



非負値テンソル分解による動的な社会・経済行動パターンの抽出

小林, 照義

(Citation)

国民経済雑誌, 218(4):31-40

(Issue Date)

2018-10-10

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/E0041647>

(URL)

<https://hdl.handle.net/20.500.14094/E0041647>



非負値テンソル分解による動的な 社会・経済行動パターンの抽出

小 林 照 義

国民経済雑誌 第 218 卷 第 4 号 抜刷

平成 30 年 10 月

非負値テンソル分解による動的な 社会・経済行動パターンの抽出

小 林 照 義

近年のデータ収集・保存技術の進展に伴い、人々の社会・経済行動に関わる膨大なデータが日々蓄積されている。一方で、種類やサイズが非常に大きいデータを解析するための新たな手法も数多く提案されており、近年のデータサイエンスの研究基盤となっている。本稿では、データ解析手法として頻繁に用いられる非負値テンソル分解（non-negative tensor factorization）の概要を解説する。また、現実の社会ネットワークデータを用いて、非負値テンソル分解によるパターン検出の実用例を示す。

キーワード 非負値テンソル分解，テンポラル・ネットワーク，
コミュニティ検出

1 はじめに

経済学のデータ分析といえば計量経済学が挙げられるが、代表的な手法である最小二乗法にみられるように、説明変数と被説明変数の関係性をあらかじめ定式化し（多くの場合は線型性を仮定）、その上でパラメータ推定によって両者の関係性を定量化しようとするアプローチが多い。これに対し、近年のいわゆるデータサイエンスで用いられる手法の多くは、説明変数と被説明変数の関係性をあらかじめ特定化することなく、変数間に内在するであろうパターンをそのまま抽出しようというアプローチをとる。

本稿では、パターン抽出の手法として近年用いられることの多い非負値テンソル分解の概要を説明する。テンソル分解は行列分解の拡張版であるが、行列分解が2次元配列を対象とするのに対してテンソル分解には次元の制約がなく、より多様なデータの分析に利用できるのが特徴である。また、情報量を増やせる分、解析精度も高くなるメリットがある。

次節では、非負値テンソル分解の手法や実装についての簡単な紹介を行う。第3節では、現実の社会ネットワークデータを用いて、非負値テンソル分解の応用例を示す。具体的には、学校における生徒間の会話ネットワークのタイムスタンプ付きデータを用いて、生徒の動的な社会行動パターンを抽出し、得られた結果の妥当性を検討する。

2 非負値テンソル分解

この節では、非負値テンソル分解 (non-negative tensor factorization: NTF) についての基本的な説明を行う。以下では、 $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ で表現される 3 階のテンソル (3 次元配列) を考える。テンソルの要素 $x_{ijk} \in \mathcal{X}$ は、例えば第 i 番目の人の曜日 j の時刻 k における消費額、といった情報を表現している。このとき、各個人の ID、曜日および時刻といった情報はモードと呼ばれる。この 3 階 (もしくは 3 モード) テンソルの情報から、各消費者の行動を分析することができる。具体的には、曜日や時刻が各消費者の消費額にどのような影響を与えるか、といった消費者の行動パターンを抽出したいという場合である。しかし、現実のデータ分析をする上では、この 3 階テンソルの次元は $I \times J \times K$ となるため、考慮しなければならない組み合わせ総数が通常莫大になってしまい、消費者の ID や日時を直接的に紐づけて意味のある情報を得ることはそのままでは難しい。そのためデータの次元をなんらかの方法で削減し、かつ有意義な情報を抽出することが必要になってくる。テンソル分解は、この次元削減を効率的に行うことができる一つの手法である。

非負値の要素しか持たないテンソルの次元削減法としてよく用いられるのが、PARAFAC (parallel factor analysis) や CANDECOMP (canonical decomposition) と呼ばれる非負値テンソル分解である (Bro and Kiers, 2003; Kolda and Bader, 2009; Lim and Comon, 2009)。非負値テンソル分解では、次元が IJK であるテンソル $\mathcal{X}$ を、 R 個のランク 1 の非負値テンソルの和として近似する。つまり以下の関係式が成立する：

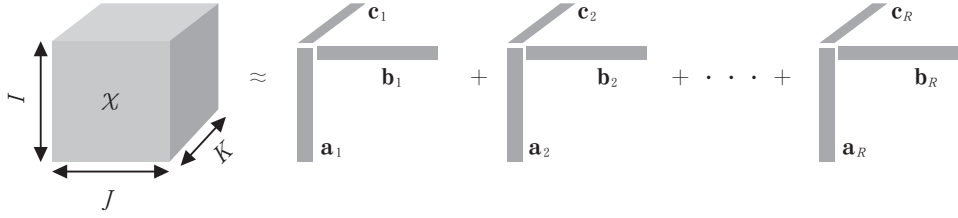
$$\mathcal{X} \approx \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \quad (1)$$

$$x_{ijk} \approx \sum_{r=1}^R a_{ir} b_{jr} c_{kr}, \quad (2)$$

ここでオペレータ $\circ$ は外積を示す。式 (1) において、 $\mathbf{a}_r \in \mathbb{R}_+^I$, $\mathbf{b}_r \in \mathbb{R}_+^J$ および $\mathbf{c}_r \in \mathbb{R}_+^K$ は r 番目のコンポーネントに属する要素ベクトルであり、それぞれのモードの基底を表す。上の消費者の例でいえば、 a_{ir} は第 i 番目の人の第 r コンポーネントへの帰属度となる。このとき、購入者のアクティビティに関する情報は $I \times R$ 行列 $\mathbf{A} \in \mathbb{R}_+^{I \times R}$ に集約されており、この行列の第 r 列が $\mathbf{a}_r$ に対応する。 $\mathbf{B} \in \mathbb{R}_+^{J \times R}$ および $\mathbf{C} \in \mathbb{R}_+^{K \times R}$ に関しても同様で、第 r 列がそれぞれ $\mathbf{b}_r$ および $\mathbf{c}_r$ に対応する。図 1 にテンソル分解の概念図を示す。

ここで重要なのが、行列 $\mathbf{A}$ における各列、すなわち各コンポーネントへの帰属性の差異が各個人の行動パターンの違いを表しているという点である。単純な例として、2 人しかいないケース ($I=2$) を考えよう。この二人がお互いに全く異なる独立な消費行動パターンをとるとすると、

図1 テンソル分解の概念図



$$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2] = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (3)$$

といった値をとる。このとき、第1番目の人の行動は全てパターン1、第2番目の人の行動はパターン2として特徴付けることができる。ランク数 R は存在する行動パターンの総数を示している。換言すれば、行列 $\mathbf{A}$ は消費行動パターンの類似性で定義されるコミュニティ構造を表しており、 R がコミュニティ数となる。このように元のテンソルを分解すると、分析者がみるべきデータはサイズがそれぞれ $I \times R$ 、 $J \times R$ および $K \times R$ となる行列 $\mathbf{A}$ 、 $\mathbf{B}$ および $\mathbf{C}$ であり、その総データサイズは $(I+J+K)R$ であるから、これは通常 IJK と比べて格段に小さい。

では、この要素行列 $\mathbf{A}$ 、 $\mathbf{B}$ 、 $\mathbf{C}$ はどのように求められるだろうか。(1)式の右辺は元々のテンソル $\mathcal{X}$ の良い近似となっている必要があるため、次のような誤差最小化問題を考えるのは自然である：

$$\min_{\mathbf{A} \geq 0, \mathbf{B} \geq 0, \mathbf{C} \geq 0} \|\mathcal{X} - \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket\|_F, \quad (4)$$

ここで $\llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket$ は Kruskal form と呼ばれるテンソル分解 (i.e., (1)式右辺) の表記法であり、 $\|\cdot\|_F$ はフロベニウスノルムを示す。ここで非負制約が付けられていることに注意されたい。

この問題 (4) を解く方法は様々あり得るが、通常は他の要素行列を所与として特定の要素行列について最小化を行い、それを他の要素行列についても繰り返す方法がとられる。ここで、 $\mathbf{X}_{(n)}$ をテンソル $\mathcal{X}$ のモード n に関して行列化 (flattening と呼ばれる) されたものとする。すなわち、 $\mathbf{X}_{(1)}$ 、 $\mathbf{X}_{(2)}$ および $\mathbf{X}_{(3)}$ は、それぞれサイズが $I \times JK$ 、 $J \times IK$ および $K \times IJ$ の行列であり、以下のように表される：

$$\mathbf{X}_{(1)} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^\top, \quad (5)$$

$$\mathbf{X}_{(2)} = \mathbf{B}(\mathbf{C} \odot \mathbf{A})^\top, \quad (6)$$

$$\mathbf{X}_{(3)} = \mathbf{C}(\mathbf{B} \odot \mathbf{A})^\top, \quad (7)$$

ここでオペレータ $\odot$ は Khatri-Rao 積を表し、 $\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1, \dots, \mathbf{a}_R \otimes \mathbf{b}_R]$ が成立する。 $\otimes$

はクロネッカー積を示す。このとき、PARAFAC 法による誤差最小化問題は次のように書ける：

$$\{\min_{\mathbf{A} \geq 0} \|\mathbf{X}_{(1)} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})^\top\|_F^2, \min_{\mathbf{B} \geq 0} \|\mathbf{X}_{(2)} - \mathbf{B}(\mathbf{C} \odot \mathbf{A})^\top\|_F^2, \min_{\mathbf{C} \geq 0} \|\mathbf{X}_{(3)} - \mathbf{C}(\mathbf{B} \odot \mathbf{A})^\top\|_F^2\}. \quad (8)$$

まず $\mathbf{B}, \mathbf{C}$ を所与として $\mathbf{A}$ の最適化を行い（1 項目）、得られた $\mathbf{A}$ と $\mathbf{C}$ を所与として $\mathbf{B}$ の最適化を行い（2 項目）、と順に続けていくことで、 $\mathbf{A}, \mathbf{B}$ および $\mathbf{C}$ の逐次最適化を行う。この方法は non-negative alternate least squares method (ANLS) と呼ばれ (Kim and Park, 2012)、実装コードは例えば Jingu Kim 氏のウェブサイト (<https://sites.google.com/site/jingukim/>) から MATLAB コードがダウンロード可能である¹⁾。

上記の PARAFAC 分解をする際には、ランクサイズ（＝コンポーネント数） R を事前に決定しておく必要がある。注意すべきは、 R が小さいと PARAFAC の近似精度が低くなるし、逆に R が大きければコンポーネントの数が過剰となり、過剰適合 (overfitting) の可能性が高まってしまう。ここでは、ランクサイズの決定法として標準的に用いられる Bro and Kiers (2003) によって提案された Core-Consistency (CC) による決定法をみていこう。

はじめに、(2) 式を次のように書き換える：

$$x_{ijk} = \sum_{n=1}^R \sum_{m=1}^R \sum_{p=1}^R \lambda_{nmp} a_{in} b_{jm} c_{kp}, \quad (9)$$

λ_{nmp} は $\{0, 1\}$ のバイナリ値をとるテンソル (superdiagonal tensor) $\mathcal{L}$ の (n, m, p) 要素であり、 $\lambda_{nmp} = \delta_{nm} \delta_{mp} \delta_{np}$ である (δ_{nm} はクロネッカーデルタで、 $n=m$ のときのみ $\delta_{nm}=1$ 、それ以外は 0 をとる)。ここで、PARAFAC よりも一般性の高い Tucker 3 と呼ばれる分解を考える (Bro and Kiers, 2003)。Tucker 3 分解は PARAFAC と異なり、基本的に各コンポーネントの数が一定ではなく各モード毎に決められ、異なるコンポーネント間についても相関を許す。したがってテンソル $\mathcal{X}$ は以下のように書ける：

$$x_{ijk} = \sum_n^{R_n} \sum_m^{R_m} \sum_p^{R_p} g_{nmp} a_{in} b_{jm} c_{kp}. \quad (10)$$

ここで g_{nmp} は Tucker 3 分解によって得られたコア・テンソル $\mathcal{Q}$ の要素であるが、 $\mathcal{Q}$ は superdiagonal ではないことに注意されたい。

Bro and Kiers (2003) のレンマによると、Tucker 3 モデルがデータに完全適合し、各要素行列が列フルランクのとき、 $\mathcal{Q}$ は $\mathcal{L}$ と等しくなる。しかし、 $\mathcal{Q}$ と $\mathcal{L}$ が大きく異なるとき、PARAFAC モデルでは適合が不十分であることになり、Tucker 3 モデルを採用する必要がある。したがって、この $\mathcal{Q}$ と $\mathcal{L}$ の差をみることで、PARAFAC モデルの適合性が Tucker 3 モデルと比べて遜色がないかどうかを確認することができる。Bro and Kiers (2003) は、Core-Consistency と呼ばれる次のような指標を提案した：

$$CC = 100 \times \left(1 - \frac{\sum_{n=1}^R \sum_{m=1}^R \sum_{p=1}^R (g_{nmp} - \lambda_{nmp})^2}{R} \right), \quad (11)$$

ここで Tucker 3 モデルについては $R_n = R_m = R_p = R$ を仮定している。CC は PARAFAC モデルがデータに完全適合していれば100となり、そうでなければより低い値をとる（負の値も取りうる）。一般には、 R の値が低いとデータへの適合度が低く、 R の値を増加させると適合度は高まるものの、コンポーネント間の相関が高まるため CC 値は低下する。そのため、PARAFAC モデルにおいては CC 値が一定値以上をとるような R を採用するのが一般的である。なお、テンソル分解は一般に分割が一意には決定されないため、以下では PARAFAC を20回行った上でその信頼区間も確認していく。CC 値の計算についての実装コードは Papalexakis and Faloutsos (2015) による計算法をベースにしている。²⁾

3 非負値テンソル分解の応用例：学校における会話ネットワーク

では実データを用いて NTF による解析を行ってみよう。ここでは、小学校や高校において形成される社会ネットワークのデータを利用する。今回用いるのは、フランスやイタリアの研究チームを中心とした「SocioPatterns プロジェクト」と呼ばれるデータ収集・研究プロジェクトによって公開されているデータである。³⁾ データの収集法は、生徒にウェアラブルセンサを身につけてもらい、お互いに一定以下の距離（1-1.5m）に接近したときにその時点の時刻（20秒刻み）と、両生徒の ID が記録される仕組みである。生徒の情報は匿名化されているが、各生徒に固有の ID が割り当てられているため、各個人の行動を追跡することが可能になっている。さらに、各生徒の所属する教室（クラス）の情報も示されている。なお、会話内容については記録されていない。このデータを解析した研究はこれまでに数多く発表されており（Cattuto et al., 2010; Stehlé et al., 2011; Gauvin et al., 2014）、社会ネットワーク研究における重要なデータとなっている。以下では、小学校のデータを“Primaryschool”、高校のデータを“Highschool”と表記する（表1）。

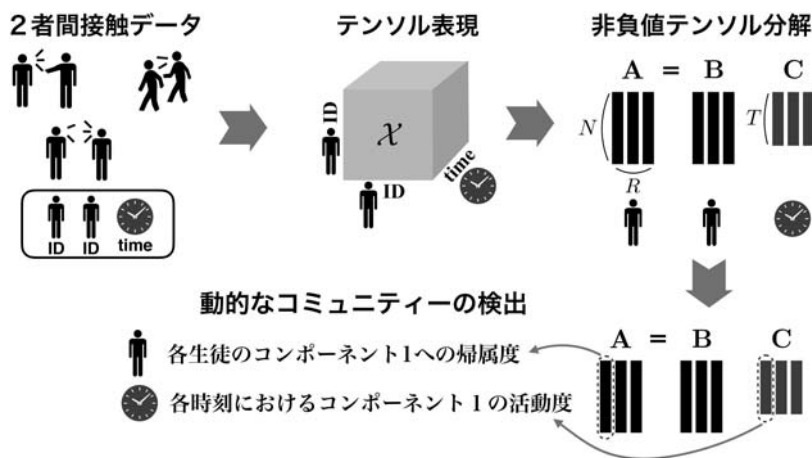
ここでは Gauvin et al. (2014) のモデルを参考に、各時刻において（教師を除く）生徒がどのような会話ネットワークを形成しているのかを分析しよう。いま、各モードがそれぞれ

表1 SocioPatterns データ

	Primaryschool	Highschool
データ期間	Oct 1-2, 2009	Dec 2-6, 2013
被験者数	232*	327
教室数	10	9
接触記録数	125,773	188,508

*教師除く。出所：<http://www.sociopatterns.org/datasets/>

図 2 学校における社会ネットワークのデータを用いたテンソル分解分析の模式図

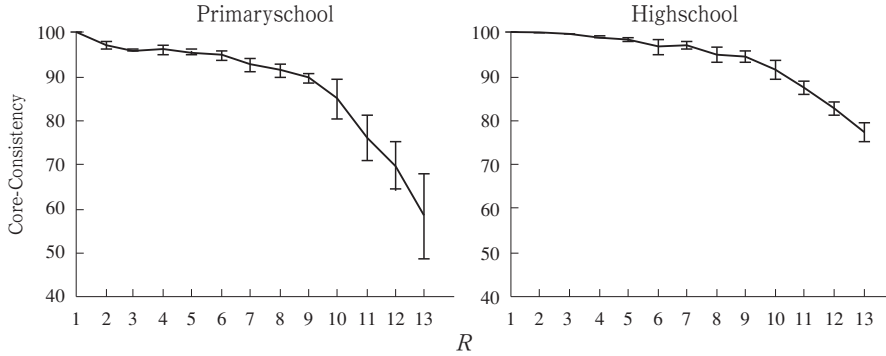


(生徒 ID, 生徒 ID, 時刻) で表される 3 階のテンソル $\mathcal{X} \in \mathbb{R}_+^{N \times N \times T}$ を考える。time window の幅は10分間とし、いずれかの日の時刻 $[t-10, t]$ (分単位) に生徒 i と生徒 j が会話をを行った場合は $x_{ijt} = x_{jit} = 1$, それ以外の場合は $x_{ijt} = x_{jit} = 0$ である。別の見方をすれば、この 3 階のテンソルは各時刻における $N \times N$ の隣接行列を T 個だけ並べたものといってもよい。なお、こうした時々刻々と変化していくネットワーク構造を持つデータは「テンポラル・ネットワーク」と呼ばれ、ネットワーク科学において主要な研究分野でもある (Holme and Saramäki, 2012; Masuda and Lambiotte, 2016)。

図 2 は今回行う分析の模式図である。SocioPatterns による 2 者間の接触データから 3 階のテンソルを作成し、それを非負値分解することで 3 つの要素行列 A, B, C が得られる。接触データは無向グラフであることから基本的には $A=B$ であり、 A の第 r 列は各生徒のコンポーネント r への帰属度を示す。例えば、もし第 i 番目の生徒がコンポーネント r に完全に属しているならば、 A の第 i 行は第 r 列のみに正の要素が入り、残りは 0 となる。また行列 C の第 r 列は各時刻 (10 分間の time window) におけるコンポーネント r の活動度を表している。

PARAFAC モデルにおいて指定すべきコンポーネントの数 R は事前にはわからないので、まずは様々な R の値について CC 値をプロットしてみる (図 3)。Primaryschool および Highschool 共に、 R と CC 値には負の相関があり、ある程度 R が大きくなると CC 値が低下する度合いが高まるとともに、そのばらつきも大きくなっていくことがわかる。ここで、Primaryschool の教室数は 10, Highschool は 9 だが、CC 値は $R > \text{教室数}$ となると急激に低下していく点に注意されたい。このことは、 $R > \text{教室数}$ となる領域で PARAFAC モデルにおいて過剰適合の可能性が高まっていることを示唆している。そこで以下では、 $R = \text{教室数}$

図3 Core-Consistency



R の各値について、PARAFAC を20回実行した結果の平均値を示している。

エラーバーは95%信頼区間を表す。

とした場合の結果を示す。

各生徒の行動パターンは PARAFAC で得られる $N \times R$ の要素行列 $\mathbf{A}$ に格納されているが、各生徒を行動パターンに応じて分類するために、 $\mathbf{A}$ の行を R 個（＝教室数）のグループ（行グループ）に分類することを考える。ここでは、分類器としてよく用いられる k-means 法を使用する。k-means 法は属性が多次元であるデータをクラスタリングする方法の一つであり、2次元でいえば平面上の点をそれらの集積度合いに応じてあらかじめ指定された数のクラスタに分割する。基本的には、各クラスタ内において重心からの距離を計算し、その距離の和が最小となるような分割を探すアルゴリズムである。まず NTF で各モードの活動パターンを抽出し、そこで得られた要素行列に対して k-means 法を適用することにより、各モードの活動パターンの分類を行うことが可能になる。実装は MATLAB 関数の kmeans を利用し、初期値を変えながら行った1000回のクラスタリングのうち最小距離を記録したものを採用した。

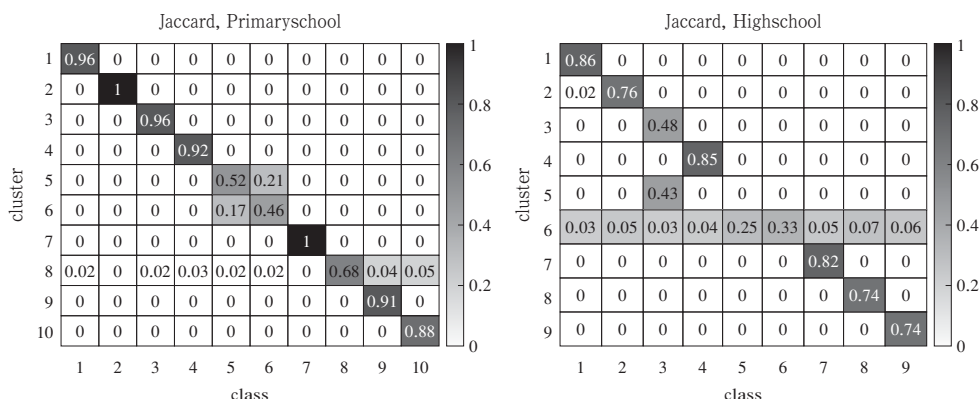
では、k-means 法で得られた要素行列 $\mathbf{A}$ の分割と、実際の学校における教室による分割は共通性があるのだろうか。これを確認するため、両者の重複度を示す指標として次のような Jaccard 指数を定義する：

$$J(u, v) = \frac{|S_u^{\text{clust}} \cap S_v^{\text{class}}|}{|S_u^{\text{clust}} \cup S_v^{\text{class}}|}, \quad (12)$$

ただし S_u^{clust} と S_v^{class} は、それぞれクラスタ u に属する生徒の集合と、実際の教室 v に属する生徒の集合である。もし PARAFAC 分解によって実際の生徒の行動パターンがうまく抽出できているのであれば、要素行列 $\mathbf{A}$ から得られるクラスタは実際のクラス分けを反映しているはずであり、 J の値は1に近くなると考えられる。

図4をみると、どちらのデータに関しても、k-means 法によるクラスタと実際の教室によ

図 4 Jaccard 指数のヒートマップ



対角要素の値が 1 に近いほど k-means で得られたクラスと実際の教室のメンバーが一致していることを示す。

る分割がかなりの程度一致していることがわかる。これは NTF によって抽出された動的な行動パターンが、同じ教室のメンバー同士においてかなり類似性が高いことを示唆している。換言すれば、クラスメート同士の行動が同期しているといえる。仮に、それぞれの教室内で独立した行動の周期を持っていれば、クラスメートの分割は完全に可能となる。しかし現実には休み時間やランチの時間、および共同授業であったり、もしくは先生のいうことを聞かずに他の教室に行ってしまう、などの理由で他の教室の生徒とも接触することがあり、それによって異なる教室間でも会話行動の周期が近くなることもある。また逆に、教室内にいてもほとんど他の生徒と話さずに過ごす生徒については、クラスメートであっても行動の同期が起こらない。Jaccard 指数が低くなっているクラスと教室の組み合わせが観察されるのは、そういった日常の不規則性が背景にあると考えられる。逆にいえば、NTF によって検出されたクラスは、実際の会話周期パターンによって定義された、会話ベースの「クラス」であり、必ずしも実際の教室による分割と一致している必要はない。しかしながら、現実には様々な会話周期の揺らぎがあるにも関わらず、クラスとクラスメートについてこれだけの一致がみられるのは、NTF のパターン抽出能力の高さによるものだろう。

ネットワーク科学においては、ネットワーク上のノード（今回は生徒）をいくつかのグループに分割するコミュニティ検出法が数多く提案されている（Lancichinetti and Fortunato, 2009; Fortunato, 2010; Lancichinetti et al., 2010）。ただし、ほとんどの手法は静的なネットワークに対して適用されるものであるのに対し、今回紹介した NTF は、いわばネットワークの動的変化を考慮したコミュニティ分割である。今回のデータセットでは実際の教室という「正しい」分割がわかっているため、NTF で検出されたパターンを真の値（ground truth）を基に評価することができ、NTF から得られた動的なコミュニティ構造が正確で

あることを示唆している。

4 お わ り に

本稿では、非負値テンソル分解の概要を解説し、社会ネットワークデータを用いた分析例を示した。いうまでもなく、今回用いたような会話ネットワークデータは一つの応用例に過ぎず、あらゆる社会・経済データに応用することが可能である。例えば、Twitter の投稿パターン (Panisson et al., 2014) やオンラインゲームの対戦パターン (Sapienza et al., 2018), 銀行間取引の動的パターン (Kobayashi et al., 2018) など、近年の応用例は数多い。また、MATLAB や Python をはじめとした実装コードも充実しており、今後経済データへの応用が広がっていくことが期待される。

注

松井暉氏 (南カリフォルニア大学博士課程) には原稿について有益なコメントを頂いた。もちろん残る誤りは全て筆者の責任である。本研究は科研費基盤 (S) 15H05729 および基盤 (C) 16K0351 の補助を受けている。

- 1) MATLAB に Tensor Toolbox をインストールすることが前提となるので注意されたい (Bader and Kolda, 2007; Bader et al., 2015)。Tensor Toolbox は <http://www.sandia.gov/~tgkolda/TensorToolbox/index-2.6.html> からダウンロード可能 (2018年 2 月22日現在)。
- 2) <http://www.cs.ucr.edu/~epapalex/code.html>
- 3) <http://www.sociopatterns.org/datasets/>

参 考 文 献

- Bader, Brett W. and Tamara G. Kolda**, “Efficient MATLAB computations with sparse and factored tensors,” *SIAM Journal on Scientific Computing*, December 2007, 30, 205–231.
- , ——— et al., “MATLAB Tensor Toolbox Version 2.6,” Available online February 2015.
- Bro, Rasmus and Henk A. L. Kiers**, “A new efficient method for determining the number of components in PARAFAC models,” *Journal of Chemometrics*, 2003, 17, 274–286.
- Cattuto, Ciro, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani**, “Dynamics of person-to-person interactions from distributed RFID sensor networks,” *PLOS ONE*, 2010, 5, e11596.
- Fortunato, Santo**, “Community detection in graphs,” *Physics Reports*, 2010, 486, 75–174.
- Gauvin, Laetitia, André Panisson, and Ciro Cattuto**, “Detecting the community structure and activity patterns of temporal networks: A non-negative tensor factorization approach,” *PLOS ONE*, 2014, 9, e860228.
- Holme, Petter and Jari Saramäki**, “Temporal networks,” *Physics Reports*, 2012, 519, 97–125.
- Kim, Jingu and Haesun Park**, “Fast nonnegative tensor factorization with an active-set-like method,” in Michrel W. Berry et al. (ed) *High-Performance Scientific Computing*, Springer, 2012, pp. 311–326.

- Kobayashi, Teruyoshi, Anna Sapienza, and Emilio Ferrara**, “Extracting the multi-timescale activity patterns of online financial markets,” *Scientific Reports*, 2018, 8, 11184.
- Kolda, Tamara G. and Brett W. Bader**, “Tensor decompositions and applications,” *SIAM Review*, 2009, 51, 455–500.
- Lancichinetti, Andrea and Santo Fortunato**, “Community detection algorithms: A comparative analysis,” *Physical Review E*, 2009, 80, 056117.
- , **Mikko Kivela, Jari Saramäki, and Santo Fortunato**, “Characterizing the community structure of complex networks,” *PLOS ONE*, 2010, 5, e11976.
- Lim, Lek-Heng and Pierre Comon**, “Nonnegative approximations of nonnegative tensors,” *Journal of Chemometrics*, 2009, 23, 432–441.
- Masuda, Naoki and Renaud Lambiotte**, *A Guide to Temporal Networks*, World Scientific Publishing, 2016.
- Panisson, André, Laetitia Gauvin, Marco Quaggiotto, and Ciro Cattuto**, “Mining concurrent topical activity in microblog streams,” in Proceedings of the 4th workshop on ‘Making Sense of Microposts’ World Wide Web Conference 2014, 2014.
- Papalexakis, Evangelos E. and Christos Faloutsos**, “Fast efficient and scalable core consistency diagnostic for the parafac decomposition for big sparse tensors,” in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference*, 2015, pp. 5441–5445.
- Sapienza, Anna, Alessandro Bessi, and Emilio Ferrara**, “Non-negative tensor factorization for human behavioral pattern mining in online games,” *Information*, 2018, 9, 66.
- Stehlé, Juliette, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, and Philippe Vanhems**, “High-resolution measurements of face-to-face contact patterns in a primary school,” *PLOS ONE*, 2011, 6, e23176.