

# 深層学習を用いた時系列データの要約と分類

山室 冨† 松原 靖子† 川畑 光希† 井手 優介† 梅田 裕平††

古川 忠延†† 大川 佳寛†† 丸橋 弘治†† 稲越 宏弥†† 櫻井 保志†

† 熊本大学大学院 自然科学研究科 〒860-8555 熊本県熊本市中央区黒髪2丁目39番1号

†† 富士通研究所 〒211-8588 神奈川県川崎市中原区上小田中4丁目1番1号

E-mail: †{saeru,kouki,yusuke.i}@dm.cs.kumamoto-u.ac.jp, {yasuko,yasushi}@cs.kumamoto-u.ac.jp,

††{umeda.yuhei,furukawa.tada,okawa.y,maruhashi.koji,inakoshi.hiroya}@jp.fujitsu.com

あらまし 本論文では、深層学習を用い、時系列データを分類するためのアルゴリズムを提案する。提案手法は、多次元センサーデータが与えられたときに、その時系列データの中から重要な特徴を発見し、それらの情報を統計的に要約し、表現する。さらに、要約情報に対して深層学習を適用することにより、時系列データを効率的かつ効果的に分類する。実データをを用いた実験では、提案手法が時系列データの中からラベル分類に関する重要パターンを抽出し、高い精度で分類できることを確認した。

キーワード 時系列データ, 特徴自動抽出, 深層学習, 分類

## 1. はじめに

時系列データは、センサーデータ [9], [12] や Web アクセス履歴 [21], [22] 等、様々なアプリケーションにおいて大量に生成されている。一般に、実際に生成される時系列データは複数のトレンドやパターンを持つことが多い。具体的な例として、ネットワーク通信のモニタリングシステムから発生するデータでは、正常と異常のパターンが存在する。人の動きを捉えたモーションキャプチャデータでは、活動内容に応じて複数の動作パターンを持つ。そうした時系列データを構成する様々なパターンの特徴を捉えることにより、各データに付与された正常/異常、ユーザ ID、習熟度等の属性情報（ラベル）を高精度に分類すること、および、分類の際に重要なパターンを知見として得ることは非常に重要な課題である。

一方、時系列データに含まれるパターンに関する情報、すなわち、パターンの変化点や種類がすべて明らかであることは稀である。収集した膨大な時系列データをすべて人手で確認し、類似パターンをラベリングすることは現実的でない。また、多くのパターン検出に関する先行研究 [13], [34] は、クラスタ数やエラーの閾値等のパラメータ設定やチューニングが必要であり、これらのパラメータが出力結果に大きな影響を与える。したがって、時系列データからのパターン検出においては、解析する時系列データの事前知識を必要とせず、ユーザの介入を必要としない手法が望ましい。

本論文では、ラベル付き大規模時系列データを対象とした分類モデルとして DEEPPLAIT を提案する。より具体的には、複数のセンサーから生成された多次元時系列データ集合が与えられたとき、(a) 各データに含まれる時系列パターンの変化点と種類を自動的に把握し、データセット全体の要約情報を抽出する。(b) 得られた要約情報を元に多層ニューラルネットワーク (Deep Neural Network: DNN) を構築することで、高い分類精

度を実現するだけでなく、分類において重要な時系列パターンを自動的に学習する。ここで、多次元時系列シーケンスにおける時系列パターンを本研究では「レジーム (regime)」と呼び、各サンプルデータに含まれるレジーム、および、大量の時系列シーケンス間で共通するレジームをデータセット全体の要約情報として抽出する。

深層学習は、多層ニューラルネットワークの学習によってデータの hochi 表現を獲得可能なことで知られ、画像認識や音声認識の分野で数多くの研究事例が報告されている [1], [4], [31]。しかしながら、既存モデルは高い分類精度を発揮する一方で、判断根拠が明らかでないという解釈性の問題を抱えている [25]。この問題を解決するために、自然言語処理では Attention [2] という技法がよく用いられる。提案モデルは、時系列データに含まれる複雑な特徴を学習することが可能な畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) と再帰型ニューラルネットワーク (Recurrent Neural Network: RNN) で構成されるネットワークに対し、Attention を導入する。Attention を用いて時系列データに含まれるレジームごとの重要度を算出することで、ラベル分類の根拠となる人の直感に合った重要パターンを明らかにする。

### 1.1 具体例

図 1 は、提案手法を用いて時系列データのラベル分類を行なったときの、あるサンプルデータから検出された時系列パターンとそれらに対する重要度の学習結果を示す。図上部のオリジナルデータは、利き手側の肘、膝にそれぞれ取り付けられた 3 軸加速度センサー、および、眼鏡型の IoT デバイスを用いてフリースローの様子と視線を捉えた 8 次元の時系列シーケンスである。提案手法は、与えられた時系列シーケンスから自動的に“ボールを放つ (レジーム #2)”, “ルーチンワーク (レジーム #3)” 等の時系列パターンを発見し、シュートの成功/失敗を分類するために重要となるレジームを学習した。各セグメント

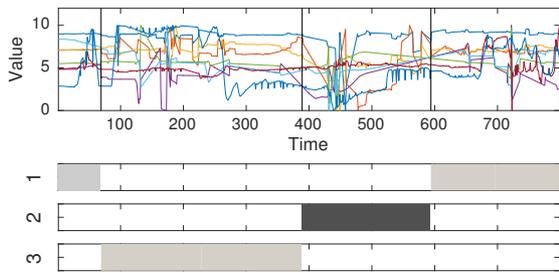


図 1: *Free-Throw* データセットにおける DEEPPLAIT のセグメンテーション結果とラベル分類に対する各セグメントの重要度。色が濃いほどシュートの成功に重要なパターンであることを示す。

の色は学習した重要度を表し、色が濃くなるほど重要であることを示す。一般的にフリースローの成功には、ボールを放つ瞬間の肘の位置と膝の屈伸に加えて放った後のフォロースルーが重要であることから、DEEPPLAIT は分類タスクにおける判断根拠となるレジームを有意に抽出したといえる。

### 1.2 本論文の貢献

提案手法は以下の特徴を持つ。

- 各シーケンスの要約情報として時系列パターン (レジーム) の種類を把握し、それぞれの適切な変化点を自動的に発見する。さらに、提案手法はパターン変化点の最適解の検出を保証する。
- データセット全体の要約情報として、大量の時系列データ間で共通する時系列パターンを検出する。
- 複数のレジームパターンを有する時系列データの特徴を効果的に学習するための多層ニューラルネットワークを構築し、高い精度で分類を行う。
- ニューラルネットワークに Attention を適用し、分類モデルの学習と同時に各時系列データにおいて分類問題に対する各時系列パターンの重要度を算出する。

本論文の構成は以下の通りである。2. 章では関連研究について述べる。3. 章では問題定義、4. 章では時系列データの要約・分類モデル、5. 章では評価実験について述べ、6. 章にて本論文をまとめる。

## 2. 関連研究

時系列データの解析に関する研究は様々な分野で進められている [5], [6], [23], [26], [27]。自己回帰モデル (AR: autoregressive model), 線形動的システム (LDS: linear dynamical systems), カルマンフィルタ (KF: Kalman filters) は代表的な技術であり、これらに基づく時系列の解析と予測手法が数多く提案されている [15], [18], [32]。時系列データからの情報抽出に関しては、Li らが文献 [17] において、欠損を含む大規模時系列シーケンス集合のためのアルゴリズムである DynaMMo を提案している。DynaMMo は LDS に基づき、時系列データのパターンを発見

し、シーケンスのセグメント化の能力を持つ。また、隠れマルコフモデル (HMM: Hidden Markov model) は音声認識を含む様々な分野において、時系列解析手法として広く利用されている。Wang ら [34] は pHMM (pattern-based hidden Markov model) を提案している。pHMM は時系列のセグメント化とクラスタリングのための動的モデルであり、時系列シーケンスをマルコフモデルに基づいて線形のセグメントに分割する能力をもつ。著者らは文献 [20] において、隠れマルコフモデルに基づく多次元時系列シーケンスのための特徴自動抽出手法を提案した。

時系列データの回帰・分類問題に対し、近年発展の目覚ましい深層学習モデルが多数報告されている [3], [19], [36]。時系列データに対する深層学習では、時間関係の特徴を抽出することができる RNN がよく用いられる。映像の連続的なフレームを時系列データとして扱う Donahue ら [10] の手法では、CNN で各フレーム画像の空間的な特徴を抽出し、それらを時系列に並べたものを RNN で学習して認識を行う。Yao ら [36] は複数のセンサーから得られる多次元時系列データの分類回帰モデルとして DeepSense を提案した。DeepSense は時系列データを微小時間隔に分割 (セグメント化) し、各セグメントから多層 CNN を用いて各センサーの軸間の特徴、およびセンサー間の特徴を抽出後、それらに RNN を適用することで高精度を達成した。

深層学習は様々な問題に対して高い精度を発揮するモデルを構築することを可能にしてきたが、ネットワークのパラメータから判断根拠を人間が理解可能な形で取り出すことが難しいという解釈性の問題を抱えている。この問題に対するひとつの解として、Bahdanau ら [2] は自然言語翻訳タスクのための Attention を提案した。Attention は、RNN Encoder によって入力系列を単一の固定長ベクトルに埋め込む際に、各隠れ層の出力ベクトルの加重平均を用いることで、機械翻訳における文脈中の重要な単語を加味して翻訳することが可能である。複数の特徴に重み付けすることができる Attention は機械翻訳以外の様々な分野で性能向上に貢献し [37]、時系列データ解析のために適用された研究事例 [8], [24], [33], [35] も多数報告されている。本研究の特徴のひとつは、時系列データをパターンごとに分割し、そうした意味のあるセグメントから抽出した特徴量に対して Attention を適用することである。

## 3. 問題定義

ここでは、本論文で必要な概念について定義を行う。

$\mathbf{X} = \{\mathbf{X}^{(k)}\}_{k=1}^K$  を  $K$  個のセンサーによって観測された長さ  $t$  の多次元時系列データとし、各センサー  $k$  は  $d^{(k)}$  個の測定軸を持つとする。本研究では、このようなラベル付き多次元時系列データの集合  $\mathcal{D} = \{\mathbf{X}_i\}_{i=1}^N$  に対する分類を行う。

本研究ではまず、単一の時系列シーケンス  $\mathbf{X}$  を  $m$  個のセグメント集合  $\mathcal{S} = \{s_1, \dots, s_m\}$  に分割する。ここで、 $s_i$  は  $i$  番目のセグメントの開始点、終了点を含み、(つまり、 $s_i = \{t_s, t_e\}$ )、各セグメントは重複がないものとする。次に、発見したセグメント集合を類似セグメントのグループ (レジーム: regime)

表 1: 主な記号と定義

記号	定義
$\mathcal{D}$	$N$ 個の時系列データ集合 $\mathcal{D} = \{\mathbf{X}_1, \dots, \mathbf{X}_N\}$
$\mathbf{X}$	長さ $t$ のラベル付き多次元時系列データ $\mathbf{X} = \{\mathbf{X}^{(k)}\}_{k=1}^K$
$\mathbf{X}^{(k)}$	センサー $k$ における $d^{(k)}$ 次元シーケンス $\mathbf{X}^{(k)} \in \mathbb{R}^{d^{(k)} \times t}$
$\mathbf{X}'$	$n$ 個のミニバッチ $\mathbf{X}' = \{\mathbf{X}_i\}_{i=1}^n \subset \mathcal{D}$
$\mathcal{X}$	深層学習の入力 $\mathcal{X} \in \mathbb{R}^{d^{(k)} \times 2f \times m}$
$\mathbf{y}$	各サンプルに付与されたラベル
$\mathcal{S}$	$m$ 個のセグメント集合 $\mathcal{S} = \{s_1, \dots, s_m\}$
$\mathcal{S}_i$	$i$ 番目のレジームに含まれるセグメント集合 $\mathcal{S}_i \subset \mathcal{S}$
$\alpha$	$m$ 個のセグメントの重要度 $\alpha = \{\alpha_1, \dots, \alpha_m\}$
$\Theta$	$r$ 個のレジームパラメータ集合 $\Theta = \{\theta_1, \dots, \theta_r\}$
$\mathcal{M}^L$	各サンプル $\mathbf{X}$ の要約情報 $\mathcal{M}^L = \{m, r, \mathcal{S}, \Theta, \alpha\}$
$\mathcal{M}^G$	$N$ 個のサンプル全体の要約情報 $\mathcal{M}^G = \{\mathcal{M}_1^L, \dots, \mathcal{M}_N^L, \Theta^G\}$

に分ける．これをローカルレジームと呼び，各レジームは統計モデル  $\theta$  で表現される．また，各サンプルから得られるレジームに加えてデータセット全体での共通パターンであるグローバルレジームを検出する．最終的に，これらの要約情報を利用した深層学習によって各時系列データを分類するとともに，時系列パターン（レジーム）の分類問題に対する重要度  $\alpha = \{\alpha_1, \dots, \alpha_m\}$  を算出する．

[定義 1] (ローカルレジーム) 多次元時系列シーケンス  $\mathbf{X}$  が与えられたとき， $\mathbf{X}$  を表現，要約する  $\mathcal{M}^L = \{m, r, \mathcal{S}, \Theta, \alpha\}$  をローカルレジームと呼び，以下で構成される．

(1) セグメントの総数  $m$  と各セグメントの位置：

$$\mathcal{S} = \{s_1, \dots, s_m\}$$

(2)  $r$  個のレジームを表現するモデルのパラメータ集合：

$$\Theta = \{\theta_1, \dots, \theta_r\}$$

(3) 各セグメントの重要度：

$$\alpha = \{\alpha_1, \dots, \alpha_m\}$$

[定義 2] (グローバルレジーム)  $\mathcal{D}$  を表現する全ての要約情報  $\mathcal{M}^G = \{\mathcal{M}_1^L, \dots, \mathcal{M}_N^L, \Theta^G\}$  をグローバルレジームとし， $\Theta^G = \{\theta_1, \dots, \theta_g\}$  をシーケンス間で共通するパターンを示す  $g$  個の統計モデルの集合とする．

したがって，本論文で扱う問題は以下のように表される．

[問題 1] 多次元時系列データ  $\mathbf{X} \subset \mathcal{D}$  が与えられたとき，

- (1) 要約情報  $\mathcal{M}^L$ ，および， $\mathcal{M}^G$  を抽出する
- (2)  $\mathbf{X}$  に含まれる各時系列パターン（レジーム）から特徴量  $\mathcal{X}$  を生成し， $\mathbf{X}$  に付与されたラベル  $\mathbf{y}$  を高精度に分類する
- (3) 各セグメントの分類問題に対する重要度  $\alpha$  を求める

結論として，本論文の目的はデータセット全体の最適な要約情報を抽出し，要約情報に基づき生成した特徴量によってそれぞれの時系列シーケンスを分類するとともに，分類問題における時系列パターンの重要度を算出することである．ここで非常に重要な課題は，(a) どのように  $\mathbf{X}$  の要約情報  $\mathcal{M}^L$  を推定するか，(b) どのようにデータセット全体のグローバルレジーム  $\mathcal{M}^G$  を決定するか，(c) 時系列パターンの重要度をパラメータに持つ分類モデルをどのように構築し，高精度に学習するかである．本論文では，多次元時系列データを要約・分類するためのアルゴリズムである DEEPPLAIT を提案する．

## 4. 提案モデル

本章では，多次元時系列データを要約・分類するためのアルゴリズムである DEEPPLAIT について述べる．

### 4.1 概要

本論文で提案する手法のフレームワークを図 2 に示す．まず初めに，与えられた多次元時系列シーケンス集合の中から Mini-batch  $\mathbf{X}'$  を取り出しセグメンテーションを施すことで各時系列シーケンスの要約情報  $\mathcal{M}^L$  を抽出する．次に，得られた  $\mathcal{M}^L$  の中から，シーケンス間で共通する時系列パターンを表す要約情報  $\mathcal{M}^G$  を検出・更新する．また，深層学習を用いたモデルに要約情報  $\mathcal{M}^L$  を利用し，各サンプルに付与されたラベル  $\mathbf{y}$  を分類すると同時に，時系列パターンの重要度  $\alpha$  を算出する．以降の節では，入力から順に深層学習モデルの詳細と

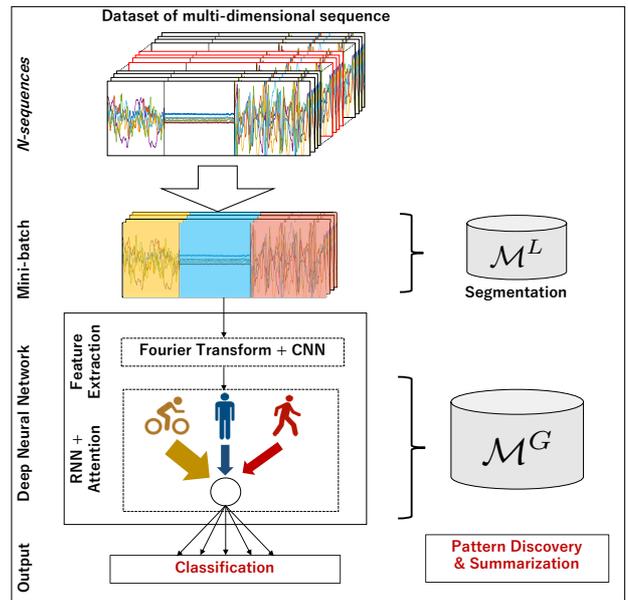


図 2: 提案モデルのフレームワーク

特徴抽出のためのアルゴリズムについて説明する．

### 4.2 時系列データからの自動特徴抽出

ここでは，多次元時系列データ  $\mathbf{X}$  が与えられたとき， $\mathbf{X}$  の要約情報を得るためのアルゴリズムについて述べる．提案手法は以下の 2 つのアルゴリズムで構成される．

- 時系列パターンの自動発見：ミニバッチ  $\mathbf{X}'$  が与えられたとき， $\mathbf{X}'$  中の各シーケンスに含まれるすべての時系列パターン，すなわち，ローカルレジーム  $\mathcal{M}^L$  を発見し，要約情報  $\{\mathcal{M}_i^L\}_{i=1}^n$  を得る．
- 共通パターン分析： $\{\mathcal{M}_i^L\}_{i=1}^n$  から，データセット全体に共通するパターンであるグローバルレジーム  $\Theta^G$  を検出し，要約情報  $\mathcal{M}^G$  を更新する．

時系列パターンの自動発見．まず，最も簡単な部分問題として，単一の時系列シーケンス  $\mathbf{X}$  が与えられたとき， $\mathbf{X}$  に含まれる時系列パターンを発見する．本研究では，与えられた時系列データに関する事前知識を必要とせず，自動的に時系列パ

---

**Algorithm 1** REGIMEASSIGNMENT ( $\mathbf{X}', \mathcal{M}^G$ )

---

**Input:** Mini-batch  $\mathbf{X}'$  and G-regimes  $\mathcal{M}^G$ **Output:** Updated G-regimes  $\mathcal{M}^G$ , i.e., L-regimes  $\{\mathcal{M}_1^L, \dots, \mathcal{M}_n^L\}$  and model parameter set  $\Theta^G$ 

```
1: /* Find L-regimes for each sequence in  $\mathbf{X}'$  */
2: for  $i = 1 : n$  do
3:    $\{m, r, \mathcal{S}, \Theta\} \leftarrow \text{AUTOPLAIT}(\mathbf{X}_i)$ ;
4:   for  $j = 1 : r$  do
5:      $\theta^* \leftarrow \arg \min_{\theta \in \Theta^G} \text{Cost}_C(\mathbf{X}_i[\mathcal{S}_j]|\theta)$ ;
6:     Compute  $\Delta \text{Cost}_C(\mathbf{X}_i[\mathcal{S}_j]|\theta_j, \theta^*)$ ; // Equation (2)
7:     if  $\text{Cost}_M(\theta_j) < \Delta \text{Cost}_C(\mathbf{X}_i[\mathcal{S}_j]|\theta_j, \theta^*)$  then
8:       Add  $\theta_j$  into  $\Theta^G$ ;
9:     else
10:       $\theta^* \leftarrow \text{MODELUPDATE}(\mathbf{X}_i[\mathcal{S}_j])$ ;
11:    end if
12:  end for
13: end for
14: return  $\{\mathcal{M}_1^L, \dots, \mathcal{M}_n^L, \Theta^G\}$ ;
```

---

ターンを発見するために、最小記述長 (Minimum description Length: MDL) に基づく手法である AutoPlait [20] を用いて解析を行う。直感的には、データがより圧縮されるようなモデルを最適とする。 $\mathbf{X}$  とモデルパラメータ集合  $\mathcal{M}^L$  が与えられたときの総符号長  $\text{Cost}_T(\mathbf{X}; \mathcal{M}^L)$  は以下の式で表される。

$$\text{Cost}_T(\mathbf{X}; \mathcal{M}^L) = \text{Cost}_M(\mathcal{M}^L) + \text{Cost}_C(\mathbf{X}|\mathcal{M}^L) \quad (1)$$

ここで、 $\text{Cost}_M(\mathcal{M}^L)$  はモデル表現コスト<sup>(注1)</sup>を示し、 $\text{Cost}_C(\mathbf{X}|\mathcal{M}^L)$  はモデルが与えられたときの  $\mathbf{X}$  の符号化コストを示す。[20] は、貪欲法によって  $\mathbf{X}$  の分割とモデル推定を繰り返し、式 (1) を最小化する  $\mathbf{X}$  の要約情報  $\mathcal{M}^L$  を自動的に推定する。

**共通パターン分析.** 続いて、各シーケンスから得られたレジームから、類似する時系列パターンを検出するためのアルゴリズム REGIMEASSIGNMENT について述べる。

具体的なアルゴリズムの流れをアルゴリズム 1 に示す。まず、 $\Theta \in \mathcal{M}^L$  中の各レジーム  $\theta_j$  に割り当てられた部分シーケンス  $\mathcal{S}_j$  に対し符号化コストが最小となるグローバルレジーム  $\theta^*$  を求める。次に、 $\theta_j, \theta^*$  それぞれで部分シーケンスを表現したときの符号化コストの差分を以下の式で求める。

$$\begin{aligned} \Delta \text{Cost}_C(\mathbf{X}_i[\mathcal{S}_j], \theta_j, \theta^*) \\ = \text{Cost}_C(\mathbf{X}_i[\mathcal{S}_j]|\theta^*) - \text{Cost}_C(\mathbf{X}_i[\mathcal{S}_j]|\theta_j) \end{aligned} \quad (2)$$

$\theta_j$  のモデル表現コストが上記のコストより小さくなると、 $\theta_j$  は新たな時系列パターンを示し、 $\theta_j$  を  $\Theta^G$  に追加する。レジームの追加が不要な場合、 $\theta_j$  は  $\theta^*$  と共通の時系列パターンを示し、 $\theta_j$  に割り当てられた部分シーケンス  $\mathcal{S}_j$  を用いて  $\theta^*$  のモデルパラメータを更新する。

---

**Algorithm 2** INPUTPRODUCER ( $\mathbf{X}, \mathcal{M}^L$ )

---

**Input:** Given sequence  $\mathbf{X}$  and L-regimes  $\mathcal{M}^L$ **Output:** Input for DNN  $\mathcal{X} = \{\mathcal{X}^{(k)}\}_{k=1}^K$ 

```
1: /* Produce Input for DNN */
2: for  $k = 1 : K$  do
3:   for  $i = 1 : m$  do
4:     Split  $\mathbf{X}^{(k)}[\mathcal{S}_i]$  into T subsequences;
5:     for  $j = 1 : T$  do
6:       Interpolate  $\mathbf{X}^{(k)}[\mathcal{S}_i][j]$  to f time stamps;
7:       Fourier Transform to  $\mathbf{X}^{(k)}[\mathcal{S}_i][j]$  along each sensor axis;
8:       Add magnitude and phase pairs into  $\mathcal{X}^{(k)}$ 
9:     end for
10:  end for
11:  Add  $\mathcal{X}^{(k)}$  into  $\mathcal{X}$ 
12: end for
13: return  $\mathcal{X} = \{\mathcal{X}^{(k)}\}_{k=1}^K$ ;
```

---

### 4.3 畳み込みニューラルネットワーク

続いて、前節のモデルで抽出したローカルレジーム  $\mathcal{M}^L$  に基づき入力  $\mathcal{X}$  を生成し、図 3 に示す多層ニューラルネットワークによる学習を行う。全結合のニューラルネットワークはデータの形状が無視されるため、画像等の空間的な情報を持つデータに対して効果的に学習が行えないという問題点があった。この課題を克服する方法として、視野脳の情報伝達システムを適用した畳み込みニューラルネットワーク (CNN) が提案された。畳み込み層では、入力データに対してフィルター (カーネル) を一定間隔でスライドさせながら積和演算を行い、入力データの局所的な特徴を抽出することができる。

本研究では CNN を用いて効果的に各時系列パターンの特徴を学習するため、 $\mathcal{X}$  の生成方法として INPUTPRODUCER (アルゴリズム 2) を提案する。具体的には、入力データ  $\mathcal{X} = \{\mathcal{X}^{(k)}\}_{k=1}^K$  を、各時系列シーケンス  $\mathbf{X}$  中の各セグメントから、非重複かつ微小な幅  $\tau$  で分割した各要素に対してフーリエ変換を適用する。周波数領域の次元数を  $f$  とし、振幅スペクトルと位相スペクトルの両方を用いると、 $\mathcal{X}^{(k)}$  は  $d^{(k)} \times 2f \times mT$  という形状になる。周波数領域は、時間領域の時系列データに比べてよりよい情報をもつため、本研究では上記の方法で生成した特徴量を入力データとする。

提案モデルの畳み込み層は、 $K$  個のセンサー別 CNN と、その出力を入力とするセンサー統合 CNN の 2 つから構成されており、3 種類の特徴/関係性を抽出することを目的としている。センサー別 CNN では、入力  $\mathcal{X}^{(k)}$  に対し、 $K$  個の CNN が入力に埋め込まれている周波数領域内での関係性と、センサーの測定軸間の関係性の 2 つの特徴を抽出することを目的としている。周波数領域の情報はたいてい隣り合う周波数で多くのローカルなパターンを有しているため、フーリエ変換後の入力値を CNN で学習する意味は大いにある。次に、センサー別 CNN で得られた  $K$  個のベクトルを連結した行列をセンサー結合 CNN の入力とする。ここではセンサー間の関係性を特徴として抽出し、得られたベクトルを時間軸方向に連結した行列を次節で紹

---

(注1): 本論文では、浮動小数点の表現コストを  $4 \times 8$  ビットとする。

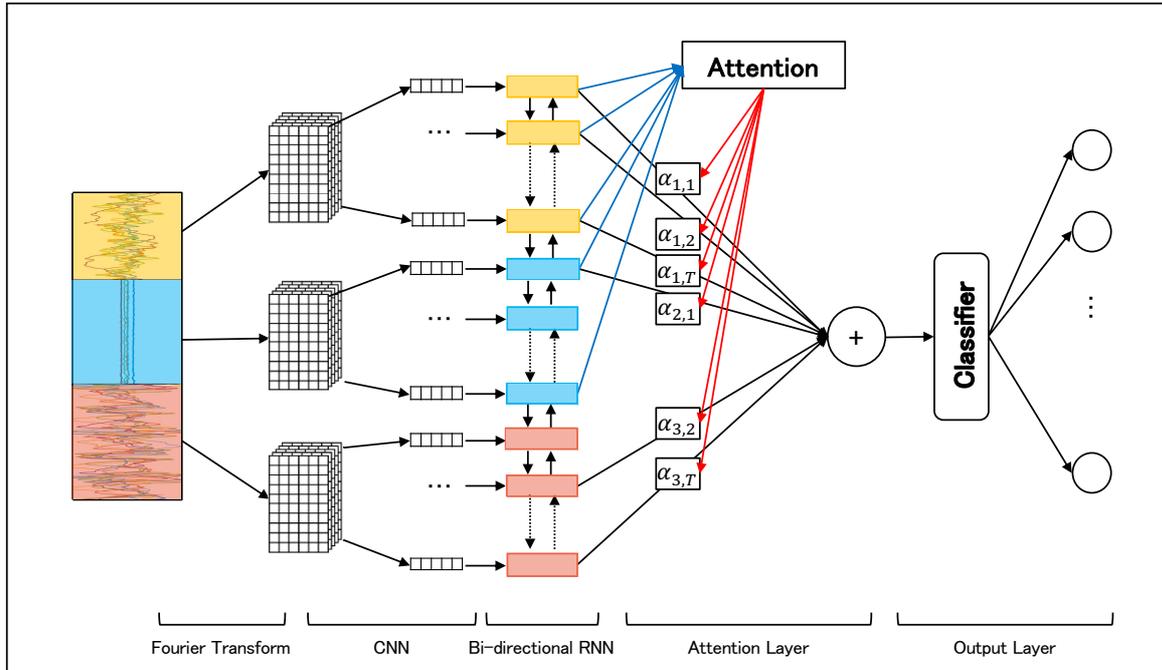


図 3: 提案モデルのネットワーク構造: (a) 特徴抽出層: 時系列データをパターンごとに分割し特徴量を生成する. (b) CNN 層: 分割された各特徴量におけるセンサー軸間, センサー間の関係性を抽出する. (c) RNN 層: CNN 層で得られた特徴の時間関係を学習する. (d) Attention 層: RNN 層で時間関係を学習するとき, 時系列パターンの重要度を加味する.

介する RNN の入力とする.

#### 4.4 再帰型ニューラルネットワーク

ここでは, CNN 層の出力に対して, 再帰型ニューラルネットワーク (Recurrent Neural Network: RNN) による時間関係の特徴抽出を行う. RNN は時系列のデータを扱うために, 任意の時刻  $t$  での出力に過去の出力を加味する. これにより, 時間関係をもつ入力に対して特徴を抽出することを可能とする. しかしながら, RNN は誤差関数, 各層間の重みに関する勾配計算を行う際に層数の増加に伴って勾配値が発散や消失してしまう恐れがある (長期依存性). この問題を克服する方法として, ネットワークへの入力を制御するために, RNN の中間層にゲートを加えた LSTM (Long Short-Term Memory) [11] が提案されている. また, LSTM と類似した性能を持ちながら, より簡潔なモデルで表現できる GRU (Gated Recurrent Unit) [7] や双方向にすることでモデリングの能力を向上させた bi-directional RNN [28] が報告されている. 本論文では RNN 層に双方向の GRU を用い, 時間関係の特徴を抽出する役割を持たせる.

#### 4.5 Attention

最後に, 前節の RNN 層の出力  $h$  に対して, Attention を用いることで, セグメントの重要度  $\alpha$  を算出する.

$$u_i = \tanh(W_a h_i + b_a) \quad (3)$$

$$\alpha_i = \frac{\exp(u_i^T u_a)}{\sum_i \exp(u_i^T u_a)} \quad (4)$$

$$v = \sum_i \alpha_i h_i \quad (5)$$

式 (4) で計算される  $\alpha_i$  のベクトル  $\alpha$  をセグメントの重要度とする. また, 式 (3) で表現されるように, RNN 層の各出力ベ

クトル  $h_i$  に対して重み  $\alpha$  の加重平均した  $v$  を Attention の出力とする. この出力  $v$  に対して線型結合の出力層を用いることで, 各サンプルに対するラベル  $y$  の出力確率を計算する.

さらに, 汎化性能の向上のため, クロスエントロピー損失関数に正則化項を加えたものをコスト関数とする. DEEPPLAIT のモデルを関数  $\mathcal{F}(\cdot)$ , 訓練データのサンプルの組を  $(\mathcal{X}, \mathbf{y})$  とすると, コスト関数  $\mathcal{L}$  は以下ようになる.

$$\mathcal{L} = \ell(\mathcal{F}(\mathcal{X}), \mathbf{y}) + \sum_j \lambda_j P_j \quad (6)$$

$\ell(\cdot)$  はクロスエントロピー損失関数,  $P_j$  は正則化項であり,  $\lambda_j$  は正則化項の寄与率を調整するパラメータである.

## 5. 評価実験

本章では, DEEPPLAIT の有効性を検証するために, 以下の項目について実データを用いて実験を行なった結果を示す.

- Q1 時系列データに対する提案手法の分類精度
- Q2 セグメンテーション結果と分類問題に対する時系列パターンの重要度
- Q3 セグメンテーションと要約情報

提案モデルは損失関数にクロスエントロピー, 最適化手法に Adam [16], 汎化性能の向上のために CNN 層には dropout 関数 [29] とバッチ正規化 [14] を適用した. 実験は 192GB のメモリ, Intel Xeon 3.5GHz の CPU と NVIDIA Quadro M5000 の GPU を搭載したマシン上で実施し, 以下のデータセットを使用した.

- HHAR: 本データセットは, Allan ら [30] が 9 人のユーザ

に対する6種類のアクティビティ(bike, sitting, standing, walking, climbStair-up, climbStair-down)の加速度, 角速度をスマートフォンを用いて収集したものである。本実験では, 時間軸方向に bike, standing, walking の順で3つのアクティビティを結合したシーケンスを3100サンプル作成した。

- *Free-Throw*: 本データセットは, バスケットボールのフリースローにおいて, 競技者の視線移動を計測する Tobii Pro グラス 2<sup>(注2)</sup> と競技者の利き手の肘と利き足の膝に装着した加速度センサー<sup>(注3)</sup> を用いて収集したものである。分類するラベルはゴールの有無で, サンプル数は500である。

### 5.1 Q1. 分類精度

はじめに, 提案モデルの分類精度を検証するため, 既存手法を用いて5分割の交差検証による比較を行った。本実験では, DEEPPLAIT のパラメータとして, CNN層のフィルター数を64, RNN層の隠れユニット数を200, dropout率を0.6に設定し, モデル学習を行った。図4は, DEEPPLAIT とRBFカーネルを利用したSVM, 6層のCNNと2層のRNNで構成されるDeepSense[36]との, 分類精度と95%信頼区間の比較である。DEEPPLAIT は *HHAR* のテストデータに対して  $94.74 \pm 0.0140\%$  の分類精度を示し, SVMと比較して2.56%の精度向上, および, より深いDNNの構造をもつDeepSenseと同等の精度を達成した。また, DEEPPLAIT はフリースローのゴールの有無を分類する *Free-Throw* のテストデータに対して  $82.71 \pm 0.0384\%$  の分類精度を示した。本データセットでは, SVMやDeepSenseと比較して分類精度の向上がみられたことから, フリースローがゴールに入るかどうかを判断する上で重要となる, 腕や膝, 視線の動きをもつ時系列パターンを捉えることは, 高い分類精度の実現に特に有効であったと言える。

### 5.2 Q2. 時系列パターンの重要度

図5に *HHAR* データセットに対してDEEPPLAITによるセグメンテーションを施し, 重要度を学習した結果の例を示す。各図の上部はオリジナルデータを示し, 下部は時系列パターンが属するレジームとパターンの開始点および終了点を示す。また, 長方形の色が濃いほど重要度が高く, 薄いほど重要度が低いことを表している。図5に示すように, DEEPPLAIT はサンプル1から  $m=3$  個のセグメント,  $r=3$  個のレジームを抽出した。各セグメントの重要度は高い順に, “bike”, “walk”, “stand” という結果になった。ユーザ間で特徴に差異がないと考えられる “stand” のラベルが付与されたセグメントと比較して, “bike” や “walk” のセグメントの重要度が高くなったという結果から, ユーザ分類タスクにおいて重要となったセグメントを有意に抽出できたと考えられる。一方, サンプル2に対しては,  $m=4$  個のセグメント,  $r=4$  個のレジームを抽出した。本サンプルでは “bike” のラベルが付与された部分シーケンス

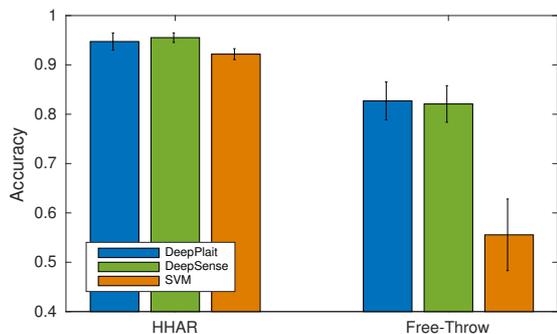
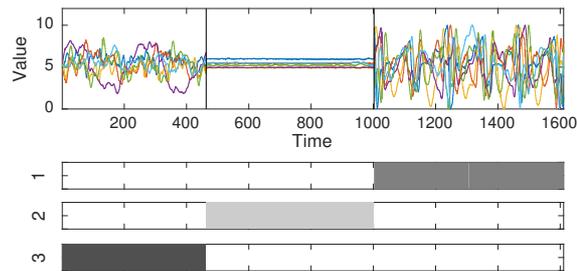
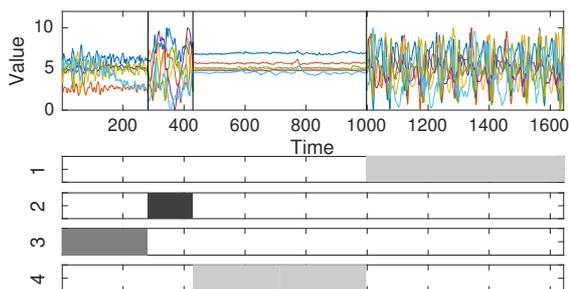


図4: 各データセットにおける分類精度の比較



(a) サンプル1



(b) サンプル2

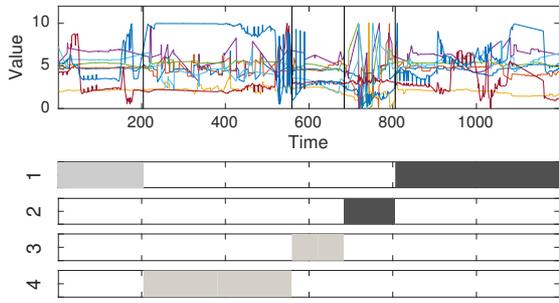
図5: *HHAR* データセットにおけるDEEPPLAITのセグメンテーション結果とラベル分類に対する各セグメントの重要度。

が2つのセグメント(セグメント#1, セグメント#2)に分割されており, 振幅の大きいセグメント#2の重要度が高くなった。これは, サンプル2のセグメント#2がサンプル1と比較して特徴的な動きを示しているため, ユーザ分類タスクにおいては個人を特定する判断基準として重要であると学習したと考えられる。

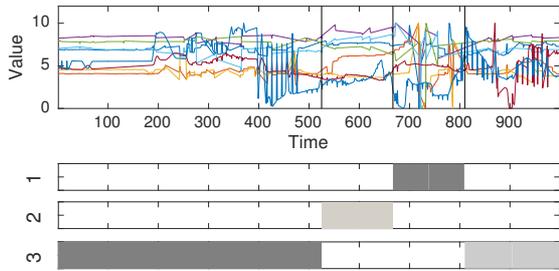
次に, *Free-Throw* データセットに対してセグメンテーションを施し, 重要度を学習した結果の例を図6に示す。DEEPPLAITは, フリースローが成功したサンプルから  $m=5$  個のセグメント,  $r=4$  個のレジームを抽出した。時系列パターンの重要度はボールを投げる瞬間のセグメント#4に加えてボールを投げた後のセグメント#5についても高くなった。一般的に, シュートの上達に重要な要素はボールを投げた後の姿勢を指すフォロースルーであると言われている。このサンプルは, バスケットボー

(注2) : <https://www.tobiiipro.com/product-listing/tobii-pro-glasses-2/>

(注3) : <https://mono-wireless.com/jp/products/TWE-Lite-2525A/>



(a) フリースローに成功したサンプル



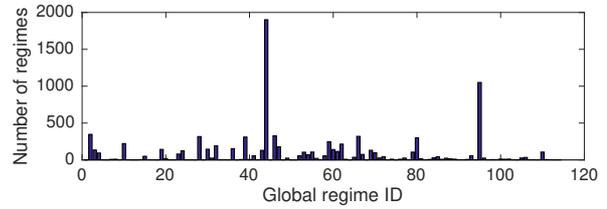
(b) フリースローに失敗したサンプル

図 6: *Free-Throw* データセットにおける DEEPPLAIT のセグメンテーション結果とラベル分類に対する各セグメントの重要度。

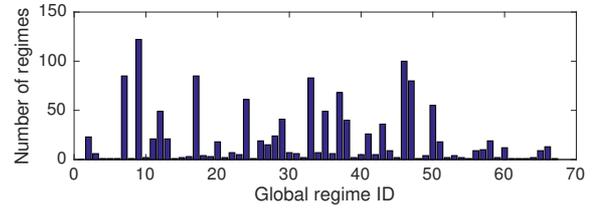
ル熟練者が投げたものであり、シュート後のフォロースルーに重要度が高く算出されていることから、DEEPPLAIT はフリースローのゴールの有無を判断する基準として有意なセグメントを抽出しているといえる。一方、フリースローに失敗したあるサンプルに対しては、 $m = 4$  個のセグメント、 $r = 3$  個のレジームを抽出した。本サンプルにおいて、DEEPPLAIT は投げる瞬間のセグメント #3 だけでなく、投げる前の準備段階であるセグメント #1 に対しても高い重要度を算出した。このサンプルはバスケットボール未経験者が投げたもので、DEEPPLAIT はフリースローに失敗した原因として投げる前に経験者が毎回行うルーチンワークの時系列パターンを抽出した。フリースローでは心理的に不安な要素を取り除くことが非常に大切であり、このサンプルではそのためのルーチンワークがうまくいっていないと考えられる。以上より、DEEPPLAIT は分類タスクにおいて判断基準となった重要な時系列パターンの学習を可能としている。

### 5.3 Q3. セグメンテーションと要約情報

図 7 は各データセットにおいて検出されたグローバルレジームに対し割り当てられたレジームの数を示す。REGIMEASSIGNMENT は、HHAR データセットから合計 114 個のグローバルレジームを発見し、データセット中の各シーケンスを表現する 8937 個のローカルレジームを 1.27% に要約する結果となった。その中で、最多の 1902 個のローカルレジームが割り当てられたグローバルレジーム #43 は、主に “bike” のラベルが付与された部分シーケンスを保持していた。したがって、類似した時



(a) HHAR



(b) *Free-Throw*

図 7: 各データセットの要約情報

系列パターンの検出/要約に成功しているといえる。一方、302 個のローカルレジームが割り当てられたグローバルレジーム #45 では、ユーザー “b” と “g” に関する “stand” の部分シーケンスを共通パターンとしていたことから、類似した時系列パターンを持つユーザーを把握することも可能である。次に、*Free-Throw* データセットにおけるグローバルレジームの総数は 67 個で、データセット中の各シーケンスから得られる 1324 個のローカルレジームを 5.06% に要約する結果となった。最多の 122 個のローカルレジームが割り当てられたグローバルレジーム #8 は、フリースローに成功したサンプルの “ボールを放つ” 部分シーケンスを多く検出していた。このことから、フリースローで得点するために “ボールを放つ” 代表的な部分シーケンスをこのグローバルレジームから把握できる。したがって、REGIMEASSIGNMENT によって検出されるグローバルレジームの要約情報、すなわち、ユーザー間、もしくは分類タスクのラベル間の類似パターンを把握することで、データセット中の代表的な時系列パターンを知見として得ることが可能となる。

## 6. むすび

本論文ではラベルつき大規模時系列データに対し深層学習を用いた要約と分類を行う手法として DEEPPLAIT を提案した。DEEPPLAIT は、与えられた時系列データに関する事前知識無しにデータセットに含まれる典型的なパターンを発見することが可能である。また、抽出した時系列データの要約情報に基づき分類問題に重要となる時系列パターンを学習する。実データを用いた評価実験では、提案手法が既存手法と比較して高い精度を発揮し、学習結果から得られた重要度を確認することで人の直感に合った重要な時系列パターンを発見できることを確認した。

## 7. 謝 辞

本研究の一部は JSPS 科研費 JP15H02705, JP17H04681, JP16K12430, JST さきがけ, 総務省 SCOPE (受付番号 162110003), 厚生労働科学研究費補助金 (H29-ICT-一般-007)

及び国立研究開発法人日本医療研究開発機構（AMED）の臨床研究等 ICT 基盤構築研究事業の助成を受けたものです。

## 文 献

- [1] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen, et al. Deep speech 2: End-to-end speech recognition in english and mandarin. In *International Conference on Machine Learning*, pages 173–182, 2016.
- [2] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. *ICLR*, 2015.
- [3] I. M. Baytas, C. Xiao, X. Zhang, F. Wang, A. K. Jain, and J. Zhou. Patient subtyping via time-aware lstm networks. In *KDD*, pages 65–74, 2017.
- [4] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov. Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*, 2016.
- [5] G. E. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice Hall, Englewood Cliffs, NJ, 3rd edition, 1994.
- [6] L. Chen and R. T. Ng. On the marriage of lp-norms and edit distance. In *VLDB*, pages 792–803, 2004.
- [7] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.
- [8] Y. G. Cinar, H. Mirisae, P. Goswami, E. Gaussier, A. Ait-Bachir, and V. Strijov. Position-based content attention for time series forecasting with sequence-to-sequence rnns. In *ICNIP*, pages 533–544, 2017.
- [9] G. De Francisci Morales, A. Bifet, L. Khan, J. Gama, and W. Fan. Iot big data stream mining. In *KDD, Tutorial*, pages 2119–2120, 2016.
- [10] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2625–2634, 2015.
- [11] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber. Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems*, 2017.
- [12] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami. Internet of things (iot): A vision, architectural elements, and future directions. *Future Gener. Comput. Syst.*, 29(7):1645–1660, 2013.
- [13] D. Hallac, S. Vare, S. Boyd, and J. Leskovec. Toeplitz inverse covariance-based clustering of multivariate time series data. In *KDD*, pages 215–223. ACM, 2017.
- [14] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, pages 448–456, 2015.
- [15] A. Jain, E. Y. Chang, and Y.-F. Wang. Adaptive stream resource management using kalman filters. In *SIGMOD*, pages 11–22, 2004.
- [16] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *ICLR*, 2015.
- [17] L. Li, J. McCann, N. Pollard, and C. Faloutsos. Dynammo: Mining and summarization of coevolving sequences with missing values. In *KDD*, 2009.
- [18] L. Li, B. A. Prakash, and C. Faloutsos. Parsimonious linear fingerprinting for time series. *PVLDB*, 3(1):385–396, 2010.
- [19] X. Ling, W. Deng, C. Gu, H. Zhou, C. Li, and F. Sun. Model ensemble for click prediction in bing search ads. In *WWW*, pages 689–698, 2017.
- [20] Y. Matsubara, Y. Sakurai, and C. Faloutsos. Autoplait: Automatic mining of co-evolving time sequences. In *SIGMOD*, pages 193–204, 2014.
- [21] Y. Matsubara, Y. Sakurai, and C. Faloutsos. Non-linear mining of competing local activities. In *WWW*, 2016.
- [22] Y. Matsubara, Y. Sakurai, B. A. Prakash, L. Li, and C. Faloutsos. Rise and fall patterns of information diffusion: model and implications. In *KDD*, pages 6–14, 2012.
- [23] S. Papadimitriou, J. Sun, and C. Faloutsos. Streaming pattern discovery in multiple time-series. In *Proceedings of VLDB*, pages 697–708, Trondheim, Norway, August–September 2005.
- [24] Y. Qin, D. Song, H. Cheng, W. Cheng, G. Jiang, and G. Cottrell. A dual-stage attention-based recurrent neural network for time series prediction. *arXiv preprint arXiv:1704.02971*, 2017.
- [25] M. T. Ribeiro, S. Singh, and C. Guestrin. Why should i trust you?: Explaining the predictions of any classifier. In *KDD*, pages 1135–1144, 2016.
- [26] Y. Sakurai, Y. Matsubara, and C. Faloutsos. Mining and forecasting of big time-series data. In *SIGMOD, Tutorial*, pages 919–922, 2015.
- [27] Y. Sakurai, Y. Matsubara, and C. Faloutsos. Mining big time-series data on the web. In *WWW, Tutorial*, pages 1029–1032, 2016.
- [28] M. Schuster and K. K. Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681, 1997.
- [29] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of machine learning research*, 15(1):1929–1958, 2014.
- [30] A. Stisen, H. Blunck, S. Bhattacharya, T. S. Prentow, M. B. Kjærsgaard, A. Dey, T. Sonne, and M. M. Jensen. Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition. In *ACM Conference on Embedded Networked Sensor Systems*, pages 127–140. ACM, 2015.
- [31] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [32] Y. Tao, C. Faloutsos, D. Papadias, and B. Liu. Prediction and indexing of moving objects with unknown motion patterns. In *SIGMOD*, pages 611–622, 2004.
- [33] B. Tong, M. Klinkigt, M. Iwayama, T. Yanase, Y. Kobayashi, A. Sahu, and R. Vennelakanti. Learning to generate rock descriptions from multivariate well logs with hierarchical attention. In *KDD*, pages 2031–2040. ACM, 2017.
- [34] P. Wang, H. Wang, and W. Wang. Finding semantics in time series. In *SIGMOD Conference*, pages 385–396, 2011.
- [35] X. Wang, L. Yu, K. Ren, G. Tao, W. Zhang, Y. Yu, and J. Wang. Dynamic attention deep model for article recommendation by learning human editors’ demonstration. In *KDD*, pages 2051–2059, 2017.
- [36] S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. Abdelzaher. DeepSense: A unified deep learning framework for time-series mobile sensing data processing. In *WWW*, pages 351–360, 2017.
- [37] S. Zhai, K.-h. Chang, R. Zhang, and Z. M. Zhang. Deepintent: Learning attentions for online advertising with recurrent neural networks. In *KDD*, pages 1295–1304, 2016.