

基于多尺度残差 FCN 的时间序列分类算法*

张雅雯^{1,2}, 王志海^{1,2}, 刘海洋^{1,2}, 曾昭博^{1,2}



¹(北京交通大学 计算机与信息技术学院, 北京 100044)

²(交通数据分析与挖掘北京市重点实验室(北京交通大学), 北京 100044)

通信作者: 刘海洋, E-mail: haiyangliu@bjtu.edu.cn

摘要: 时间序列数据广泛存在于我们的生活中,吸引了越来越多的学者对其进行深入的研究. 时间序列分类是时间序列的一个重要研究领域,目前已有上百种分类算法被提出. 这些方法大致分为基于距离的方法、基于特征的方法以及基于深度学习的方法. 前两类方法需要手动处理特征和人为选择分类器,而大多数的深度学习的方法属于端到端的方法,并且在时间序列分类问题中表现出不错的分类效果. 但是,目前基于深度学习的方法很少能够针对时间序列数据中时间尺度选择的问题对网络进行改进,在网络结构方面,很少将网络进行融合,从而更好地发挥各自的优势. 为了解决这两类问题,提出一种多尺度残差全卷积神经网络(MRes-FCN)结构,用来处理时间序列问题. 该结构主要分为数据预处理阶段、全卷积网络与残差网络结合的阶段. 为了评价该方法的性能,在 UCR 的 85 个公开数据集上进行了实验,与基于距离的方法、基于特征的方法和基于深度学习的方法分别进行了比较. 实验结果表明:所提出的方法较其他方法而言具有很好的性能,在多个数据集上优于大多数方法.

关键词: 时间序列; 时间序列分类; 残差网络; 全卷积神经网络; 深度学习

中图法分类号: TP18

中文引用格式: 张雅雯, 王志海, 刘海洋, 曾昭博. 基于多尺度残差 FCN 的时间序列分类算法. 软件学报, 2022, 33(2): 555-570. <http://www.jos.org.cn/1000-9825/6142.htm>

英文引用格式: Zhang YW, Wang ZH, Liu HY, Zeng ZB. Time Series Classification Algorithm Based on Multiscale Residual Full Convolutional Neural Network. Ruan Jian Xue Bao/Journal of Software, 2022, 33(2): 555-570 (in Chinese). <http://www.jos.org.cn/1000-9825/6142.htm>

Time Series Classification Algorithm Based on Multiscale Residual Full Convolutional Neural Network

ZHANG Ya-Wen^{1,2}, WANG Zhi-Hai^{1,2}, LIU Hai-Yang^{1,2}, ZENG Zhao-Bo^{1,2}

¹(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

²(Beijing Key Laboratory of Traffic Data Analysis and Mining (Beijing Jiaotong University), Beijing 100044, China)

Abstract: Time series data widely exists in daily lives, attracting more and more scholars to conduct in-depth research on it. Time series classification is an important research field of time series, and hundreds of classification algorithms have been proposed. These methods are roughly divided into distance-based methods, feature-based methods, and deep learning-based methods. The first two types of methods require manual processing of features and artificial selection of classifiers, and most deep learning-based methods are end-to-end methods and show good classification results in time series classification problems. Nevertheless, the current deep learning-based methods are rarely able to improve the network for the problem of time scale selection in time series data, and rarely integrate the network in terms of network structure to better leverage their respective advantages. In order to solve these two kinds of problems, this study proposes a multi-scale residual full convolutional neural network (MRes-FCN) structure to deal with time series problems. The structure is mainly divided into the data preprocessing stage, the stage of combining the full convolutional network and the residual network. In order to evaluate the performance of this method, this study conducted experiments on 85 public data sets of UCR and compared them with

* 基金项目: 北京市自然科学基金(4182052); 国家自然科学基金(61672086, 61702030, 61771058)

收稿时间: 2020-03-10; 修改时间: 2020-05-21, 2020-07-20; 采用时间: 2020-09-01

distance-based methods, feature-based methods, and deep learning-based methods. Experiments show that the proposed method has better performance than other methods, and it is better than most methods on multiple data sets.

Key words: time series; time series classification; residual network; full convolutional neural network; deep learning

时间序列是一组随着时间的推移所产生的数据, 比如金融领域的股票数据、医学领域的心电图以及气象领域的温度数据等. 近 20 多年来, 时间序列分类问题一直被认为是数据挖掘领域具有挑战性的问题之一^[1,2], 目前已有上百种时间序列分类算法被提出^[3]. 除了可以直接观测到的序列, 还有越来越多的研究者聚焦在将图像、音频、视频等转化为时间序列进行研究, 比如将农作物的遥感数据转化为时序数据^[4,5], 从手势视频中提取每一帧的图像, 进行手势识别^[6]以及将婴儿啼哭的声音当作时间序列数据^[7]等. 另外, 在涵盖时间序列数据集最多的 UCR 数据集中^[3,8], 时间序列数据种类的多样性(大致可分为图像数据、光谱数据、仿真数据、运动数据、传感器数据、心电信号)也说明时间序列数据与我们的生活密切相关.

目前, 深度学习方法在各个方面都取得了不错的效果. 2015 年, 卷积神经网络在图像识别方面的性能达到人类的水平^[9]. 在自然语言处理领域, 机器翻译^[10]、学习单词嵌入和文本分类^[11,12]等方面的研究也取得了不错的效果. 2017 年, Wang 等人提出的残差网络用于时间序列的分类算法堪比传统的 37 个分类器集成的 COTE 方法^[13].

传统的时间序列分类算法主要是基于距离的动态时间规整方法和基于符号化的特征工程方法, 这些方法有些需要进行人为的手工处理. 随着基于深度学习的时间序列分类算法研究的深入, 越来越多的端到端的时间序列分类算法被提出. 但是目前, 现有的基于深度学习的时间序列分类算法仍然存在一定的问题.

- 首先, 由于公开的 UCR 时间序列数据集数据较小, 深度学习方法容易出现过拟合问题;
- 其次, 大多数方法没有考虑时间尺度的问题. 一条时间序列有时需要根据全局变化选取特征, 有时更关注局部的子序列, 比如 shapelets 方法;
- 最后, 卷积神经网络面对一个问题: 随着层数的加深, 分类效果会出现饱和. 虽然残差网络可以解决这个问题, 但是前期没有任何特征提取的数据没有必要再次传入网络.

针对以上基于深度学习的时间序列分类算法存在的问题, 本文提出了多尺度残差全卷积网络(MRes-FCN). 本文的主要贡献如下:

- (1) 在构建网络之前先对时间序列数据进行预处理, 对数据进行多滑度的平滑滤波和采样处理, 这样做的目的不仅可以增强数据, 同时能够解决时间尺度的选择问题;
- (2) 网络结构将全卷积神经网络和残差网络相结合, 首先通过 3 层卷积神经网络找到具有辨别性的特征, 然后加入 3 个残差块进一步进行特征选择和分类. 这样可以使得前面的全卷积神经网络找到相对有用的特征, 同时, 网络结构中的残差块可以防止后期因为层数过多而出现分类效果下降的问题;
- (3) 通过在 UCR 上的 85 个数据集进行训练和测试, 本文与现有的基于距离的方法、基于特征的方法和基于深度学习的方法进行了多种方式的比较, 对本文提出的网络结构进行了整体评估.

本文第 1 节是相关工作, 详尽介绍目前时间序列分类算法的研究现状, 包括现有的突出的方法以及方法的优缺点等. 第 2 节主要介绍本文的主要贡献点: 首先对时间序列进行数据转换, 然后将全卷积神经网络和残差网络相结合对时间序列数据进行特征提取和分类. 第 3 节主要介绍实验环境、实验的数据集、实验设置以及结果. 通过实验对现有的比较突出的方法进行比较, 从多个方面验证本文方法的可行性和性能, 并讨论实验结果和造成结果的原因. 第 4 节对全文的研究进行总结和展望. 总结全文的工作, 分析目前存在的不足和以后可以改进的地方, 并且提出一些对于未来研究方向的建议.

1 相关工作

时间序列分类问题是指通过学习给定类标的时间序列训练集, 从而生成时间序列的分类器, 对新输入的

未标定的时间序列, 输出其类标. 目前, 对于时间序列分类的方法大致可以分为 3 类: 基于距离度量的方法、基于特征提取的方法、基于深度学习的方法.

基于距离度量的方法是一种传统的时间序列分类方法. 这类方法是结合距离函数使用的最近邻(NN)分类器^[14]. 动态时间规整(DTW)已被证明是一个非常好的方法^[3]. Bagnall 在其一篇综述中比较全面地介绍了基于非深度学习方法的时间序列分类算法, 文中表示, 基于具有不同距离度量的最近邻(NN)分类器的集成效果优于任何一个单个的分类器, 基于最近邻的分类器中很多方法都具有数据转换阶段, 即时间序列数据被转换为一个新的特征空间. 这种转换也被 Bagnall 应用到 COTE 中^[15], 实际上, COTE 就是 35 个分类器的集合. Lines 等人在此基础上基于分级投票系统提出了基于转换的集合的时间序列分类的分层投票集合(HIVE-COTE)^[16], 该方法采用新的分级结构和概率投票, 运行时训练 37 个分类器, 并且要对每个超参数进行交叉验证, 当对 UCR/UEA 包含的 85 个数据集进行测试时, HIVE-COTE 被认为是目前最先进的时序分类算法^[3].

基于特征提取的方法主要有词袋表示和符号化方法. 根据文本表示提出的词袋方法(BoW)^[17], Lin 等人提出了应用于时间序列分类的模式袋(bag of patterns, BOP)模型^[18]. BOP 模型首先使用符号近似估计(symbolic aggregate approximation, SAX)^[19]将时间序列转化为一系列的单词集合, 然后用直方图统计单词出现的频率用于分类. 自从 Ye 等人提出了辨别性子序列之后, 它们可以以不同的方式使用^[20]. 例如, 它们可以被看作字典, 每个 shapelet 都被看作一个单词, 再采用词袋表示法进行分类. Lines 构造了特征向量, 使得每个特征的值都是时间序列中任意位置与对应 shapelet 之间的最小距离^[21]. Shapelet 方法的一个缺点是, 它需要在较大的空间中进行大量的判别 shapelets 搜索. 为了避免需要尝试大量 shapelet 候选项, Grabocka 等人提出了与分类器共同学习一些相同大小的 shapelets 的方法. 然而, 他们的方法只能提供线性分离能力^[22]. Senin 等人提出了符号近似估计向量空间模型(symbolic aggregate approximation-vector space model, SAXVSM)^[23]. SAXVSM 对每一个类生成单词频率向量, 并且使用逆文档频率对频率向量进行加权, 通过单词所属的类别找到辨别性子序列, 但是所发掘的辨别性子序列并不能发现子序列形状上的不同. 此外, Schäfer 等人提出的 BOSS(bag of SFA symbols)模型利用离散傅里叶变换将时间序列转化到频域进行分析^[24], 这样更容易发掘频率信息.

随着深度学习在机器视觉、机器翻译、自然语言处理等领域取得的成功, 越来越多的基于深度学习的时间序列算法被提出. Cui 等人提出了多尺度卷积神经网络(MCNN), 在卷积网络前进行了数据转换, 但是该方法的分类效果并不是很理想^[25]. Wang 等人提出了深层的多层感知机(MLPs)、完全卷积神经网络(FCNs)和残差网络(ResNet)用于时间序列分类, 这些方法直接将原始数据传入到网络结构中, 没有考虑数据本身存在的问题^[13]. Serrà 等人提出了两种近似于编码器技术的方法: 第 1 种方法是在目标数据集上以端到端的方式从头开始训练模型; 而第 2 种方法是在原数据集上预先训练相同的架构, 然后在目标数据集上微调它^[26]. 第 1 种方法直接使用原始数据, 相比于第 2 种方法的迁移学习, 后一种方法准确率更高. Karim 提出了长短时记忆全卷积网络(LSTM-FCN), 将 LSTM 与 FCN 融合, 在网络结构中引入了注意机制, 能够了解到决策过程. 由于小而短序列数据容易过拟合, 以及大而长的序列数据很难找到长期依赖关系, 所以该方法还对多变量的时间序列数据进行维度洗牌^[27]. Fawaz 将 MLP、FCNs、ResNet、Encoder、MCNN、t-LeNet、MCDLNN、Time-CNN、TWIESN 这 9 种方法分别应用到 85 个 UCR/UEA 数据集和 12 个 MTS 数据集上^[28]. 通过对比实验结果发现: ResNet 是唯一能够达到与 COTE 相似性能的算法, 是 9 种方法中 TSC 效果最好的. 随着基于深度学习的时间序列分类算法的发展, 也有一些学者将这些方法进行了具体的应用. Han 提出了一种 ML-ResNet 的 MI 检测和定位方法, 并且针对 5 种类型的 MI 定位, 取得了很高的准确率^[29]. 该方法实际上是 3 个并行的残差网络结构, 并且这类方法主要针对于海量的心电图数据, 对于其他类型的时间序列数据泛化性能比较差. Ramzan 针对阿尔茨海默病(AD)的分类, 详细研究了 18 层的残差网络的分类效果, 并从分类精度、召回率、F1-measure、AUC 和 ROC 曲线等方面对模型的性能进行评估^[30]. 该方法仅对残差网络结构进行了加深.

综上所述, 基于特征的方法需要人为地进行特征处理, 而基于深度学习的方法没有着重解决过拟合问题以及时间尺度的选择. 另外, 大多数方法使用原始的网络结构, 没有将网络进行融合, 从而更好地发挥各自的优势.

2 基于多尺度残差全卷积神经网络的时间序列分类算法

本文的时间序列数据直接采用 UCR 数据集里分好的训练集和测试集. 时间序列是指按照时间排序的一系列实值数据, 某一时刻的实值记作 t_i , 其中, $i \in \{1, 2, \dots, n\}$, 则时间序列表示为 $T = \{t_1, t_2, \dots, t_n\}$. 实际上, 任何一组有序的数据均可称为时间序列.

时间序列分类是指给定一个时间序列数据集 D , 通过训练数据集得到一个分类器, 对于测试集中的时间序列数据, 能够将每一条时间序列映射到所属的类标. 本文通过对网络进行训练, 找到最优的参数, 从而确定好网络, 再利用测试集来验证本文所提出的算法.

下面主要介绍本文中提出的具体算法, 该算法主要分为两个阶段: 数据预处理阶段、全卷积与残差网络结合阶段.

2.1 数据预处理阶段

时间序列分类问题的难点主要体现在时间尺度的选择和高频噪声的影响上, 而基于深度学习的方法又面临着过拟合问题, 所以关键是有针对性地解决上述问题, 从而提升分类精度. Cui 等人的方法同样采用了数据转换阶段, 但是由于采用卷积神经网络, 分类效果不是很好^[25]. UCR 数据集数量级较低, 通过对数据进行预处理能够加大数据量, 从而更好地训练网络.

数据处理阶段包括 3 个方面.

(1) 保持原有数据

为了保留部分数据的真实性, 所以数据处理阶段会选择保留原有数据. 此时的时间序列数据如式(1)所示:

$$T' = \{t_1, t_2, \dots, t_n\} \quad (1)$$

(2) 平滑滤波处理

滤波是滤除掉信号中特定波段频率的方法, 是抑制和防止干扰信号的一项重要措施. 常见的滤波操作有限幅滤波法(又称为程序判断滤波法)、中位值滤波法、算术平均滤波法、递推平均滤波法(又称为滑动平均滤波法)、中位值平均滤波法(又称为防脉冲干扰平均滤波法)、限幅平均滤波法等. 本文采用多滑度的平滑滤波法, 该滤波法通过滤除掉信号中特定波段频率来有效地抑制和防止干扰信号. 在平滑滤波中, 平滑度与平滑的效果直接相关, 平滑度越大, 平滑的效果越好. 但平滑度过大会使数据丢失掉有用信息, 因此需合理选择平滑度. 给定一个输入时间序列, 本文使用具有不同窗口大小的移动平均值生成具有不同平滑度的多个新时间序列.

对于窗口大小为 l 的滤波器, 对第 i 个数据到第 $i+l-1$ 的时间序列数据进行滤波, 滤波处理如公式(2)所示, 原时间序列如公式(3)所示, 那么滤波处理后的数据如公式(4)所示:

$$\bar{t} = \frac{t_i + t_{i+1} + \dots + t_{i+l-1}}{l} \quad (2)$$

$$T = \{t_1, t_2, \dots, t_i, t_{i+1}, \dots, t_{i+l-1}, \dots, t_n\} \quad (3)$$

$$T' = \{t_1, t_2, \dots, \bar{t}, \bar{t}, \dots, \bar{t}, \dots, t_n\} \quad (4)$$

(3) 数据采样处理

一条时间序列包含的信息多少不等, 时间尺度有大有小, 时间序列的长期特征反映总体, 短期特征反映局部变化. 能够找到合适的时间尺度是一个关键问题. 比如我们要研究太阳花适宜生长的环境, 一天 24 小时内, 白天早 6 点-晚 6 点的光照数据才是我们最关心的问题. 比如一个人的心电图数据, 可能就早上 6 点、下午 2 点和晚上 10 点这 3 个时间点的数据更有参考价值. 通过对数据进行不同尺度的切片处理, 会更好获得有用的信息. 本文所采用的切片方式是对数据进行采样, 设采样率为 k , 那么新的时间序列如公式(5)所示:

$$T' = \left\{ t_k, t_{2k}, \dots, t_{\lfloor \frac{n}{k} \rfloor} \right\} \quad (5)$$

同时, 采样率能够有不同的取值, $k=1, 2, 3, \dots$. 通过对数据进行不同的采样, 可以将数据切分成多个序列,

每个序列通过一层卷积神经网络后, 将会初步学到类似于 *shpelets* 的子序列, 后续的网络结构中将会找到更完善的特征, 从而能够更好地对时间序列进行分类。

数据经过以上 3 个方面的并行处理之后, 会得到多个不同的时间序列数据. 将多个不同的时间序列数据分别进行卷积再进行拼接, 得到的时间序列数据再传入网络结构中。

数据预处理阶段的设置不仅可以解决时间序列数据的时间尺度的选择问题, 得到该条时间序列的多个尺度的时间序列数据, 还可以尽可能多地得到原时间序列的各种变体. 得到的数据再进行局部卷积池化, 可以对特征进行进一步筛选. 该阶段的参数设置如表 1 中的数据预处理阶段所示. 其中, *filters* 代表卷积核(滤波器)的数目, 换言之就是输出的维度; *kernal_size* 代表卷积核的规模, 即一个卷积核的大小。

表 1 实验参数

网络结构阶段	具体部分	<i>filters</i>	<i>kernel_size</i>
数据处理阶段	原数据部分	256	由拼接好的
	滤波部分	256	时间序列
	采样部分	256	大小而定
全卷积与残差网络相结合阶段	第 1 层卷积	128	8
	第 2 层卷积	256	5
	第 3 层卷积	128	3
	(残差块 1)第 1 层卷积	64	8
	(残差块 1)第 2 层卷积	64	5
	(残差块 1)第 3 层卷积	64	3
	(残差块 2)第 1 层卷积	128	8
	(残差块 2)第 2 层卷积	128	5
	(残差块 2)第 3 层卷积	128	3
	(残差块 3)第 1 层卷积	128	8
(残差块 3)第 2 层卷积	128	5	
(残差块 3)第 3 层卷积	128	3	

2.2 全卷积与残差网络结合阶段

本文具体的网络结构采用将全卷积神经网络与残差网络相结合的方式. 预处理后的数据先要经过 3 层的全卷积神经网络进行特征提取, 之后再经过残差网络进行特征提取. 这样结合的原因是: 全卷积神经网络能够有效地找到辨别性的特征, 从而更好地分类; 残差网络能够防止网络层数过深导致的分类效果出现饱和的情况. 下面将分别介绍两个网络和全卷积与残差网络结合阶段。

2.2.1 全卷积网络

全卷积神经网络(FCN)相比于卷积神经网络, 就是将网络结构中的全连接层全部转化为卷积层. 全卷积神经网络在图像的语义分割中的效果几乎是最好的^[31]. Wang 等人首次提出用全卷积神经网络(FCN)对时间序列数据进行分类^[13]. 该网络在 UCR 单变量数据集中心电图数据上, 取得了比残差网络更好的效果. 全卷积神经网络的网络结构如图 1 所示。

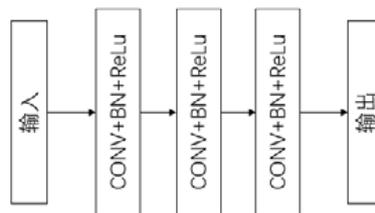


图 1 全卷积网络结构

2.2.2 残差网络

一般情况下, 随着神经网络的深度增加, 模型的复杂度越高, 算法在一定程度上的效果应该越好. 但是到达某个程度之后就会出现饱和, 即随着深度的增加, 分类效果不再明显增强. 于是, 就有学者提出了残差网

络,以抑制梯度下降.

与卷积神经网络不同的是,残差网络多了一个残差单元.具体结构如图 2 所示.输入数据 x 后,在残差网络中,会有一个残差项 $F(x)$,此时学习到的特征为 $x+F(x)$.所以,残差网络关键的问题是拟合出残差项 $F(x)$.当残差项为 0 时,对网络没有任何影响,实际上就是一个恒等映射;当残差项不为 0 时,残差项的设置相当于一个卷积神经网络,能够学习到有利于分类的特征,从而拥有更好的性能.

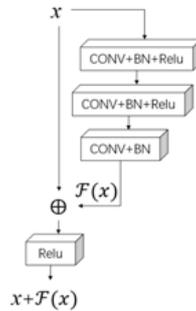


图 2 残差单元

残差网络在目标检测和其他机器视觉领域中取得了很好的效果.在 Fawaz 的基于深度学习的时间序列分类算法综述中提到, Wang 等人提出的残差网络(ResNet)在大多数时间序列数据集中取得了不错的分类效果^[13,28].然而本文发现:仅仅是单纯用残差网络对时间序列进行分类往往容易出现过拟合现象,训练数据的准确率往往能达到 100%,而在测试数据上的准确率有时仅有 80%+.图 3 展示了 InlineSkate 数据通过残差网络进行分类过程的损失函数值的变化. InlineSkate 数据是在专业的速滑比赛测试中收集的,对于此数据,每个时间序列表示一个运动周期内脚踝的角度测量值.从图 3 可以看出:训练数据损失函数的值几乎为 0,而测试数据损失函数的值为 2.残差网络的层数较深,需要的数据量往往比较大. UCR 提供的时间序列的数据比较小,本文采用数据处理方式不仅是为了时间尺度和高频噪声问题,也是为了增强数据,从而更好地训练网络.

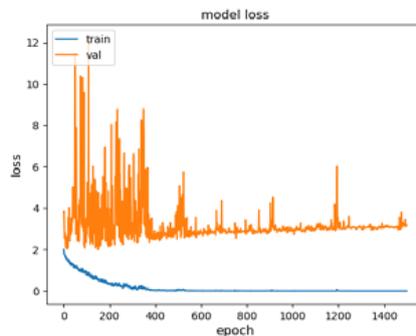


图 3 损失值对比

2.2.3 全卷积与残差网络结合阶段

全卷积与残差网络结合阶段是由 3 个卷积块和 3 个残差块组成的网络结构,如图 4 所示.前面处理后的数据先要经过 3 个卷积块进行特征提取,之后再经过残差网络进行特征提取和分类.该阶段的参数沿用了 Wang 等人提出的 FCN 和 ResNet 取得的分类效果最好的参数设置^[13],具体的参数设置如表 1 的全卷积与残差网络相结合阶段所示.

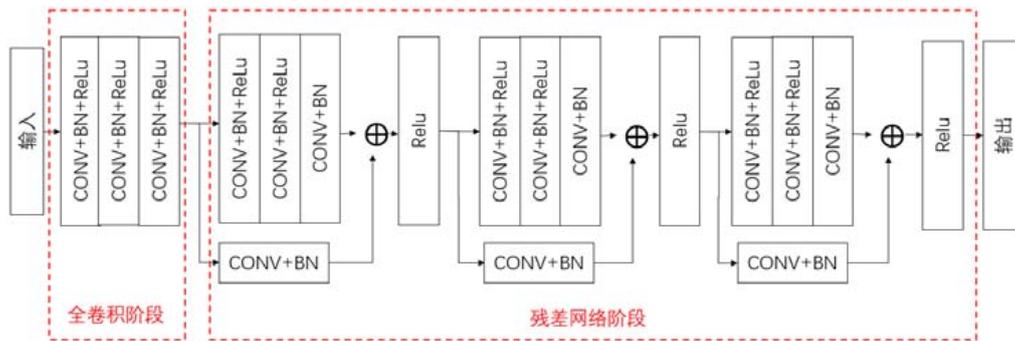


图4 全卷积与残差网络结合结构

3 个卷积块的结构是每一个卷积块由卷积层(CONV)、批归一化(BN)和激活函数 ReLU 构成。批归一化(BN)的方法可以有效地减少梯度消失, 加快收敛速度; 同时允许使用更大的学习率, 防止过拟合。该阶段每一个卷积块内的卷积核数目和卷积核大小的参数设置如表 1 的卷积与残差网络相结合阶段的第 1 行-第 3 行所示。

3 个残差块的结构是每一个残差块由两个完整的卷积块(CONV+BN+ReLU)经过 CONV 和 BN 之后与当前卷积块的输入经过 CONV 和 BN 之后求和再进行 ReLU 激活构成。该阶段每一个残差块的卷积核数目和卷积核大小的参数设置如表 1 的卷积与残差网络相结合阶段的第 2 行-第 4 行所示。

2.3 多尺度残差全卷积神经网络(MRes-FCN)

本文所提出的多尺度残差全卷积神经网络(MRes-FCN)主要包括两个主要部分: 数据预处理阶段、全卷积与残差网络相结合阶段。具体的网络结构如图 5 所示。

- (1) 数据预处理阶段: 该阶段通过对数据进行并行处理, 进而解决时间序列数据面临的两个问题: 时间尺度的选择和噪声的干扰。并行处理包括保留原始数据, 形成原始数据的恒等映射; 对数据进行切片处理, 来解决时间尺度的选择问题; 对数据进行平滑滤波处理, 来解决高频扰动和噪声问题。同时, 每一部分获得数据之后, 会分别经过一层卷积, 这样会从每一段时间序列中学习到的粗略的子序列。在后续的网络结构中, 将会学习到更具判别性的序列;
- (2) 全卷积与残差网络相结合阶段: 该阶段主要是将数据预处理后的 3 类数据传入到 3 层的全卷积网络进行特征提取。为了防止分类效果达到饱和, 再传入具有 3 个残差块的残差网络中, 最后进行全局的平均池化操作和 Softmax。

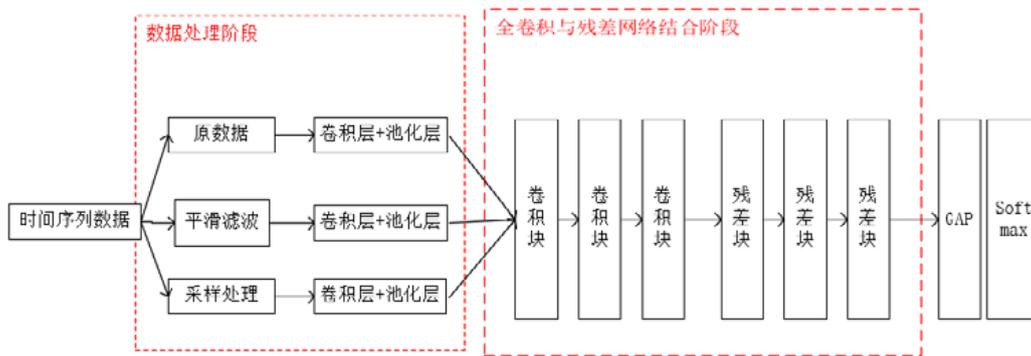


图5 多尺度残差全卷积神经网络结构

3 实验

在本节中, 本文对 UCR 的 85 个数据集进行了广泛的实验, 并通过多种方式来评价本文所提出的多尺度

残差全卷积神经网络在时间序列数据上的分类效果.

3.1 实验环境

本文的实验环境基于 Ubuntu18.04 系统, NVIDIA 显卡是 GeForce GT 730, 与之对应的驱动版本采用 418.56, CUDA 版本采用 10.1. 我们的实验代码基于 Python3.6, 采用深度学习库 TensorFlow 和 Keras.

3.2 数据集介绍

本实验使用的数据均来源于公开的 UCR 数据集(http://www.cs.ucr.edu/~eamonn/time_series_data/). UCR 数据集可以分为 7 类: 图像轮廓 IMAGE、传感器读数 SENSOR、动作捕捉 MOTION、光谱仪 SPECTRO、心电图 ECG、电气设备 DEVICE 和模拟数据 SIMULATED. 本实验在 UCR 提出的 85 个数据集的训练数据集上对网络结构进行训练, 在测试数据集上进行测试实验, 数据集信息参见表 2, 包括最大的(最小的)训练数据集、测试数据集、数据集长度以及类的数目. 其中, 表 2 中的“规模”是指数量多少或者长度的大小.

表 2 数据集信息

名称	最大或最小	数据集名称	类型	规模
训练实例数	Min	DiatomSizeReduction	IMAGE	16
	Max	ElectricDevices	DEVICE	8 926
测试实例数	Min	BeetleFly	IMAGE	20
	Max	StarlightCurves	SENSOR	8 236
数据长度	Min	ItalyPowerDemand	SENSOR	24
	Max	HandOutlines	IMAGE	2 709
类的数量	Min	31 个数据集 size 为 2, 涵盖所有类型		
	Max	ShapesAll	IMAGE	60

3.3 实验设置及结果

本文将在测试集中多次测试的结果与现有算法进行对比来验证所提方法的性能. 下面, 我们从参数分析、对比实验和性能分析这 3 个方面进行介绍.

3.3.1 参数分析

网络结构中有两个关键性的参数 *epochs* 和 *learning rate* (*lr*). Cui 等人提出的 MCNN 方法采用的是 *epochs*=200, *learning rate*=0.1^[25]. 经过实验, 本文发现, 这两个参数的设定并不能很好地训练网络. 图 6 展示了 *epochs*=200 时, 针对 HandOutlines 数据集算法的损失值.

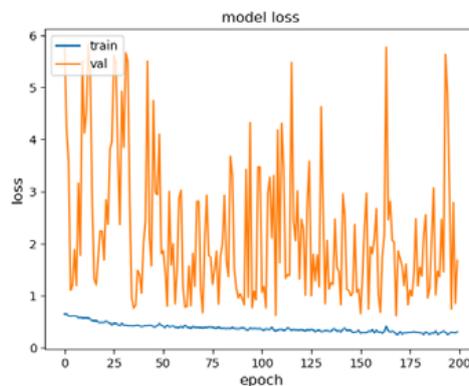


图 6 HandOutlines 数据集上的运行结果

很明显, 200 的迭代次数还没有训练好网络, 训练好的网络损失值应该是保持平稳的. 所以, 本文选取了 14 个数据集对 *lr* 和 *epochs* 的设置进行实验, 数据集信息见表 3. 这 14 个数据集分别对应 UCR 数据集的不同类型, 每一类包含两个数据集, 数据集的长度选择的是长度间隔 100 左右的数据, 96 到 1 639 不等, 二分类数据集居多, 但也包含类数量为 4、5、7、8、50 的数据集.

表 3 参数设置数据集信息

数据集	类型	训练大小	测试大小	长度	类数量
ElectricDevices	DEVICE	8 926	7 711	96	7
Computers	DEVICE	250	250	720	2
ECG5000	ECG	500	4 500	140	5
CinCECGtorso	ECG	40	1 380	1 639	4
50words	IMAGE	450	455	270	50
DiatomSizeReduction	IMAGE	16	306	345	4
GunPoint	MOTION	50	150	150	2
UWaveGestureLibraryAll	MOTION	896	3 582	945	8
ItalyPowerDemand	SENSOR	67	1 029	24	2
Lightning2	SENSOR	60	61	637	2
TwoPatterns	SIMULATED	1 000	4 000	128	4
Mallat	SIMULATED	55	2 345	1 024	8
Ham	SPECTRO	109	105	431	2
OliveOil	SPECTRO	30	30	570	4

首先, 对学习率 lr 进行分析. 本文所选择的参数大部分来自于已有方法所使用的参数: Wang 等人提出的多层感知机(MLP)使用的是 $1^{[13]}$; Cui 等人提出的多尺度卷积神经网络使用的是 $0.1^{[25]}$; Zheng 等人的多通道卷积神经网络(MCDCNN)采用 $0.01^{[32]}$; Wang 等人提出的 FCN 和 ResNet 均采用 $0.001^{[13]}$; 而自编码选用的是 $0.000\ 01^{[26]}$; 本文又补充了 $0.000\ 1$. 图 7 展示了在 14 个数据集上, lr 分别取值 1 、 0.1 、 0.01 、 0.001 、 $0.000\ 1$ 和 $0.000\ 01$ 时, 本文所提出方法的分类准确率. 由图中不同颜色的折线可以明显看出: lr 取值为 1 、 0.1 、 0.01 明显落后于另外的取值. 对于难以区分的 0.001 、 $0.000\ 1$ 和 $0.000\ 01$, 本文求取了每一种取值在这些数据集上的分类准确率的平均值, 它们分别是 0.783 、 0.801 、 0.800 , 所以 $lr=0.0001$ 是较好的参数.

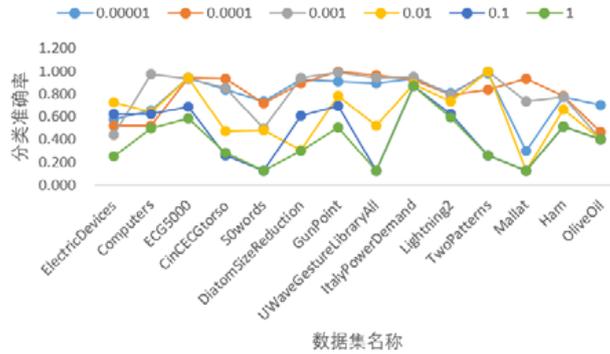


图 7 不同取值的学习率(lr)对应的分类准确率

其次, 对迭代次数 $epochs$ 进行分析. 本文所选择的参数大部分来自于已有方法所使用的参数: 100 来自于自编码方法^[26]; Cui 等人提出的 MCNN 采用的是 $200^{[25]}$; $1\ 000$ 来自于 t-LeNet^[33]; $1\ 500$ 来自于 Wang 等人提出的 ResNet^[13]; FCN^[13]和 time-CNN^[34]均使用的是 $2\ 000$. 图 8 展示了在 14 个数据集上, $epochs$ 分别取值 100 、 200 、 $1\ 000$ 、 $1\ 500$ 和 $2\ 000$ 时, 本文所提出方法的分类准确率.

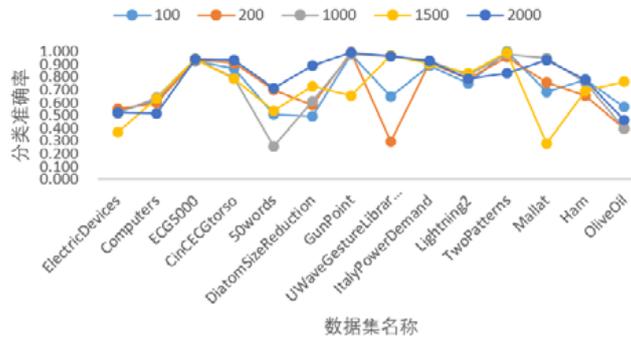


图 8 不同取值的 $epochs$ 对应的分类准确率

由图中不同颜色的折线可以看出: 当 $epochs=2000$ 时, 即蓝色的折线, 其分类准确率略高于其他取值. 本文求取了每一种取值在这些数据集上的分类准确率的平均值, 它们分别是 0.733、0.719、0.754、0.722、0.801, 所以 $epochs=2000$ 是较好的参数.

3.3.2 对比实验

目前, 针对时间序列的分类有 3 类算法: 基于距离的方法、基于特征的方法和基于深度学习的方法. 本文将分别与这 3 类方法进行比较.

本文提出的 MRes-FCN 与比较流行的和分类效果较好的基于距离的时间序列分类算法进行比较, 其中包括 DTW^[35]、DDTW^[35]、MSM^[36]、WDTW^[37]、CID_DTW^[38]、LWDTW^[39]、shapeDTW^[40]. 图 9 显示了包括本文方法在内的 9 种方法的临界差分图. 本文提出的方法比现有的基于距离的方法更加优越.

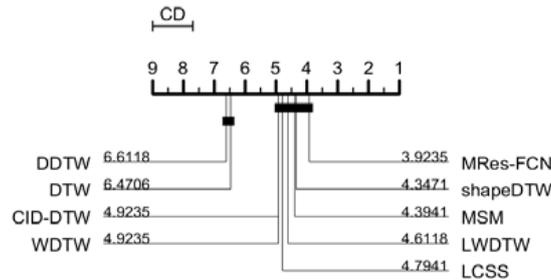


图 9 基于距离的分类器的临界差分图

与基于特征的时间序列分类方法进行比较. 本文尝试使用目前比较新的提取特征的方法与本文的 MRes-FCN 进行分析, 包括符号集合近似向量空间模型(SAXVSM)^[23]、快速 shapelet(FS)^[41]、学习 shapelet(LS)^[22]、模式包(BOP)^[18]、时间序列森林(TSF)^[42]、时间序列特征包(TSBF)^[43]. 图 10 展示了这 7 类方法的错误率的临界差分图, MRes-FCN 处于这些方法的首位, 说明本文的方法较基于特征的分类器而言, 分类效果较好, 而且并不需要特征工程那么多繁琐的步骤, 是端到端的分类器.

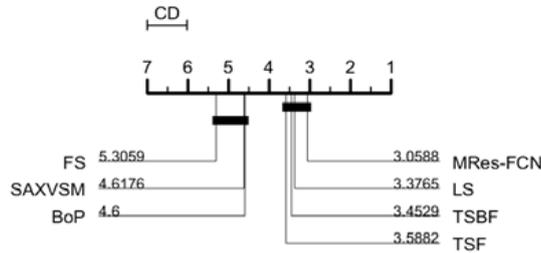


图 10 基于特征的分类器的临界差分图

本文还与基于深度学习的时间序列分类算法(多层感知机 MLP、全卷积神经网络 FCN、残差网络 ResNet^[13]、多尺度卷积神经网络 MCNN^[25]、自编码器 Encoder^[26]、使用卷积神经网络进行的数据扩充网络 t-LeNet^[33]、多通道深度卷积神经网络 MCDCNN^[32]、时间-卷积神经网络 time-CNN^[34]、长短时记忆全卷积神经网络^[27])进行比较, 其中, 各类深度学习算法的超参数均沿用原作者所使用的超参数, 具体见表 4. 其中, $epochs$ 是指一个训练集训练多少次, $batchsize$ 是批大小. 传统的梯度下降法是遍历一遍所有的数据集, 再计算损失函数以更新参数, 但是这样做使得计算量很大. 现在大多数学者选择批梯度下降方法, 把数据分成若干份, 按份来进行梯度下降, 从而更新参数, 分成多少份就取决于批大小. $Learning\ rates$ 是学习率, 在进行网络结构优化过程中, 大多数方法选择梯度下降法使得损失函数达到最小. 而学习率就是在梯度下降算法中的迭代步长. $Decay$ 是权重衰减因子, 随着迭代次数越来越多, 在损失函数收敛到全局最优时会出现左右来回摆荡的情况, 要么过大, 要么过小. 这主要是因为学习率过大造成的, 所以在有些方法中通过引入权重衰减因子对学习

率进行缩小.

表 4 深度学习算法的超参数

Method	epochs	batch	Learning rate	Decay
MLP	5 000	16	1.0	0
FCN	2 000	16	0.001	0
time-CNN	2 000	16	1	0
MCNN	200	256	0.1	0
ResNet	1 500	16	0.001	0
MCDCNN	120	16	0.01	0.000 5
Encoder	100	12	0.000 01	0.005
t-LeNet	1 000	256	0.01	0
LSTM-FCN	2 000	128	1	0
MRes-FCN	2 000	256	0.000 1	0

首先, 本文将 Encoder、t-LeNet、Time-CNN、MCDCNN、MCNN、MLP 以及本文的 MRes-FCN 这 7 类方法的分类错误率绘制了临界差分图, 如图 11 所示. 从图 11 中可以明显地看到, 本文的方法(MRes-FCN)比 Encoder、t-LeNet、Time-CNN、MCDCNN、MCNN 以及 MLP 的分类效果要好.

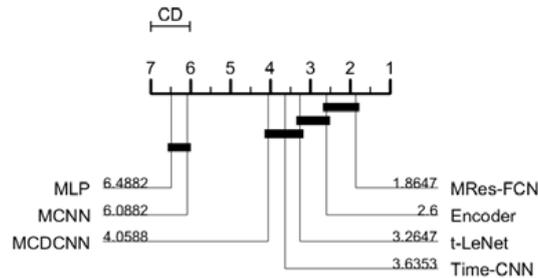


图 11 基于深度学习方法的临界差分图

为了更进一步地分析这些方法, 图 12 展示了这些方法较 MRes-FCN 算法的相对准确率. 一条红色的直线将一个坐标分成 3 部分: 直线上、直线上方的上三角和直线下方的下三角. 直线上的点代表该方法与 MRes-FCN 的分类准确率是一样的, 上三角区域内的点代表 y 轴的方法(6 种方法之一)分类准确率更高, 下三角区域的点代表 x 轴的方法(MRes-FCN)分类准确率更高. 相比于 t-LeNet 和 MCNN 方法, MRes-FCN 以压倒性的优势战胜这两种方法, 这两个方法均只有一个数据点与本文的方法分类效果一致, 落在了直线上, 而其他的数据点都落在下三角区域. 对于 MLP、time-CNN、Encoder 和 MCDCNN 方法, 大部分的数据点落在了下三角区域, 只有 10–20 个数据点落在了上三角, 这说明 MRes-FCN 明显优于 MLP、time-CNN、Encoder 和 MCDCNN.

其次, 由于本文与 ResNet、FCN 以及 LSTM-FCN 这 3 种方法具有一定的关联性, 所以将这 3 种方法与本文的方法在 22 个数据集上进行了更详细的分析, 结果见表 5. 从表 5 可以看出: 本文方法在部分数据集上与 ResNet、LSTM-FCN 以及 FCN 方法达到了一样的分类效果; 而对于一些数据集, 本文方法明显优于 ResNet、LSTM-FCN 以及 FCN 方法, 比如 DiatomSizeReduction 数据集和 Wine 数据集. DiatomSizeReduction 数据集属于 IMAGE 类型, 训练实例数有 16 个, 测试实例数有 306 个, 实例长度是 345, 有 4 个类别, 分别是不同种类的硅藻. Wine 数据集属于 SPECTRO 类型, 训练实例数有 57 个, 测试实例数有 54 个, 实例长度是 234, 有 2 个类别, 代表通过光谱仪获得的 2 种不同的葡萄酒. 图 13 展示了这两个数据集和 Coffee 数据集受到数据预处理过程中池化层的影响的变化情况. 可以明显地看出: DiatomSizeReduction 数据集和 Wine 数据集受数据预处理影响比较大, 而 coffee 数据集不受数据预处理的影响. 池化层实际上也是对数据作进一步采样, 所以这表明本文的方法针对于受时间尺度影响较大的数据集具有更好的效果.

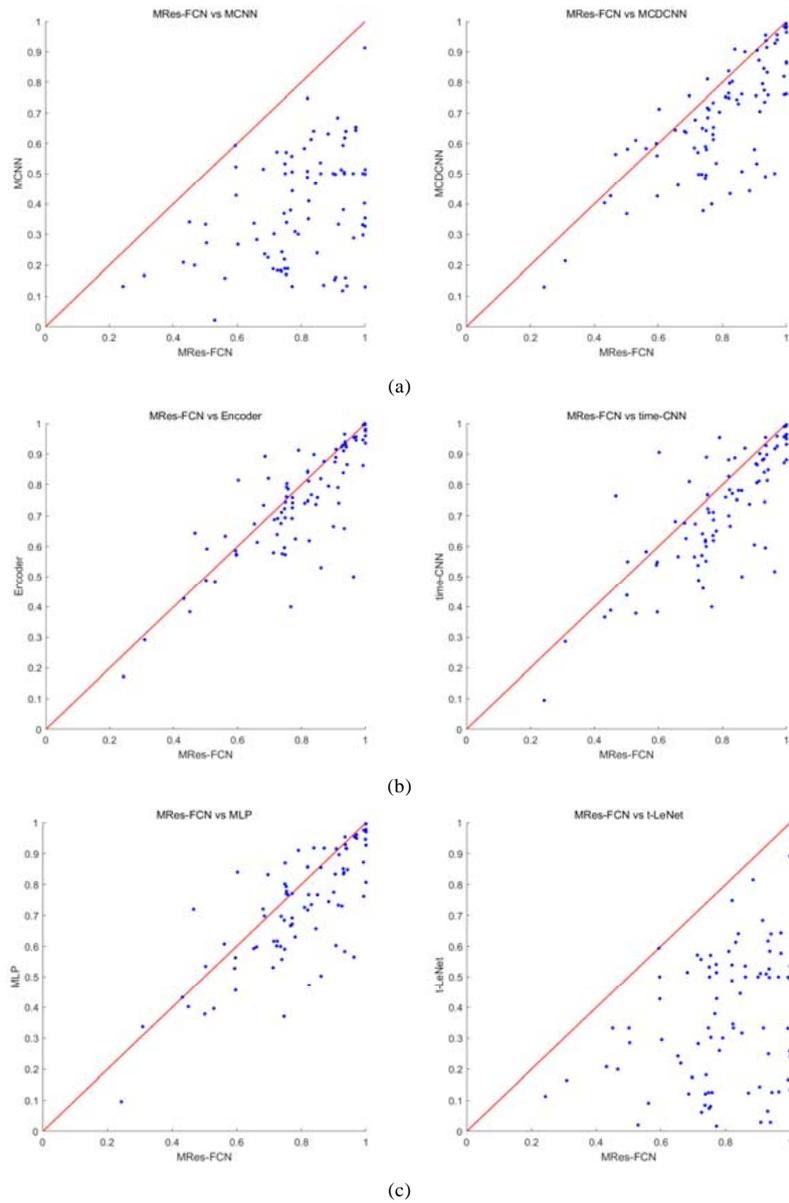


图 12 基于深度学习的分类器与 MRes-FCN 的相对准确率

表 5 实验结果

数据集	类型	MRes-FCN	ResNet	LSTM-FCN	FCN
Coffee	SPECTRO	1.000	1.000	1.000	1.000
DiatomSizeReduction	IMAGE	0.892	0.301	0.306	0.313
DistalPhalanxOutlineAgeGroup	IMAGE	0.845	0.717	0.748	0.710
DistalPhalanxOutlineCorrect	IMAGE	0.810	0.771	0.758	0.760
DistalPhalanxTW	IMAGE	0.715	0.665	0.707	0.690
Earthquakes	SENSOR	0.820	0.712	0.728	0.727
ECG200	ECG	0.940	0.874	0.874	0.889
ECGFiveDays	ECG	1.000	0.975	0.991	0.987
FordA	SENSOR	0.925	0.920	0.924	0.904
GunPoint	MOTION	1.000	0.991	0.998	1.000
Ham	SPECTRO	0.771	0.757	0.711	0.718

表 5 实验结果(续)

数据集	类型	MRes-FCN	ResNet	LSTM-FCN	FCN
MiddlePhalanxOutlineAgeGroup	IMAGE	0.713	0.569	0.563	0.553
Plane	SENSOR	1.000	1.000	1.000	1.000
ProximalPhalanxTW	IMAGE	0.823	0.780	0.791	0.767
ShapeletSim	SIMULATED	0.861	0.779	0.766	0.724
Trace	SENSOR	1.000	1.000	0.250	1.000
UWaveGestureLibraryAll	MOTION	0.964	0.860	0.918	0.817
UWaveGestureLibraryY	MOTION	0.686	0.670	0.679	0.639
Wine	SPECTRO	0.963	0.744	0.675	0.587
WordSynonyms	IMAGE	0.661	0.622	0.661	0.564

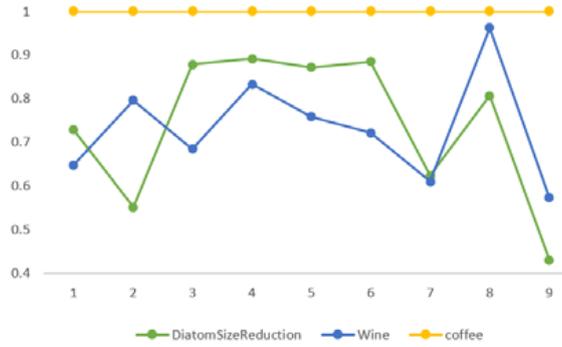


图 13 不同数据集受到数据预处理阶段影响的分类准确率

3.3.3 性能分析

本文方法主要分为数据预处理阶段和全卷积网络与残差网络结合阶段，我们采用消融法对本文方法进行实验分析。实验结果如图 14 所示，这 4 种方法依次是去掉数据预处理阶段的只有全卷积网络与残差网络结合阶段的 Res-FCN、只有残差网络的 ResNet、只有 3 个卷积层的全卷积的 FCN 以及本文的多尺度残差全卷积神经网络的 MRes-FCN。可见：在这 12 个数据集上，MRes-FCN 方法，即将数据预处理和全卷积网络与残差网络结合都包括在内的方法，具有更好的分类效果；而 Res-FCN 方法是这 4 种方法中最差的，说明数据预处理过程对网络的分类效果能够产生很大的影响。

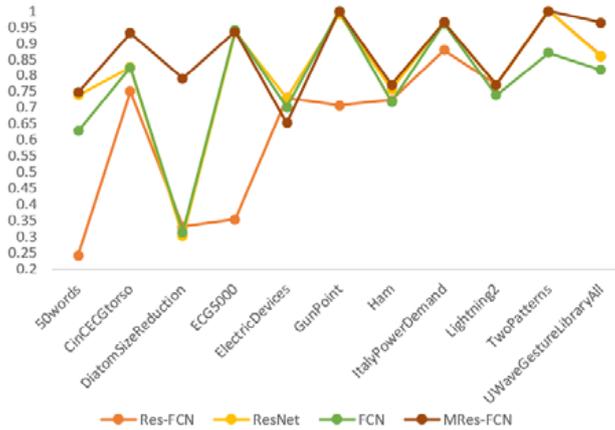


图 14 消融法得到的实验结果

另外，本文对所提出的方法在数据集上的表现进行了探究，见表 6。表 6 展示了针对 UCR 的每一类数据集，这 9 种方法的分类准确率的累计和。本文的 MRes-FCN 方法在所有类型的数据集上均优于 CNN、MCNN、MCDCNN、MLP、t-LeNet 和 Encoder 方法；在 SIMULATED、SPECTRO 和 ECG 数据集上，明显优于 FCN

方法; 在 ECG 数据集上, 优于 ResNet 方法. 另外, 由表 5 可以看到: 本文方法在 ECG200、ECGFiveDays 数据集上比任何现有的方法都更具优势.

表 6 实验结果

Data type	MRes-FCN	ResNet	FCN	CNN	MCNN	MCDCNN	MLP	t-LeNet	Encoder
DEVICE 6	3.771	4.377	4.342	3.339	2.312	2.918	2.781	2.076	3.335
ECG 7	6.648	6.501	6.549	6.004	2.960	5.847	6.249	2.528	6.466
IMAGE 29	22.175	23.141	22.962	21.050	10.414	20.120	21.29	10.206	21.731
MOTION 15	10.236	10.756	10.477	8.345	4.477	7.831	9.007	3.909	9.725
SENSOR 16	13.286	13.723	13.493	12.298	7.251	11.590	11.87	7.019	12.718
SIMULATED6	5.464	5.588	5.356	4.959	2.200	4.831	5.016	1.913	4.921
SPECTRO 7	5.855	6.034	5.550	5.248	2.889	4.840	5.497	3.107	4.936

4 总结与展望

本文提出的方法通过与基于距离的方法、基于特征的方法和基于深度学习的方法进行比较, 证明本文方法有较高的分类准确率. 通过该方法自身的比较, 说明数据预处理阶段和全卷积网络与残差网络阶段均能够对分类产生影响, 本文的两个阶段的结合能够提升分类效果.

总的来说, 我们提出的方法具有一定的分类准确率. 但是, 本文提出的方法数据处理工作比较多, 平滑滤波也很浪费内存, 这样操作加大了时间和空间复杂度. 另外, 基于深度学习的时间序列分类算法的解释性一直是一个有争议的问题. 虽然深度学习算法的准确率往往比较高, 但是很多学者依旧致力于研究非深度学习算法, 这些算法的决策过程研究者可以很容易理解, 也能够容易给出合理的解释. 所以, 本文后续研究工作在于:

- (1) 对于深度学习算法用于时间序列分类问题, 在注意提升其准确率的同时, 也应注重提升其可解释性;
- (2) 本文方法仅在通用数据集上进行了研究验证, 后续还需进一步进行专门的应用研究;
- (3) 随着机器视觉的发展与成功, 除了采用对时间序列数据进行操作之外, 是否可以考虑直接对时间序列图片进行研究, 这样就将时间序列数据问题转化为图像问题.

References:

- [1] Yang Q, Wu X. Ten challenging problems in data mining research. *Int'l Journal of Information Technology & Decision Making*, 2006, 5(4): 597–604.
- [2] Esling P, Agon C. Time-series data mining. *ACM Computing Surveys*, 2012, 45(1): 1–34.
- [3] Bagnall A, Lines J, Bostrom A, *et al.* The great time series classification bake off: A review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, 2017, 31(3): 606–660.
- [4] Ping YP, Zang SY. Crop classification based on MODIS time series and phenological characteristics. *Journal of Natural Resources*, 2016, 31(3): 503–513 (in Chinese with English abstract).
- [5] Du BJ, Zhang J, Wang ZM, *et al.* Crop classification using Sentinel-2A NDVI time series and object-oriented decision tree method. *Journal of Geoinformatics*, 2019, 21(5): 740–751 (in Chinese with English abstract).
- [6] Huang DF, Yang DJ. Dynamic gesture recognition based on improved ND-DTW algorithm. *Electronic Technology*, 2017, 30(3): 37–40 (in Chinese with English abstract).
- [7] Xie X, Zhang LQ, Wang J. Application of residual network in cry recognition of infants. *Journal of Electronics and Information Technology*, 2019, 41(1): 233–239 (in Chinese with English abstract).
- [8] Chen H, Tang F, Cohn A, *et al.* Model metric co-learning for time series classification. In: *Proc. of the 21st Int'l Joint Conf. on Artificial Intelligence (IJCAI 2015)*. 2015. 3387–3394.
- [9] Szegedy C, Liu W, Jia Y, *et al.* Going deeper with convolutions. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2015. 1–9.
- [10] Srivastava N, Hinton G, Krizhevsky A, *et al.* Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 2014, 15(1): 1929–1958.

- [11] Le Q, Mikolov T. Distributed representations of sentences and documents. In: Proc. of the Int'l Conf. on Machine Learning. 2014. 1188–1196.
- [12] Goldberg Y. A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research*, 2016, 57: 345–420.
- [13] Wang Z, Yan W, Oates T. Time series classification from scratch with deep neural networks: A strong baseline. In: Proc. of the 2017 Int'l Joint Conf. on Neural Networks (IJCNN). IEEE, 2017. 1578–1585.
- [14] Lines J, Bagnall A. Time series classification with ensembles of elastic distance measures. *Data Mining and Knowledge Discovery*, 2015, 29(3): 565–592.
- [15] Bagnall A, Lines J, Hills J, *et al.* Time-series classification with COTE: The collective of transformation-based ensembles. *IEEE Trans. on Knowledge and Data Engineering*, 2015, 27(9): 2522–2535.
- [16] Lines J, Taylor S, Bagnall A. Hive-Cote: The hierarchical vote collective of transformation-based ensembles for time series classification. In: Proc. of the 16th IEEE Int'l Conf. on Data Mining (ICDM). IEEE, 2016. 1041–1046.
- [17] Harris ZS. Distributional structure. *Word*, 1954, 10(2-3): 146–162.
- [18] Lin J, Khade R, Li Y. Rotation-invariant similarity in time series using bag-of-patterns representation. *Journal of Intelligent Information Systems*, 2012, 39(2): 287–315.
- [19] Lin J, Keogh E, Wei L, *et al.* Experiencing SAX: A novel symbolic representation of time series. *Data Mining and Knowledge Discovery*, 2007, 15(2): 107–144.
- [20] Ye L, Keogh E. Time series shapelets: A new primitive for data mining. In: Proc. of the 15th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2009. 947–956.
- [21] Lines J, Davis LM, Hills J, *et al.* A shapelet transform for time series classification. In: Proc. of the 18th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2012. 289–297.
- [22] Grabocka J, Schilling N, Wistuba M, *et al.* Learning time-series shapelets. In: Proc. of the 20th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. ACM, 2014. 392–401.
- [23] Senin P, Malinchik S. Sax-VSM: Interpretable time series classification using sax and vector space model. In: Proc. of the 13th IEEE Int'l Conf. on Data Mining. IEEE, 2013. 1175–1180.
- [24] Schäfer P. The BOSS is concerned with time series classification in the presence of noise. *Data Mining and Knowledge Discovery*, 2015, 29(6): 1505–1530.
- [25] Cui Z, Chen W, Chen Y. Multi-scale convolutional neural networks for time series classification. 2016. <https://arxiv.org/pdf/1603.06995.pdf>
- [26] Serrà J, Pascual S, Karatzoglou A. Towards a universal neural network encoder for time series. In: Proc. of the Artificial Intelligence Research and Development: Current Challenges. 2018. 120–129.
- [27] Karim F, Majumdar S, Darabi H, *et al.* LSTM fully convolutional networks for time series classification. *IEEE Access*, 2017, 6(99): 1662–1669.
- [28] Fawaz HI, Forestier G, Weber J, *et al.* Deep learning for time series classification: A review. *Data Mining and Knowledge Discovery*, 2019, 33(4): 917–963.
- [29] Han C, Shi L. ML-ResNet: A novel network to detect and locate myocardial infarction using 12 leads ECG. *Computer Methods and Programs in Biomedicine*, 2020, 185: 105–138.
- [30] Ramzan F, Khan MUG, Rehmat A, *et al.* A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks. *Journal of Medical Systems*, 2020, 44(2): 37.
- [31] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 3431–3440.
- [32] Zheng Y, Liu Q, Chen E, *et al.* Time series classification using multi-channels deep convolutional neural networks. In: Proc. of the Int'l Conf. on Web-age Information Management. Cham: Springer, 2014. 298–310.
- [33] Guennec AL, Malinowski S, Tavenard R. Data augmentation for time series classification using convolutional neural networks. In: Proc. of the ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data. Riva Del Garda, 2016. 1–8.
- [34] Zhao B, Lu H, Chen S, *et al.* Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, 2017, 28(1): 162–169.

- [35] Keogh EJ, Pazzani MJ. Derivative dynamic time warping. In: Proc. of the 2001 SIAM Int'l Conf. on Data Mining. Society for Industrial and Applied Mathematics, 2001. 1–11.
- [36] Stefan A, Athitsos V, Das G. The move-split-merge metric for time series. IEEE Trans. on Knowledge and Data Engineering, 2012, 25(6): 1425–1438.
- [37] Jeong YS, Jeong MK, Omiaomu OA. Weighted dynamic time warping for time series classification. Pattern Recognition, 2011, 44(9): 2231–2240.
- [38] Batista GE, Keogh EJ, Tataw OM, *et al.* CID: An efficient complexity-invariant distance for time series. Data Mining and Knowledge Discovery, 2014, 28(3): 634–669.
- [39] Yuan J, Douzal-Chouakria A, Yazdi SV, *et al.* A large margin time series nearest neighbour classification under locally weighted time warps. Knowledge and Information Systems, 2019, 59(1): 117–135.
- [40] Zhao JP, Itti L. shapeDTW: Shape dynamic time warping. 2016. <https://arxiv.org/pdf/1606.01601.pdf>
- [41] Rakthanmanon T, Keogh E. Fast shapelets: A scalable algorithm for discovering time series shapelets. In: Proc. of the 2013 SIAM Int'l Conf. on Data Mining. Society for Industrial and Applied Mathematics. 2013. 668–676.
- [42] Deng H, Runger G, Tuv E, *et al.* A time series forest for classification and feature extraction. Information Sciences, 2013, 239: 142–153.
- [43] Baydogan MG, Runger G, Tuv E. A bag-of-features framework to classify time series. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2013, 35(11): 2796–2802.

附中文参考文献:

- [4] 平跃鹏, 臧淑英. 基于 MODIS 时间序列及物候特征的农作物分类. 自然资源学报, 2016, 31(3): 503–513.
- [5] 杜保佳, 张晶, 王宗明, 等. 应用 Sentinel-2A NDVI 时间序列和面向对象决策树方法的农作物分类. 地球信息科学学报, 2019, 21(5): 740–751.
- [6] 黄东方, 杨晶东. 基于改进 ND-DTW 算法的动态手势识别. 电子科技, 2017, 30(3): 37–40.
- [7] 谢湘, 张立强, 王晶. 残差网络在婴幼儿哭声识别中的应用. 电子与信息学报, 2019, 41(1): 233–239.



张雅雯(1996—), 女, 硕士生, 主要研究领域为机器学习, 数据挖掘.



刘海洋(1987—), 男, 博士, 讲师, CCF 专业会员, 主要研究领域为数据挖掘, 机器学习.



王志海(1975—), 男, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为数据挖掘, 机器学习.



曾昭博(1998—), 男, 学士, 主要研究领域为文本分类, 主动学习.