# 時系列テンソルデータのための将来予測

小幡　紘平†　　川畑　光希†　　松原　靖子†　　櫻井　保志†

† 大阪大学 産業科学研究所 産業科学 AI センター
E-mail: †obata88@sanken.osaka-u.ac.jp

**あらまし**　本論文では，大規模時系列テンソルデータのためのパターン検出手法である DDNF について述べる．DDNF は，(sensor, location, time）の三つ組で構成される時系列テンソルデータに対し，sensor 間の直接的な依存関係を表す下層ネットワークと，location 間の直接的な依存関係を表す上層ネットワークから構成される階層ネットワークによってモデル化された重要なパターンを発見し，将来予測を実現する．具体的に，提案手法は，(a）時系列テンソルデータから下層／上層ネットワークに基づいた解釈性の高いクラスタを発見する．(b）計算量は入力データのサイズに対して線形である．そして，(c）ネットワークを用いた高精度な将来予測を実現する．人工データを用いたクラスタリング精度評価実験では最新の既存手法と比較して DDNF が大幅な精度向上を達成していることを明らかにした．また，実データを用いた実験では提案手法が解釈性の高いクラスタを発見し，予測精度の向上を達成していることを確認した．

**キーワード**　時系列テンソル，特徴抽出，将来予測，グラフィカルラッソ

## 1　ま え が き

車両走行センサ [1]，生体信号 [2], [3]，株価に代表される金融データ [4] など，様々なアプリケーションにおいて時系列データが生成される．多くの場合，これらのデータは，タイムスタンプが付与された複数のセンサ値が複数の地域（またはユーザー）から取得される．本論文では，二つの属性とタイムスタンプから構成されるデータを時系列テンソルと呼ぶ．データ解析にとって，最も重要な課題の一つとして，将来予測に有用な解釈可能なパターン（クラスタ）の発見がある．例えば，大気成分濃度を正確に予測するためには，センサ間の依存関係と，センサの種類を考慮した地域間の依存関係によって特徴付けられた時間変化するパターンを発見することが重要である．本研究では，時系列テンソルの依存関係を表現する要約情報を抽出することで，複数のセンサが互いにどのように相関するのか，また，それぞれのセンサの振る舞いを考慮した地域間の相関関係を明らかにするパターンを発見する．さらに，パターンを用いた将来予測を達成する．

本論文で扱う問題は以下のとおりである．

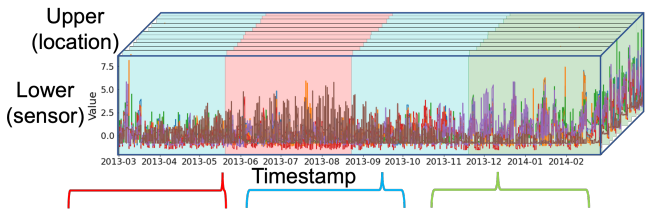
**問題：**$p$ 個のセンサ，$d$ の地域，$n$ 個のタイムスタンプから構成される時系列テンソル $\mathcal{X} \in \mathbb{R}^{p \times d \times n}$ が与えられたとき，

- セグメントと分割点を発見する
- ネットワークに基づいたセグメントをグループ化し，類似パターン（クラスタ）を発見する
- 時系列テンソルを予測する

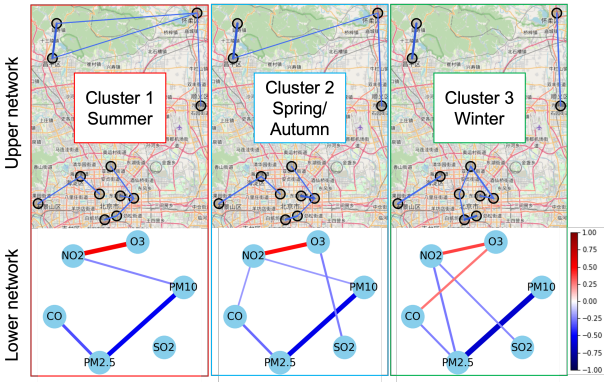
上記の問題を達成するには，隣接点が同じクラスタに属する傾向があるという，シーケンスの時間類似性を考慮する必要がある．また，データについて事前知識があることは稀であるため，クラスタ数を指定せずにクラスタを発見することも必要であり，得られたクラスタの解釈性が高いことも望ましい．DTW [5] などの従来のクラスタリング手法は，値の距離由来の指標に依存するものが多く，クラスタの解釈が困難である．クラスタリング手法の中には，変数間の関係性を考慮することで，解釈可能なクラスタを見つけることに重点を置いた手法もある [6], [7]．しかし，クラスタ数の自動決定，時系列テンソルの扱い，クラスタを利用した将来予測の達成には至っていない．時系列テンソルは複雑なデータ構造を持つため，PARAFAC [8] は変数間の関係を犠牲にし，テンソルを低ランクのコアテンソルと行列の集合に分解する．近年のグラフニューラルネットワーク（GNN）[9] の進歩により，時系列データからより良い予測結果が得られるグラフ構造を構築することができるが，これらの多くはデータから一つの静的グラフを発見するため，クラスタを発見することはできない．

そこで本研究では，時系列テンソルから解釈性の高いクラスタを発見し予測に用いる新しい手法として DDNF を提案する．提案手法は時系列テンソルを二つの疎なネットワーク，下層／上層ネットワーク，から構成される階層ネットワークによってモデル化する．ネットワークはグラフ構造をしており，変数がノードを，二つの変数間の直接的な依存関係がエッジを表す．ここで，値に対して影響力の大きい次元を下位変数（センサ），もう一方を上位変数（地域）と呼ぶ．下層ネットワークは下位変数間の関係を表し，上層ネットワークは下位変数の差を考慮した上位変数間の関係を表す．提案手法は時系列テンソルからネットワークが異なるパターン／クラスタを発見し，それらのグラフを予測に利用する．

**具体例.** 図 1 は大気データと DDNF の出力結果例である．このデータは $p = 6$ 個のセンサから取得された大気成分濃度が $d = 12$ の地域から取得された $n = 8760$ ポイントのデータである．図 1（a）は時系列テンソルのクラスタリング結果であり．同一のクラスタに含まれるセグメントは同一の色で表現されている．DDNF はテンソルを 4 つのセグメントと 3 つのク




(a) 3rd-order tensor




(b) Lower networks and upper networks

図 1　(a) 大気データにおける DDNF のクラスタリング結果．(b) 地図にプロットされた上層ネットワークと下層ネットワーク．各測定地点が上層ネットワークのノードを形成している．各測定地点で測定された物質が下層ネットワークのノードを形成している．変数間の関係値はエッジの太さと色で表されている．

ラスタで表現している．DDNF は事前知識を必要とせず，大気データを季節によって分類している．図 1（b）は各クラスタの下層／上層ネットワークを示しており，ノードが変数，エッジの太さと色で変数間の偏相関係数を表している．提案手法が発見した 3 つクラスタはそれぞれ異なるネットワークを持つことがわかる．ノードが地図上にプロットされた上層ネットワークでは，近い場所同士のノードにしかエッジが存在しないことがわかる。7 章では提案手法がネットワークをテンソルの将来予測に利用していることを示す．

### 1.1　本論文の貢献

DDNF は以下の特長を持つ．

(a)　DDNF は類似したネットワークをもつ時系列パターン（クラスタ）の個数と種類を把握し，データから解釈性の高いクラスタを発見する

(b)　DDNF はクラスタリング精度と予測精度において，最新の既存手法と比較し性能向上を達成している

(c)　提案したクラスタリングアルゴリズムは入力データの長さに対して線形である．つまり，長期間，高次元のデータに適応可能である

## 2　背　　景

### 2.1　関 連 研 究

関連研究は以下の 3 つに分類される．

**パターン発見.** 時系列データの解析に関する研究はさまざまな分野で進められている [10] [11] [12]．中でも，時系列サブシーケンスのクラスタリングはデータを理解するために有用である．時系列データの教師なしクラスタリングの代表的な技術である，DTW（Dynamic Time Warping）[5] と K-menas は値の距離由来の指標に基づいたクラスタリングを行い，データの構造よりも実値を比べることに焦点を置いている．Li ら [13] が提案した，DynaMMo は線形動的システム（LDS: Linear Dynamical System）に基づく手法で欠損を含む大規模時系列データ集合から時系列のパターンを発見できる．Wang ら [14] による pHMM（pattern-based hidden Markov model）は隠れマルコフモデル（HMM: Hidden Markov model）に基づく手法であり，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る．Mastubara ら [15] は多階層 HMM モデルを使ったパラメータフリーの手法として AutoPlait を提案している．これらの手法は，時系列の複雑な動的パターンを表現する能力はあるが，その一方で，ネットワーク構造を考慮していないため，クラスタの解釈には困難が伴う．

**時系列ネットワーク推定.** 時系列情報を加味したネットワーク推定は経済データ，生体信号データの解析手法として研究されている [16]．グラフィカルラッソ [17] は静的なネットワーク推定手法であり，損失関数に $\ell_1$ 正則化項を加味することで解釈が容易なスパースなネットワーク構造が推定できる [18]．Hallac ら [19] は文献 [17] に時系列情報を考慮したネットワーク推定手法である TVGL（Time Varying Graphical Lasso）を提案し，Harutyunyan ら [20] は文献 [21] を改良した共分散行列推定手法として T-CorEx を提案した．Tomasi ら [22] は文献 [23] を時系列データに適応し，潜在状態を考慮した動的なネットワーク構造を推定する手法である，LTGL（Latent variable Time-varying Graphical Lasso）を提案した．これらの手法は，ネットワーク構造の時系列変化をモデル化しており，前後のネットワーク構造を比較することで変化点の検知は可能だが，クラスタリングする能力はない．ネットワーク構造を基にしたクラスタリング手法として Hallac ら [6] が提案した，TICC（Toeplitz Inverse Covariance-based Clustering）と Tozzo ら [7] が提案した，TAGM（Time Adaptive Gaussian Model）がある．TICC はマルコフランダムフィールド（MRF: Markov Random Field）とテプリッツ行列を用い変数間に内在する関係を捉える手法であり，TAGM は HMM と混合ガウスモデル（GMM: Gaussian Mixture Model）を融合した手法である．これらの手法は各サブシーケンスのネットワーク構造に応じたクラスタを発見する．これにより，クラスタに解釈性を持たせ，従来のクラスタリング手法では発見できなかったパターンを発見することができる．しかしながら，これらの手法はクラスタ数を指定する必要があり，さらには時系列テンソルのクラスタリングに適していない．

**時系列予測.** 時系列データの将来予測は重要な課題であり，さまざまな研究がされている．近年，深層学習を用いた手法が非線形な特徴を捉える能力の高さから，よく研究されている．LSTM [24] は時系列データから非線形性と時間依存性を捉える手法である．TCN [25] と WaveNet [26] は畳み込み層により，時

系列データの時間的な相互作用を潜在状態にして捉える手法である．いくつかの深層学習由来の手法 [27] は変数間の依存性を捉えることができる．特に近年，GNN 由来の手法がグラフ畳み込み層により，大きな成功を収めている [28], [29]．しかしながら，これらの手法は変数間の関係を表すグラフを入力として必要とする．グラフを持たない時系列データに対応するため，予測に有用なグラフを生成する GNN が研究されている [9], [30], [31]．これらの手法はグラフがパターンに従って動的に変化するにも関わらず，一つの静的なグラフを使用する．結果として，動的なパターンと予測に有用なグラフを同時に発見することは依然として困難な課題である．さらに，これまでのところ，時系列テンソルに対するグラフに着目した研究はない．

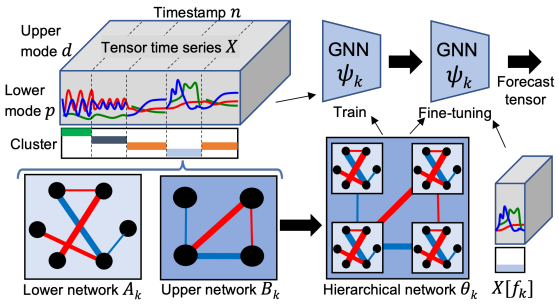



図 2　DDNF は時系列テンソルから分割点とクラスタを発見する．各クラスタはグラフィカルラッソによって特徴づけられらスパースな下層ネットワークと上層ネットワークから構成される．DDNF は下層ネットワークと上層ネットワークを階層ネットワークとして利用することにより，GNN を用いた時系列テンソルの予測を可能にする．

## 3　問 題 定 義

ここでは本論文で取り組む問題について詳細に定義する．$\mathcal{X} \in \mathbb{R}^{p \times d \times n}$ を $p$ と $d$（センサと地域）と長さ $n$ から構成される時系列テンソルとする．また，$\mathcal{X}$ のベクトル化として $X \in \mathbb{R}^{pd \times n}$，$X = \{X_1, X_2, \ldots, X_n\}$ を用いる．まず，$X$ の静的なネットワーク（$pd$ 変数間の関係）を推定する典型的な手法について述べる，そして，$\mathcal{X}$ のより複雑なネットワークを表現する．

### 3.1　グラフィカルラッソ

グラフィカルラッソは高い解釈性を持つグラフ表現を推定できる．多変量データ $X$ が与えられたとき，グラフィカルラッソは $\ell_1$ 正則化項により，精度行列とも呼ばれるスパースな逆共分散行列 $\theta \in \mathbb{R}^{pd \times pd}$, を推定する．グラフ表現から $pd$ 変数間の条件付き独立を理解できる．($\theta_{i,j} = 0$ のとき，変数 $i$ と $j$ は他の全ての変数値を与えられたとき，条件付き独立である.) 具体的には，以下の式を最適化する：

$$\text{minimize}_{\theta \in S^p_{++}} \lambda ||\theta||_{od,1} - ll(X, \theta), \tag{1}$$

$$ll(X, \theta) = \sum_{i=1}^{n} \Big\{ -\frac{1}{2}(X_i - \mu)^T \theta (X_i - \mu) + \frac{1}{2} \log \det\theta - \frac{pd}{2} \log(2\pi) \Big\}, \tag{2}$$

ただし，$\theta$ は正定値対称行列（$S^p_{++}$）である．$||\cdot||_{od,1}$ は対角成分を除いた $\ell_1$ ノルムである．$ll(X, \theta)$ は対数尤度関数であり $\mu$ は $X$ の平均である．正則化ハイパーパラメータ $\lambda \geqq 0$ により損失関数と $\ell_1$ 正則化項のバランスを調整することで，スパース性を制御する．式 1 は凸最適化問題であり，本論文では交互方向乗数法（ADMM）[32] を用いることで高速に解く．ADMM は凸最適化問題が大域解に収束することを保証する．

### 3.2　DDNF 問題

実データの $\mathcal{X}$ はそれぞれが異なる関係／ネットワークを持つ複数のパターンを含むため，一つの静的ネットワークでは表現しきれない．そこで本研究では，動的ネットワークに基づくテンソル予測問題に取り組む．$\mathcal{X}$ が $m$ 個のセグメントに分割され，それぞれが $K$ 個の動的ネットワーク（クラスタ）のうち一つに属するとする．$cp$ を $m$ 個のセグメントの開始点の集合とする，$cp = \{cp_1, cp_2, ..., cp_m\}$．$i$ 番目のセグメントの $X$ は $X_{cp_i:cp_{i+1}}$ （$cp_{m+1} = n + 1$）と表す．$n$ ポイントを $K$ に割り当てる，クラスタ割り当て集合を $\mathcal{F} = \{f_1, f_2, \ldots, f_K\}$ とする．$f_k \subset \{1, 2, \ldots, n\}$ であり，クラスタ $k$ に属する全てのデータは $X[f_k] \subset \mathcal{X}$ となる．また，各クラスタのパターンを表現するモデルパラメータ集合を $\Theta = \{\theta_1, \theta_2, \ldots, \theta_K\}$ とし，$\theta_k \in \mathbb{R}^{pd \times pd}$ は $X[f_k]$ の変数間の関係を要約表現する疎なガウス逆共分散行列である．よって，クラスタ情報集合は $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_K\}$ とする，$\mathcal{M}_k = \{f_k, \theta_k\}$．最適な $\mathcal{M}$ は $\mathcal{X}$ における新たな特徴を表すため，$\mathcal{M}$ を明示的にテンソル予測に使用する．クラスタ観点の回帰モデルを $\Psi = \{\Psi_1, \Psi_2, \ldots, \Psi_K\}$ とする．よって，本研究の課題は以下のように定義される．

［問題 1］　時系列テンソル $\mathcal{X}$ が与えられたとき，

- クラスタ情報集合 $\mathcal{M} = \{\mathcal{M}_k\}_{k=1}^K$
- 回帰パタメータ集合 $\Psi = \{\Psi_k\}_{k=1}^K$
- 最適なセグメント数 $m$，クラスタ数 $K$

を発見する．

## 4　提案モデル

本章では解釈性の高いクラスタを発見し，予測に利用する DDNF を提案する．図 2 は提案手法のモデル図である．まず，提案手法のモデルである階層ネットワーク $\theta$ について詳細に述べる．続いて，$\mathcal{M}$ を使用したテンソル予測について述べる．そして，クラスタ数を決定する基準を定義する．

### 4.1　二階層グラフィカルラッソ

$\mathcal{M}$ を発見する問題は以下のように書ける．

$$\arg\min_{\Theta, \mathcal{F}} \sum_{k=1}^{K} \lambda ||\theta_k||_{od,1} - ll(X[f_k], \theta_k). \tag{3}$$

$K$ と $\mathcal{F}$ が与えられたとき，$\theta_k$ をどのように定義すれば良いのだろうか．一般的なグラフィカルラッソで定義される $\theta_k$ は，べ

クトル化したテンソルの全ての変数（$pd$）に対して関係を推定する．しかしながら，この表現は二つの非時間次元の関係性をそれぞれ表現するには無駄が多い．過剰に複雑なネットワークの形成を避けるため，本研究では，変数間の関係性を二つの側面から捉えることで階層ネットワーク $\theta_k$ を表現する．

階層ネットワーク $\theta$ は，下位変数の依存関係を表す下層ネットワーク $A \in \mathbb{R}^{p\times p}$ （$a_{i,j}$ は変数 $i$ と $j$ の関係値を表す）と，上位変数の依存関係を表す上層ネットワーク $B \in \mathbb{R}^{d\times d}$ から構成される．下位階層では $p$ 個の変数（センサ）間に共通の関係性を捉えることを目的とするため，$A$ はデータ全体を表現したものとなり共有される．一方，上位階層では $p$ 個の変数の違いを加味した上で，$d$ 個の変数（地域）間の関係性を捉える．そして，$\theta$ の表現を簡潔にするため，ある下位変数は下位変数と上位変数が共に異なる変数と関係しないという制約を加える．よって，$\theta$ は $pd \times pd$ のテプリッツ行列となり，以下のように表される．

$$\theta = \begin{pmatrix} A & C_{1,2} & C_{1,3} & \cdots & & C_{1,p} \\ C_{2,1} & A & C_{2,3} & \ddots & & \vdots \\ C_{3,1} & C_{3,2} & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & C_{p-2,p-1} & C_{p-2,p} \\ \vdots & & \ddots & C_{p-1,p-2} & A & C_{p-1,p} \\ C_{p,1} & \ddots & \ddots & C_{p,p-2} & C_{p,p-1} & A \end{pmatrix},$$

$C_{i,j} = b_{i,j} \in \mathbb{R}^{p\times p}$ は対角成分が $b_{i,j} \in B$ (i.e.,$C_{i,j} = b_{i,j}\cdot\delta_{i.j}$) の対角行列であり，$\delta_{i.j}$ はクロネッカーのデルタである．

時系列テンソルから $\theta$ の要素である $A$ と $B$ を推定するために．グラフィカルラッソを拡張する．各クラスタにおいて上位変数は同じ $A$ を持つ．それゆえ，$A$ は以下の式を最適化することで得られる：

$$\text{minimize}_{A\in S^p_{++}} \lambda||A||_{od,1} - ll_A(\mathcal{X}, A), \tag{4}$$

$$ll_A(\mathcal{X}, A) = \sum_{j=1}^{d}\sum_{i=1}^{n}\Big\{ -\frac{1}{2}(x_{:,j,i} - \mu)^T A(x_{:,j,i} - \mu) + \frac{1}{2}\log\det A - \frac{p}{2}\log(2\pi)\Big\}, \tag{5}$$

$x_{:j:}$ は $\mathcal{X}$ の上位変数 $j$ におけるデータである．$\mu$ は下位変数の平均である．$B$ における制約のため，対数尤度 Eq. (2) を拡張し，各下位変数が同じ関係値を持つが，異なる平均を持つようにする．具体的には，以下の式を最適化する：

$$\text{minimize}_{B\in S^p_{++}} \lambda||B||_{od,1} - ll_B(\mathcal{X}, B), \tag{6}$$

$$ll_B(\mathcal{X}, B) = \sum_{j=1}^{p}\sum_{i=1}^{n}\Big\{ -\frac{1}{2}(x_{j,:,i} - \mu_j)^T B(x_{j,:,i} - \mu_j) + \frac{1}{2}\log\det B - \frac{d}{2}\log(2\pi)\Big\}, \tag{7}$$

$x_{j::}$ は $\mathcal{X}$ の下位変数 $j$ におけるデータであり，$\mu_j \in \mathbb{R}^1$ は $x_{:j:}$ の平均である．Eq. (4)，Eq. (6) は凸最適化問題であり，ADMM によって解かれる．

### 4.2 予測フレームワーク

GNN とファインチューニングによりクラスタ情報集合 $\mathcal{M}$ を利用する，テンソル予測のフレームワークについて述べる．提案手法は $K$ 個の回帰モデル $\Psi = \{\Psi_k\}_{k=1}^K$ を構築し，データの性質に基づき最適なモデルを選択することで予測する．例えば，データに季節性が存在する場合は，テストデータと同じ季節の $\Psi_k$ を使用する（後述の大気データ）．データの特徴が明らかでない場合は，最新のクラスターの $\Psi_k$ を使用する（株価データ）．$\mathcal{M}_k$ が与えられたとき，$\Psi_k$ を構築する．グラフ構造 $\theta_k$ を利用するため，$\Psi_k$ として GNN を使用する．$\theta_k$ は訓練とファインチューニングにおいて共通して入力する．訓練において $X$ を入力データとして使用する．ファインチューニングでは $X[f_k]$ を入力データとして使用することで，異なるネットワークを持つデータを取り除く．

### 4.3 特徴抽出とデータ圧縮

ネットワークに基づいたクラスタリングがされた $\mathcal{M}$ を得ることは，精良な $\Psi$ の構築につながる．最適な分割点とクラスタ数 $K$ の発見のために，最小記述長（MDL: minimum description length）の概念を用いる．MDL は情報理論に基づくモデル選択基準の一つであり，直感的には，データをより圧縮できれば良いモデルとみなすことができる．本論文の目的を解決するために，新しい符号体形をグラフィカルラッソモデルに対して定義する．

ここでは，時系列テンソルを表現するための符号化スキームを導入する．簡潔に表すと，MDL を用いてデータを表現するために必要な $\theta$ の最小数を求めることを目標とする．時系列テンソル $\mathcal{X}$ が与えられたときのモデル $\mathcal{M}$ のよさは次の式で表現できる：

$$\langle \mathcal{X}; \mathcal{M}\rangle \;=\; \alpha\cdot\langle\mathcal{M}\rangle + \langle\mathcal{X}|\mathcal{M}\rangle, \tag{8}$$

ここで，$\langle\mathcal{M}\rangle$ は $\mathcal{M}$ を表現するためのコストを示し，$\langle\mathcal{X}|\mathcal{M}\rangle$ は $\mathcal{M}$ が与えられたときの $\mathcal{X}$ の符号化のコストを示す．ハイパーパラメータ $\alpha > 0$ はモデル表現コストと符号化コストのバランスを調整し，モデルの複雑さを制御する．

#### 4.3.1 モデル表現コスト

モデル $M$ の表現コストは以下の要素の総和から構成される．

- クラスタの総数 $K$：$\log^*(K)$ [1]
- 各クラスタの観測値数：$\sum_{k=1}^K \log^*(|f_k|)$
- 各クラスタの平均値 $p\times 1$ と $d\times 1$：$\sum_{k=1}^K((p+d)\times c_F)$
- 各クラスタの下層ネットワーク $p\times p$ と上層ネットワーク $d\times d$：$\sum_{k=1}^K |A_k|_{\neq 0}(2\log(p) + c_F) + \log^*(|A_k|_{\neq 0}) + |B_k|_{\neq 0}(2\log(d) + c_F) + \log^*(|B_k|_{\neq 0})$

ここで，$|\cdot|_{\neq 0}$ は行列の非 0 要素の数を，$c_F$ は浮動小数点のコストを示す．[2]

---

1：ここで，$\log^*$ は整数のユニバーサル符号長を表す．

2：本論文では $4\times 8$ ビットとする．

#### 4.3.2 データ記述長

先述のとおり，本論文では $\theta$ を用いて $\mathcal{X}$ のパターンを表現するが，ここで重要なのは，推定したモデルが $\mathcal{X}$ を正しく表現しているかを判断する指標の導入である．ハフマン符号 [33] を用いた情報圧縮では，$\mathcal{M}$ が与えられた際の $\mathcal{X}$ の符号化コストを負の対数尤度を用いて表現する．本論文では精度行列を用いる代わりに分散共分散行列を用い，モデルの差異を正確に捉える．

$$\langle \mathcal{X}|\mathcal{M}\rangle = \sum_{k=1}^{K} ll_A(X[f_k], A_k^{-1}) + ll_B(X[f_k], B_k^{-1}).$$

本論文の次の目標は上記のコスト関数 $\langle \mathcal{X}; \mathcal{M}\rangle$ を最小化するようなクラスタ情報集合 $\mathcal{M}$ を発見することである．

## 5 提案アルゴリズム

前章では各クラスタを表現する階層ネットワークについて記述した．ここで重要なことは，どのようにして予測につながる最適な分割点とクラスタ割り当てを発見するのかである．本章では DDNF を実現するスケーラブルなアルゴリズムを提案する．アルゴリズム 1 に全体の処理の流れを示す．時系列テンソル $\mathcal{X}$ を与えられたとき，以下の二つの部分アルゴリズムにより式（8）を最小化する．

- CutPointDetector ：セグメント数 $m$ とその分割点 $cp$ を発見する．
- ClusterDetector ：クラスタ数 $K$ とクラスタ情報集合 $\mathcal{M}$ を発見する．

そして，DDNF は最終的なクラスタ情報集合 $\mathcal{M}$ を利用しテンソル予測を行う（4.2 章）．

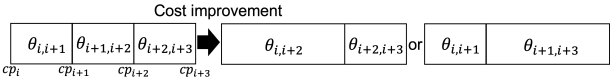
### 5.1 CutPointDetector

最初の目標はデータについての事前情報なしで $\mathcal{X}$ を $m$ 個のセグメント（パターン）に分割することである．組合せ爆発を防ぐため，分割統治法を基にした CutPointDetector により最適な分割点を探索する．

CutPointDetector は小サイズのセグメント集合に分割した $\mathcal{X}$ を，隣接セグメントが同じクラスタに属する傾向を利用し，隣接セグメントを MDL コストが減少する方向にマージする問題を再帰的に解くことで分割点を発見する．図 3 に例を示す．$\mathbf{w} = \{w_i\}_{i=1}^{m}$ はハイパーパラメータで，初期セグメントのサイズの集合（ひと月の日数ごと，小さい定数，など）とする．$i$ 番目のセグメントの MDL コストは $\langle \mathcal{X}; \theta_{i:i+1}, \{j\}_{j=cp_i}^{cp_{i+1}-1}\rangle$ となる．図 3（a）のように 3 つのセグメントが与えられたとき，中央のセグメントを片端のセグメントのどちらかとマージ（図 3（b）（c））するかを判断する．マージすることで MDL コストが 3 つのセグメントの合計から減少する場合，不要な分割点を省き，新しいセグメントに対して $\theta$ を推定する．この処理を全てのセグメントにおいて繰り返すことで，$m$ は減少し収束する．

### 5.2 ClusterDetector

DDNF はコスト関数 $\langle \mathcal{X}; \mathcal{M}\rangle$ が減少する間，$K = 1, 2, \ldots, m$ を増加させることで，最適なクラスタ数 $K$ を発見する．MDL コストには，$\mathcal{F}$ と $\Theta$ を推定する必要があり，片方の結果にはもう一方の結果が影響を与えるため，ClusterDetector では EM アルゴリズムにより最適化を行う．E ステップでは，データ記述長 $\langle \mathcal{X}|\mathcal{M}\rangle$ を最小化する $\mathcal{F}$ を計算する．具体的には，$i$ 番目のセグメントについて以下の式を解く．




(a) Original cp (b) Left merge (c) Right merge

図 3 CutPointDetector で比較される 3 つの分割点の候補の概略図．これらの分割点の候補の MDL コスト式（8）を比較する．

$$\arg\min_{k\in\{1,\ldots,K\}} \langle \mathcal{X}|\theta_k, \{j\}_{j=cp_i}^{cp_{i+1}-1}\rangle, \tag{9}$$

$i$ 番目のセグメントのインデックス $\{j\}_{j=cp_i}^{cp_{i+1}-1}$ を最適な割り当て先である $f_k \in \mathcal{F}$ に挿入する．M ステップでは，各クラスタについて割り当てられたデータ $X[f_k]$ のモデルパラメータ $\theta$ の要素である $A$ と $B$ を式（4）と式（6）により求める．ここで，$\Theta$ の初期化は適当に行う．結果として，$\Psi$ を推定するのに適切な $\mathcal{M}$ を得ることができる．

### 5.3 理論的な分析

［補助定理 1］ 提案手法の計算コストは $O((p+d)n)$ である．

［証明 1］ DDNF の計算コストの大部分は CutPointDetector のイテレーション回数と，全ての $A$ と $B$ を推定する計算コストによる．最終的に分割点の数が 0 個になる場合を考える．全ての $A$ と $B$ を推測するための合計計算コストは，それぞれの計算コストの和であるため，$A$ の場合について述べる．$n \gg p$ のとき，全ての $A$ を推定する計算コストは $O(pn)$ である．CutPointDetector のイテレーション回数は，セグメント数が一つずつ減少する場合は $|\mathbf{w}|$ 回であるが，これは起こりそうにない．イテレーションごとにセグメント数が半減していく場合は $\log_2 |\mathbf{w}|$ 回である．$n \gg \log_2 |\mathbf{w}|$ であるため，$B$ の推定を除いた計算コストは $O(pn)$ となる．$B$ の推定は $A$ と同等のため，DDNF の計算コストは $O((p+d)n)$ となる．

## 6 評 価 実 験

本章では，人工データに対する DDNF のクラスタリング精度と計算コストの検証を行う．クラスタリング精度比較に用いられる典型的な実データはネットワーク構造に基づいた正解ラベルが与えられていない．一方で，人工データでは明確なネットワーク構造のあるデータが生成可能で，ネットワーク構造に基づいたクラスタリング精度の比較が可能である．

**人工データ生成.** 人工データの生成，実験設計は文献 [6], [34] に従った．多変量正規分布 $X \sim N(0, \theta^{-1})$ に従う時系列テンソル $\mathbb{R}^{p\times d\times n}$ をランダムグラフに基づき生成した．ネットワーク構造に基づいたクラスタリング精度を評価するため，$K$ 個の各

**Algorithm 1** DDNF($\mathcal{X}, \mathbf{w}$)

**Input:** 3rd-order tensor $\mathcal{X}$ and initial segment sizes set $\mathbf{w}$
**Output:** Cluster parameters $\Theta$ and cluster assignments $\mathcal{F}$
Initialize $cp$ with $\mathbf{w}$;
**repeat**
  $cp = \textsc{CutPointDetector}(\mathcal{X}, cp)$;
**until** $cp$ is stable;
$K = 1$; Initialize $\Theta = \{\theta_1\}$; $\mathcal{F} = \{\{1, \ldots, n\}\}$;
Compute $\langle \mathcal{X}; \Theta, \mathcal{F} \rangle$;
**repeat**
  $K = K + 1$; Initialize $\Theta$ for $K$ clusters;
  $\{\Theta, \mathcal{F}\} = \textsc{ClusterDetector}(\mathcal{X}, cp, K)$;
  Compute $\langle \mathcal{X}; \Theta, \mathcal{F} \rangle$;
**until** $\langle \mathcal{X}; \Theta, \mathcal{F} \rangle$ converges;
Estimate $\Psi_k$ with $\mathcal{X}$ for each $k = 1, \ldots, K$;
Fine-tune $\Psi_k$ with $X[f_k]$ for each $k = 1, \ldots, K$;
**return** $\{\mathcal{M}, \Psi\}$

クラスタの平均値は $\vec{0}$ とした．以下の手順で，各クラスタの $\theta$ を作成した [34].

（1） 下層ネットワーク $A \in \mathbb{R}^{p \times p}$ を Erdős-Rényi モデルに従って作成する．全ノードペアについて，確率 20% でエッジを形成する．

（2） $A$ の選ばれたエッジについて，$a_{ij} \sim Uniform([-0.6, -0.3] \cup [0.3, 0.6])$ を設定する．また，$A$ は対称行列 $a_{ij} = a_{ji}$ とする．

（3） 上層ネットワーク $B \in \mathbb{R}^{d \times d}$ を $A$ と同じ手順で作成する．

（4） テプリッツ行列 $D \in \mathbb{R}^{pd \times pd}$ を $A$ と $B$ によって作成する．

（5） $\theta$ を正定値行列とするために，$\theta = D + (0.1 + |c|)I$ とする．$c = \lambda_{\min}(D)$ は $D$ の最小固有値で，$I$ は $pd \times pd$ の単位行列である．

**評価指標.** 次のような異なるセグメントの組み合わせの 4 つのデータセットについて実験を行った [6] （“1,2,1”，“1,2,3,2,1”，“1,2,3,4,1,2,3,4”，“1,2,2,1,3,3,3,1”）．それぞれのデータセットにつき 10 回実験を行い，$macro\text{-}F_1$ スコアの平均と標準偏差を記録した．$macro\text{-}F_1$ スコアは，適合率（Precision）と再現率（Recall）の調和平均を各クラスタについて求め，平均したもので，1 に近い値は高いクラスタリング精度を意味する．[3]

**比較手法.** 与えられた人工データに対する提案手法のクラスタリング精度を検証するために，最新の時系列クラスタリング手法と比較する．TICC [6]，および TAGM [7] はネットワーク構造に基づいたクラスタリングを行う手法である．TICC にはスパース性を制限するハイパーパラメータを $\lambda = 0.1$ とし，隣接ポイントを同じクラスタに割り当てない際の罰則コストであるハイパーパラメータ $\beta = 0, \ldots 1000$ と変化させ，訓練データにおいて最良の結果を示した値を選択した．TAGM にはスパース性を制限するハイパーパラメータ $\lambda = 0.1$ とした．さらに，これらの手法はクラスタ数の指定が必要であるため，正しいクラスタ数を与え実験した．AutoPlait [15] は多階層 HMM ベースの自動クラスタリングアルゴリズムである．なお，これらの手法はテンソルを扱えないため，ベクトル化されたデータ $X$ が与えられた．DDNF には $\alpha$ と $\mathbf{w}$ のハイパーパラメータがある．実験では全ての初期セグメントサイズを一定 $w_i$ (s.t., $i = 1, \ldots, m$) とした．訓練データを用いたパラメータチューニングでは，$w_i = 4$ とし，$\alpha$ を $p$ と $d$ に応じて変化させた．$\alpha$ が決定されると，$\mathbf{w}$ をコスト関数 Eq. (8) により決定できる．テストデータを用いて $w_i = 4, 8, 16, 32$ と変化させ，最小コスト Eq. (8) を示した結果を採用した．

[3] : $macro\text{-}F_1 = \frac{1}{K} \sum_i^K \frac{1}{1/presicion_i + 1/recall_i}$

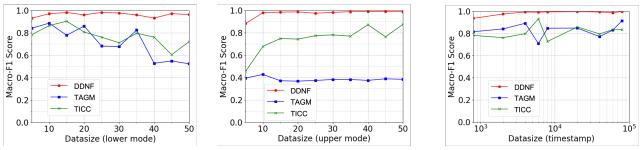
表 1　4 つの異なるデータセットにおける DDNF と比較手法の $macro\text{-}F_1$ スコアによるクラスタリング精度（高いほど高精度）

| Datasize | Pattern | DDNF | TICC | TAGM | AutoPlait |
|---|---|---|---|---|---|
| p=5, d=1 | 1,2,1 | **0.972** | 0.872 | 0.827 | 0.400 |
| | 1,2,3,2,1 | **0.960** | 0.907 | 0.793 | 0.190 |
| | 1,2,3,4,1,2,3,4 | **0.936** | 0.788 | 0.843 | 0.100 |
| | 1,2,2,1,3,3,1 | **0.986** | 0.907 | 0.814 | 0.182 |
| p=5, d=10 | 1,2,1 | **0.988** | 0.711 | 0.524 | 0.400 |
| | 1,2,3,2,1 | **0.992** | 0.696 | 0.432 | 0.190 |
| | 1,2,3,4,1,2,3,4 | **0.977** | 0.678 | 0.427 | 0.100 |
| | 1,2,2,1,3,3,1 | **0.996** | 0.655 | 0.457 | 0.182 |

**多種類の人工データにおけるクラスタリング精度.** クラスタリング精度を $macro\text{-}F_1$ スコアで比較した結果を表 1 に示す．それぞれのセグメントについて $p = 5, d = 1, 10, n = 100$ とした (例 “1,2,1” では $n = 300$). 本手法が全てのデータセットにおいて最も高い平均精度を記録した．AutoPlait はネットワーク構造を考慮しないため，クラスタを発見できなかった．TICC と TAGM は正しいクラスタ数を与えられたにも関わらず，$d = 1$ の場合においても，平均精度について DDNF と比較して 9% 以上低かった．これらの結果は提案手法が式（8）により，異なるネットワークを持つセグメントを発見し，正しくクラスタに割り当てることができることを示している．

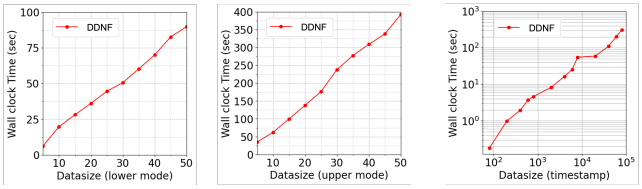
**上位/下位変数を変化させたときのクラスタリング精度.** 本手法の符号化コスト関数はデータの次元数による影響を受ける．“1,2,3,4,1,2,3,4” を例に取り $p$（$p = 5 \sim 50, d = 1, n = 800$）と $d$（$p = 5, d = 5 \sim 50, n = 800$）を変化させ，精度への影響を評価した. 図 4（a）（b）は次元数に対する $macro\text{-}F_1$ スコアをプロットした結果である. DDNF が全ての値において比較手法を上回っていることがわかる．TICC と TAGM は $p$ の増加により精度が低下したが，DDNF は $\alpha$ により高精度を維持している．TICC と TAGM はテンソルを扱えないため，図 4（b）における精度が相対的に低い．

**サンプル数を変化させたときのクラスタリング精度.** サンプル数の増加に対して精度を維持することは大規模時系列データを扱うにあたり重要である．”1,2,3,4,1,2,3,4”を例に取り $n$（$p = 5, d = 1, n = 800 \sim 80000$）を変化させ，精度への影響を評価した．図 4（c）は $n$ に対する $macro\text{-}F_1$ スコアをプロットした結果である．DDNF は精度において他手法を上回っており，$n$ が大きいほど高い精度を示している．これは $w_i$ が変化



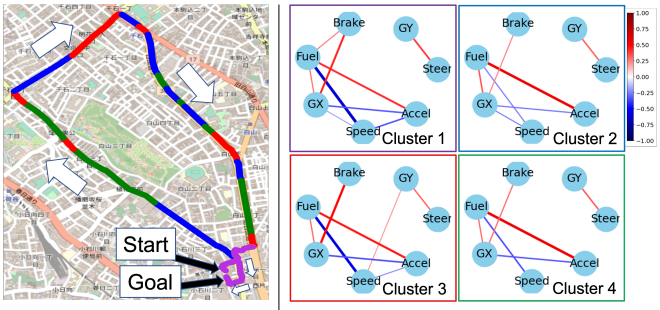

(a) vs. lower mode (b) vs. upper mode (c) vs. number of samples

図 4 DDNF の（a）下位変数 $p$, (b）上位変数 $d$, (c）サンプル数 $n$ に対するクラスタリング精度（$macro\text{-}F_1$ スコア）.




(a) vs. lower mode (b) vs. upper mode (c) vs. number of samples

図 5 DDNF の（a）下位変数 $p$, (b）上位変数 $d$, (c）サンプル数 $n$ に対する計算コスト.




(a) DDNF result on map (b) 下層ネットワーク

図 6 車両走行センサデータコース A についてのクラスタリング結果：（a） クラスタ割り当てを地図上にプロットした結果. . 同一色が同一クラスタを表す. (i.e., #紫 → 狭路, #青 → 高速度域, #赤 → 減速, #緑 → 中速度域) . (b）各クラスタの下層ネットワーク.

させることで各セグメントのネットワークを正確に推定できるからである.

**提案手法の計算コスト.** ここでは，提案手法の計算コストについて検証する. 補助定理 1 では DDNF の計算コストが $O((p+d)n)$ であることを示した. 図 5 に “1,2,3,4,1,2,3,4” の下位変数 $p$, 上位変数 $d$, サンプル数 $n$ を変化させたときの DDNF の計算コストを示す. 補助定理 1 において示した通り，DDNF は提案アルゴリズムにより，入力データサイズ $p, d, n$ に対して線形である.

## 7 ケーススタディ

本章では，実データを対象とした実験により，DDNF の予測精度と DDNF が意味のあるネットワークを発見可能であることを示す.

**車両走行センサデータ.**3 つの異なる状況のコースから取得された車両走行センサデータを用いる. コース A は市街地を走行したデータで（$n = 3241$m），コース B は高架下の車通りコースのデータで（$n = 2593$m），コース H は高速道路を走行したデータ（$n = 4000$m）である. 7 つのセンサーの値が 1m ごとに計測された：(ブレーキ，速度，前後加速度，左右加速度，アクセル開度，ハンドル角，燃費）それぞれ 40 周回あり，28 周回を訓練，4 周回を検証，8 周回をテストに使用した.

**大気データ.** 中国北京の $d = 12$ の地域から計測された $p = 6$ 個の大気成分濃度センサについて 2013 年 3 月 1 日から 2016 年 2 月 29 日まで 1 時間間隔で取得された（$n = 8760$hours）大気データを使用する. それぞれ 1 年分のデータを使って，訓練，検証，テストを行う.

**評価指標.** それぞれのデータセットにつき各手法で実験を 10 回行い，平均絶対誤差（MAE）と平均二乗誤差（MSE）の平均を記載した.

**比較手法.** 以下の比較手法を用いて実験を行った.

- Linear：全結合層.
- LSTM [24]：エンコーダに LSTM，出力層に全結合層を使用.
- seq2seq：LSTM をエンコーダとデコーダ共にに使用.
- TCN [25]：Causal dilated convolution と残差ブロックによる手法.
- WaveNet_STN [35]：WaveNet にを時間空間正規化を追加した手法.
- MTGNN [9]：多次元時系列データから予測に最適な隣接行列を生成する，GNN ベースの予測手法.

各手法について “+C” と手法名の後に記載したものは DDNF によるクラスタリング結果 $\mathcal{F}$ を加えた手法である. 提案手法は MTGNN を $\Psi$ として使用し，$\Theta$ を事前グラフとして与えた.

コスト関数におけるハイパーパタメータ $\alpha$ はどれだけの数のパターンを発見するかを決定する. 実データには正解ネットワークが存在しないため，検証データにおいて最良の結果を示した $\alpha$ を採用した. 車両走行センサデータでは同じ座標におけるデータは同じクラスタに属するとし，$w_i = 1$（コース A：$\alpha = 0.7$，コース B：$\alpha = 0.75$，コース H：$\alpha = 1.5$）とした. また，ドライバーは同時に走行しないため，上層ネットワークは使用しなかった. 大気データでは，同じ日時におけるデータは同じクラスタに属するとし，$w_i = 24$（1day），$\alpha = 1.7$ とした.

**予測精度.**DDNF による時系列テンソルの予測精度を検証した. 車両走行センサデータと大気データにおける結果は表 2 に示した. 最良の比較手法に対する DDNF の改善率が Improvements に記されている. DDNF が全てにおいて最も良い結果を示したことがわかる. これは DDNF がクラスタ割り当てと各クラスタの変数間の依存関係を示すグラフが与えられたからである. 次に，データを分割して学習する効果を “+C” との比較で検証した. 結果から，変数間の依存関係が異なるデータセットに分割して学習する有効性がわかる.

**解釈性.**DDNF におけるクラスタリング結果が理にかなっていることを示す. 大気データについての結果は 1 章（図 1）で述べた. 図 6 は車両走行センサデータコース A についてのクラスタリング結果である. 図 6（a）は地図上にプロットされたクラスタリング結果であり，図 6（b）は各クラスタの下層ネット

表 2　車両走行センサデータと大気データにおける精度比較．“+C” は提案手法によるクラスタリング結果 $\mathcal{F}$ が使用された手法を表す．

| Models | Course A | | | | Course B | | | | Course H | | | | Air quality | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 m | | 10 m | | 5 m | | 10 m | | 5 m | | 10 m | | 1 hour | | 6 hour | | 12 hour | |
| | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE |
| Linear | 0.339 | 0.419 | 0.472 | 0.649 | 0.344 | 0.384 | 0.479 | 0.629 | 0.147 | 0.078 | 0.242 | 0.182 | 0.194 | 0.097 | 0.448 | 0.445 | 0.536 | 0.617 |
| Linear+C | 0.324 | 0.395 | 0.457 | 0.617 | 0.337 | 0.377 | 0.471 | 0.616 | 0.146 | 0.077 | 0.240 | 0.179 | 0.177 | 0.087 | 0.436 | 0.427 | 0.529 | 0.607 |
| LSTM | 0.274 | 0.299 | 0.405 | 0.517 | 0.270 | 0.272 | 0.388 | 0.464 | 0.135 | 0.067 | 0.224 | 0.164 | 0.201 | 0.122 | 0.444 | 0.488 | 0.547 | 0.683 |
| LSTM+C | 0.269 | 0.288 | 0.397 | 0.494 | 0.268 | 0.271 | 0.386 | 0.461 | 0.133 | 0.066 | 0.221 | 0.162 | 0.200 | 0.122 | 0.445 | 0.491 | 0.541 | 0.661 |
| seq2seq | 0.261 | 0.280 | 0.394 | 0.504 | 0.260 | 0.261 | 0.373 | 0.460 | 0.133 | 0.066 | 0.219 | 0.163 | 0.152 | 0.080 | 0.447 | 0.492 | 0.560 | 0.714 |
| seq2seq+C | 0.256 | 0.268 | 0.382 | 0.482 | 0.255 | 0.258 | 0.367 | 0.453 | 0.130 | 0.065 | 0.215 | 0.161 | 0.151 | 0.079 | 0.441 | 0.483 | 0.554 | 0.699 |
| TCN | 0.287 | 0.334 | 0.418 | 0.554 | 0.284 | 0.297 | 0.408 | 0.496 | 0.137 | 0.069 | 0.229 | 0.165 | 0.240 | 0.153 | 0.414 | 0.415 | 0.514 | 0.594 |
| TCN+C | 0.279 | 0.313 | 0.407 | 0.521 | 0.282 | 0.293 | 0.403 | 0.488 | 0.136 | 0.068 | 0.225 | 0.162 | 0.241 | 0.153 | 0.411 | 0.411 | 0.511 | 0.596 |
| WaveNet_STN | 0.263 | 0.290 | 0.396 | 0.506 | 0.253 | 0.257 | 0.374 | 0.457 | 0.127 | 0.063 | 0.216 | 0.155 | 0.141 | 0.071 | 0.386 | 0.377 | 0.490 | 0.539 |
| WaveNet_STN+C | 0.257 | 0.273 | 0.385 | 0.478 | 0.248 | 0.252 | 0.369 | 0.450 | 0.125 | 0.062 | 0.209 | 0.152 | 0.141 | 0.071 | 0.386 | 0.376 | 0.485 | 0.536 |
| MTGNN | 0.250 | 0.274 | 0.381 | 0.487 | 0.246 | 0.252 | 0.361 | 0.440 | 0.124 | 0.063 | 0.206 | 0.154 | 0.143 | 0.074 | 0.382 | 0.379 | 0.482 | 0.531 |
| MTGNN+C | 0.246 | 0.268 | 0.373 | 0.465 | 0.239 | 0.244 | 0.356 | 0.431 | 0.121 | 0.062 | 0.203 | 0.151 | 0.141 | 0.073 | 0.382 | 0.373 | 0.480 | 0.534 |
| DDNF | **0.243** | **0.254** | **0.369** | **0.455** | **0.237** | **0.240** | **0.353** | **0.425** | **0.120** | **0.061** | **0.202** | **0.149** | **0.136** | **0.069** | **0.379** | **0.369** | **0.475** | **0.522** |
| Improvements | +3.0% | +7.4% | +3.2% | +6.7% | +3.4% | +4.8% | +2.2% | +3.4% | +3.2% | +3.7% | +1.7% | +3.0% | +3.2% | +3.2% | +1.0% | +2.1% | +1.5% | +1.8% |

ワークである．同一色は同一クラスタを表している．図 6（a）のクラスタ 1 はアクセル，ブレーキシステムに関するエッジが他のクラスタよりも多い．これはこのパターンが繊細な速度調整が必要な狭路で観測されることを示し，地図からもそれがわかる．一方で，図 6（a）のクラスタ 3 はブレーキと前後加速度に太いエッジが形成されている．地図を確認することで，このパターンが信号の前に多く観測されていることがわかる．

## 8 むすび

本論文では，下層ネットワークと上層ネットワークにより特徴付けられた解釈性の高いクラスタを検出し予測に利用する手法として，DDNF を提案した．DDNF は時系列テンソルを二つのネットワークによる階層構造に基いてモデル化することで解釈性の高いクラスタを発見できる．人工データと実データを用いた実験により，DDNF は最新の既存手法と比べてより高いクラスタリング精度と予測精度を持ち，解釈性の高いネットワーク構造を持つクラスタを発見することを示した．さらに，提案手法のアルゴリムは入力データに対して線形であり，長期間，高次元のデータに適応可能であることを示した．

**謝辞** 本研究の一部は JSPS 科研費，JP20H00585，JP21H03446，JP22K17896，国立研究開発法人情報通信研究機構委託研究 NICT 03501，総務省 SCOPE JP192107004，JSTAIP 加速課題 JPMJCR21U4，ERCA 環境研究総合推進費 JPMEERF20201R02，の助成を受けたものです．

## 文　献

[1] Miyajima, C., Nishiwaki, Y., Ozawa, K., Wakita, T., Itou, K., Takeda, K. and Itakura, F.: Driver modeling based on driving behavior and its evaluation in driver identification, *IEEE*, Vol. 95, No. 2, pp. 427–437 (2007).

[2] Hirano, S. and Tsumoto, S.: Cluster analysis of time-series medical data based on the trajectory representation and multiscale comparison techniques, *ICDM*, IEEE, pp. 896–901 (2006).

[3] Monti, R. P., Hellyer, P., Sharp, D., Leech, R., Anagnostopoulos, C. and Montana, G.: Estimating time-varying brain connectivity networks from functional MRI time series, *NeuroImage*, Vol. 103, pp. 427–443 (2014).

[4] Chiang, T. C., Jeon, B. N. and Li, H.: Dynamic correlation analysis of financial contagion: Evidence from Asian markets, *Journal of International Money and Finance*, Vol. 26, No. 7, pp. 1206–1228 (2007).

[5] Berndt, D. J. and Clifford, J.: Using Dynamic Time Warping to Find Patterns in Time Series, *Knowledge Discovery in Databases: Papers from the 1994 AAAI Workshop, Seattle, Washington, USA, July 1994. Technical Report WS-94-03*, pp. 359–370 (1994).

[6] Hallac, D., Vare, S., Boyd, S. P. and Leskovec, J.: Toeplitz Inverse Covariance-Based Clustering of Multivariate Time Series Data, *KDD*, pp. 215–223 (2017).

[7] Tozzo, V., Ciech, F., Garbarino, D. and Verri, A.: Statistical Models Coupling Allows for Complex Local Multivariate Time Series Analysis, *KDD*, pp. 1593–1603 (2021).

[8] Kolda, T. G. and Bader, B. W.: Tensor decompositions and applications, *SIAM review*, Vol. 51, No. 3, pp. 455–500 (2009).

[9] Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X. and Zhang, C.: Connecting the dots: Multivariate time series forecasting with graph neural networks, *KDD*, pp. 753–763 (2020).

[10] Kumar, S., Zhang, X. and Leskovec, J.: Predicting Dynamic Embedding Trajectory in Temporal Interaction Networks, *Proceedings of the 25th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM (2019).

[11] Wen, Q., Gao, J., Song, X., Sun, L., Xu, H. and Zhu, S.: RobustSTL: A robust seasonal-trend decomposition algorithm for long time series, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, pp. 5409–5416 (2019).

[12] Hallac, D., Bhooshan, S., Chen, M. H., Abida, K., Sosic, R. and Leskovec, J.: Drive2Vec: Multiscale State-Space Embedding of Vehicular Sensor Data, *21st International Conference on Intelligent Transportation Systems, ITSC 2018, Maui, HI, USA, November 4-7, 2018* (Zhang, W., Bayen, A. M., Medina, J. J. S. and Barth, M. J., eds.), IEEE, pp. 3233–3238 (2018).

[13] Li, L., McCann, J., Pollard, N. S. and Faloutsos, C.: DynaMMo: mining and summarization of coevolving sequences with missing values, *KDD*, pp. 507–516 (2009).

[14] Wang, P., Wang, H. and Wang, W.: Finding semantics in time series, *SIGMOD*, pp. 385–396 (2011).

[15] Matsubara, Y., Sakurai, Y. and Faloutsos, C.: AutoPlait: Automatic Mining of Co-evolving Time Sequences, *SIGMOD* (2014).

[16] Mohan, K., Chung, M., Han, S., Witten, D., Lee, S.-I. and Fazel, M.: Structured learning of Gaussian graphical models, *Advances in neural information processing systems*, Vol. 25 (2012).

[17] Friedman, J., Hastie, T. and Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso, *Biostatistics*, Vol. 9, No. 3, pp. 432–441 (2008).

[18] Tomasi, F., Tozzo, V., Verri, A. and Salzo, S.: Forward-Backward Splitting for Time-Varying Graphical Models, *Proceedings of the Ninth International Conference on Probabilistic Graphical Models*

(Kratochvíl, V. and Studený, M., eds.), Proceedings of Machine Learning Research, Vol. 72, PMLR, pp. 475–486 (2018).

[19] Hallac, D., Park, Y., Boyd, S. P. and Leskovec, J.: Network Inference via the Time-Varying Graphical Lasso, *KDD*, pp. 205–213 (2017).

[20] Harutyunyan, H., Moyer, D., Khachatrian, H., Steeg, G. V. and Galstyan, A.: Efficient Covariance Estimation from Temporal Data, *arXiv preprint arXiv:1905.13276* (2019).

[21] Steeg, G. V., Harutyunyan, H., Moyer, D. and Galstyan, A.: *Fast Structure Learning with Modular Regularization*, Curran Associates Inc. (2019).

[22] Tomasi, F., Tozzo, V., Salzo, S. and Verri, A.: Latent Variable Time-varying Network Inference, *KDD*, pp. 2338–2346 (2018).

[23] Chandrasekaran, V., Parrilo, P. A. and Willsky, A. S.: Latent variable graphical model selection via convex optimization, *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1610–1613 (2010).

[24] Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural computation*, Vol. 9, No. 8, pp. 1735–1780 (1997).

[25] Bai, S., Kolter, J. Z. and Koltun, V.: An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling, *arXiv:1803.01271* (2018).

[26] Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A. and Kavukcuoglu, K.: Wavenet: A generative model for raw audio, *arXiv preprint arXiv:1609.03499* (2016).

[27] Shih, S.-Y., Sun, F.-K. and Lee, H.-y.: Temporal pattern attention for multivariate time series forecasting, *Machine Learning*, pp. 1421–1441 (2019).

[28] Yu, B., Yin, H. and Zhu, Z.: Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting, *arXiv preprint arXiv:1709.04875* (2017).

[29] Zheng, C., Fan, X., Wang, C. and Qi, J.: Gman: A graph multi-attention network for traffic prediction, *Proceedings of the AAAI conference on artificial intelligence*, pp. 1234–1241 (2020).

[30] Wu, Z., Pan, S., Long, G., Jiang, J. and Zhang, C.: Graph Wavenet for Deep Spatial-Temporal Graph Modeling, *IJCAI*, p. 1907–1913 (2019).

[31] Shang, C., Chen, J. and Bi, J.: Discrete graph structure learning for forecasting multiple time series, *arXiv preprint arXiv:2101.06861* (2021).

[32] Boyd, S. P., Parikh, N., Chu, E., Peleato, B. and Eckstein, J.: Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, *Found. Trends Mach. Learn.*, Vol. 3, No. 1, pp. 1–122 (2011).

[33] Böhm, C., Faloutsos, C., Pan, J.-Y. and Plant, C.: Ric: Parameter-free noise-robust clustering, *ACM Transactions on Knowledge Discovery from Data (TKDD)*, Vol. 1, No. 3, pp. 10–es (2007).

[34] Mohan, K., London, P., Fazel, M., Witten, D. and Lee, S.-I.: Node-Based Learning of Multiple Gaussian Graphical Models, *J. Mach. Learn. Res.*, Vol. 15, No. 1, p. 445–488 (2014).

[35] Deng, J., Chen, X., Jiang, R., Song, X. and Tsang, I. W.: ST-Norm: Spatial and Temporal Normalization for Multi-variate Time Series Forecasting, *KDD*, pp. 269–278 (2021).