

根据 2.2 节中的分析, 整个数据集产生的 shapelet 个数达到 $O(nm^2)$ 的量级, 非常庞大, 由于评估一个 shapelet 的时间开销也不容小觑, 所以单位评估时间被这个数量放大之后, 时间开销爆炸性增长。

然而, 我们发现, 如果我们使用“shapelet 转换”的方式配合定点浮动距离 (式 (3)) 来构造分类器的话, 我们可以通过固定 shapelet 的长度来大幅缩减 shapelet 候选集的规模, 接下来我们就来说明这件事。

我们看到“shapelet 转换”的第一步是从候选 shapelet 中选择判别性强的若干个 shapelet 出来, 第二步是基于这些选择出来的 shapelet 构造特征空间, 将原始数据转换至特征空间, 第三步是在特征空间中对转换后的数据进行分类。

在这个过程中, 我们能够发现最后对数据分类所倚赖的关键是构建特征空间的 shapelet 的判别性。而 shapelet 的判别性体现在, 和该 shapelet 同类的时间序列, 与该 shapelet 的距离要足够小, 而反之则要与该 shapelet 的距离足够大。而根据转换的方式 (参见 2.3 节), 当原始数据被转换到特征空间之后, 这就体现在和 shapelet 同类的时间序列, 转换后在该 shapelet 对应的坐标轴上的范数要比较小, 而反之则对应的范数要比较大。

所以我们看到, 在特征空间中分类的关键依据, 其实是原始数据被转换到特征空间之后的范数。如果有两个不同的特征空间, 原始数据被转换到它们之中后拥有相同的空间范数, 那最终两个特征空间中的分类依据就是相同的, 分类效果也会大同小异, 在某种程度上, 我们可以认为这两个特征空间是等价的。

现在假设原来的 shapelet 候选集是 A , 如果我们找到一个 A 的很小的子集 B , 使得: 从 A 里面找出的任意一组 shapelet P , 任意原始数据 x 被转换到 P 构造的特征空间中的范数记录为 $\|x\|_P$, B 中都存在对应的一组 shapelet Q , 原始数据 x 被转换到 Q 构造的特征空间中的范数记录为 $\|x\|_Q$, 且任意 x , 都有 $\|x\|_P \approx \|x\|_Q$, 即这两个特征空间是等价的。那么这样的 B 显然具备构造等价特征空间的能力。而如此一来, 我们就能够抛弃原来巨大的候选 shapelet 集 A , 而只选用它的很小的子集 B , 这样时间开销会得到大幅降低。

我们现在就来证明这件事, 即存在上述那样一个小子集 B , 其具备构造同等特征空间的 shapelet。此事关键在于证明如下的定理:

定理 1. 将一个 shapelet s 分割成若干段 $C=\{s_1, s_2, \dots, s_n\}$, 使得它们按序拼接起来构成完整的 s , 我们称这样的 C 为 s 的一个划分集。对任意时间序列 x , 其通过定点距离 (式 (2)) 转换至 s 构造的特征空间中的欧氏范数记录为 $\|x\|_s$, 而其通过定点距离转换至 C 构造的特征空间中的欧氏范数记录为 $\|x\|_C$, 则我们有 $\|x\|_s = \|x\|_C$ 。

证明: 不妨设 $s=(s_i, s_{i+1}, \dots, s_{i+k})$, 其中 i 是 s 在原始时间序列中的起始位置, $k+1$ 为其长度。令 t_j 为 C 中 s_j 的终止位置, 且规定 $t_0=i-1$, $t_n=i+k$, 那么我们有 $s_j=(s_{t_{j-1}+1}, s_{t_{j-1}+1}, \dots, s_{t_j})$ 。对于任意时间序列 x , 我们设 $x=(x_1, x_2, \dots, x_m)$, 显然 $m>i+k$ 。

则根据 shapelet 转换的规则, 我们有:

$$\begin{aligned} \|x\|_C &= \left\| \left(\sqrt{\sum_{p=t_0+1}^{t_1} (x_p - s_p)^2}, \dots, \sqrt{\sum_{p=t_{n-1}+1}^{t_n} (x_p - s_p)^2} \right) \right\|_2 \\ &= \sqrt{\sum_{p=i}^{t_1} (x_p - s_p)^2 + \sum_{q=2}^n \sum_{p=t_{q-1}+1}^{t_q} (x_p - s_p)^2} \\ &= \sqrt{\sum_{p=i}^{t_n} (x_p - s_p)^2} = \sqrt{\sum_{p=i}^{i+k} (x_p - s_p)^2} = \|x\|_s \end{aligned} \quad (4)$$

证毕。

根据上述定理我们发现, 在当前叙述背景下, 一个 shapelet 构造的特征空间, 和它的划分集构造的特征空间可以看成是等价的。尽管最后我们用作转换的距离度量方式是定点浮动距离 (式 (3)) 而不是定点距离 (式 (2)), 但是定点浮动距离只是定点距离的松弛版本, 所以最后转换后的数据的范数与定理 1 中的范数并不会相差太远, 这样我们依旧有 $\|x\|_C \approx \|x\|_s$ 。所以在定点浮动距离作为转换坐标的计算公式的前提下, 我们得到 shapelet 构造的特征空间和它们的划分集构造的特征空间是近似的。

而另一方面, 每一个 shapelet 都可以划分为若干短的 shapelet, 所以我们只需要保留 shapelet 候选集里足够短的 shapelet, 就足够我们找到好的 shapelet 来构造好的特征空间了。如此一来, 我们只需要选取长度为某个固定小数值的的所有 shapelet 来作为 shapelet 候选集即可, 这个数值一般取 3 或者 4 即可, 我们把这样的 shapelet 称作为“微局部特征”。我们将长度设置为 3 或者 4 是一种折衷, 当长度设置比 3 更小的时候, 这些短 shapelet 将失去统计意义, 因为时序数据是连续的数据,

在某个时间点的值并不能构成统计意义上的特征,它们提供的分类信息也因而不具备高可信度.而当长度比4还大时,将无法覆盖某些短 shapelet,比如长度为4的 shapelet,失去这些 shapelet 会对分类结果造成影响.由这些微局部特征构成的集合正是我们需要寻找的原候选集的小子集.对于 n 个长度为 m 的时间序列构建的数据集,我们构建的 shapelet 候选集只有 $O(nm)$ 的规模,而不再是 $O(nm^2)$ 的量级.

3.3 shapelet 判别性的评价指标

在具备 shapelet 候选集后,需要从中选取判别性强的 shapelet 来作为构建特征空间的一组基底.

容易知道,选取 shapelet 需要量化 shapelet 的判别性能,目前普遍采用的做法是用最佳信息增益来量化 shapelet 的判别性能.该做法如算法1所示.

算法1. 计算 shapelet 的信息增益

输入: shapelet s , data set $T=(X, Y)$
输出: prime information gain of s

- 1) For each time series (x, y) in T
- 2) Calculate $D(s, x)$, namely distance between s and x ;
- 3) Depict $D(s, x)$ with its label y in a real line r ;
- 4) End for
- 5) Find each possible segmentation in real line r to build C ;
- 6) Set $prime_information_gain=0$;
- 7) For each segmentation c in C
- 8) Calculate information gain of c as g ;
- 9) If $g > prime_information_gain$
- 10) $Prime_information_gain=g$;
- 11) End if
- 12) End for
- 13) Return $prime_information_gain$;

从该算法中可以看到,如果要计算一个 shapelet 的最佳信息增益,则需要计算所有“分割”的信息增益再挑出里面最大的.这样做非常耗时,尤其当数据集里面数据比较多时,则可能的“分割”的数目指数增长,评价一个 shapelet 的代价变得相当大.

为了避免这个问题,我们决定采用广义雷利熵来作为 shapelet 的判别性能的评价指标.对于一个实数轴上的二分类问题,我们假设两类数据的集合分别为 P 和 Q ,则广义雷利熵的计算公式如下式所示:

$$GRQ(P, Q) = \frac{|\mu(P) - \mu(Q)|}{\sigma^2(P) + \sigma^2(Q)} \quad (5)$$

式中, $\mu(\cdot)$ 是均值函数, $\sigma^2(\cdot)$ 是方差函数.

利用广义雷利熵来作为判别指标后,我们有效避开了寻找最佳分割的过程,不再需要计算每种分割的

信息增益来寻找最佳信息增益了,这种耗时的操作因此也被去除了.因为 shapelet 的评估是一个单位操作,所以在单位操作上带来的时间节省,被操作次数放大后,会得到非常可观的优化效果.如此一来,评估一个 shapelet 的算法如下所示.

算法2. 计算 shapelet 的判别性能(广义雷利熵)

输入: shapelet s , data set $T=(X, Y)$
输出: prime information gain of s

- 1) For each time series (x, y) in T
- 2) Calculate $D(s, x)$ by formula (3);
- 3) Depict $D(s, x)$ with its label y in a real line r ;
- 4) End for
- 5) Calculate general Rayleigh quotient grq of points in r ;
- 6) Return grq ;

3.4 特征空间中的分类器

正如我们2.3节中所述,使用“shapelet 转换”的方式必须承担转换后的数据在特征空间中的分布不定性这一代价,所以要想保证分类精度,需要在转换后的特征空间中训练多个不同类型的分类器,保证不遗漏可能的数据分布.但是这样做的时间代价相当高.

自然地,我们想要避免这种操作来降低时间开销.由于我们无法保证特征空间中的数据具有某种特定的分布特性,所以我们只能用不同类型的分类器把所有可能的分布特性都考虑进去.而造成这种现象的根源在于选择 shapelet 的环节.我们确实选择了最具判别性的 shapelet 来作为特征空间的基底,但是却忽略了它们的组合效应.在经典的基于 shapelet 的算法中,他们选择最佳信息增益作为判别指标,这使得在“shapelet 转换”后,特征空间的数据在每个坐标轴上的投影具有非常好的线性可分性,但是在每个坐标上的分量具有很好的线性可分性,并不能保证在整个空间上具备很好的线性可分性,这是问题的关键所在.

所以若我们可以保证原始数据被转换到相应的特征空间之后,具备线性可分性这一分布特性.我们就能使用单个的 SVM 去替代集成的分类器而避免训练集成分类器这一耗时操作.

我们现在断言,如果我们使用广义雷利熵作为 shapelet 的评价指标,那么高得分的 shapelet 能够保证特征空间中数据的线性可分性,这依赖于如下的定理:

定理2. 对于一个时序数据的二分类问题, shapelet 的广义雷利熵与特征空间中数据的线性可分性存在正

相关。

证明: 我们假设选择了 k 个 shapelet 构造好了一个特征空间, 那我们要证明的是, 将其中某个 shapelet s_1 替换为拥有更大广义雷利熵的 shapelet s_2 , 数据的线性可分性会提高. 我们假设原始数据中有两个类的时间序列, 根据中心极限定理, 它们通过这 k 个 shapelet 转换到特征空间之后实际上形成了两个随机向量 A, B , 且 A 和 B 服从球形正态分布。

我们自然地按如下方式来定义这两类数据的线性可分性:

$$LS = \max_{\|\vec{w}\|=1, \vec{w} \in R^k} P((A-B) \cdot \vec{w} > 0) \quad (6)$$

式中, LS 是线性可分性, $P(\cdot)$ 是概率函数. 我们现规定 $\mu(\cdot)$ 是均值函数, 返回随机向量的均值向量; $\sigma^2(\cdot)$ 为随机向量的协方差矩阵的对角线函数, 它返回由协方差矩阵的对角线构成的向量; $[\cdot]^2$ 是平方函数, 对向量中每一维度的数据取平方来得到一新向量. 根据 A, B 服从球形高斯分布, 我们有:

$$(A-B) \cdot \vec{w} \sim N((\mu(A)-\mu(B)) \cdot \vec{w}, (\sigma^2(A)+\sigma^2(B)) \cdot [\vec{w}]^2) \quad (7)$$

式中, $N(\cdot, \cdot)$ 为正态分布的符号. 结合式 (7), 根据线性可分性的定义以及 \vec{w} 的取值范围 (式 (6)), 对于最佳的 \vec{w} , 我们显然可以得到如下关系:

$$(\mu(A) - \mu(B)) \cdot \vec{w}^* \geq 0 \quad (8)$$

$$(\mu_i(A) - \mu_i(B)) \cdot \vec{w}_i^* \geq 0 \quad 1 \leq i \leq k \quad (9)$$

式中, \vec{w}^* 表示其为使式 (6) 取得最大值的最佳 \vec{w} .

根据广义雷利熵的计算式 (式 (5)), 我们若证明任意 $0 < i < k+1$, $|\mu_i(A) - \mu_i(B)|$ 与式 (6) 中的线性可分性成正相关以及 $\sigma_i^2(A) + \sigma_i^2(B)$ 与其成负相关, 则可完成定理的证明。

但这件事并不难, 假设我们已经拥有最佳的 LS 值 (式 (6)), 现在将某个特定的 $\mu_i(A) - \mu_i(B)$ 替换为拥有更大绝对值的 $\mu_i(A') - \mu_i(B')$, 保持其它数值不变. 我们不妨假设 $(\mu_i(A') - \mu_i(B')) \cdot \vec{w}_i^* > 0$, 因为若其小于 0, 我们只需将相应的 \vec{w}_i^* 取成相反数即可. 由于 $(\mu_i(A') - \mu_i(B')) \cdot \vec{w}_i^* > (\mu_i(A) - \mu_i(B)) \cdot \vec{w}_i^* \geq 0$, 其他项不变, 所以 $(\mu(A') - \mu(B')) \cdot \vec{w}^* > (\mu(A) - \mu(B)) \cdot \vec{w}^* \geq 0$. 再根据正态分布的特性及式 (8), 我们有:

$$LS' \geq P((A' - B') \cdot \vec{w}^* > 0) > P((A - B) \cdot \vec{w}^* > 0) = LS \quad (10)$$

这样即证明了 $|\mu_i(A) - \mu_i(B)|$ 与式 (6) 中的线性可分性成正相关, 同样地方法可以证明与 $\sigma_i^2(A) + \sigma_i^2(B)$ 式

(6) 中的线性可分性成负相关, 此处不再赘述。

证毕。

定理 2 为特征空间中的转换后的数据的线性可分性这一分布特性提供了理论支撑, 所以在我们提出的新算法中, 当原始时间序列被转换到特征空间之后, 我们可以放心地使用单个的 SVM 来执行分类, 而不再需要训练集成分类器, 这一结果极大优化了时间性能。

4 算法框架与实验结果

4.1 主体算法

我们在此节给出最终的算法框架, 如算法 3 和算法 4 所示。

算法 3. 基于微局部特征的时序数据二分类算法 (分类器构造)

输入: shapelet length L , binary data set $T=(X, Y)$, shapelet number N , relaxation factor r and l

输入: a priority queue of N shapelet, a SVM classifier

初始化: a null priority queue Q , a null set S , a null vector set V

- 1) For each time series (x, y) in T
- 2) Capture all the L -length shapelet in (x, y) and put them in S ;
- 3) End for
- 4) For each L -length shapelet (s, z) in S
- 5) Input (s, z) and T to Algorithm 2, and get a return value grq ;
- 6) Add (s, z) to Q , and set its priority as grq ;
- 7) End for
- 8) $Q=N$ -length subqueue of Q ;
- 9) For each time series (x, y) in T
- 10) Transform (x, y) to a N -dimensional label vector (v, y) by calculate distance with shapelet in Q by formula (3);
- 11) Add (v, y) to V ;
- 12) End for
- 13) Train a SVM svm in V ;
- 14) Return Q, svm .

算法 4. 基于微局部特征的时序数据二分类算法 (数据分类)

输入: shapelet queue Q , SVM classifier svm , data x

输出: class label of x

- 1) Transform x to the N -dimensional vector v by calculate distance with shapelet in Q by formula (3);
- 2) Use svm to classify v , get a label y ;
- 3) Return y .

4.2 实验数据集

为了公平起见, 我们选择 Bagnall 等人在其工作中所使用的数据集^[1]. 他们在相关研究工作中精心筛选数据集以及各种算法, 并做出了比较公平公正的对比, 他们所使用的数据集也被选作时序数据分类算法社区的

标准数据集^[10]. 先对数据集做如下介绍:

Ham, 火腿光谱图数据, 通过对光谱图进行分类来判断火腿的种类, 训练集 109 个数据, 测试集 105 个数据, 数据长度 431.

MPOC, 全称 Middle Phalanx Outline Correct, 手指中部骨节的 X 光投影轮廓图. 科学家根据该数据来判断人们所处的年龄阶段, 训练集 600 个数据, 测试集 291 个数据, 数据长度 80.

Eq, 全称 Earthquakes, 用传感器捕捉的地震波数据, 用来判断近期内是否会有地震发生, 数据来自于北加利福尼亚地震研究中心. 训练集 322 个数据, 测试集 139 个数据, 数据长度 512.

Herring, 鲑鱼的耳石轮廓, 该数据用于生物多样性研究, 通过耳石轮廓对应的时序数据来判定鲑鱼生活的地区. 训练集 64 个数据, 测试集 64 个数据, 数据长度 512.

IPD, 即 Italy Power Demand, 意大利人民不同季度生活用电时序数据, 不同类别的时序数据对应不同季度的用电水平. 训练集 67 个, 测试集 1029 个, 数据长度 24.

Wine, 葡萄酒的光谱图, 光谱图上不同种类的时序数据对应不同种类的葡萄酒. 训练集 57 个数据, 测试集 54 个数据, 数据长度 234.

用于做实验的数据集来自于实际应用的各方各面, 包括天文地理, 衣食住行等多个领域, 也从侧面反映了时序数据有着广泛的应用.

4.3 对比算法与算法超参

本文针对基于 shapelet 的时序分类算法进行分析与改进, 旨在提升算法的分类精度和降低算法的时间开销, 所选对比算法为基于 shapelet 的时序分类算法中的优秀算法, 介绍如下:

FS, 该算法专注于时间开销, 是现有的基于 shapelet 的算法中平均时间开销比较低的, 但是其牺牲了部分精度, 采用近似的方法选取 shapelet^[7].

LS, 是用学习的方式获取 shapelet 的代表算法, 通过把获取判别性 shapelet 这一难题转换为优化问题, 并用梯度下降来求得最优 shapelet, 具有非常不错的时间开销和分类精度^[8].

ST, 是当今主流的基于 shapelet 的时序数据分类算法, 它通过 shapelet 将原始数据转换至特征空间, 在特征空间里训练集成分类器进行分类, 拥有较高的时

间复杂度, 但是分类精度属于领域内的顶尖^[12].

COTE, 是基于 shapelet 的集成算法中的集大成者, 除了集成基于 shapelet 的时序数据分类算法, 也集成了其它类型的时序数据分类算法, 因此是四个对比算法中时间复杂度最高的, 同时也是分类精度最好的^[13].

值得一提的是, 在 Bagnall 等人的工作中, ST 和 COTE 不仅是基于 shapelet (局部特征) 的时序数据分类算法中的最好的, 也是所有时序数据分类算法中分类精度最好的^[1]. 详情请见图 1^[1].

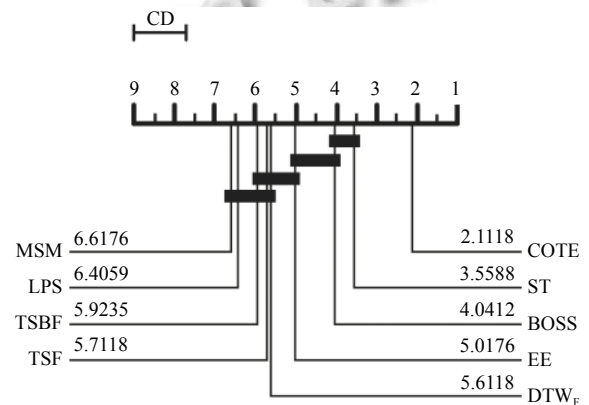


图 1 最佳算法在 UCI 数据集上的分类精度平均排名^[1]

另外, 我们提出的基于微局部特征的时序数据二分类算法, 它的英文名称为“mini-shapelet based algorithm”, 简称为“MS”. MS 的超参选择集如表 1 所示.

表 1 MS 的超参取值

MS 的超参	取值范围
左右松弛因子 l, r	{3, 4}
Shapelets 的长度 L	{3, 4}
选择的 shapelet 个数 N	{10, 50, 100, 250, 500, 1000, 1500}

4.4 分类性能对比

关于分类性能, 我们延用时序分类算法社区里面的硬指标, 即分类精度对比以及分类精度的排名对比, 具体的对比数据如表 2, 表 3 以及图 2 所示.

表 2 分类精度对比

数据集	FS	LS	ST	COTE	MS
Ham	0.677	0.832	0.808	0.805	0.838
MPOC	0.716	0.822	0.815	0.801	0.897
Eq	0.747	0.742	0.737	0.747	0.871
Herring	0.558	0.628	0.653	0.632	0.875
IPD	0.909	0.952	0.953	0.970	0.978
Wine	0.794	0.524	0.926	0.904	1.000

表3 分类精度排名对比

数据集	FS	LS	ST	COTE	MS
Ham	5	2	3	4	1
MPOC	5	2	3	4	1
Eq	2	4	5	2	1
Herring	5	4	2	3	1
IPD	5	4	3	2	1
Wine	4	5	2	3	1
Average	4.3	3.5	3.0	3.0	1.0

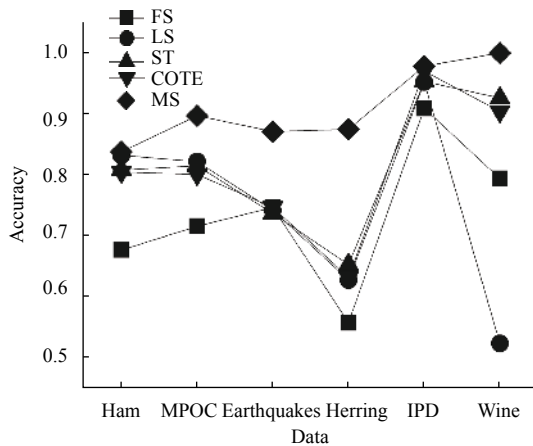


图2 算法分类精度点线对比图

从表2以及表3中可以看出,在所有测试数据集上面,基于微局部特征的分类算法(MS)都取得了最优的分类精度,并且在某些数据集上还具有非常明显的分类精度优势,比如Herring和Eq这两个数据集。综合来看,相对于其它4种经典的基于shapelet的分类算法,基于微局部特征的分类算法显而易见地具有最佳的分类能力。而在Wine这个数据集上,它甚至取得了无任何错误的分类结果,体现了其极强的分类能力。因而综上所述,基于微局部特征的时序数据分类算法相较于4种对比算法有明显的分类精度优势。

从图2中我们可以观察到四种对比算法分类性能不太稳定,都有大幅度的波动。而相比之下,基于微局部特征的分类算法(MS)具有较为稳定的分类性能,其折线波动较小,相对来说比较平稳。所以基于微局部特征的时序数据分类算法在分类表现上稳定性更好,且具有稳定的分类性能优势。

4.5 时间开销对比

关于时间开销,因为5个算法都是使用eager learning的方式进行分类,而且其主要的时间都用在构造分类器上,所以我们主要比较它们的构建分类器的时间。我们在同样的硬件条件(内存:8GB;CPU:2.5GHz)

及软件条件下(OS:Win10;platform:JAVA)进行实验,具体数据如表4及表5所示。

表4 时间消耗对比(单位:秒)

数据集	FS	LS	ST	COTE	MS
Ham	182.8	883.5	51.1	2.8E+05	8.0
MPOC	14.8	133.5	1.3E+05	2.5E+05	36.9
Eq	1.3E+03	1.9E+03	1.5E+03	9.1E+05	110.2
Herring	118.4	678.8	15.3	1.8E+05	7.0
IPD	0.1	99.1	16.7	88.3	0.5
Wine	5.1	111.1	11.2	4.3E+04	1.5

表5 时间消耗排名对比(升序)

数据集	FS	LS	ST	COTE	MS
Ham	3	4	2	5	1
MPOC	1	3	4	5	2
Eq	2	4	3	5	1
Herring	3	4	2	5	1
IPD	1	5	3	4	2
Wine	2	4	3	5	1
Average	2.0	4.0	2.8	4.8	1.3

从表4和表5中可以观察到,除了MPOC和IPD这两个数据集外,基于微局部特征的分类算法(MS)在剩余数据集上都具有最小的时间开销,而在MPOC和IPD这两个数据集上,也仅次于FS这一算法,但是FS在MPOC和IPD上的精度远不如基于微局部特征的时序数据分类算法。再结合表5中的平均排名,我们可以认为相对于其它4个对比算法,在保持最高分类精度的同时,基于微局部特征的时序数据分类算法具有最佳的时间性能。

5 结语

本文针对当前基于局部特征的时序数据分类算法中存在的问题与挑战,在充分的理论依据的支撑下,使用缩减候选集,调整判别性评定指标,修改距离度量以及替换集成分类器4项技术设计了高效实用的新型算法。该基于微局部特征的时序数据分类算法在实验数据集上表现出良好的分类性能和时间性能。通过实验对比,也证明了其对当前研究领域内存在的分类精度不足以及时间开销过高等问题有不错的改进。

参考文献

- 1 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. [doi: 10.1007/s10618-016-

- 0483-9]
- 2 Ye LX, Keogh E. Time series shapelets: A novel technique that allows accurate, interpretable and fast classification. *Data Mining and Knowledge Discovery*, 2011, 22(1–2): 149–182. [doi: [10.1007/s10618-010-0179-5](https://doi.org/10.1007/s10618-010-0179-5)]
 - 3 He Q, Dong Z, Zhuang FZ, *et al.* Fast time series classification based on infrequent shapelets. *Proceedings of the 11th ICMLA International Conference on Machine Learning and Applications*. Boca Raton, FL, USA. 2012. 215–219.
 - 4 Deng HT, Runger G, Tuv E, *et al.* A time series forest for classification and feature extraction. *Information Sciences*, 2013, 239: 142–153. [doi: [10.1016/j.ins.2013.02.030](https://doi.org/10.1016/j.ins.2013.02.030)]
 - 5 Zakaria J, Mueen A, Keogh E, *et al.* Accelerating the discovery of unsupervised-shapelets. *Data Mining and Knowledge Discovery*, 2016, 30(1): 243–281. [doi: [10.1007/s10618-015-0411-4](https://doi.org/10.1007/s10618-015-0411-4)]
 - 6 Baydogan MG, Runger G. Time series representation and similarity based on local autopatterns. *Data Mining and Knowledge Discovery*, 2016, 30(2): 476–509. [doi: [10.1007/s10618-015-0425-y](https://doi.org/10.1007/s10618-015-0425-y)]
 - 7 Rakthanmanon T, Keogh EJ. Fast shapelets: A scalable algorithm for discovering time series shapelets. *Proceedings of the 2013 SIAM International Conference on Data Mining*. Austin, TX, USA. 2013. 668–676.
 - 8 Grabocka J, Schilling N, Wistuba M, *et al.* Learning time-series shapelets. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA. 2014. 392–401.
 - 9 Hou L, Kwok JT, Zurada JM. Efficient learning of timeseries shapelets. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. Phoenix, AZ, USA. 2016. 1209–1215.
 - 10 Hills J, Lines J, Baranauskas E, *et al.* Classification of time series by shapelet transformation. *Data Mining and Knowledge Discovery*, 2014, 28(4): 851–881. [doi: [10.1007/s10618-013-0322-1](https://doi.org/10.1007/s10618-013-0322-1)]
 - 11 Guido RC. Fusing time, frequency and shape-related information: Introduction to the discrete shapelet transform's second generation (DST-II). *Information Fusion*, 2018, 41: 9–15. [doi: [10.1016/j.inffus.2017.07.004](https://doi.org/10.1016/j.inffus.2017.07.004)]
 - 12 Bostrom A, Bagnall A. Binary shapelet transform for multiclass time series classification. *Madria S, Hara T. Big Data Analytics and Knowledge Discovery*. Cham: Springer, 2015. 257–269.
 - 13 Bagnall A, Lines J, Hills J, *et al.* Time-series classification with COTE: The collective of transformation-based ensembles. *IEEE Transactions on Knowledge and Data Engineering*, 2015, 27(9): 2522–2535. [doi: [10.1109/TKDE.2015.2416723](https://doi.org/10.1109/TKDE.2015.2416723)]