

# 時間変化する shapelets を学習する時系列分類手法

山口 晃広<sup>†</sup> 植野 研<sup>†</sup> 鹿島 久嗣<sup>††</sup>

<sup>†</sup> 株式会社東芝 研究開発センター システム AI ラボラトリー 〒 212-8582 川崎市幸区小向東芝町 1

<sup>††</sup> 京都大学 大学院情報学研究科 知能情報学専攻 〒 606-8501 京都府京都市左京区吉田本町

E-mail: [†akihiro5.yamaguchi@toshiba.co.jp](mailto:†akihiro5.yamaguchi@toshiba.co.jp)

あらまし IoT の活用として機械学習による時系列インスタンスのクラス分類手法が研究されている。特に近年、分類器を学習しながら shapelets と呼ばれる分類に用いる波形パターンを同時に学習する手法が、分類性能が高く説明性もあるため注目を集めている。本研究では、被験者の慣れ・季節性・機械の劣化などの影響による shapelets の変化を予測する技術を提案する。本提案では、時系列インスタンスとそれらが観測されたタイミングやクラスラベルから、shapelets と分類器だけでなく shapelets の変化を予測する回帰モデルも同時に学習するように数理最適化問題を定式化する。また本定式化から既製の回帰ソルバをブラックボックスとして利用できる解法を導出する。実験では被験者の慣れ・季節性・機械の劣化がある各産業分野への適用事例で有効性を示す。

キーワード 時系列分類, Shapelet, 波形変形

## 1 はじめに

IoT の活用に向けて、機械学習による時系列データのクラス分類手法の研究が進められている。2 クラス分類の場合は、学習時に正または負のクラスラベルが付与された複数の時系列インスタンスを正例または負例として与えて、テスト時に未知の時系列インスタンスのクラスラベルを予測する問題となる。一般のクラス分類とは異なり時系列分類では、時間軸における順序関係や波形パターンの形状が重要であり、時系列データのノイズや波形パターンの出現位置のズレなどを扱える時系列分類に特化した手法が提案されている [1]。

時系列分類手法の中で、時系列分類に有効な複数の波形パターン (shapelets) を発見することで分類器を学習する研究に注目が集まっている [2, 3]。これらの shapelets 手法では、分類に有効な特徴は長い時系列全体ではなく少数の短い部分時系列に表れるというアイデアに基づく。特長として、学習が終わればテスト時の分類は高速であり、高い分類性能を達成する。更に、分類に用いる波形パターンを専門家に提示できるため説明性がある。特に、医療・インフラ・製造などの産業分野では、専門家は時系列分類に有効な波形パターンに着目して分析する機会が多いため、専門家に受け入れられやすい。

Shapelets 手法は、学習データから切り出される多数の部分時系列を探索することで shapelets を発見するアプローチから始まったが [3]、近年 shapelets を任意の波形パターンとして学習する shapelets 学習法が提案されている [2]。このような shapelets 学習法では、確率的勾配降下法 (SGD) により shapelets と分類器の両方を同時に学習する。これにより、計算量を削減し正解率を向上し [2]、F 値や Area Under the Curve (AUC) などの分類性能を向上する手法も提案された [4, 5]。

本来、時系列インスタンスはそれを取ったタイミングに依存してその形状を変化させる場合が多い。図 1 の例に示すように、日ごとに交通機関の利用者数を 1 時間間隔で時系列インスタンスとして取得した場合、夏から冬への季節の変化に従って

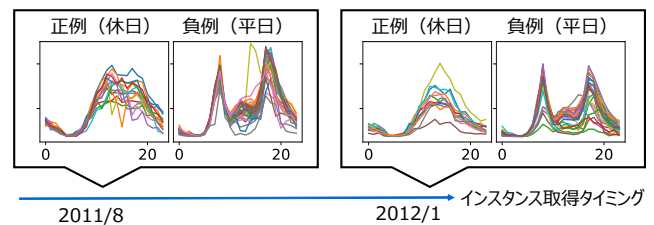


図 1 取得タイミングにより変化する時系列インスタンス (毎日の交通機関の利用者数が季節変化する例)。

その利用状況は明らかに変化している。そのため、shapelets も時系列インスタンスを取ったタイミングに従って変化するはずである。しかしながら、従来の shapelets 手法では、古典的な時系列分類問題に基づき、時系列インスタンスはその取得タイミングに依存しないことを前提としており、インスタンス取得タイミングに従って変形する shapelets を扱うことができない。

インスタンス取得タイミングに従って変形する shapelets を学習し予測することは、分類性能の向上だけでなく、以下のような産業応用において専門家の理解を促進することもできる。

**例 1.** 被験者が短期間に一定間隔で同じ試行を繰り返した結果をセンシングしてデータセットとして収集した場合、被験者の慣れによって時系列インスタンスの形状が試行期間中に変化する [6, 7]。事象関連電位 (ERP) では、脳と刺激の相互作用を理解するための分類問題に加えて短期間の被験者の慣れが同時に重要な問題となる [8, 9, 10]。分類に有効な波形パターンが被験者の慣れによってどのように変形するかを発見できれば、専門家は機械学習の分類結果をより納得して受け入れられる。

**例 2.** 交通機関の利用は季節の影響を受けやすい [11]。図 1 の例のように交通インフラの整備を進めるために休日と平日の利用状況の違いを調べる場合、分析するデータは日単位の利用状況を表す時系列インスタンスとなる。休日と平日との違いを表す利用状況のパターンを発見するだけでなくそれらの季節変化までも発見できれば、運用計画や意思決定にそれらの知識を活用できる [12]。

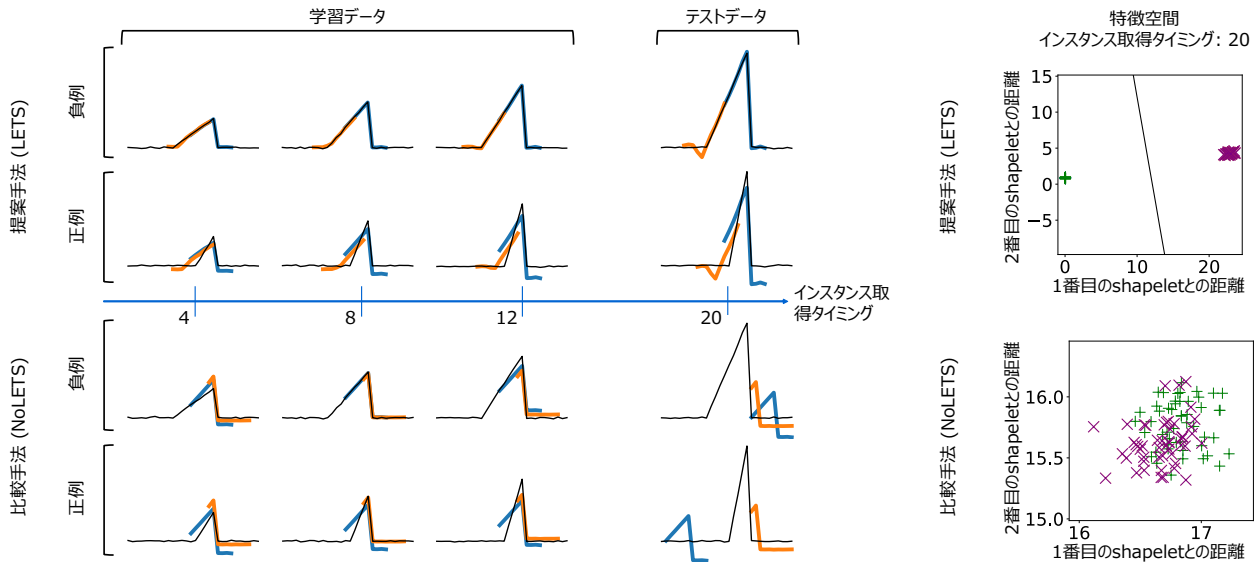


図2 インスタンス取得タイミング 20 において、提案手法（上段）では shapelets の変形を予測できるため（左）、特徴空間でも正例と負例を正しく分離でき分類性能を大幅に向上させる（右）。

**例 3.** 一般に寿命のある機械は、安定期を過ぎると機械の稼働回数に応じて劣化していく [6]。この劣化は、機械に取り付けられたセンサによって波形の歪みとして現れることがある。一方、機械が稼働している状態や周辺環境をセンサで認識することは、機械工学などの分野で重要な課題である [13]。専門家は、機械が稼働している状態を区別できる波形パターンを発見できるだけでなくその波形パターンがどのように劣化していくかを発見できれば、効果的・効率的な機械の設計に役立てられる。

本研究では、分類器と shapelets に加えて shapelets の変化を予測する回帰モデルを同時に学習する shapelets 学習法 “LETS” を提案する。LETS では、時系列インスタンスとその取得タイミング、及びクラスラベルが与えられると、学習データに含まれない取得タイミングに対しても、学習した回帰モデルを用いた外挿や内挿により shapelets の変形を予測することができる。我々の知る限り、時系列インスタンスの取得タイミングに従って変形できる shapelets（可変 shapelets）を扱うのは、本研究が初めてである。以下では、可変 shapelets を予測した場合の効果について簡単な例を用いて説明する。

### 1.1 提案するアプローチの有効性を示す例

図 2 の黒線は上に尖った波形周辺を拡大した時系列インスタンスである。上に尖った波形の傾きは負例では正例よりも緩やかに上昇するため、上に尖った波形に着目することで正例と負例を正しく分類できる。しかし、この上に尖った波形はインスタンスの取得タイミングに従って高さが線形に増加していく。学習時には、インスタンス取得タイミングは 4 から 12 の間で一様にランダムに選ぶ。この例では、インスタンス取得タイミングが 20 のときの可変 shapelets を予測する効果を確認する。

従来の shapelets 学習法のように可変 shapelets を予測できない比較手法 “NoLETS” と提案手法 LETS とを比較する<sup>1</sup>。

1：可変 shapelets を予測する効果を比較評価するため、比較手法 NoLETS では提案手法 LETS から shapelets の変形を予測する機能を除外する。

LETS と NoLETS のいずれも、図 2(左) の太青線と太橙線で描いた 2 つの shapelets を学習した。Shapelets は時系列データにベストマッチングした箇所に描かれている。インスタンス取得タイミングが 20 のとき、提案手法 LETS だけが、上に尖った波形パターンの高さが徐々に上昇することを学習しその結果可変 shapelets を正しく予測できていることが分かる。

図 2(右) では、インスタンス取得タイミングが 20 のときについて、各時系列インスタンスを 2 つの shapelets への距離<sup>2</sup>として定義される 2 次元特徴ベクトルとして特徴空間へマッピングしている。正例と負例をそれぞれ “+” と “x” としてプロットしている。この結果から、提案手法 LETS では全てのインスタンスを正しく分類できる (AUC=100%) のに対し、比較手法 NoLETS では分類できない (AUC=58%) という大きな性能差が生じる。

NoLETS と同様に、[2, 4, 5] などの従来の shapelets 手法でも可変 shapelets を予測できないため、インスタンス取得タイミングが 20 のとき shapelets を上に尖った波形に正しくマッチングできない。その結果、AUC も大きく劣化する。それゆえ、この単純な例では可変 shapelets を予測することが分類性能においても有効であることが分かった。

### 1.2 本研究の主な貢献

本論文の主な貢献を以下に示す。なお、本内容は基本的に [14] に準じ、詳しくは [14] を参照いただきたい。

- 従来手法が shapelets の時間変化を考慮できないのに対して、本研究では初めて shapelets の時間変化を考慮する。
- Shapelets と分類器だけでなく shapelets の変化を予測する回帰モデルも同時に学習するという新しい技術課題をもたらし、それを 1 つの数値最適化問題として定式化する。
- 本最適化において、理論的かつ実践的な解法を導出し、既製の回帰ソルバを用いて回帰モデルを学習できることを示す。
- 医療・インフラ・製造の各産業分野への適用事例で提案手法 LETS の有効性を示す。

2：時系列インスタンスと shapelet との距離は 3.2 節で定義する。

## 2 準備

### 2.1 記号と定義

時系列インスタンスを2クラス  $\mathcal{Y} := \{1, -1\}$  に分類する問題を扱う。学習データのインスタンスの数を  $I$  個として、 $i$  番目のインスタンスは時系列インスタンス  $\mathbf{t}_i$  とクラスラベル  $y_i$  とインスタンス取得タイミング  $p_i$  の3つ組とする。つまり、学習データ  $\mathcal{T}$  は  $\{(\mathbf{t}_1, p_1, y_1), (\mathbf{t}_2, p_2, y_2), \dots, (\mathbf{t}_I, p_I, y_I)\}$  である。従来の shapelets 手法とは異なり、 $i$  番目の時系列インスタンスが取得されたタイミング  $p_i \in \mathbb{R}$  を考慮する。

インスタンス取得タイミングの意味はアプリケーションに依存する。例えば、例1における被験者の反応の慣れであれば試行回数として、例2における交通機関の利用状況の季節性であればある時刻からの経過時間として、例3における機械の劣化であれば機械の稼働回数として、扱うことができる。

図2に示すように、時系列インスタンスと shapelet はそれぞれ長い数列と短い数列として表される。長さ  $Q$  の時系列データセットを  $\mathbf{T} \in \mathbb{R}^{I \times Q}$  とし、shapelets の個数と長さをそれぞれ  $K$  と  $L$  とする。従来の shapelets との違いを明確にするため、インスタンス取得タイミングに応じて shapelets が変形しないか変形しうるかで、不変 shapelets と可変 shapelets を区別する：

**定義 1.**  $k = 1, 2, \dots, K$  及び  $l = 1, 2, \dots, L$  に対して、 $k$  番目の不変 shapelet  $\mathbf{c}_k$  とは、インスタンス取得タイミング  $p$  に  $\mathbf{c}_k$  の  $l$  番目の値  $c_{k,l}$  が依存しない shapelet である。 $K$  個の不変 shapelets を  $\mathbf{C} \in \mathbb{R}^{K \times L}$  と記述する。

**定義 2.**  $k = 1, 2, \dots, K$  及び  $l = 1, 2, \dots, L$  に対して、 $k$  番目の可変 shapelet  $\mathbf{s}_k(p)$  とは、インスタンス取得タイミング  $p$  に  $\mathbf{s}_k(p)$  の  $l$  番目の値  $s_{k,l}(p)$  が依存する shapelet である。インスタンス取得タイミング  $p$  における  $K$  個の可変 shapelets を  $\mathbf{S}(p) \in \mathbb{R}^{K \times L}$  と記述する。

可変 shapelets の変形の仕方はアプリケーションに依存する。そのため、後述の定理2で理論的に保証されるように本研究では任意の回帰モデルを利用できる。ここでは例1-3をもとに3つの変形パターンをあげるが、本研究はこれらの例に限定されず、回帰モデルで予測できればどのような変形パターンであっても良い。

**例 4.** 線形的変形は、インスタンス取得タイミングに従って shapelet の各点が線形回帰モデルに従って変化する。これは、可変 shapelet の変化のトレンドを発見するのに有効である。

**例 5.** 周期的変形は、インスタンス取得タイミングに従って shapelet の各点が周期的に変化する。これは、季節などの影響による可変 shapelets の変化を捉えるのに有効である。

**例 6.** 複数モードの変形は、shapelet がインスタンス取得タイミングの区間によって異なる回帰モデルに従って変化する。これは、可変 shapelets の変形に対して安定期と不安定（劣化）期など複数のモードを考慮する必要がある場合に有効である。

古典的な時系列分類の問題設定とは異なり、本研究では学習データとテストデータにインスタンス取得タイミングが含まれる。学習データに含まれるインスタンス取得タイミングの値の集

合を学習インスタンス取得タイミング集合  $\mathcal{P}$  と記述する。学習インスタンス取得タイミング集合  $\mathcal{P}$  は  $|\mathcal{P}| \leq I$  であるため有限集合である。なお、 $p \in \mathcal{P}$  にわたって、 $\mathbf{S}(p)$  や  $\mathbf{s}_k(p)$  や  $s_{k,l}(p)$  の集合をとったものをそれぞれ  $\mathbf{S}(\mathcal{P})$  や  $\mathbf{s}_k(\mathcal{P})$  や  $s_{k,l}(\mathcal{P})$  と記述する。

本研究では、可変 shapelets  $\mathbf{S}(\mathcal{P})$  とその変化を予測する回帰モデルを学習し、 $\mathcal{P}$  に含まれないインスタンス取得タイミングの値  $p \notin \mathcal{P}$  でも可変 shapelets  $\mathbf{S}(p)$  の変化を予測する。

### 2.2 従来の shapelets 学習法

本節では文献[2]の定式化を説明する。インスタンスの取得タイミングに従来の shapelets は依存しないため、本節ではそれらを不変 shapelets  $\mathbf{C}$  として扱う。各時系列インスタンスに対して長さ  $L$  の部分時系列は  $J := Q - L + 1$  個ある。 $j$  番目の部分時系列  $(t_{i,j}, t_{i,j+1}, \dots, t_{i,j+L-1})$  と  $\mathbf{c}_k$  とのユークリッド距離を測り、 $j = 1, 2, \dots, J$  の中で最小のものを  $\mathbf{t}_i$  と  $\mathbf{c}_k$  との距離と定義する。

$$\tilde{x}_{i,k}(\mathbf{c}_k, \mathbf{t}_i) := \min_{j=1,2,\dots,J} \frac{1}{L} \sum_{l=1}^L (t_{i,j+l-1} - c_{k,l})^2. \quad (1)$$

$i$  番目のインスタンスに対する特徴ベクトル  $\tilde{\mathbf{x}}_i \in \mathbb{R}^K$  を  $(\tilde{x}_{i,1}, \tilde{x}_{i,2}, \dots, \tilde{x}_{i,K})$  で定義する。バイアス項を含めた分類器の重み  $\tilde{\mathbf{w}} \in \mathbb{R}^{K+1}$  が与えられて、次の線形モデルを用いる。

$$\hat{y}_i(\tilde{\mathbf{x}}_i, \tilde{\mathbf{w}}) := \tilde{w}_0 + \sum_{k=1}^K \tilde{w}_k \tilde{x}_{i,k}. \quad (2)$$

クラスラベルが  $\mathcal{Y} = \{1, -1\}$  であることに注意し<sup>3</sup>、 $\hat{y}_i$  とシグモイド関数  $\sigma$  を介してロジスティック損失関数を用いる。

$$\begin{aligned} \mathcal{L}(y_i, \hat{y}_i) &:= -\frac{1+y_i}{2} \ln(\sigma(\hat{y}_i)) - \frac{1-y_i}{2} \ln(1-\sigma(\hat{y}_i)), \\ \sigma(\hat{y}_i) &:= \left(1 + e^{-\hat{y}_i}\right)^{-1}. \end{aligned} \quad (3)$$

最終的には、特徴ベクトル  $\tilde{\mathbf{X}}$  を介して不変 shapelets  $\mathbf{C}$  と分類器の重み  $\tilde{\mathbf{w}}$  とを同時に学習する最適化問題として定式化される。

$$\underset{\mathbf{C} \in \mathbb{R}^{K \times L}, \tilde{\mathbf{w}} \in \mathbb{R}^{K+1}}{\text{minimize}} \quad \sum_{i=1}^I \mathcal{L}(y_i, \hat{y}_i) + \alpha \tilde{\Omega}(\tilde{\mathbf{w}}). \quad (4)$$

ここで、 $\alpha \geq 0$  は正則化パラメータであり、 $\tilde{\Omega}(\tilde{\mathbf{w}})$  は次で定義される分類器の重み  $\tilde{\mathbf{w}}$  における  $l_2$  正則化である。

$$\tilde{\Omega}(\tilde{\mathbf{w}}) := \sum_{k=1}^K w_k^2. \quad (5)$$

この非凸最適化問題には制約式は無くインスタンスごとに目的関数を分解できることからSGDアルゴリズムを用いて解かれる。

しかしながら、従来の shapelets 手法では、時系列インスタンスが取得タイミング  $p$  に依存しないことを前提としており、可変 shapelets  $\mathbf{S}(p)$  を予測できない。

## 3 提案手法の定式化

提案する数理最適化問題は、分類器の重みと可変 shapelets と可変 shapelets の変化を予測する回帰モデルとをそれらの相

3:  $y_i = -1$  を  $y_i = 0$  と置換することで式(3)は[2]の定式化と等価となる。

互依存を考慮して同時に学習するように定式化される。この最適化問題は、可変 shapelets と回帰モデルとを関連付ける等式制約を持つ。以降では、3.1 節でその等式制約をまず定式化し、3.2 節で制約付き最適化問題として定式化する。

### 3.1 可変 shapelets と回帰モデルとを関連付ける等式制約

本節では、可変 shapelets とその変化を予測する回帰モデルとを関連付けるための等式制約を定式化する。  $k=1,2,\dots,K$  において、学習インスタンス取得タイミング  $p \in \mathcal{P}$  における可変 shapelet  $\mathbf{s}_k(p)$  の重心を不変 shapelet  $\mathbf{c}_k$  とみなす。

$$c_{k,l}(s_{k,l}(\mathcal{P})) := \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} s_{k,l}(p). \quad (6)$$

回帰モデルのパラメータ（以降、回帰パラメータ）を  $\mathbf{a}_k$  とする。また、  $s_{k,l}(p)$  と  $c_{k,l}$  の差を予測するために、インスタンス取得タイミング  $p$  と shapelet の長さに対するインデックス  $l$  とを 2 つの従属変数とする回帰モデル  $f_k(p,l;\mathbf{a}_k)$  を導入する。つまり、以下のように可変 shapelet  $\mathbf{s}_k(p)$  と不変 shapelet  $\mathbf{c}_k$  との差を回帰モデル  $f_k$  として定義する。

$$f_k(p,l;\mathbf{a}_k) = s_{k,l}(p) - c_{k,l}. \quad (7)$$

回帰モデルの予測誤差を無くすように次の等式制約を定義する。

$$d_k(\mathbf{s}_k(\mathcal{P}), \mathbf{a}_k) = 0, \quad (8)$$

$$d_k(\mathbf{s}_k(\mathcal{P}), \mathbf{a}_k) := \sum_{l=1}^L \sum_{p \in \mathcal{P}} (f_k(p,l;\mathbf{a}_k) - (s_{k,l}(p) - c_{k,l}))^2.$$

等号制約式 (8) は式 (7) を満たす必要十分条件となる。

具体的な回帰パラメータ  $\mathbf{a}_k$  の意味は回帰モデル  $f_k$  の種類に依存する。例えば、  $f_k$  が線形回帰であれば  $\mathbf{a}_k$  は回帰係数となり、  $f_k$  がカーネルリッジ回帰であれば重みベクトルとそれらの係数となる。しかし、4.1 節で後述するように、既製の回帰ソルバを用いて回帰パラメータ  $\mathbf{a}_k$  を最適化できることが示されるため、本定式化において特定の回帰パラメータ  $\mathbf{a}_k$  を意識する必要がない。

次節で後述するように、等号制約式 (8) は、回帰パラメータら  $\mathbf{A}$  と可変 shapelets  $\mathbf{S}(\mathcal{P})$  とを同時に学習するように LETS の数理最適化問題に組み込まれる。

### 3.2 数理計画問題の定式化

本節では、提案手法 LETS の数理計画問題を定式化する。まず可変 shapelets とインスタンス取得タイミングによる特徴ベクトルと分類器を定義し、次に数理最適化問題としてまとめる。

式 (1) における従来の特徴ベクトルの代わりに、  $i=1,2,\dots,I$  及び  $k=1,2,\dots,K$  に対して、  $i$  番目の時系列インスタンス  $\mathbf{t}_i$  と  $k$  番目の可変 shapelet  $\mathbf{s}_k(p_i)$  とを用いて、特徴量  $x_{i,k}$  は次のような最小ユークリッド距離で定義される。

$$x_{i,k}(\mathbf{s}_k(p_i), \mathbf{t}_i) := \min_{j=1,2,\dots,J} \frac{1}{L} \sum_{l=1}^L (t_{i,j+l-1} - s_{k,l}(p_i))^2. \quad (9)$$

ここで、特徴量  $x_{i,k}(\mathbf{s}_k(p_i), \mathbf{t}_i)$  は非負であり、特徴ベクトルは  $\mathbf{x}_i := (x_{i,1}, x_{i,2}, \dots, x_{i,K})$  となる。分類器の重み  $\mathbf{w} \in \mathbb{R}^{K+2}$  を用いて、従来の式 (2) を次の式に換える。

従来の分類器の重みの正則化のみを用いた場合

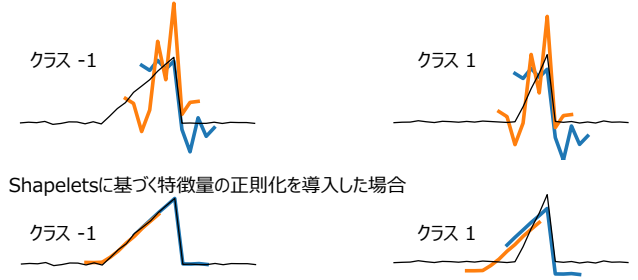


図3 Shapelets に基づく特徴量の正則化による効果。

$$\hat{y}_i(\mathbf{x}_i, p_i, \mathbf{w}) := w_0 + w_{K+1} p_i + \sum_{k=1}^K w_k x_{i,k}. \quad (10)$$

式 (10) を介して式 (3) のロジスティック損失関数  $\mathcal{L}$  を用いる。

従来の shapelets 学習法では式 (5) において  $\ell_2$  正則化のみを導入していたが、shapelets の説明性を保持するために、更に shapelets に基づく特徴量の正則化を導入する。まず、これを導入する動機を明らかにする。図3はインスタンス取得タイミングが6の場合で1.1節(図2)と同じ時系列インスタンスを表している。図3(上)の青太線と橙太線は、式(5)で用いられる古典的な  $\ell_2$  正則化のみを用いて学習した2つの可変 shapelets  $\mathbf{S}(6)$  を表している。この図より  $\mathbf{S}(6)$  は元の時系列波形に似ていないことが分かる。一方、図3(下)の青太線と橙太線は、shapelets に基づく特徴量の正則化を導入して学習した2つの可変 shapelets  $\mathbf{S}(6)$  を表している。このように、  $\mathbf{S}(6)$  は片方のクラスの元の時系列波形に似ていることが分かる。以下の性質は、時系列インスタンス  $\mathbf{t}_i$  と可変 shapelet  $\mathbf{s}_k(p_i)$  との距離を縮めること（つまり、shapelets に基づく特徴量  $x_{i,k}$  を縮小すること）は、  $y_i w_k < 0$  の条件下で分類性能に悪影響を及ぼさないことを保証する [14]。

**定理 1.** 任意の  $i=1,2,\dots,I$  及び  $k=1,2,\dots,K$  に対して、  $i$  番目の時系列インスタンス  $\mathbf{t}_i$  と  $k$  番目の可変 shapelet  $\mathbf{s}_k(p_i)$  との距離は、式 (9) で定義される特徴量と一致することに注意する。  $\mathbf{t}_i$  と  $\mathbf{s}_k(p_i)$  の距離が縮まるとき次の3ケースがある：

- (a)  $y_i w_k < 0$  の場合、式 (3) の分類損失は減少する。
- (b)  $y_i w_k > 0$  の場合、式 (3) の分類損失は増加する。
- (c)  $w_k = 0$  の場合、式 (3) の分類損失は変化しない。

定理 1 を考慮して、式 (5) の代わりに本定式化における正則化では分類器の重み  $\mathbf{w}$  と shapelets に基づく特徴ベクトル  $\mathbf{X}$  との両方を以下のように縮小する。

$$\Omega(\mathbf{w}, \mathbf{X}, \mathbf{y}) := \sum_{k=1}^K w_k^2 + \sum_{k=1}^K \beta_k \sum_{i=1}^I x_{i,k} \mathbb{I}(y_i w_k < 0). \quad (11)$$

ここで、インジケータ関数  $\mathbb{I}(\cdot)$  は指数が真のとき 1 を返し偽のとき 0 を返し、式 (9) より  $x_{i,k} \geq 0$  である。正則化パラメータ  $\beta_k > 0$  は shapelets があるクラスの時系列に似ていないことを罰する程度を調整するもので自動的かつ動的に調整される [14]。

最終的には、学習データ  $\mathcal{T}$  が与えられて、LETS の定式化は可変 shapelets  $\mathbf{S}(\mathcal{P})$ 、分類器の重み  $\mathbf{w}$ 、及び回帰パラメータら  $\mathbf{A}$  を以下のように制約付きで同時に最適化する。

$$\begin{aligned} & \underset{\substack{\mathbf{S}(\mathcal{P}) \in \mathbb{R}^{K \times L \times \mathcal{P}}, \mathbf{w} \in \mathbb{R}^{K+2} \\ \mathbf{A} \in \mathcal{A}^K}}{\text{minimize}} \sum_{i=1}^I \mathcal{L}(y_i, \hat{y}_i) + \alpha \Omega(\mathbf{w}, \mathbf{X}, \mathbf{y}) \\ & \text{s.t.} \quad \text{equality constraints in Eq. (8).} \end{aligned} \quad (12)$$

ここで、 $\mathcal{A}$  は回帰パラメータ集合  $\mathbf{A}$  を含むような集合であり、分類損失関数  $\mathcal{L}$  は式 (9) と (10) とを介して式 (3) と同じであり、 $\Omega$  は式 (11) と同じ正規化である。式 (12) の実行可能解は式 (8) の等式制約を満たす。そのため、学習が完了すれば式 (7) で可変 shapelets  $\mathbf{S}(p)$  を予測できる。加えて、[5] の Proposition 4.1 と同様に、可変 shapelets は次の性質を持つ。

**性質 1.** 可変 shapelets  $\mathbf{S}(\mathcal{P})$  は式 (12) の問題を最適化する  $K$  個の shapelets である。その目的関数から、LETS は式 (3) の分類損失と式 (11) の正則化とを削減するように可変 shapelets を学習する。定理 1 と式 (11) から、 $i=1,2,\dots,I$  及び  $k=1,2,\dots,K$  において  $y_i w_k < 0$  のとき、この損失とこの正則化はそれぞれ式 (9) における距離  $x_{i,k}$  を縮める方向に働く。すなわち、 $k$  番目の可変 shapelet は、式 (9) の距離において  $w_k < 0$  及び  $w_k > 0$  の場合にそれぞれ正例と負例の時系列インスタンスに類似する傾向がある。そのため、 $k$  番目の可変 shapelet の属するクラスは、 $w_k < 0$  なら 1 として  $w_k > 0$  なら -1 とする。

## 4 提案手法の解法

前節で提案した式 (12) は非凸の目的関数と非線形の等式制約から構成されている。有効な解法アプローチとして、拡張ラグランジュ法を用いて式 (12) を等価な無制約最適化問題に変換する。

$$\begin{aligned} & \underset{\substack{\mathbf{S}(\mathcal{P}) \in \mathbb{R}^{K \times L \times \mathcal{P}}, \mathbf{w} \in \mathbb{R}^{K+2} \\ \mathbf{A} \in \mathcal{A}^K, \boldsymbol{\lambda} \in \mathbb{R}^K}}{\text{arg min}} \sum_{i=1}^I \mathcal{L}(y_i, \hat{y}_i) + \alpha \Omega(\mathbf{w}, \mathbf{X}, \mathbf{y}) \\ & + \langle \boldsymbol{\lambda}, \mathbf{d}(\mathbf{S}(\mathcal{P}), \mathbf{A}) \rangle + \frac{\mu}{2} \|\mathbf{d}(\mathbf{S}(\mathcal{P}), \mathbf{A})\|_2^2. \end{aligned} \quad (13)$$

ここで  $k=1,2,\dots,K$  に対して、 $\mathbf{d}(\mathbf{S}(\mathcal{P}), \mathbf{A})$  は、式 (8) の  $d_k(\mathbf{s}_k(\mathcal{P}), \mathbf{a}_k)$  が  $k$  番目の要素となる  $K$  次元ベクトルであり、 $\boldsymbol{\lambda} \in \mathbb{R}^K$  はラグランジュ乗数であり、 $\mu \geq 0$  は罰則パラメータであり、 $\langle \cdot, \cdot \rangle$  は内積を表し、 $\|\cdot\|_2$  は  $\ell_2$  ノルムを表す。

変数を 2 つのブロックに分け、1 つ目のブロックを回帰パラメータ集合  $\mathbf{A}$  とし、2 つ目のブロックを可変 shapelets  $\mathbf{S}(\mathcal{P})$  と分類器の重み  $\mathbf{w}$  とする。式 (13) を効率的に解くため ADMM [15] を用いて、以下の手順を反復する。

- (1) 以下の部分問題を解き回帰パラメータら  $\mathbf{A}$  を更新する<sup>4</sup>。

$$\underset{\mathbf{A} \in \mathcal{A}^K}{\text{arg min}} \langle \boldsymbol{\lambda}, \mathbf{d}(\mathbf{S}(\mathcal{P}), \mathbf{A}) \rangle + \frac{\mu}{2} \|\mathbf{d}(\mathbf{S}(\mathcal{P}), \mathbf{A})\|_2^2. \quad (14)$$

- (2) 以下の部分問題を解き可変 shapelets  $\mathbf{S}(\mathcal{P})$  と分類器の重み  $\mathbf{w}$  を更新する。

$$\begin{aligned} & \underset{\mathbf{S}(\mathcal{P}) \in \mathbb{R}^{K \times L \times \mathcal{P}}, \mathbf{w} \in \mathbb{R}^{K+2}}{\text{arg min}} \sum_{i=1}^I \mathcal{L}(y_i, \hat{y}_i) + \alpha \Omega(\mathbf{w}, \mathbf{X}, \mathbf{y}) \\ & + \langle \boldsymbol{\lambda}, \mathbf{d}(\mathbf{S}(\mathcal{P}), \mathbf{A}) \rangle + \frac{\mu}{2} \|\mathbf{d}(\mathbf{S}(\mathcal{P}), \mathbf{A})\|_2^2. \end{aligned} \quad (15)$$

- (3) ラグランジュ乗数  $\boldsymbol{\lambda}$  を  $\boldsymbol{\lambda} + \mu \mathbf{d}(\mathbf{S}(\mathcal{P}), \mathbf{A})$  として更新する。次の 2 つの節では、式 (14) と (15) の部分問題の解き方を述べる。

4:  $\sum_{i=1}^I \mathcal{L}(y_i, \hat{y}_i) + \alpha \Omega(\mathbf{w}, \mathbf{X}, \mathbf{y})$  は変数  $\mathbf{A}$  に依存しないので、式 (14) ではこれらの項を式 (13) から除外している。

### 4.1 回帰パラメータ集合 $\mathbf{A}$ の学習

本節では、式 (14) の最適化問題が本質的には回帰モデルの標準的な最適化問題に変換できることを示す。この事実、標準的な回帰モデルの学習から得られた回帰パラメータが式 (14) の最適な回帰パラメータでもあることを保証する。つまり、既製の回帰ソルバをそのまま用いて式 (14) の最適化問題を解ける。これを保証するために以下の定理を導出する [14]。

**定理 2.**  $k=1,2,\dots,K$  に対して、次の最適化問題を最小化する回帰パラメータ  $\mathbf{a}_k$  は、式 (14) の最適化問題も最小化する。

$$\begin{aligned} & \underset{\mathbf{a}_k \in \mathcal{A}}{\text{arg min}} d_k(\mathbf{s}_k(\mathcal{P}), \mathbf{a}_k), \\ & d_k(\mathbf{s}_k(\mathcal{P}), \mathbf{a}_k) := \sum_{l=1}^L \sum_{p \in \mathcal{P}} (f_k(p, l; \mathbf{a}_k) - (s_{k,l}(p) - c_{k,l}))^2. \end{aligned} \quad (16)$$

それぞれ  $s_{k,l}(p) - c_{k,l}$  と  $f_k(p, l; \mathbf{a}_k)$  とを  $k$  番目の回帰モデル  $f_k$  における観測値と予測値とみなすと、式 (16) の最適化問題は、 $p \in \mathcal{P}$  と  $l=1,2,\dots,L$  にわたって  $s_{k,l}(p) - c_{k,l}$  と  $f_k(p, l; \mathbf{a}_k)$  の 2 乗誤差の和を最小化しているだけである。したがって、式 (16) は標準的な回帰問題の最適化と本質的に同じである。

### 4.2 可変 shapelets $\mathbf{S}(\mathcal{P})$ と分類器の重み $\mathbf{w}$ の学習

部分問題の式 (15) は非凸最適化問題のままであるが、その目的関数はインスタンスごとの目的関数  $G_i$  に分解できる。

$$\begin{aligned} G_i(\mathbf{S}(\mathcal{P}), \mathbf{w}) & := \mathcal{L}(y_i, \hat{y}_i) + \alpha \Omega_i(\mathbf{w}, \mathbf{x}_i, y_i) \\ & + \frac{1}{I} \langle \boldsymbol{\lambda}, \mathbf{d}(\mathbf{S}(\mathcal{P}), \mathbf{A}) \rangle + \frac{\mu}{2I} \|\mathbf{d}(\mathbf{S}(\mathcal{P}), \mathbf{A})\|_2^2. \end{aligned} \quad (17)$$

ここで、 $\Omega_i$  は式 (11) をインスタンスごとに分解した式である。

$$\Omega_i(\mathbf{w}, \mathbf{x}_i, y_i) := \frac{1}{I} \sum_{k=1}^K w_k^2 + \sum_{k=1}^K \beta_k x_{i,k} \mathbb{I}(y_i w_k < 0). \quad (18)$$

SGD を用いて  $\mathbf{S}(\mathcal{P})$  と  $\mathbf{w}$  を学習する。可変 shapelets  $\mathbf{S}(\mathcal{P})$  に対する  $G_i$  の勾配は、微分の連鎖律を用いて次のように書き下せる。

$$\begin{aligned} \frac{\partial G_i(\mathbf{S}(\mathcal{P}), \mathbf{w})}{\partial s_{k,l}(p)} & = \frac{\partial \mathcal{L}(y_i, \hat{y}_i)}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial x_{i,k}} \frac{\partial x_{i,k}}{\partial s_{k,l}(p)} \\ & + \alpha \frac{\partial \Omega_i(\mathbf{w}, \mathbf{x}_i, y_i)}{\partial x_{i,k}} \frac{\partial x_{i,k}}{\partial s_{k,l}(p)} \\ & + \frac{2}{I} \lambda_k z_{k,l}(p) + \frac{2\mu}{I} z_{k,l}(p) \sum_{l'=1}^L \sum_{p' \in \mathcal{P}} z_{k,l'}(p')^2. \end{aligned} \quad (19)$$

ここで、 $z_{k,l}(p) := s_{k,l}(p) - c_{k,l} - f_k(p, l; \mathbf{a}_k)$  である。分類器の重み  $\mathbf{w}$  に対する  $G_i$  の勾配も、同様に次のように書き下せる。

$$\frac{\partial G_i(\mathbf{S}(\mathcal{P}), \mathbf{w})}{\partial w_k} = \frac{\partial \mathcal{L}(y_i, \hat{y}_i)}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial w_k} + \alpha \frac{\partial \Omega_i(\mathbf{w}, \mathbf{x}_i, y_i)}{\partial w_k}. \quad (20)$$

式 (19) と (20) の右辺の各勾配は、従来の shapelets 学習法と同様のアプローチで個別に導出する。特に、勾配  $\partial x_{i,k} / \partial s_{k,l}(p)$  については、[5, 16, 17] と同様に劣勾配を用いて導出することで時系列長に対する計算量を線形オーダーまで削減できる [14]。

## 5 適用事例

例 1-3 で挙げた 3 つの産業分野への適用事例を用いて、可変 shapelets を予測することへの有効性を示す。

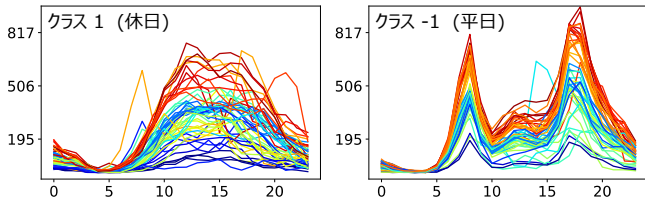


図4 自転車シェアリングの時系列インスタンスの例.

## 5.1 実験設定

$k=1,2,\dots,K$  に対して,  $k$  番目の shapelet の分類への寄与度を分類器の重みの絶対値  $|w_k|$  で測る. 一般性を損なうことなく, shapelets のインデックスは分類器の重みの絶対値の降順にソートする (つまり,  $|w_1| \geq |w_2| \geq \dots \geq |w_K|$ ). 可変 shapelets を扱える従来手法は存在しないため, 予測された可変 shapelets の変化が専門家の知識や物理現象と一致するかを検証する.

また, AUC は分類器の閾値に依存せず均衡/不均衡データのいずれでも適切な分類性能指標であり, 提案手法 LETS と以下のベースライン手法との AUC も比較する. LTS [2] は, 古典的な時系列分類において高い正解率を達成する [2, 18, 19]. LTSSFS [5] は, 教師有り特徴選択を介して初期 shapelets を選定し self-paced learning で更に学習することで2クラス分類において高い AUC を達成する. CSLTS [4] は, 本研究の実験データセットの一部にも含まれるような不均衡な2クラス分類を扱うためのコスト考慮型学習を LTS に導入する. NoLETS は, 可変 shapelets の純粋な導入効果を比較評価するために, LETS から可変 shapelets の変形を予測する機能を除外する. LTS や LTSSFS や CSLTS などの従来 shapelets 手法ではインスタンス取得タイミングを入力情報として活用できない. この潜在的な不公平さを解消するため, NoLETS は式 (10) を介してインスタンス取得タイミングを考慮した分類器を学習する.

## 5.2 事例: 季節変化の影響を受ける交通機関の利用

交通期間の利用において, 自転車シェアリングの利用者数における日別の推移は季節変化の影響を受ける時系列インスタンスである. 米国では温暖な季節に比べて冬季の自転車シェアリングシステムの予約者は少ない [11]. 一方で, 自転車再配置には時間的・空間的にコストがかかるため, これらのシステムでは適切な自転車配置の戦略が重要であり, 特に休日と平日とでは利用パターンが異なる [12]. そのため, 季節の影響を考慮しながら1日の利用パターンの違いを発見することで, 意思決定者がより良い運用計画を立てやすくなる.

使用する自転車シェアリングの時系列データは, 2011年と2012年の2年間にわたりワシントン D.C. 周辺で収集されたレンタルバイクの1時間ごとの利用台数であり [20], UCI の公開データセットに含まれる<sup>5</sup>. この元となる時系列データから日単位で時系列インスタンスを取得する. 学習データは2011/1/1から2012/9/30までで, テストデータは2012/10/1から2012/12/31までである. 学習用インスタンスは569個でテスト用インスタンスは86個であり, 残りの76日は欠損している. 休日と平日をそれぞれクラス1とクラス-1として分類

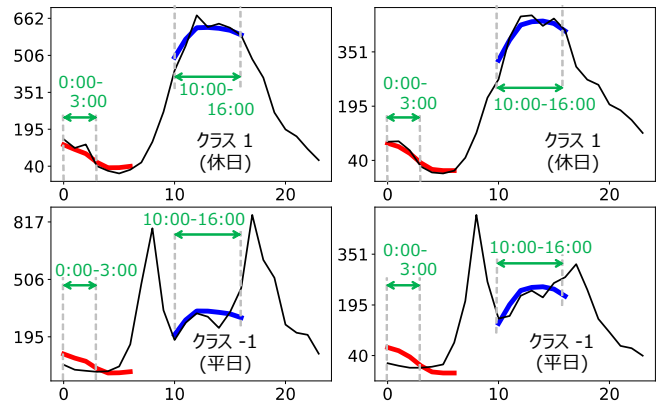


図5 自転車シェアリングにおける予測結果. 左:2012/10, 右:2012/12.

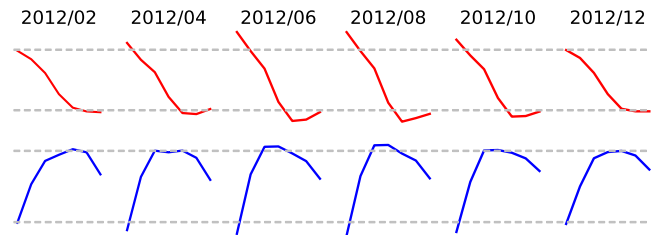


図6 自転車シェアリングにおける可変 shapelets の変化.

する. 時系列インスタンスの幾つかを図4に示す. 2011/1/1からの経過時間をインスタンス取得タイミング  $p$  とする. 季節効果を考慮するため, 例5の周期的変形を予測することを狙う. LETS では, Mackay 周期カーネルを用いたカーネルリッジ回帰モデルを採用し, 周期パラメータを1年に設定する.

図5(上)と(下)は, それぞれクラス1(休日)とクラス-1(平日)の時系列インスタンスを黒線で描き, それらにベストマッチングする位置に可変 shapelets を重ね描きしている. 以降, 1番目と2番目に貢献度の高い可変 shapelets (つまり,  $s_1$  と  $s_2$ ) を赤太線と青太線で描く. 分類器の重み  $w_1$  と  $w_2$  の符号は両方とも負であり, 可変 shapelets  $s_1$  と  $s_2$  はいずれもクラス-1よりもクラス1の時系列データに類似していることが分かる. この結果は性質1と一致する. 図5(左)と(右)は, インスタンス取得タイミング  $p$  がそれぞれ2012/10と2012/12のときの結果である. 最も貢献度の高い可変 shapelet  $s_1(p)$  は, 休日の深夜0時から早朝3時頃における利用者数が平日の場合よりも多いという事実を捉えている. 2番目に貢献度の高い可変 shapelet  $s_2(p)$  は, 休日の利用者数は平日のように2つの鋭いピークを持つのではなく10時から16時頃に緩やかな1つのピークを持つことを捉えている. テストデータの AUC は, LETS と NoLETS と LTS と LTSSFS と CSLTS とでそれぞれ 97%, 92%, 91%, 91%, 73% である. これらの結果から, 可変 shapelets は休日と平日の分類に有効な利用パターンを捉えられることが分かる.

図6は, 2012/2/16から2012/12/16にかけてインスタンス取得タイミング  $p$  が変化したときの可変 shapelets  $s_1(p)$  と  $s_2(p)$  を示している. 可変 shapelets  $s_1(p)$  と  $s_2(p)$  が表す利用パターンは, 冬季になるとその変化が緩やかになり自転車利用者数が減少する. この結果は [11] の報告とも一致する. 更に, この現象は貢献度の最も高い可変 shapelet  $s_1(p)$  に特によく現れている. 図5(上)から, 2012/10と2012/12の休日の  $s_1(p)$  の変化は, 冬

5: <https://archive.ics.uci.edu/ml/datasets/bike+sharing+dataset>

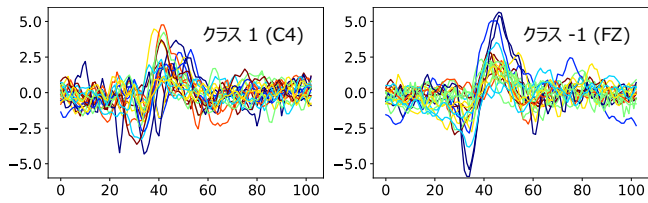


図 7 脳波における時系列インスタンスの例.

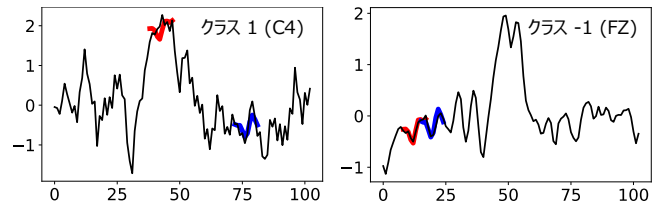


図 9 脳波における可変 shapelets の予測結果.

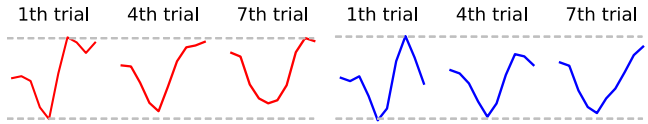


図 8 脳波における可変 shapelets の変化.

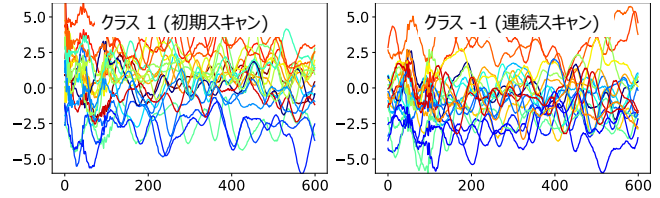


図 10 機械のローラー劣化における時系列インスタンスの例.

の深夜 0 時から早朝 3 時頃にかけての自転車の利用者数が冬季に減少することを捉えており、恐らく殆どの利用者は真冬の休日の早朝に自転車に乗りたくないため、これは納得できる結果である。本節の実験評価から、LETS により発見された可変 shapelets は休日と平日を区別する利用パターンを捉えるだけでなく季節変化による利用パターンの変化も捉えられることを確認できた。

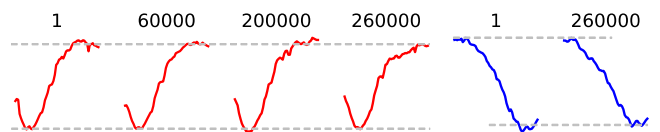


図 11 機械のローラー劣化における可変 shapelets の変化.

### 5.3 事例：被験者の慣れの影響を受ける脳波の測定

事象関連電位 (ERP) は認知刺激に関する非侵襲な脳波の時系列データである。同じ刺激が短い間隔で繰り返されると、脳内反応は短期間の試行で慣らされ減衰していく。この短期間の慣れが ERP 波形を減衰させることが知られている [8, 9]。また、異なる頭皮領域間の ERP を区別する波形パターンを理解することは、脳領域と刺激の相互作用を理解し頭皮に装着する電極を効率的に選択するための手がかりとなる [10]。したがって本研究では、異なる頭皮領域 (チャンネル) 間の ERP を分類する波形パターンを可変 shapelets として発見し、被験者の短期間の慣れによる可変 shapelets の変化も発見することを狙う。

使用する ERP 時系列データ<sup>6</sup>は、被験者の短期間の慣れによって生じる ERP 波形の歪みを研究するために用意された [8]。頭皮領域の C4 チャンネルと FZ チャンネルで収集された時系列インスタンスをそれぞれクラス 1 とクラス -1 として分類する。被験者は 19–31 歳の健常者 32 名 (女性 14 名) であり、各自 7 回同じ実験を試行し短期間の慣れを評価した。32 名を 28 名 (インスタンス数としては  $28 \times 7 \times 2$  個) と 4 名 (インスタンス数としては  $4 \times 7 \times 2$  個) に分けて、それぞれ学習データとテストデータとする。時系列インスタンスの幾つかを図 7 に示す。各被験者の試行回数をインスタンス取得タイミング  $p$  とする。試行回数に対する可変 shapelets の変化のトレンドを発見するため、例 4 の線形的変形を予測することを狙い、LETS では線形回帰モデルを採用する。

図 8 は、1 回目から 7 回目まで (つまり、 $p=1, 2, \dots, 7$ ) の試行における 1 番目と 2 番目に貢献度の高い可変 shapelets  $s_1(p)$  と  $s_2(p)$  の変化を示している。試行回数  $p$  が増加すると各可変 shapelets の振幅は減衰している。この結果は、被験者の短期間の慣れに関する専門家の知見と一致する [8, 9]。

図 9 (左) と (右) は、それぞれクラス 1 とクラス -1 のテスト用被験者の時系列インスタンスを黒線で描き、それらにベスト

マッチングする位置に可変 shapelets を重ね描きしている。分類器の重み  $w_1$  と  $w_2$  の符号は共に正であり、可変 shapelets のいずれもクラス 1 よりもクラス -1 の時系列データに類似していることが分かる。この結果は性質 1 と一致する。可変 shapelets はクラス 1 の波形の方がクラス -1 の波形よりも細かな変動があるという違いを捉えている。テストデータの AUC は、LETS と NoLETS と LTS と LTSSFS と CSLTS とでそれぞれ 79%, 70%, 65%, 68%, 62% である。本節の実験評価から、LETS はチャンネルの分類に有効な可変 shapelets を発見するだけでなく被験者の短期間の慣れに起因する可変 shapelets の変化を期待通り予測できることを確認できた。

### 5.4 事例：ローラー劣化の影響を受ける機械の稼動状態

スキャナやコピー機やプリンタなどの機械においてローラーは紙を搬送する重要な部品である。ローラーは何度も使用するうちにほこりの堆積や摩耗により徐々に滑りやすくなることが知られており、この種の機械は安定期を過ぎると動作が不安定になる。また、ローラーにかかる負荷は稼動状況によって異なる。機械に取り付けた安価なセンサからこのような現象を発見できれば、機械設計の改善やプロアクティブな保守に役立てることができる。

我々はスキャナのローラーの近くに低コストの加速度センサ (MMA7361) を取り付けて 26 万枚の紙をスキャンしてローラーを劣化させた。25 万枚の紙をスキャンする前後で、学習データとテストデータを分ける。初期スキャンと連続スキャンをそれぞれクラス 1 とクラス -1 として稼動状況を分類する。時系列インスタンスの幾つかを図 10 に示す。スキャンの回数をインスタンス取得タイミング  $p$  とする。安定期と不安定期に対して例 6 の複数モードの変形を考慮し、LETS では線形回帰木モデルを採用する<sup>7</sup>。

図 11 は、スキャンの回数  $p$  が 1 回から 260,000 回まで変化

6 : <https://osf.io/8wj3s/>

7 : 線形回帰木とは葉に線形回帰モデルを持つ決定木であり、関数の連続性が保証されないため拡張版 LETS を用いる [14]。

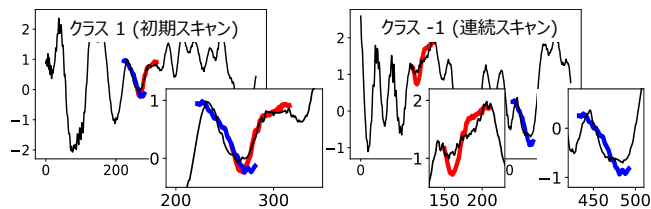


図 12 機械のローラー劣化における可変 shapelets の予測結果.

したときに予測された可変 shapelets  $s_1(p)$  と  $s_2(p)$  を示している。スキャン回数  $p$  が増加するとローラーのスリップによりこれらの可変 shapelets は平坦になっていく。特に  $s_1(p)$  は  $s_2(p)$  よりも明瞭に変化している。一方で、20 万枚を超えたあたりからスキャナはメンテナンスを促す警告を発しており、 $s_1(p)$  が変形する速さは 20 万枚を超える前より後の方がより速くなっていることが分かる。これらの結果から、可変 shapelets の変化が機械的な知見と一致していることが分かる。

図 12 (左) と (右) は、それぞれクラス 1 (初期スキャン) とクラス -1 (連続スキャン) のテスト用時系列インスタンスを黒線で描き、それらにベストマッチングする位置に可変 shapelets を重ね描きしている。分類器の重み  $w_1$  と  $w_2$  の符号は共に負であり、可変 shapelets のいずれもクラス -1 よりもクラス 1 の時系列データに類似していることが分かる。この結果は性質 1 と一致する。初期スキャン時の摩擦は連続スキャン時よりも強い。1 番目の可変 shapelets は初期スキャンのピークが連続スキャンのものより鋭いことを捉えており、この現象と一致する。テストデータの AUC は、LETS と NoLETS と LTS と LTSSFS と CSLTS とでそれぞれ 73%, 64%, 60%, 64%, 61% である。本節の実験評価から、LETS はこの稼動状況の分類に有効な可変 shapelets を発見でき、可変 shapelets とのその変化は機械的な知見と一致することを確認できた。

## 6 ま と め

本研究では、インスタンス取得タイミングに従って形状を変化させられる可変 shapelets を導入した。提案手法 LETS では、任意のインスタンス取得タイミング  $p$  における可変 shapelets  $S(p)$  を予測するため、可変 shapelets  $S(P)$  とその変化を予測する回帰モデル  $f$  と分類器  $w$  を同時に学習する。ADMM に基づき導出された解法では既製の回帰ソルバを用いてこの回帰モデル  $f$  を学習できる。インフラ・医療・製造の各産業分野への適用事例により本手法の有効性を確認した。

## 文 献

[1] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh. “Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures”. In: *Proc. VLDB Endow.* (2008).

[2] J. Grabocka, N. Schilling, M. Wistuba, and L. Schmidt-Thieme. “Learning Time-series Shapelets”. In: *KDD*. ACM, 2014.

[3] L. Ye and E. Keogh. “Time Series Shapelets: A New Primitive for Data Mining”. In: *KDD*. ACM, 2009.

[4] S. Roychoudhury, M. Ghalwash, and Z. Obradovic. “Cost Sensitive Time-Series Classification”. In: *ECML PKDD*. Springer, 2017.

[5] A. Yamaguchi and K. Ueno. “Learning Time-series Shapelets via Supervised Feature Selection”. In: *SDM*. SIAM, 2021.

[6] Y. Zhu, M. Imamura, D. Nikovski, and E. Keogh. “Introducing Time Series Chains: A New Primitive for Time Series Data Mining”. In: *Knowl. Inf. Syst.* (2019).

[7] M. Imamura, T. Nakamura, and E. Keogh. “Matrix Profile XXI: A Geometric Approach to Time Series Chains Improves Robustness”. In: *KDD*. ACM, 2020.

[8] F. Mancini, A. Pepe, A. Bernacchia, G. Di Stefano, A. Mouraux, and G. D. Iannetti. “Characterizing the Short-Term Habituation of Event-Related Evoked Potentials”. In: *eNeuro* (2018).

[9] O. Hernandez, R. Garcia-Martinez, and V. Monteon. “Alcohol effects on the P2 component of auditory evoked potentials”. In: *Anais da Academia Brasileira de Ciencias* (2014).

[10] F. Lotte. “A Tutorial on EEG Signal Processing Techniques for Mental State Recognition in Brain-Computer Interfaces”. In: *Guide to Brain-Computer Music Interfacing*. Springer, 2014.

[11] R. P. Godavarthy and A. R. Taleciani. “Winter bikesharing in US: user willingness, and operator’s challenges and best practices”. In: *Sustainable cities and society* (2017).

[12] S. Reiss and K. Bogenberger. “Validation of a Relocation Strategy for Munich’s Bike Sharing System”. In: *Transportation Research Procedia* (2016).

[13] “Learning from accelerometer data on a legged robot”. In: *IFAC Proceedings Volumes* (2004).

[14] A. Yamaguchi, K. Ueno, and H. Kashima. “Learning Evolvable Time-series Shapelets”. In: *ICDE*. IEEE, 2022 (to appear).

[15] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. “Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers”. In: *Found. Trends Mach. Learn.* (2011).

[16] A. Yamaguchi, S. Maya, K. Maruchi, and K. Ueno. “LTSpAUC: Learning Time-series Shapelets for Optimizing Partial AUC”. In: *SDM*. SIAM, 2020.

[17] A. Yamaguchi, S. Maya, and K. Ueno. “RLTS: Robust Learning Time-series Shapelets”. In: *ECML PKDD*. Springer, 2020.

[18] L. Hou, J. T. Kwok, and J. M. Zurada. “Efficient Learning of Timeseries Shapelets”. In: *AAAI*. AAAI Press, 2016.

[19] X. Li and J. Lin. “Evolving Separating References for Time Series Classification”. In: *SDM*. SIAM, 2018.

[20] H. Fanaee-T and J. Gama. “Event labeling combining ensemble detectors and background knowledge”. In: *Progress in Artificial Intelligence* (2013).