テンソル分解を用いた関係データのモデリングとその応用

林 浩平 †

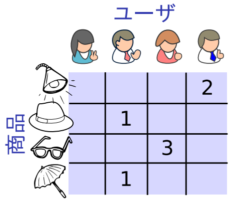
† 奈良先端科学技術大学院大学

2011 年 11 月 10 日
第 14 回 IBIS workshop 企画セッション
「関係データとテンソル分解」

テンソルによる関係データの表現

複数種類の物事に関する「関係」の集まり

ペアに関する関係　⇔　行列

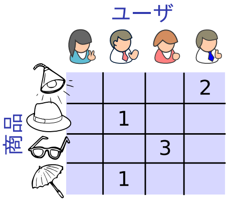


行列

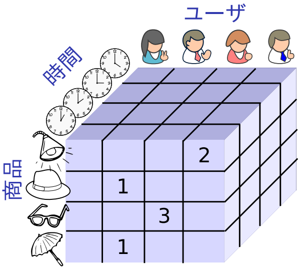
テンソルによる関係データの表現

複数種類の物事に関する「関係」の集まり

ペアに関する関係 ⇔ **行列**
3 **つ組に関する関係** ⇔ 3 **次元配列**



行列

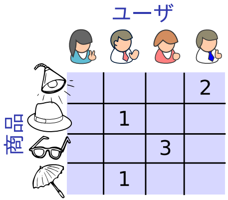


3次元配列

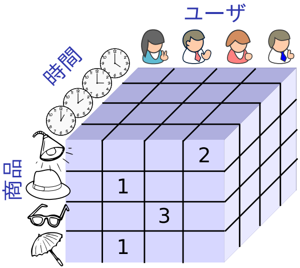
テンソルによる関係データの表現

複数種類の物事に関する「関係」の集まり

ペアに関する関係	⇔	行列	テンソル
3 つ組に関する関係	⇔	3 次元配列	
⋮			
L つ組に関する関係	⇔	L 次元配列	



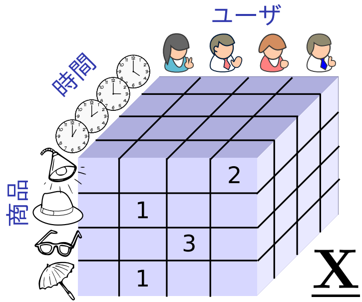
行列



3次元配列

欠損値を含むテンソルとして表現可能

テンソルに関する用語



- テンソルの各軸を**モード**と呼ぶ
- 各モードの要素数を**次元**と呼ぶ
- モードの数を**次数**と呼ぶ
  - 特別な場合として，2 次のテンソル=行列

テンソル表現の問題点

関係データのテンソル表現：一般的に大規模かつ高次元

- $1,000$ ユーザ $\times$ $1,000$ 商品 $\times$ $1,000$ 時間 $=$ 計 $1,000,000,000$ 個の関係

直接の取り扱いは困難

テンソル分解を用いたアプローチ

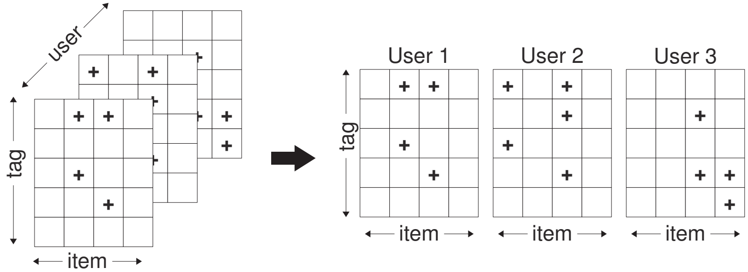
アウトライン

1. はじめに：関係データとテンソル
2. テンソル分解の応用事例
3. 古典的なテンソル分解
4. テンソル分解の確率的解釈
5. まとめ，今後の展開，参考文献

アウトライン

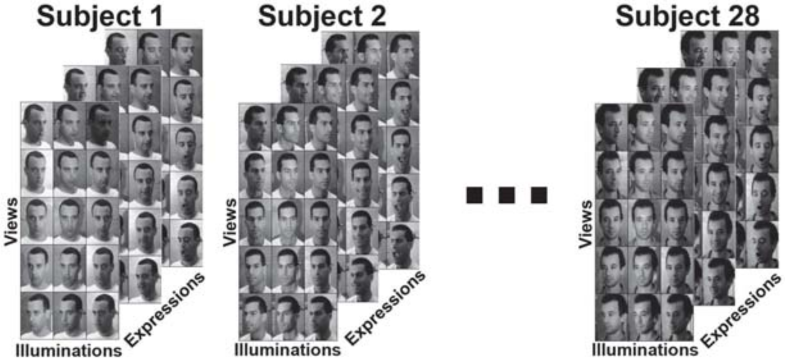
1. はじめに：関係データとテンソル
2. テンソル分解の応用事例
3. 古典的なテンソル分解
4. テンソル分解の確率的解釈
5. まとめ，今後の展開，参考文献

Pairwise Interaction Tensor Factorization [Rendle+ WSDM'10]



- データ：複数ユーザから付与された音楽へのタグ
  - タグの例："alternative", "rock", "heavy metal", ...
  - ユーザ × 商品 × タグ
- できること：各ユーザの趣向に特化したタグ推薦

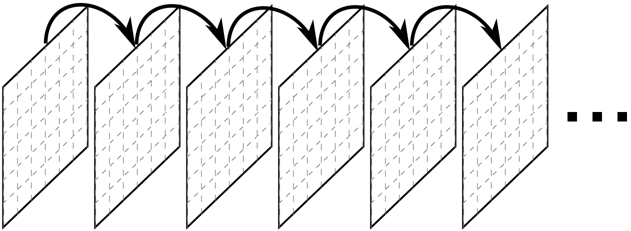
Tensor Face [Vasilescu+ ECCV'02][1]



- データ：顔画像データ
  - 人 × 角度 × 光源位置 × 表情
- できること：より精緻な特徴抽出

[1] Original image appeared in: Mørup Applications of Tensor Decomposition in Data Mining and Machine Learning, NIPS workshop TKML2010

Dynamic Tensor Analysis [Sun+ KDD'06]



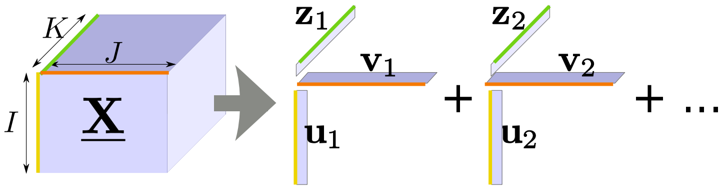
- データ：時変なネットワーク
  - 例：共著ネットワーク (DBLP)
  - 著者 × 著者 × 時間
- できること：未来の予測，異常検知

アウトライン

1. はじめに：関係データとテンソル
2. テンソル分解の応用事例
3. 古典的なテンソル分解
    - PARAFAC
    - Tucker 分解
4. テンソル分解の確率的解釈
5. まとめ，今後の展開，参考文献

PARAFAC [Harshman 1970]

$\underline{\mathbf{X}}$ をランク 1 テンソルの和に分解



$$\underline{\mathbf{X}} \simeq \sum_{q=1}^{Q} \mathbf{u}_q \circ \mathbf{v}_q \circ \mathbf{z}_q$$

- $\mathbf{u}_q, \mathbf{v}_q, \mathbf{z}_q (q = 1, \ldots, Q)$：各モードの基底
  - Q：基底の数

PARAFAC (Cont.)

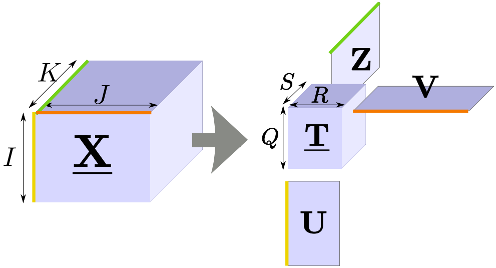
特異値分解 (SVD) **の素直な拡張**
—しかしその性質は行列の場合とは大きく異なる

1. **ランクの計算**
   - **行列：**SVD **で計算可能**
   - **テンソル：**NP **困難**

2. **完全な復元に必要な基底の数**
   - $I \times J$ **行列：**$\min(I, J)$ **個の基底で十分**
   - $I \times J \times K$ **テンソル：**$\min(I, J, K)$ **個の基底では不十分な場合が存在**

Tucker 分解 [Tucker 1966]

$\underline{\mathrm{X}}$ をコアテンソル $\underline{\mathrm{T}}$ と因子行列 $\mathrm{U}, \mathrm{V}, \mathrm{Z}$ に分解



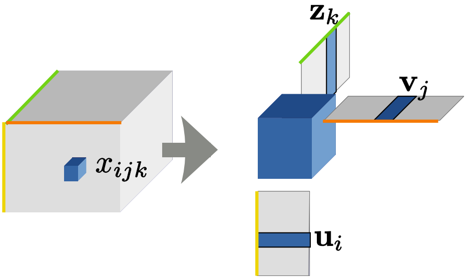
$$x_{ijk} \simeq \sum_{q=1}^{Q} \sum_{r=1}^{R} \sum_{s=1}^{S} t_{qrs} u_{iq} v_{jr} z_{ks}$$

- $\mathrm{U}, \mathrm{V}, \mathrm{Z}$：各モードの基底
- $\underline{\mathrm{T}}$：座標

$\underline{\mathrm{T}}$ が対角テンソルのとき PARAFAC と一致

Tucker 分解 [Tucker 1966]

$\underline{\mathrm{X}}$ をコアテンソル $\underline{\mathrm{T}}$ と因子行列 $\mathrm{U}, \mathrm{V}, \mathrm{Z}$ に分解



$$x_{ijk} \simeq \sum_{q=1}^{Q} \sum_{r=1}^{R} \sum_{s=1}^{S} t_{qrs} u_{iq} v_{jr} z_{ks}$$

- $\mathrm{U}, \mathrm{V}, \mathrm{Z}$：各モードの基底
- $\underline{\mathrm{T}}$：座標

$\underline{\mathrm{T}}$ が対角テンソルのとき PARAFAC と一致

Tucker分解の計算法

二乗誤差を最小化

$$\underset{\underline{\mathbf{T}},\mathbf{U},\mathbf{V},\mathbf{Z}}{\operatorname{argmin}} \sum_{(i,j,k)\in\Omega}^{K} ||x_{ijk} - y_{ijk}||^2$$

$$\text{s.t.} \quad y_{ijk} = \sum_{q=1}^{Q}\sum_{r=1}^{R}\sum_{s=1}^{S} t_{qrs}u_{iq}v_{jr}z_{ks},$$

$$\mathbf{U}^\top\mathbf{U} = \mathbf{I}, \mathbf{V}^\top\mathbf{V} = \mathbf{I}, \mathbf{Z}^\top\mathbf{Z} = \mathbf{I}$$

- Ω：観測した要素の添字集合
- 一般的に凸ではない

代表的な計算法：Higher Order SVD [Lathauwer+ 2000]

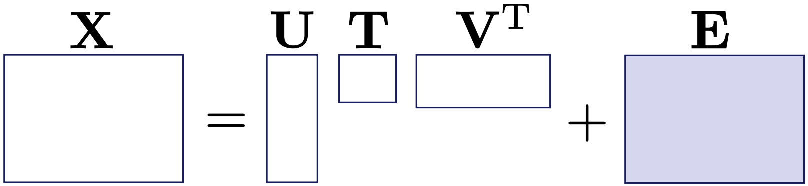
- $\underline{\mathbf{X}}$ の各モードに関して特異値分解 (SVD) を行う

アウトライン

1. はじめに：関係データとテンソル
2. テンソル分解の応用事例
3. 古典的なテンソル分解
4. テンソル分解の確率的解釈
5. まとめ，今後の展開，参考文献

Tucker分解の確率的解釈

簡単のため 2 次のテンソル (=行列) の場合を考える



- $E_{ij} \sim N(0, \sigma^2)$：**観測ノイズ**

このモデルの対数尤度は Tucker 分解の目的関数と等価

$$\ln p(\mathbf{X} \mid \mathbf{U}, \mathbf{V}, \mathbf{T}) = \frac{1}{2\sigma^2} ||\mathbf{X} - \mathbf{Y}||^2_{\mathrm{Fro}} + \mathrm{const.}$$

$$\text{where} \quad \mathbf{Y} = \mathbf{U}\mathbf{T}\mathbf{V}^\top$$

- $\mathbf{Y}$：$\mathbf{X}$ **の背後にある真の行列**

Tucker 分解の確率的解釈 (Cont.)

T に対し平均 0 分散 1 のガウス事前分布を仮定

- $p(\mathbf{T}) \sim \prod_{qr} N(t_{qr}|0,1)$

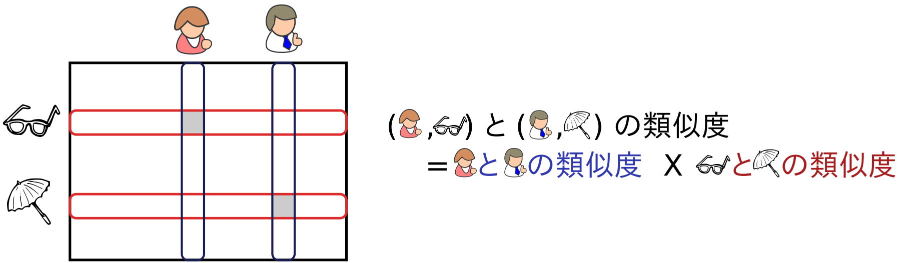
このとき Y の周辺分布 (平均的な振舞い) は平均 0 の行列ガウス分布となる [Chu & Ghahramani AISTAT'09]

$$\int p(\mathbf{Y}|\mathbf{T},\mathbf{U},\mathbf{V})p(\mathbf{T})\mathrm{d}\mathbf{T} = MN(\mathbf{Y}|\mathbf{U}\mathbf{U}^\top, \mathbf{V}\mathbf{V}^\top)$$

$$MN(\mathbf{Y}|\boldsymbol{\Sigma},\boldsymbol{\Omega}) \propto \exp\left(-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Omega}^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right]\right)$$

- $\boldsymbol{\Sigma} \in \mathbb{R}^{I\times I}, \boldsymbol{\Omega} \in \mathbb{R}^{J\times J}$ **：行，列の共分散行列**
  - **Y の全要素間の共分散 $(IJ \times IJ)$ を $O(I^2+J^2)$ のパラメータで表現**
- $\mathrm{cov}[y_{ij}, y_{kl}] = \Sigma_{ik}\Omega_{jl}$

Y の共分散 = Y の要素間の類似度 (近さ)

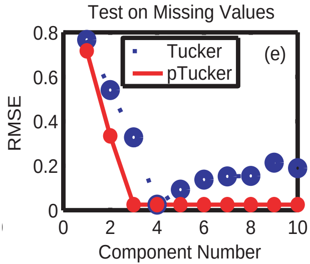


Tucker 分解：

- 要素間の類似度を各モードごとの類似度の組合せで表現
- 各類似度行列を低ランク制約のもとで推定

確率モデル化の利点

- ベイズ推論による正則化



[Chu & Ghahramani] の実験結果より

- 系統的な拡張が可能
  - 観測ノイズの変更，カーネル化，混合モデル化，⋯

アウトライン

1. はじめに：関係データとテンソル
2. テンソル分解の応用事例
3. 古典的なテンソル分解
4. テンソル分解の確率的解釈
5. まとめ，今後の展開，参考文献

まとめ

- **関係データ** = **複数の物事の組に関するデータ**
- **関係データはテンソルとして表現可能**
- **テンソルの圧縮表現としての** PARAFAC，Tucker **分解**
- Tucker **分解の確率モデル化**

今後の展開

テンソル分解の確率モデルとしての発展

- 数学的基盤はまだまだ未整備
  - 行列で成り立つことがテンソルで成り立つとは限らない
- 実用上は確率モデルを用いたアプローチが有望

大規模データの取り扱い

- 複数の変数が絡みあってるため，確率勾配法だと収束が遅くなる
- データのスパース性をどう扱うかが重要

参考文献 (サーベイ論文など)

- Kolda & Bader. Tensor decompositions and applications. SIAM Review, 2009.
  - **理論的にしっかりしたサーベイ**
- Mørup. Applications of tensor factorizations and decompositions in data mining, Wiley Interdisciplinary Reviews, 2011.
  - **最近の応用事例の紹介 (脳情報など)**
  - NIPS2010 **ワークショップのスライド：** `http://csmr.ca.sandia.gov/~dfgleic/tkml2010/slides/morup.pdf`
- Cichocki et al. Nonnegative Matrix and Tensor Factorizations. Wiley (Book), 2009
  - **非負以外の場合についても詳しい**
- Hoff. Latent Factor Models for Relational Arrays and Network Data. NIPS tutorial, 2010.
  - **テンソルの確率モデルに関する紹介**