间序列预测研究

Time Series Prediction Based on Improved Differential Evolution and Echo State Network

自动化学报. 2021, 47(7): 1589–1597 <https://doi.org/10.16383/j.aas.c180549>

延迟深度回声状态网络及其在时间序列预测中的应用

Delayed Deep Echo State Network and Its Application on Time Series Prediction

自动化学报. 2020, 46(8): 1644–1653 <https://doi.org/10.16383/j.aas.c180264>

基于不确定性的多元时间序列分类算法研究

张旭¹ 张亮² 金博³ 张红哲¹

摘要 多元时间序列 (Multivariate time series, MTS) 分类是许多领域中的重要问题, 准确的分类结果可以有效地帮助决策. 当前的 MTS 分类算法在个体的表征学习阶段难以自动建模多元变量之间复杂的交互关系, 并且无法评估分类结果的可信度, 这会导致模型性能受限, 以及缺乏具备统计意义的可靠性解释. 本文提出了一种基于不确定性的多元时间序列分类算法, 变分贝叶斯共享图神经网络, 即 VBSGNN (Variational Bayes shared graph neural network). 首先通过图神经网络 (Graph neural network, GNN) 提取多元变量之间的交互特征, 然后利用贝叶斯神经网络 (Bayesian neural network, BNN) 为预测过程引入了不确定性. 最后在 10 个公开 MTS 数据集上进行了算法实验, 并与当前提出的 7 类算法进行了比较, 结果表明 VBSGNN 可有效学习多元变量之间的交互关系, 提升了分类效果, 并使得模型具备一定的可靠性评估能力.

关键词 多元时间序列预测, 图神经网络, 不确定性, 随机变分推断, 贝叶斯神经网络

引用格式 张旭, 张亮, 金博, 张红哲. 基于不确定性的多元时间序列分类算法研究. 自动化学报, 2023, 49(4): 790–804

DOI 10.16383/j.aas.c210302

Uncertainty-based Multivariate Time Series Classification

ZHANG Xu¹ ZHANG Liang² JIN Bo³ ZHANG Hong-Zhe¹

Abstract Multivariate time series (MTS) classification is an important problem in many fields. Accurate classification results can help realize effective decision-making. Current MTS classification algorithms fail to automatically capture the complex interactions among multiple variables in the representation learning stage, and lack confidence evaluation of the classification results. This causes limited model performance and a shortage of statistically meaningful model reliability explanation. In this paper, an uncertainty-aware multivariate time series classification algorithm, namely VBSGNN (Variational Bayes graph neural network), is proposed. First, the multivariate interaction features are extracted automatically via a graph neural network module. Then, the uncertainty is introduced into the prediction process based on the Bayesian neural network, which finally gives the prediction results some level of confidence. Finally, the algorithm experiments are carried out on 10 publicly available MTS datasets, and compared with seven state-of-the-art benchmarks. The results show that the proposed algorithm VBSGNN can efficiently learn the interactions between multiple variables, improve the classification results, and obtains confidence evaluation ability.

Key words Multivariate time series prediction, graph neural network (GNN), uncertainty, stochastic variational inference, Bayesian neural network (BNN)

Citation Zhang Xu, Zhang Liang, Jin Bo, Zhang Hong-Zhe. Uncertainty-based multivariate time series classification. *Acta Automatica Sinica*, 2023, 49(4): 790–804

收稿日期 2021-04-11 录用日期 2021-11-02

Manuscript received April 11, 2021; accepted November 2, 2021

国家自然科学基金 (61772110), 辽宁省教育厅科学研究经费 (LJKZ1045), 上海市卫生和计划生育委员会科研课题 (20184Y0247) 资助

Supported by National Natural Science Foundation of China (61772110), Scientific Research Project of the Education Department of Liaoning Province (LJKZ1045), and Scientific Research Project of Shanghai Health and Family Planning Commission (20184Y0247)

本文责任编辑 金连文

Recommended by Associate Editor JIN Lian-Wen

1. 大连理工大学机械工程学院 大连 116086 2. 东北财经大学国际商学院 大连 116023 3. 大连理工大学创新创业学院 大连 116086

1. Mechanical Engineering College, Dalian University of Technology, Dalian 116086 2. International Business College, Dongbei University of Finance and Economics, Dalian 116023 3. School of Innovation and Entrepreneurship, Dalian University of Technology, Dalian 116086

医疗器械的信号记录等多元时间序列, 也称为 MTS (Multivariate time series), 数据广泛存在于现实生活中, 有效利用这些数据可以提高生产效率及经济收益^[1-3]. 尤其是多元时间序列分类, 也称为 MTSC (Multivariate time series classification), 问题受到了广泛的关注^[4]. 多元时间序列分类模型被用于基于脑电图, 也称为 EEG (Electroencephalogram) 或心电图, 也称为 ECG (Electrocardiogram) 的疾病预测^[5] 和运动识别^[6-7] 等.

目前学者们提出了许多 MTS 分类算法. 例如基于最近邻算法的动态时间规划^[8] 和基于特征的隐藏单元逻辑回归模型^[9] 成功应用于许多 MTS 基准数据集的分类预测, 但是这些方法需要经过大量的

数据预处理和特征工程设计过程. 最近, 基于深度学习的方法^[4, 10-11]在时间序列分类任务中取得了较好的效果. 这些方法可以通过卷积或递归操作有效地学习 MTS 的低维个体特征表示, 但是无法对 MTS 变量之间复杂的交互关系进行建模, 从而影响最终的模型效果. 虽然毛文涛等^[12]提出了基于异常序列剔除的多变量时间序列结构化预测方法, 考虑了多元变量时间的依赖关系, 但是该方法依赖于复杂的特征处理过程.

图是一种特殊的数据形式, 由不同的节点(实体)组成, 它描述了不同节点之间的关系. 图神经网络, 又称为 GNN (Graph neural network), 可以聚合图数据中相邻节点的特征信息, 获得表达能力更强的实体表征^[13-14]. 多元时间序列数据也可以从图结构角度来处理, 可以将多元时间序列中的变量视为图结构中的节点, 并且它们通过一定的交互关系产生连接. 因此, 利用图神经网络对多元时间序列数据进行建模是一种既考虑变量之间的交互关系又同时保留相应原始时间轨迹的有效方法^[15-16]. 由于多元时间序列数据并没有显式的图结构, 因而如何建立合适的图结构是首先需要解决的问题. 有的方法依赖于计算复杂度较高的自注意力机制建立图结构^[15], 有的依赖于精心设计的多层神经网络来建立图结构^[16], 还有的利用传统距离度量(例如点积、欧氏距离等)建立图结构的形式^[17], 时间和空间复杂度较高^[16]. 目前有学者提出了“共生变量对”^[18]的思想, 即在多元时间序列中出现在同一个批量样本和时间步上两个非零变量为“共生变量对”, 两个变量之间会互相影响, 初始化的影响程度为两个特征在批量样本所有共同出现时间点上的特征之和. 基于该思想构建的多元时间序列数据图结构经过一次非线性优化后在预测任务中取得了很好的效果. 该轻量级方法启发我们以此方式建立初始化的图结构, 然后利用一层神经网络进一步优化. 进而可以应用于目前比较流行的图卷积计算方式 GraphSage^[19], GIN^[20], GAT^[21]等. 同时, 现有的多元时间序列分类预测模型仅基于输出的概率分数来确定预测类别, 这种预测分数往往是模型过度自信产生的结果, 并不可靠^[22], 该类模型没有考虑预测过程中的不确定性, 无法表达每个样本判定结果的不确定程度. 这种不确定性已被证实带来灾难性的结果^[23]. 因而多元时间序列分类模型需要能够评估对分类结果的不确定性程度. 在提供模型的不确定性估计方面, 具备严谨数学逻辑的贝叶斯神经网络, 又称为 BNN (Bayesian neural network)^[22, 24]取得了不错的效果. BNN 模型主要对参数的分布进行建模, 通过参数

分布的不确定性来估计预测结果的不确定性^[25].

综上所述, 针对目前 MTS 算法难以自动建模多元变量之间复杂的交互关系以及模型未考虑不确定性的评估等问题, 本文首先基于“共生变量对”的思想构建初始化的图结构, 利用全连接神经网络自动学习(微调)多元变量对之间显性的关系, 然后利用图神经网络的消息传播与聚合机制提取多元变量隐性的复杂关系, 得到考虑了多元变量之间高阶交互关系的初始样本表征信息. 最后, 基于学习到的初始样本表征信息, 本文利用贝叶斯神经网络将预测过程中的不确定性引入模型, 使得最终的分​​类结果可进行不确定性度量. 本文的创新点总结如下:

1) 本文提出一种端到端的基于不确定性的多元时间序列分类框架变分贝叶斯共享图神经网络, 即 VBSGNN (Variational Bayes shared graph neural network). 该框架利用图神经网络建模多元变量之间的交互关系, 然后将融合了高阶隐性与显性关系的表征向量送入贝叶斯神经网络, 建模不确定性. 框架中核心的图神经网络模块和不确定性度量模块串联而非耦合, 易于扩展, 是一种轻量级的框架.

2) 根据贝叶斯神经网络的计算结果, 本文提出基于样本内部类别相对不确定性的“相对方差”的不确定性度量方式来识别潜在的容易造成误判断的不确定性样本.

3) 本文基于多个 MTS 基准数据集进行了大量的算法实验, 并与当前的主要多元时间序列分类算法进行了比较, 实验结果表明我们提出的算法模型取得了较好的分类效果, 证明了所提出的图结构学习方法、模型预测过程中提出的不确定性度量等方法是有​​效的.

1 基于图结构的多变量交互关系学习

该模块旨在建立以多元时间序列变量为节点的图结构. 首先基于“共生变量对”思想得到一批 MTS 数据的共享初始图结构, 进一步基于神经网络学习得到对于分类任务具有潜在意义的(依据任务本身的目标损失函数, 通过梯度下降法来微调, 并对其进行稀疏过滤操作, 以降低噪声)自适应图结构, 在此基础上学习更好的样本表征, 进而得到更好的预测结果. 具体如下所示.

1.1 构建初始化图结构

给定 MTS 数据 T , $T \in \mathbf{R}^{s \times t \times m}$, s 代表样本维度, t 代表时间维度, m 代表多变量维度, 对应的标签为 Y , $Y \in \mathbf{R}^{s \times 1}$. $D = \{T_i, Y_i\}_{i=1}^s$ 即为输入数据集,

定义 $\mathcal{F} \in \mathcal{T}$ 表示抽取的批量样本. 定义每个变量为图结构中的一个节点, 即共 m 个节点, 每个变量时间维度上的序列特征即为每个节点上的特征, 即每个节点上有 t 个特征. 首先基于“共生变量对”的思想将 MTS 映射至二维张量, 即得到一个批量数据的对称加权共享邻接矩阵 A , $A \in \mathbf{R}^{m \times m}$. 该矩阵可以描述多变量之间的交互关系, 构成了初始化的图结构, 即 $A = (V, E)$. 该图结构由节点和边组成, V, E 分别表示节点和边的集合, 其中节点数量等于 MTS 数据中的变量个数 m , 边数量等于“共生变量对”的数量. 每个节点 $v \in V$ 都对应 MTS 数据中的某个变量, 每条边 $e \in E$ 都对应“共生变量对”之间的边连接, 即表示 $\langle v, u \rangle$ 或 $\langle u, v \rangle$ 之间的连接, $u, v \in V$. 边之间的权重 f 大小为“共生变量对” $\langle u, v \rangle$ 变量值的和, 即如下公式:

$$f(u, v) = \sum_{i=0}^{s-1} \sum_{j=0}^{t-1} T_{i,j,u} + T_{i,j,v} \quad (1)$$

若两个变量没有同时出现在一个批量样本和时间步上, 即没有构成“共生变量对”, 那么图结构中不会为这两个变量之间建立边连接, 即 $\langle u, v \rangle \notin E$, $f(u, v) = 0$.

当 \mathcal{F} 表示单个样本, 即 $s = 1$ 时, 得到的是单个样本 (类别) 的初始图结构, 由于一次仅输入单个样本, 这种方法用于模型训练效率低; 当 \mathcal{F} 表示多个样本, 即 $s > 1$ 时, 构建的图结构 A 是多个样本共用的初始图结构, 由于批量样本输入, 训练模型效率高, 本文采取的方式是批量样本输入.

1.2 基于神经网络学习图结构

基于式 (1) 得到的对称加权邻接矩阵 A 描绘了初始状态下变量之间的边连接关系、权重大小, 是有权重的无向图. 由于在本文数据集上并不存在没有构成“共生变量对”的情况, 即任意两个特征之间都存在边连接, 若特征 (节点) 数量多, 得到的图结构是过于稠密的, 存在噪声大、训练效率慢等问题, 并且考虑变量同时出现不等价于存在相互影响关系的问题, 需要通过线性变换函数与激活函数来进一步学习边之间的连接关系以及权重大小, 得到更适用描述变量之间关系的最终图结构, 同时利用激活函数对边进行自适应裁剪, 即如下公式:

$$A_{h+1} = \sigma(W_{\bar{u}} \cdot A_h + b_{\bar{u}}), A_{h+1} \geq 0 \quad (2)$$

其中 $\sigma(\cdot)$ 表示 ReLU 激活函数, $W_{\bar{u}}$ 和 $b_{\bar{u}}$ 为可学习参数, A_{h+1} 表示迭代第 $h+1$ 次后的邻接矩阵. 当初始图结构为批量输入样本得到的共享图结构时,

训练模型效率高, 它捕获了样本不同类别之间的共性, 进一步被送入 GNN, 根据每个节点上时序特征的差异可聚合有效的特征信息, 进而可用于分类或贝叶斯网络的建模.

另外, 基于扰乱顺序的批量训练样本, 经过神经网络的迭代优化可学习到用于提取多元变量交互关系的权重参数, 此时通过输入批量样本可以基于该参数得到两种图结构. 即输入同一类别的批量样本得到单类别图结构和输入多个不同类别的批量样本得到多类别共享图结构. 前者可用于研究单类别的强相关特征; 后者可用于研究多类别的共同强相关特征, 均可用于特征筛选等操作. 共享图结构所描绘的较重要特征包含了单类别图结构中的部分强相关特征, 下文图结构可视化一节中也证实了这一点.

为了进一步降低噪声影响、提高模型性能、鲁棒性以及改善训练效率, 通过设定一个百分位权重阈值 a 来过滤学习到的特征 (节点) 数量较多的图结构, 达到进一步稀疏化邻接矩阵的目的:

$$A_{i,j} = \begin{cases} A_{i,j}, & A_{i,j} > a \\ 0, & A_{i,j} \leq a \end{cases} \quad (3)$$

经过神经网络迭代优化、裁剪后的矩阵为非对称矩阵, 即有权重的有向图. 式中 i 和 j 分别表示一条边的源节点和目标节点, 若 $A_{i,j}$ 为 0 则视为两个变量之间没有边连接, 忽略相互影响关系, 否则视为存在边连接且有较强相互影响关系, 边上的权重大小即为 $A_{i,j}$ 元素值, 定义该图结构为 $G = (V, E, A)$, 其中 V 表示图结构中的节点, E 表示图结构中的边, A 表示有权重的邻接矩阵.

2 基于图神经网络的特征提取

2.1 TAGCN 图神经网络

现有的基于谱域图卷积的 GNN 在运算时只采用特定大小的卷积核进行特征提取, 易导致信息丢失^[26-27]. TAGCN^[27] 卷积核设定类似于 GoogleNet^[28], 每一个卷积层都有大小不同的卷积核提取特征, 针对时间序列数据可以捕获更多的时间序列特征, 因此在 MTS 的背景下, 提取出的特征分类效果相对于前文提到的 GraphSage, GIN 和 GAT 具有一定的优势, 在后续实验中也将对 TAGCN 与它们的性能.

2.2 TAGCN 特征提取原理

对于 TAGCN 的第 n 个隐藏层, 输入为图结构

G 及相应节点特征数据或第 $n-1$ 个隐藏层的输出, 作为第 n 个隐藏层输入的所有节点上的第 k 个特征数据可定义为 $x_k^n \in \mathbf{R}^{m_n}$, $k=1, 2, \dots, t$. 特征向量 x_k^n 的具体组成由所得图结构 G 中的节点索引确定. 定义 $G_{k,f}^n \in \mathbf{R}^{m_n \times m_n}$ 为第 f 个图卷积核, 图卷积本质上是矩阵向量的乘积运算, 例如 $G_{k,f}^n x_k^n$ 则表示一次图卷积.

TAGCN 图卷积运算时首先将邻接矩阵归一化, 使其特征值均在单位圆内以此保证卷积运算的稳定性, 进一步基于图卷积核的平移不变性原理^[27] 可得:

$$G_{k,f}^n = \sum_{c=0}^C g_{k,f,c}^n \bar{A}^c \quad (4)$$

其中 $g_{k,f,c}^n$ 为多项式图卷积核系数, C 为超参数, 表示 C 个不同尺寸的卷积核, \bar{A}^c 为归一化后的邻接矩阵. 进一步, C 个不同尺寸的图卷积核在图结构数据上提取不同尺度的特征, 并进行线性组合, 即得第 n 个隐藏层输出的特征映射:

$$y_f^n = \sum_{k=1}^{t_n} G_{k,f}^n x_k^n + b_f^n \mathbf{1}_{m_n} \quad (5)$$

其中 b_f^n 为可学习参数, $\mathbf{1}_{m_n}$ 是维度为 m 且均为 1 的向量, $G_{k,f}^n x_k^n$ 表示在具有任意拓扑结构图上的一次有效图卷积操作.

最后经过激活函数与线性变换得最终特征, 定义为 V . 进一步在堆叠的图神经网络之间以及最后的线性层之后加上“Dropout”层^[29] 以抑制过拟合, 增强模型鲁棒性. 最后将批量提取的特征送入分类层得到交叉熵损失:

$$\mathcal{L}_{CE} = \frac{1}{\mathcal{F}} \sum_{\tilde{s}} - \sum_{u=1}^U y_{\tilde{s}u} \ln(p_{\tilde{s}u}) \quad (6)$$

其中 CE 表示 Cross Entropy (交叉熵), \mathcal{F} 为批量样本个数, U 为单个样本类别总数, $y_{\tilde{s}u}$ 为指示变量. 若第 \tilde{s} 个样本的预测标签与真实标签相同, 则 $y_{\tilde{s}u} = 1$ 否则 $y_{\tilde{s}u} = 0$. $p_{\tilde{s}u}$ 表示第 \tilde{s} 个样本属于第 u 个类别的概率. 图结构学习与特征提取器通过 Adam^[30] 进行优化. 整个算法流程如图 1 所示.

3 基于贝叶斯神经网络的不确定性建模

本节基于变分推断求解了贝叶斯后验, 为模型提供了不确定性, 并且第一次提出考虑了样本内部类别相对不确定性的“相对方差”不确定性衡量方式. 另外, 本文没有直接将贝叶斯加在 TAGCN 上, 而是利用其提取出来的特征进行贝叶斯建模, 第一个好处是计算更加简便, 因为受时序特征维度的影响, TAGCN 的参数量会有比较大的波动, 时序特征维度很大时求解贝叶斯网络会更加复杂; 第二个好处是框架的解耦性好, 因为当我们直接替换特征提取模型时仍可以方便地赋予模型不确定性和可解释性.

3.1 贝叶斯神经网络

贝叶斯神经网络 (BNN) 由于学习到的是参数的分布, 可以基于此获得预测结果的置信区间以及对模型输出的不确定性建模, 从而捕获不确定性样本. 同时, 权重参数不是固定参数, 因此可抑制过拟合. 相对于深度学习这种黑盒模型, 贝叶斯统计具有很好的鲁棒性和可解释性^[24-25, 31].

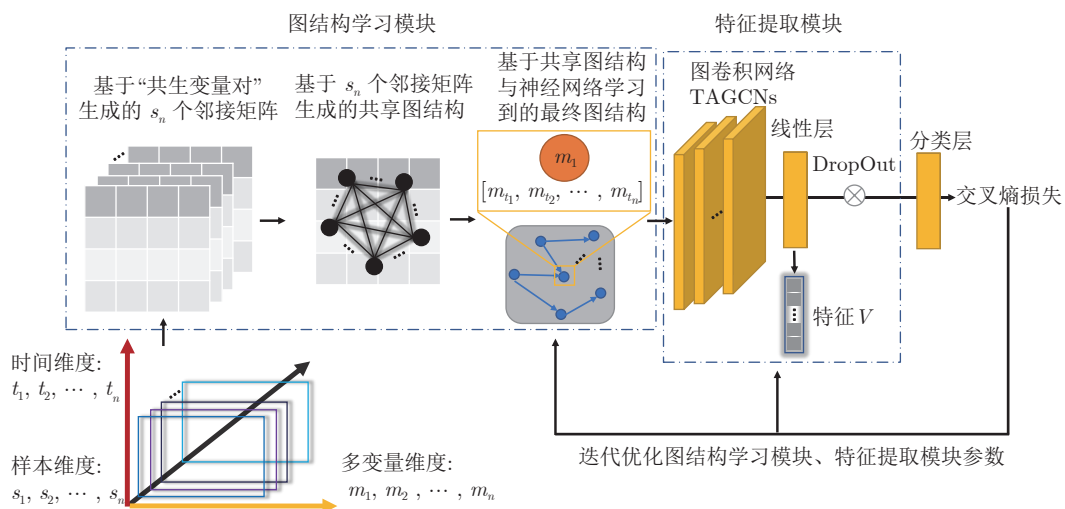


图 1 MTS 数据特征提取模型架构与优化流程

Fig.1 Feature extraction model architecture and optimization process of MTS data

MTS 数据经特征提取模型可得初始的特征向量 $V = (v_1, v_2, v_3, \dots, v_s)$. BNN 可以通过权重参数的后验分布来评估模型的不确定性. 根据贝叶斯法则可得最终的预测分布:

$$P(y_*|V_*, V) = \int P(y_*|V_*, \theta) P(\theta|V) d\theta \quad (7)$$

其中 θ 为 BNN 的权重参数, $P(\theta)$ 是参数的先验分布.

$$P(\theta|V) = \frac{P(V|\theta) P(\theta)}{P(V)} \quad (8)$$

其中 $P(\theta|V)$ 是后验分布, $P(V|\theta)$ 是似然函数, $P(V)$ 为边缘似然. 假设 θ 具有 $\{\theta_1, \theta_2, \theta_3, \dots, \theta_n\}$ 个参数, 则

$$P(V) = \int_{\theta_1} \dots \int_{\theta_n} P(V, \theta) d\theta_1 \dots d\theta_n \quad (9)$$

$P(V)$ 是一个难以求解的多重积分, 从而导致后验分布难以获得解析解, 因此也无法求得最终的预测分布.

3.2 随机变分推断

随机变分推断, 也称为 SVI (Stochastic variational inference), 可以将复杂的贝叶斯模型应用于大规模的数据集, 该方法已经在很多实践中被证明能够有效地解决各种学习和推理问题^[32]. 因此本文采用 SVI 来求解贝叶斯网络的后验分布. 假设后验分布 $P(\theta|V)$ 用一个简单的变分分布 $q(\theta|\phi)$ 来近似. 通常采用 KL 散度来衡量两个分布的近似程度^[32], 因此可构造如下优化问题:

$$\min_{\phi} KL(q(\theta|\phi) || P(\theta|V)) \quad (10)$$

通过求解使得两个分布距离最小的变分分布参数 ϕ 来得到近似的后验分布, 而后验分布未知, 因此将其转换为证据下界, 又称为 ELBO (Evidence lower bound) 优化问题来处理^[32], 基于上述公式进一步推导得到:

$$\begin{aligned} KL(q(\theta|\phi) || P(\theta|V)) &= E_{q(\theta|\phi)} \ln q(\theta|\phi) - \\ &E_{q(\theta|\phi)} \ln P(\theta, V) + \ln P(V) - \\ &ELBO(q) + \ln P(V) \end{aligned} \quad (11)$$

进一步可得:

$$\begin{aligned} \ln P(V) &= KL(q(\theta|\phi) || P(\theta|V)) + \\ ELBO(q) &\geq ELBO(q) \end{aligned} \quad (12)$$

由于 $KL(q(\theta|\phi) || P(\theta|V)) \geq 0$, 且 $\ln P(V)$ 是常数, 因此最小化 KL 散度的优化问题等价于最大化 $ELBO(q)$, 因此原先的优化目标函数由 KL 散度替换为 ELBO. 联合贝叶斯法则和式 (11) 推导过程还可得:

$$\begin{aligned} ELBO(q) &= E_{q(\theta|\phi)} [\ln P(V|\theta)] - \\ &KL(q(\theta|\phi) || P(\theta)) \end{aligned} \quad (13)$$

式中 $E_{q(\theta|\phi)} [\ln P(V|\theta)]$ 即为基于变分后验分布的重建似然函数; $KL(q(\theta|\phi) || P(\theta))$ 为变分分布与先验分布的 KL 散度, 其中变分分布 $q(\theta|\phi)$ 是优化目标. 联合式 (12) 和 (13) 即可得最终目标等式 (ELBO):

$$\begin{aligned} \ln P(V) &\geq E_{q(\theta|\phi)} [\ln P(V|\theta)] - \\ &KL(q(\theta|\phi) || P(\theta)) \end{aligned} \quad (14)$$

式 (14) 可通过任意的随机梯度下降法进行优化, 因为似然函数 $P(V|\theta)$ 可以基于 MTS 数据实例分解得到. 即为:

$$P(V|\theta) = \prod_{i=1}^p P(V_i|\theta_i) \quad (15)$$

此时可得最终目标函数:

$$\mathcal{L} = \mathcal{L}_{CE} + \eta \sum_{V \in \mathcal{F}} (-ELBO_V) \quad (16)$$

其中 $\sum_{V \in \mathcal{F}} ELBO_V$ 为基于一个批量样本及其特征 V 与蒙特卡洛采样法估计得到的 ELBO 值, 取负是为了将最大化 ELBO 的问题转换为最小化负的 ELBO, 以便与最小化交叉熵损失共同优化; η 是为了平衡交叉熵损失与 ELBO 的超参数.

同时为了高效地求解 ELBO 的优化问题, 基于 MFVI (Mean field VI) 假设来处理参数的优化, 即假设变分后验分布是完全可分解的分布^[33]. 定义 $q(\theta|\phi) = N(\mu, \sigma^2)$, $P(\theta) = N(0, 1)$, 其中 $\phi = \{\mu, \sigma^2\}$. 通过蒙特卡洛估计法即可近似估计 $E_{q(\theta|\phi)} [\ln P(V|\theta)]$, 而式 (14) 中的 KL 散度是基于两个高斯分布定义的, 因此可获得解析解, 基于重参数采样法即可不断迭代优化变分分布的参数, 得到目标后验分布 $P(\theta|V)$.

3.3 预测结果的不确定性估计

给定基于测试集 MTS 数据提取的特征 V , 最终的目标是得到预测分布 $P(y_*|V_*, V)$:

$$P(y_*|V_*, V) = \frac{1}{K} \sum_{k=1}^K P(y_*|V_*, \theta_*) \quad (17)$$

其中, $\theta_* \sim q(\theta|\phi)$. 最后基于输入特征 V 采用蒙特卡洛采样法即可计算得预测分布的均值 $E_{p(y_*)}[y_*]$ 和方差 $\text{Var}_{p(y_*)}[y_*]$

该方差 Var 为每个样本对于即将预测的所有类别的方差. 对所得方差数据取以 e 为底的对数, 压缩变量尺度, 使得数据更加稳定, 削弱数据的异方差性. 所得每个类别的方差如下所示, s 代表样本

维度, l 代表样本对应标签的维度.

$$\text{Var}_s = [\ln(\text{var}_{s_1}), \ln(\text{var}_{s_2}), \dots, \ln(\text{var}_{s_l})] \quad (18)$$

式中 $\ln(\text{var}_{s_l})$ 指的是对第 s 个样本的第 l 个类别的方差取 e 为底的对数.

进一步可得每个样本的不确定得分. 本文提出了基于每个样本所有类别的相对方差来评估样本的不确定性的概念. 若某个样本所有类别的最高方差和最低方差之差的值较大, 则认为该样本的不确定性较大, 可以筛选出来留给人类专家评判, 同时降低误判率, 提高预测效果. 不确定性得分计算方式如下:

$$u_score_r^s = (\text{Var}_s)_{\max} - (\text{Var}_s)_{\min} \quad (19)$$

式中, Var_s 是 $1 \times l$ 维的向量, l 表示该样本的类别总数, $(\text{Var}_s)_{\max}$ 、 $(\text{Var}_s)_{\min}$ 分别表示第 s 个样本所有类别中方差最大者和最小者, 该式表示第 s 个样本的不确定得分.

另外, 基于式 (18) 还可以得到别的不确定得分的衡量方式, 如基于“最大方差”的思想, 将当前样本每个类别的方差直接累加和作为该样本的不确定性得分, 定义其为 $u_score_r^s$; 还可以通过“最大方差”与“相对方差”简单加权结合的衡量方式, 如 $u_score_{rr}^s = \lambda \times u_score_r^s + u_score_r^s$, λ 用于控制最大方差的影响程度, 但是我们发现基于单独“最大方差”或当前简单加权组合融入了“最大方差”的衡量方式效果都不理想, 由于篇幅原因, 我们在下文展示了部分对比实验结果. 本文最终使用所提出的基于“相对方差” $u_score_r^s$ 的不确定性得分衡量方式, 具体的实验结果将在下文展示. 整个算法流程如算法 1 所示. 整个随机变分推断流程如图 2 所示.

算法 1. VBSGNN 方法

1) **Input:** 数据集 $D = \{T_i, Y_i\}_{i=1}^s$
VBSGNN 模型参数 $\{\bar{w}, \bar{b}, \theta, \phi\}$

2) **for** $i = 1$ **to** epochs **do**

3) 取一个批量 MTS 样本 $\mathcal{F} \subset T$

4) 提取得到特征 $V = [v_1, v_2, v_3, \dots, v_n]$

5) 通过 Y_i 、 V 、全连接分类层、式 (6) 计算交叉熵损失 \mathcal{L}_{CE}

6) 通过随机变分推断 (SVI) 训练变分 BNN 模型

7) **While** the $ELBO$ not converged **do**

8) 从 $q(\theta|\phi)$ 生成 K 个样本 $\bar{\phi}$

9) **for** 每条特征 $V \in \mathcal{F}$ **do**

10) 从 $P(\theta)$ 生成 K 个样本 θ_V

11) 基于式 (14) 及蒙特卡洛法采样的样本 $(\bar{\phi}, \theta_V)$ 估计 $ELBO_V$ 的值

12) **end**

13) $ELBO = \sum_{V \in \mathcal{F}} ELBO_V$

14) 计算总损失 $\mathcal{L} = \mathcal{L}_{CE} + \eta(-ELBO)$

基于 \mathcal{L} 进行反向传播, 通过随机梯度下降法更新 VBSGNN

参数 $\{\bar{w}, \bar{b}, \theta, \phi\}$

15) **end**

16) **end**

17) **Output:** 参数 $\{\bar{w}^*, \bar{b}^*, \theta^*, \phi^*\}$

4 仿真与结果分析

4.1 实验设置

数据集. 为了验证所提算法的有效性, 选取了 10 个公开的静态多元时间序列分类数据集¹ 进行仿真分析, 所选数据集的概要如表 1 所示. AtrialFibrillation 中多元变量分别指两个通道的心房颤动心电图信号, 预测任务是判别当前心房颤动属于终止性心房颤动还是非终止性心房颤动. FingerMovements 中多元变量分别指 28 个 500 毫秒长的脑电图通道, 预测任务是根据脑电图信号判别当前即将

¹ 数据获取地址: <http://timeseriesclassification.com>

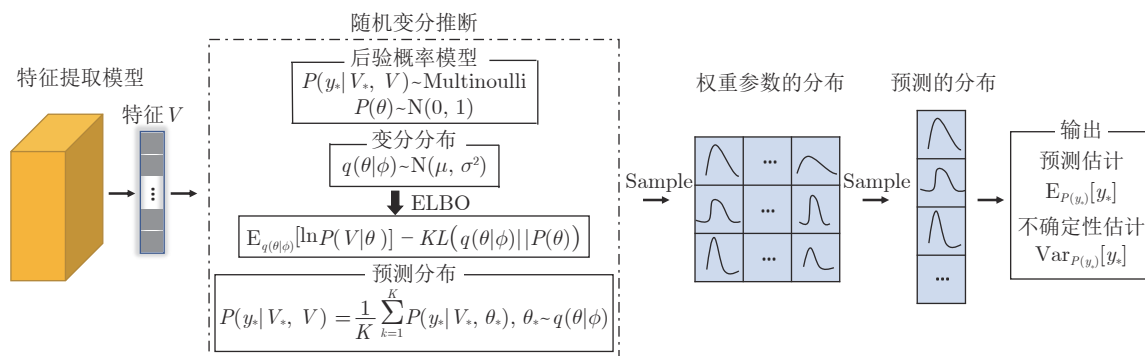


图 2 随机变分推断流程

Fig.2 The process of stochastic variational inference

表 1 实验中使用的 10 个数据集概要
Table 1 Summary of the 10 UEA datasets used in experimentation

	名称	训练集大小	测试集大小	多变量维度	时间维度	类别个数
AF	AtrialFibrillation	15	15	2	640	3
FM	FingerMovements	316	100	28	50	2
HMD	HandMovementDirection	160	74	10	400	4
HB	Heartbeat	204	205	61	405	2
LIB	Libras	180	180	2	45	15
MI	MotorImagery	278	100	64	3000	2
NATO	NATOPS	180	180	24	51	6
PD	PenDigits	7494	3498	2	8	10
SRS2	SelfRegulationSCP2	200	180	7	1152	2
SWJ	StandWalkJump	12	15	4	2500	3

进行左手动作还是右手动作. HandMovementDirection 中多元变量分别指 10 个脑电图通道, 预测任务是根据脑电图信号判别当前手腕的可能运动方向. Heartbeat 中多元变量分别指频谱图中 61 个不同频带, 预测任务是判别当前心音记录者属于健康人或心脏病患者. Libras 中多元变量分别指经过手部运动视频前处理得到的手部质心位置的 X 和 Y 坐标, 预测任务是基于时间序列 X 和 Y 坐标判别当前手部动作的类别. MotorImagery 中多元变量分别指 64 个脑皮层电图信号, 预测任务是判别大脑所想象的运动类别是左手小指运动还是舌头部位运动. NATOPS 中多元变量分别指由传感器记录得到的手部、手腕、肘部、拇指的 X 、 Y 、 Z 坐标, 每个部位又分左、右两种 (如左手腕和右手腕), 预测任务是判别当前 24 个时序坐标所描述的动作类别. PenDigits 中多元变量分别指由笔在数字屏幕上写数字 0~9 而接触生成的 X 和 Y 坐标, 预测任务是判别当前手写数字的类别. SelfRegulationSCP2 中多元变量分别指 7 个脑电图通道, 预测任务是判别慢皮层电位的变化类别, 即正向或负向变化. StandWalkJump 中多元变量分别指频谱图中的 4 个不同频带, 预测任务是判别进行的活动类别, 即站立、行走或跳跃.

模型参数与实验细节. 根据各自数据集的时间序列长度的不同, TAGCN 隐藏层参数选取范围为 {32, 64, 128, 256, 512, 1024}, 全连接层隐藏层参数选取范围为 {16, 32, 64, 128, 256, 512}. 模型优化器为 Adam, 特征提取部分模型学习率选取范围为 $\{1 \times 10^{-5}, 1 \times 10^{-4}, 1 \times 10^{-3}\}$, 变分贝叶斯部分模型学习率选取范围为 $\{1 \times 10^{-4}, 1 \times 10^{-3}, 6 \times 10^{-3}, 1 \times 10^{-2}\}$. 损失函数中的超参数 η 设定为 2×10^{-5} . 每个数据集的实验结果都是在当前数据

集上同样参数下独立运行了 5 次后取得的最佳结果. 所有实验都是基于 Python3.7、Pytorch1.7.1 进行的, 并且均在相同的软件环境下实现, 即 Intel (R) Core (TM) CPU I7-10700 @ 2.90 GHz 16 核.

对比算法. 基于公开的多元时间序列数据集与已经提出的 7 类多元时间序列分类算法进行了比较, 所有实验结果均从测试集上取得. MTSC 算法中 ED 和 ED(norm)^[4] 为基于欧氏距离的 MTS 最近邻分类器, 按数据是否标准划分为两种. DTW_I, DTW_I(norm)^[4] 为基于与维度无关的动态时间弯曲距离 MTS 最近邻分类器, 按数据是否标准划分为两种. DTW_D, DTW_D(norm)^[34] 为基于与维度相关的动态时间弯曲距离 MTS 最近邻分类器, 是 DTW_I 的变体, 按数据是否标准化分为两种. WEASEL+MUSE^[4] 是基于模式袋的最新 MTS 分类算法. HIVE-COTE^[35] 是一种基于分层投票系统的 MTS 分类算法. MLSTM-FCN^[10] 是最新通用的多元时间序列分类的深度学习框架. 该模型由一个长短期记忆网络层、堆叠的卷积神经网络层以及一个压缩和激发模块组成, 以产生潜在的特征用于分类. TapNet^[4] 结合了传统和深度学习方法的优点, 是一种基于长短期记忆网络层和堆叠卷积神经网络层的注意力原型网络. MTPool^[17] 是一种基于图池化方法的 MTS 分类算法, 根据不同的池化方法, 又分为 MTPool-M, MTPool-D, MTPool-S, MTPool-One, MTPool-Corr 几种变体算法. 本文提出的特征提取模块算法, 即共享图神经网络 SGNN (Shared graph neural network), 未加变分贝叶斯推断, 包括基于 GraphSage^[19] 的 SGNN-S, 基于 GIN^[20] 的 SGNN-I, 基于 GAT^[21] 的 SGNN-A, 基于 TAGCN^[28] 的 SGNN-T. 变分贝叶斯共享图神经网络, 即 VBSGNN 是在 SGNN-T 的基础上加了变分贝叶斯推断的算法.

4.2 分类性能评价

表 2 列出了不同 MTS 分类算法的准确率对比结果, 将每个数据集的最佳准确率进行了黑体加粗显示, “Wins” 一列表示基于当前算法表现优于或等于其他算法最佳准确率的数据集个数. 本文构建的图结构搭配四种先进的图神经网络算法的结果如表 2 所示, 观察 SGNN-S、SGNN-I、SGNN-A 的结果可以发现, 由于数据集以及不同图网络卷积方式的差异, 不同的图卷积算法表现出的性能也有所差异, 例如 SGNN-S 在 FM 数据集上表现较好, 而在 HMD 数据集上效果一般, 三种算法的差异不大, 但是各个算法的效果都要好于表 2 中其他大部分先进的算法. 说明本文所构建的图结构对于不同的图卷积算法的适应性比较好. 而 SGNN-T 的效果最好, 这是因为 MTS 数据的特性很大程度上由时间序列决定, 而 TAGCN 可以通过不同感受野的卷积核来提取不同时间序列长度的特征, 不同感受野的特性使得 TAGCN 可以较好的处理时间序列数据, 提取表征能力更强的特征, 因而最终分类效果好.

总体而言, 本文提出的算法具有较好的分类效果. 从表 2 可以发现, 未加变分贝叶斯推断的 SGNN

系列算法整体表现优于其他算法, 说明所提的图结构学习以及特征提取方法是有效的. 在不同的数据集和算法实验中发现, 变分贝叶斯的效果严重依赖于特征提取器的效果, 因此加了变分推断后的 SGNN-S、SGNN-I、SGNN-A 的效果略好于未加之前且各自效果差异不大, 但由于在 MTS 数据背景下原始图卷积方式提取的特征效果没有 TAGCN 好, 因此效果都差于加了变分推断的 SGNN-T. 并且受篇幅限制, 本文仅展示基于加了变分推断的 SGNN-T (即 VBSGNN) 进行的分类、不确定性评估等实验结果. VBSGNN 预测过程引入了不确定性并可以抑制过拟合、提高模型鲁棒性, 在较难数据集 (其他算法最高准确率低于 0.8 的数据集) 上的效果都有了进一步的提升. 虽然在基础准确率较高的数据集如 LIB、NATO、PD 上, 从准确率角度来看表现有所下降, 但也要优于其他基准算法, 同时变分贝叶斯推断带来的模型可解释性如样本不确定性度量等是非常有意义的.

4.3 VBSGNN 预测分布的方差可视化

基于 LIB 数据集, 分别使用加了变分贝叶斯的 VBSGNN 和未加贝叶斯的 SGNN-T 通过蒙特

表 2 在 10 个公开数据集上的不同算法准确率对比
Table 2 Accuracy of different algorithms on 10 public datasets are compared

算法	数据集										
	AF	FM	HMD	HB	LIB	MI	NATO	PD	SRS2	SWJ	Wins
ED	0.267	0.519	0.279	0.620	0.833	0.510	0.850	0.973	0.483	0.333	0
DTW _I	0.267	0.513	0.297	0.659	0.894	0.390	0.850	0.939	0.533	0.200	0
DTW _D	0.267	0.529	0.231	0.717	0.872	0.500	0.883	0.977	0.539	0.200	0
ED(norm)	0.200	0.510	0.278	0.619	0.833	0.510	0.850	0.973	0.483	0.333	0
DTW _I (norm)	0.267	0.520	0.297	0.658	0.894	0.390	0.850	0.939	0.533	0.200	0
DTW _D (norm)	0.267	0.530	0.231	0.717	0.870	0.500	0.883	0.977	0.539	0.200	0
WEASEL+MUSE	0.400	0.550	0.365	0.727	0.894	0.500	0.870	0.948	0.460	0.267	0
HIVE-COTE	0.133	0.550	0.446	0.722	0.900	0.610	0.889	0.934	0.461	0.333	1
MLSTM-FCN	0.333	0.580	0.527	0.663	0.850	0.510	0.900	0.978	0.472	0.400	0
TapNet	0.333	0.470	0.338	0.751	0.878	0.590	0.939	0.980	0.550	0.133	0
MTPool-M	0.533	0.504	0.486	0.742	0.828	0.560	0.928	0.978	0.550	0.533	0
MTPool-D	0.400	0.530	0.459	0.737	0.811	0.600	0.944	0.977	0.550	0.533	0
MTPool-S	0.400	0.590	0.473	0.722	0.811	0.540	0.889	0.983	0.539	0.667	0
MTPool-One	0.400	0.570	0.405	0.717	0.833	0.540	0.889	0.970	0.539	0.600	0
MTPool-Corr	0.400	0.590	0.419	0.722	0.828	0.560	0.904	0.973	0.550	0.600	0
MTPool	0.467	0.620	0.432	0.742	0.861	0.630	0.904	0.983	0.600	0.667	0
SGNN-S	0.600	0.650	0.541	0.741	0.889	0.600	0.961	0.984	0.589	0.600	2
SGNN-I	0.533	0.550	0.514	0.741	0.883	0.640	0.933	0.974	0.572	0.600	1
SGNN-A	0.533	0.560	0.500	0.751	0.878	0.560	0.961	0.980	0.550	0.600	0
SGNN-T	0.600	0.640	0.608	0.756	0.889	0.630	0.978	0.985	0.600	0.733	7
VBSGNN	0.667	0.680	0.622	0.776	0.872	0.680	0.972	0.984	0.622	0.733	9

卡洛采样法生成预测分布, 得到样本各类别方差, 并基于雷达图进行了可视化, 如图 3 所示. 由图 3(a) 可以发现 SGNN-T 对于样本各类别不确定性建模效果很差, 各类别不确定性几乎均处于同一水平, 难以基于此筛选出不确定样本; 而图 3(b) VBSGNN 得到的样本各类别的方差波动性大, 说明对于每个样本的不同类别不确定性都进行了建模, 为基于不确定性得分进行样本筛选来改善预测以及基于方差进行模型的不确定性估计奠定了基础.

4.4 利用不确定得分改善预测效果

从表 1 可以发现个别数据集中存在样本不平衡情况, 为了考虑这种不平衡现象以及从多指标角度充分评估不确定性得分计算方式的有效性, 引入除了准确率 (Accuracy) 之外的其他几个指标:

$$F1\ Score = \frac{2 \times (prec \times rec)}{prec + rec} \quad (20)$$

其中, $prec$ 为精确率 (Precision), rec 为召回率 (Recall).

$$Cohen's\ Kappa = \frac{\rho_o - \rho_e}{1 - \rho_e} \quad (21)$$

式 (21) 可以用来有效衡量模型对二分类或多分类的性能, 对于不平衡数据集更加适用. ρ_o 代表观察到的符合的比例, ρ_e 是由于随机性产生的符合比例. $Cohen's\ Kappa$ 值越大, 代表模型性能越好.

$$PRAUC = \sum_{k=1}^n prec(k) \Delta rec(k) \quad (22)$$

$PRAUC$ 表示精确率、召回率曲线下方的面积,

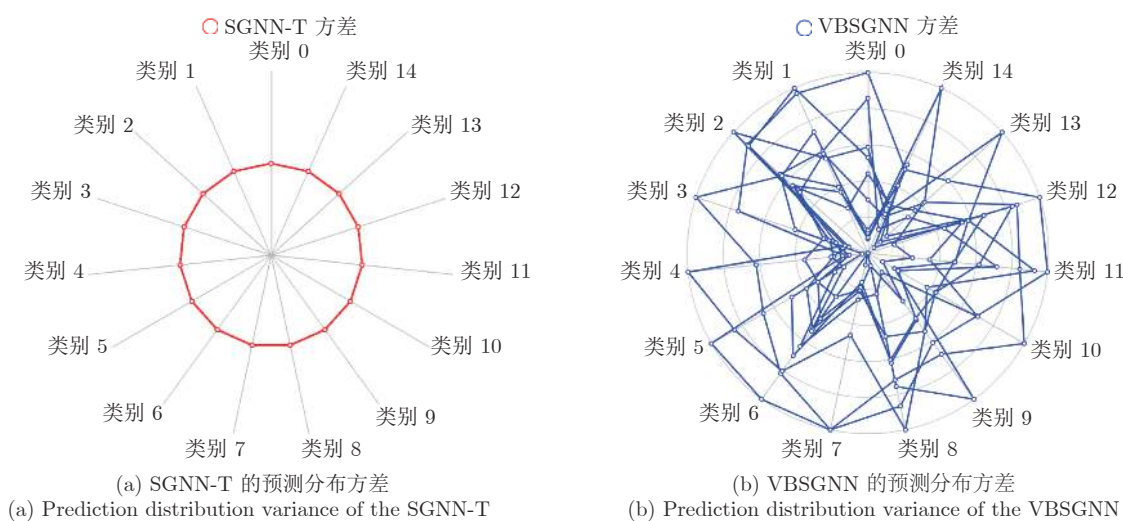


图 3 SGNN-T 和 VBSGNN 的预测分布方差对比

Fig. 3 Variance comparison of prediction distribution between SGNN-T and VBSGNN

常用于评价模型的性能.

根据表 2 数据集实验结果, 以准确率指标衡量数据集的难预测程度, 据此选取了 7 个较难分类的数据集 (准确率低于 0.8), 基于 VBSGNN 和不确定得分剔除得分较高的样本再进行分类预测, 并比较不剔除数据以及剔除 10%、20%、30% 数据量下的各项评估指标差异. 基于上文第 3.3 节提出的 $u_score_{\tau}^s$, $u_score_{\tau\tau}^s$ 和 $u_score_{\tau}^s$ 计算不确定得分进行实验, 发现前两个衡量方式效果并不理想, 由于篇幅原因, 我们仅在此放上基于 $u_score_{\tau}^s$ 和 AF、HMD、HB 数据集的实验结果. 选择这三个数据集是因为它们比较有代表性, 测试集大小 (15、74、205) 跨度大, 分类种类 (3、4、2) 也有一定跨度.

整体的实验结果如图 4 所示, 基于 $u_score_{\tau}^s$ 和 $u_score_{\tau\tau}^s$ 的两组实验的参数设置与环境都一样, 不同的仅仅是不确定得分的计算方式. 可以发现在小数据集 (AF) 上区分不是很明显. 而在 HB、HMD 数据集上区分显著, 对于 $u_score_{\tau}^s$, 以未剔除样本的各指标值为参照, 剔除 10%、20%、30% 后的各指标值几乎都明显处于剔除 0% 数据的各指标值下方, 仅个别指标略微在其上方; 而基于“相对方差” $u_score_{\tau}^s$ 的各剔除量下的各指标值几乎都明显在剔除 0% 的各指标值上方. 这说明随着数据集增大, 基于“最大方差” $u_score_{\tau}^s$ 的不确定得分衡量方式对于分类预测的改善效果变得很差, 远低于基于“相对方差”的改善效果, 也说明基于“最大方差”样本筛选存在过多误筛选情况, 即剔除了许多模型可以分类正确的样本, 因此所提基于“相对方差”的不确定性衡量是相对更有效的. 对于“最大方差”表现不好的原因之一可能是基于最大方差求出来的不确定

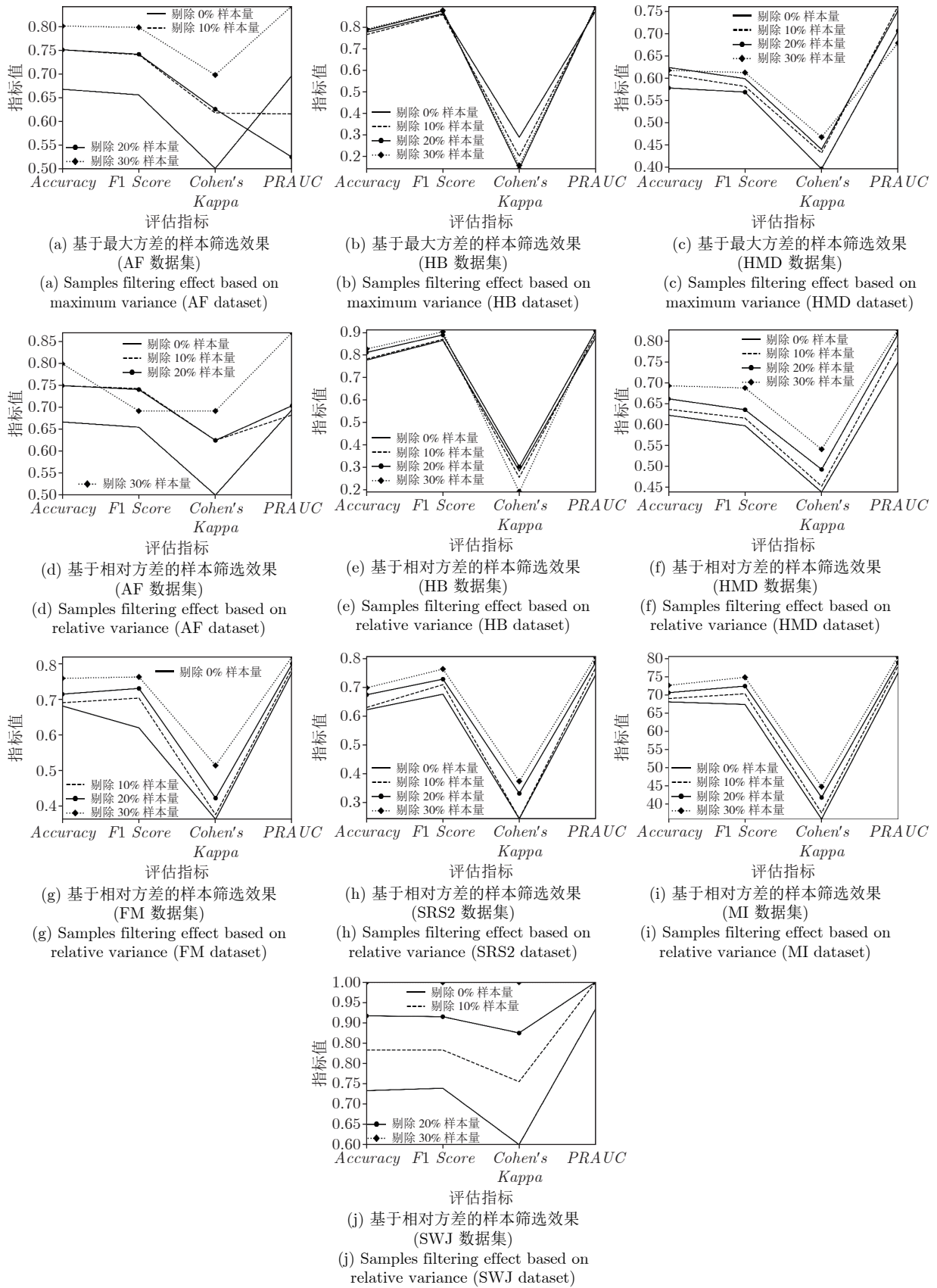


图 4 基于 VBSGNN 不确定得分改善预测效果评估

Fig. 4 Evaluation of improving prediction effect based on VBSGNN uncertainty score

得分引入了过多的噪声, 因此衡量效果不好. 实际应用时可以将不确定得分高的样本取出留给人类专家评判, 降低模型误判率, 提高模型效益.

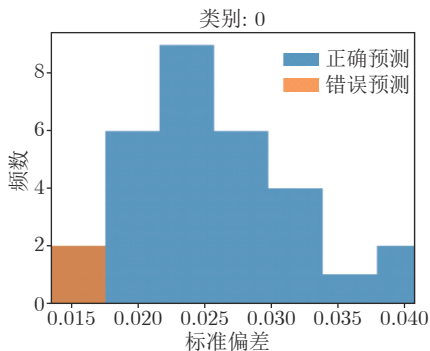
4.5 利用不确定得分改善预测效果

基于式 (17) 可输出每个样本的预测分布 $P(y_*|V_*, V)$, 从而求得相应分布的均值、标准差, 可增加模型的可解释性, 基于 NATO 数据集进行了模型的不确定性估计. 如图 5 所示, 图中不确定范围即预测分布均值的置信区间, 通过均值加减两倍标准差得到, 表明所属类别预测概率值可能的波动范围. 基于对模型的不确定性估计, 可以研究每个类别预测正确以及预测不正确的样本标准偏差情况, 进而判断模型对于特定样本的预测情况. 例如图 5 中类别 0 与类别 2 的预测分布直方图中, 对于标准偏差较小且预测错误的样本, 说明模型对于该样本形成了较为稳定的“错误预测”, 意味着该样本可能是需要剔除的“问题样本”或者需要让人类专家人工评判. 结合对预测均值的不确定估计, 可以衡量模型对于样本中某个类别的预测质量. 例如类别 5 对应的 30 个样本, 模型都进行了准确预测, 且预测均值的不确定性范围也较小, 说明模型对该类样本有着可靠的预测性能. 类别 3 虽然不确定性范

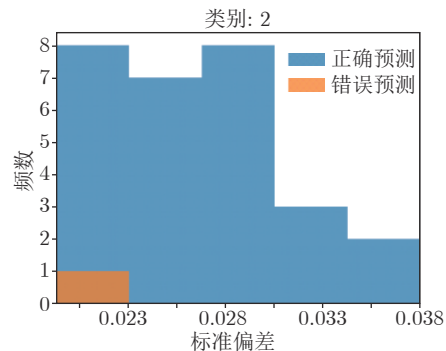
围较大, 但是模型同样都给出了准确预测, 说明模型的鲁棒性较好, 对于该类样本的预测效果也是可靠的.

4.6 图结构可视化

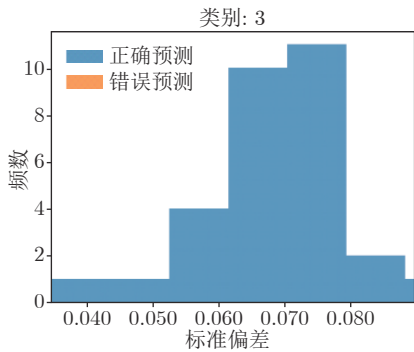
该实验在 NATO 数据集上进行, 多变量名称如表 3 所示. 由上文 1.2 节可知, VBSGNN 可以学习到两种图结构, 一种是单类别图结构, 一种是多类别共享图结构. 实验中发现, 通过设定邻接矩阵的权重阈值对边进行过滤使得图结构变得稀疏, 有利于抑制噪声、过拟合现象, 改善了预测性能; 同时由于稀疏性, 提高了模型的鲁棒性以及训练速度. 图结构中每个节点对应一个变量, 神经网络每学习到一对边结构的同时会得到一个目标节点的权重, 同一个目标节点的权重值会被不断累加, 节点越大则表示该变量被赋予的权重越大, 围绕该变量建立的边结构越稠密, 影响范围越大, 该变量对于最终预测任务也越重要, 如图 6 所示. 图 7(a) 为基于式 (3) 过滤了 70% 边数量的单类别图结构, 所描述的是类别为“折叠手臂”姿势的多变量之间的图结构关系, 对于“折叠手臂”类别的样本, 与之强相关的变量有左手尖 Z 坐标、右拇指 X 坐标、右肘部 Z 坐标、左肘部 Z 坐标、左肘部 X 坐标、右拇指 Y 坐标、



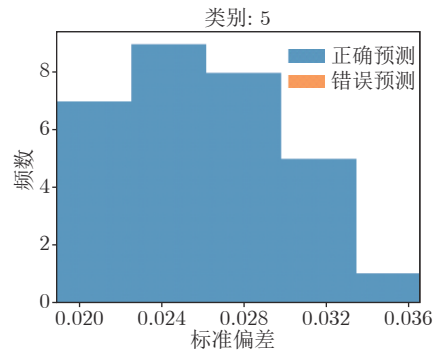
(a) NATO 数据集所有样本中类别 0 的预测标准偏差
(a) Predicted standard deviation of category 0 in all samples of NATO dataset



(b) NATO 数据集所有样本中类别 2 的预测标准偏差
(b) Predicted standard deviation of category 2 in all samples of NATO dataset



(c) NATO 数据集所有样本中类别 3 的预测标准偏差
(c) Predicted standard deviation of category 3 in all samples of NATO dataset



(d) NATO 数据集所有样本中类别 5 的预测标准偏差
(d) Predicted standard deviation of category 5 in all samples of NATO dataset

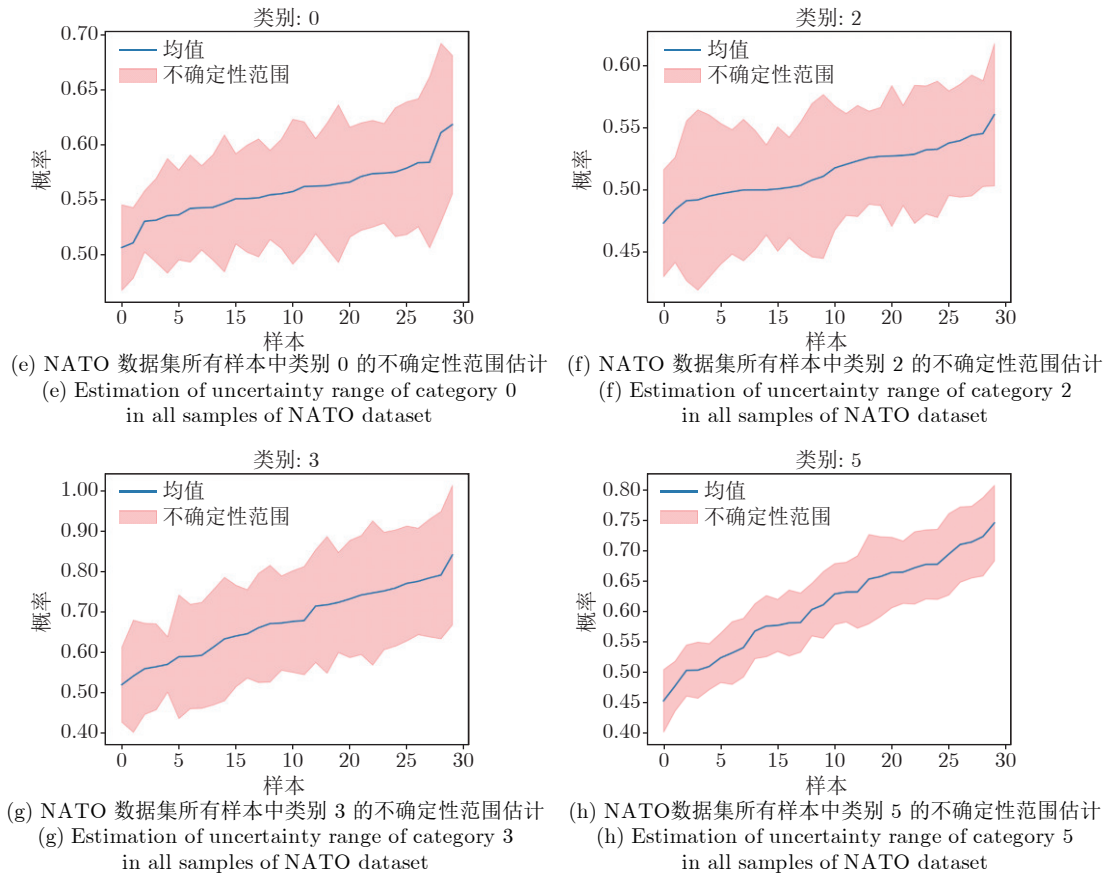


图 5 基于 VBSGNN 的模型不确定性估计 (NATO 数据集)

Fig.5 Model uncertainty estimation based on the VBSGNN (NATO dataset)

表 3 NATO 图结构中 24 个节点对应的变量名称

Table 3 Corresponding variable names of 24 nodes in graph structure based on NATO dataset

手部传感器变量	肘部传感器变量	手腕传感器变量	拇指传感器变量
节点 0: 左手尖 X 坐标	节点 6: 左肘部 X 坐标	节点 12: 左手腕 X 坐标	节点 18: 左拇指 X 坐标
节点 1: 左手尖 Y 坐标	节点 7: 左肘部 Y 坐标	节点 13: 左手腕 Y 坐标	节点 19: 左拇指 Y 坐标
节点 2: 左手尖 Z 坐标	节点 8: 左肘部 Z 坐标	节点 14: 左手腕 Z 坐标	节点 20: 左拇指 Z 坐标
节点 3: 右手尖 X 坐标	节点 9: 右肘部 X 坐标	节点 15: 右手腕 X 坐标	节点 21: 右拇指 X 坐标
节点 4: 右手尖 Y 坐标	节点 10: 右肘部 Y 坐标	节点 16: 右手腕 Y 坐标	节点 22: 右拇指 Y 坐标
节点 5: 右手尖 Z 坐标	节点 11: 右肘部 Z 坐标	节点 17: 右手腕 Z 坐标	节点 23: 右拇指 Z 坐标

右肘部 X 坐标, 说明这几个变量对于区分是否是“折叠手臂”姿势具有较好的效果。

图 7(b) 所示为基于批量数据学到的图结构, 该图结构描述了对于整体 6 个手势类别而言, 利于最终分类预测的强相关变量, 从该共享图结构可以发现, 对于整体分类任务而言, 强相关变量有右手尖 X 坐标、左手尖 Y 坐标、右手腕 X 坐标, 左手腕 Z 坐标、左肘部 Z 坐标、右手尖 Z 坐标、左手尖 Z 坐标、右拇指 X 坐标、右肘部 Z 坐标、右拇指 Y 坐标、右拇指 Z 坐标、左拇指 Z 坐标、左手腕 Y 坐标、右肘部 X 坐标。基于该共享图结构可以进行变量筛

选, 降低数据维度、提高模型训练效率。该共享图结构中学习到的重要特征包含了单类别图结构中的部分强相关特征, 如左手尖 Z 坐标、右拇指 X 坐标等。

5 结论

本文提出了基于不确定性的多元时间序列分类算法框架。基于该框架可学习单类别与多类别图结构, 实验结果可以为特征分析提供很好的可解释性; 我们的框架赋予了模型不确定性度量能力, 并首次提出考虑了样本内部类别相对不确定性的“相对方差”不确定性衡量方式, 基于该方法筛选不确定性

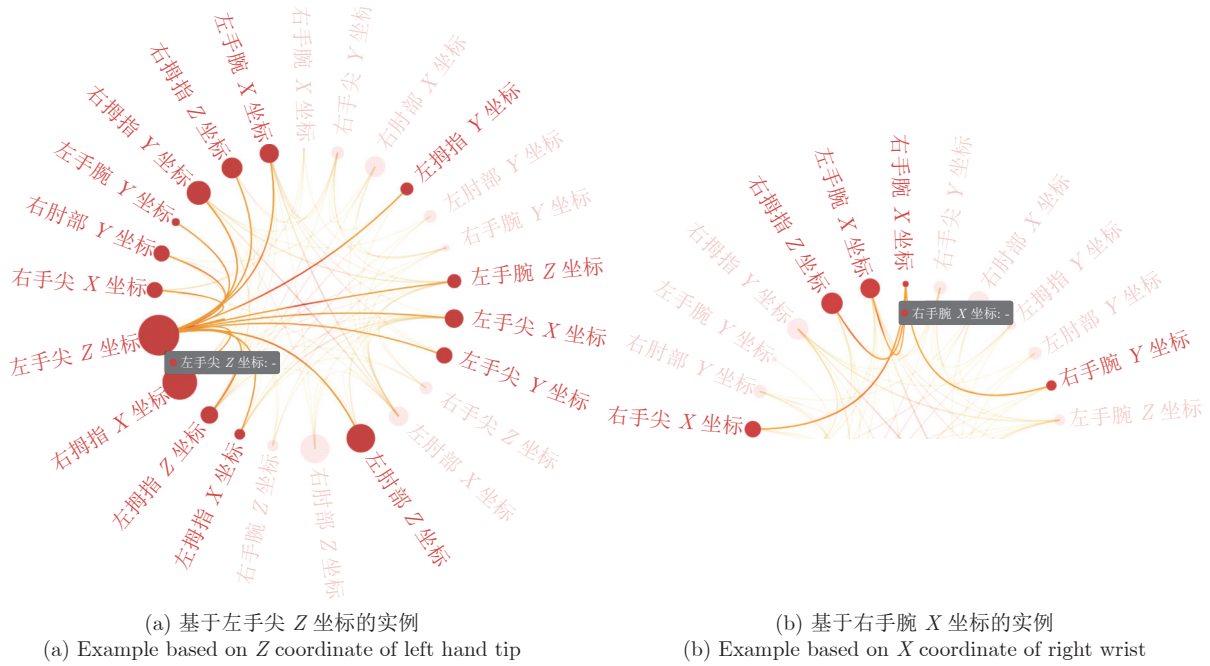


图 6 节点大小与边连接的关系 (NATO 数据集)

Fig.6 The relationship between node size and edge connection (NATO dataset)

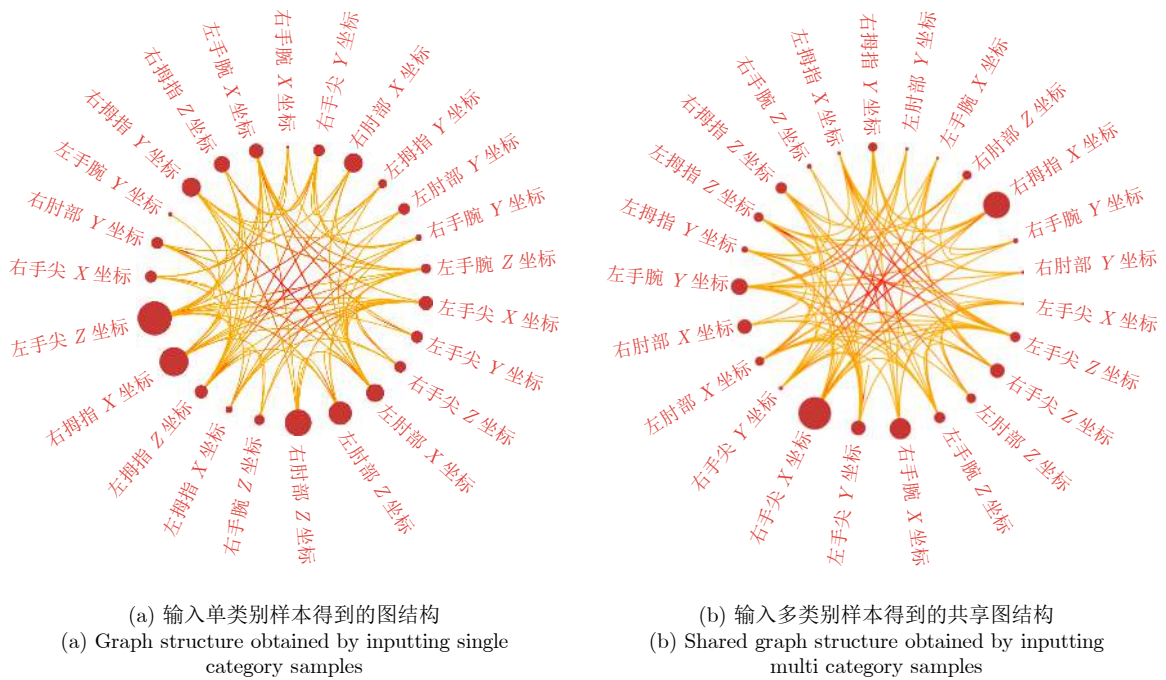


图 7 神经网络学习到的单类别与多类别共享图结构 (NATO 数据集)

Fig.7 Single class and multi class shared graph structures learned by neural networks (NATO dataset)

样本有效地改善了预测效果; 该框架是一个“轻量化”框架, 由于是利用 TAGCN 提取出来的特征进行贝叶斯建模, 因此具有计算简捷、解耦性好等特点. 与此同时, 我们的工作也存在一些局限性. 首先

本文所提算法仅适用于静态时间序列数据, 而难以用于随时间延续而增长的流式时间序列预测任务. 未来可以引入迁移学习以满足流式数据的快速计算需求. 其次, 本文提出了基于“相对方差”的概念来

衡量样本的不确定性,但目前缺乏理论上深入的分析,未来可以考虑引入互信息的角度衡量不确定性。

References

- 1 Zhang Xi-Lai, Zhao Jian-Hui, Cai Bo. Prediction model with dynamic adjustment for single time series of PM_{2.5}. *Acta Automatica Sinica*, 2018, **44**(10): 1790–1798
(张熙来, 赵俭辉, 蔡波. 针对 PM_{2.5} 单时间序列数据的动态调整预测模型. *自动化学报*, 2018, **44**(10): 1790–1798)
- 2 Xu Ren-Chao, Yan Wei-Wu, Wang Guo-Liang, Yang Jian-Cheng, Zhang Xi. Time series forecasting based on seasonality modeling and its application to electricity price forecasting. *Acta Automatica Sinica*, 2020, **46**(6): 1136–1144
(徐任超, 阎威武, 王国良, 杨健程, 张曦. 基于周期性建模的时间序列预测方法及电价预测研究. *自动化学报*, 2020, **46**(6): 1136–1144)
- 3 Keogh E, Chu S, Hart D, Pazzani M. Segmenting time series: A survey and novel approach. *Data Mining in Time Series Databases*. Singapore: World Scientific, 2004. 1–21
- 4 Zhang X C, Gao Y F, Lin J, Lu C T. TapNet: Multivariate time series classification with attentional prototypical network. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI, 2020. 6845–6852
- 5 Wang X, Gao Y F, Lin J, Rangwala H, Mittu R. A machine learning approach to false alarm detection for critical arrhythmia alarms. In: Proceedings of the IEEE 14th International Conference on Machine Learning and Applications (ICMLA). Miami, USA: IEEE, 2015. 202–207
- 6 Minnen D, Starner T, Essa I, Isbell C. Discovering characteristic actions from on-body sensor data. In: Proceedings of the 10th IEEE International Symposium on Wearable Computers. Montreux, Switzerland: IEEE, 2006. 11–18
- 7 Rakthanmanon T, Keogh E. Fast shapelets: A scalable algorithm for discovering time series shapelets. In: Proceedings of the 2013 SIAM International Conference on Data Mining. Austin, USA: SIAM, 2013. 668–676
- 8 Seto S, Zhang W Y, Zhou Y C. Multivariate time series classification using dynamic time warping template selection for human activity recognition. In: Proceedings of the 2015 IEEE Symposium Series on Computational Intelligence. Cape Town, South Africa: IEEE, 2015. 1399–1406
- 9 Pei W J, Dibeklioglu H, Tax D M J, Van Der Maaten L. Multivariate time-series classification using the hidden-unit logistic model. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, **29**(4): 920–931
- 10 Karim F, Majumdar S, Darabi H, Harford S. Multivariate LSTM-FCNs for time series classification. *Neural Networks*, 2019, **116**: 237–245
- 11 Zheng Y, Liu Q, Chen E H, Ge Y, Zhao J L. Time series classification using multi-channels deep convolutional neural networks. In: Proceedings of the 15th International Conference on Web-Age Information Management. Macau, China: Springer, 2014. 298–310
- 12 Mao Wen-Tao, Jiang Meng-Xue, Li Yuan, Zhang Shi-Guang. Structural prediction of multivariate time series through outlier elimination. *Acta Automatica Sinica*, 2018, **44**(4): 619–634
(毛文涛, 蒋梦雪, 李源, 张仕光. 基于异常序列剔除的多变量时间序列结构化预测. *自动化学报*, 2018, **44**(4): 619–634)
- 13 Wu Z H, Pan S R, Chen F W, Long G D, Zhang C Q, Yu P S. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, **32**(1): 4–24
- 14 Scarselli F, Gori M, Tsoi A C, Hagenbuchner M, Monfardini G. The graph neural network model. *IEEE Transactions on Neural Networks*, 2009, **20**(1): 61–80
- 15 Cao D F, Wang Y J, Duan J Y, Zhang C, Zhu X, Huang C R, et al. Spectral temporal graph neural network for multivariate time-series forecasting. In: Proceedings of the 34th Advances in Neural Information Processing Systems. arXiv: 2103.07719
- 16 Wu Z H, Pan S R, Long G D, Jiang J, Chang X J, Zhang C Q. Connecting the dots: Multivariate time series forecasting with graph neural networks. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Virtual Event: ACM, 2020. 753–763
- 17 Duan Z H, Xu H Y, Wang Y Y, Huang Y D, Ren A N, Xu Z B, et al. Multivariate time-series classification with hierarchical variational graph pooling. arXiv preprint arXiv: 2010.05649, 2020.
- 18 Spadon G, Hong S D, Brandoli B, Matwin S, Rodrigues-Jr J F, Sun J M. Pay attention to evolution: Time series forecasting with deep graph-evolution learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, DOI: 10.1109/TPAMI.2021.3076155
- 19 Hamilton W L, Ying R, Leskovec J. Inductive representation learning on large graphs. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017. 1025–1035
- 20 Xu K, Hu W H, Leskovec J, Jegelka S. How powerful are graph neural networks. In: Proceedings of the 7th International Conference on Learning Representations. New Orleans, USA: ICLR, 2019.
- 21 Veličković P, Cucurull C, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. In: Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada: ICLR, 2018.
- 22 Tran D, Dusenberry M W, Van Der Wilk M, Hafner D. Bayesian layers: A module for neural network uncertainty. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2019. 14660–14672
- 23 Kendall A, Gal Y. What uncertainties do we need in Bayesian deep learning for computer vision. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: 2017. 5575–5585
- 24 Zhang C, Büttepage J, Kjellström H, Mandt S. Advances in variational inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, **41**(8): 2008–2026
- 25 Zhang X L, Qian B Y, Cao S L, Li Y, Chen H, Zheng Y F. INPREM: An interpretable and trustworthy predictive model for healthcare. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Virtual Event: ACM, 2020. 450–460
- 26 Ranjan E, Sanyal S, Talukdar P P. ASAP: Adaptive structure aware pooling for learning hierarchical graph representations. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI, 2020. 5470–5477
- 27 Du J, Zhang S H, Wu G H, Moura J M F, Kar S. Topology adaptive graph convolutional networks. arXiv: 1710.10370, 2018.
- 28 Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015. 1–9
- 29 Hinton G E, Srivastava N, Krizhevsky A. Improving neural networks by preventing co-adaptation of feature detectors. *Computer Science*, 2012, **3**(4): 212–223

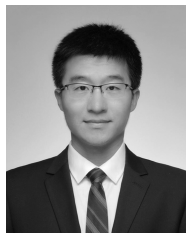
- 30 Kingma D P, Ba J. Adam: A method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA: ICLR, 2015. 1–15
- 31 Ji Wen-Qiang. Research on Data-driven Remaining Useful Life Prediction Method Based on Deep Learning and Uncertainty Quantification [Master thesis], University of Science and Technology of China, China, 2020
(季文强. 基于深度学习和不确定性量化的数据驱动剩余寿命预测方法研究 [硕士学位论文], 中国科学技术大学, 中国, 2020)
- 32 Hoffman M D, Blei D M, Wang C, Paisley J W. Stochastic variational inference. *Journal of Machine Learning Research*, 2013, **14**(1): 1303–1347
- 33 Zhang A Y, Zhou H H. Theoretical and computational guarantees of mean field variational inference for community detection. *The Annals of Statistics*, 2020, **48**(5): 2575–2598
- 34 Shokoohi-Yekta M, Wang J, Keogh E J. On the non-trivial generalization of dynamic time warping to the multi-dimensional case. In: Proceedings of the 2015 SIAM International Conference on Data Mining. Vancouver, Canada: SIAM, 2015. 289–297
- 35 Bagnall A, Flynn M, Large J. A tale of two toolkits, report the third: On the usage and performance of HIVE-COTE v1.0. arXiv preprint arXiv: 2004.06069, 2020.



张 旭 大连理工大学机械工程学院硕士研究生. 主要研究方向为机器学习, 数据挖掘与应用.

E-mail: zhangxu1@mail.dlut.edu.cn
(ZHANG Xu Master student at the Mechanical Engineering College, Dalian University of Technology.

His research interest covers machine learning, data mining and applications.)



张 亮 东北财经大学国际商学院讲师. 主要研究方向为多元时间序列挖掘, 医疗健康大数据. 本文通信作者.

E-mail: liang.zhang@dufe.edu.cn

(ZHANG Liang Lecturer at the International Business College, Dongbei University of Finance and Economics.

His research interest covers multivariate time series mining and healthcare big data analytics. Corresponding author of this paper.)



金 博 大连理工大学创新创业学院教授. 主要研究方向为信息检索, 数据挖掘和智能计算.

E-mail: jinbo@dlut.edu.cn

(JIN Bo Professor at the School of Innovation and Entrepreneurship, Dalian University of Technology.

His research interest covers information retrieval, data mining, and intelligent computing.)



张红哲 大连理工大学机械工程学院副教授. 主要研究方向为工业大数据的挖掘与应用.

E-mail: zhanghongzhe@dlut.edu.cn

(ZHANG Hong-Zhe Associate professor at the Mechanical Engineering College, Dalian University of

Technology. Her research interest covers industrial big data mining and application.)