2014 年大会速報論文特集

Short Paper

# 接続行列埋め込みに基づく複数種類の多項関係の同時予測

# Simultaneous Higher-order Relation Prediction via Collective Incidence Matrix Embedding

則 のぞみ
Nozomi Nori
京都大学
Kyoto University
nozomi@ml.ist.i.kyoto-u.ac.jp

ボレガラ ダヌシカ
Danushka Bollegala
リヴァプール大学
University of Liverpool
danushka.bollegala@liverpool.ac.uk

鹿島 久嗣
Hisashi Kashima
京都大学
Kyoto University
kashima@i.kyoto-u.ac.jp

**keywords:** higher-order relational data, multi-relational data, incidence matrix, nonlinear dimensionality reduction

**Summary**

We propose a prediction method for higher-order relational data from multiple sources. The high-dimensional property of higher-order relations causes problems associated with sparse observations. To cope with this problem, we propose a method to integrate higher-order relational data from multiple sources. Our target task is the simultaneous decomposition of higher-order, multi-relational data, which corresponds to the simultaneous decomposition of multiple tensors. However, we transform each tensor into an incidence matrix for the corresponding hypergraph and apply a nonlinear dimensionality reduction technique that results in a generalized eigenvalue problem guaranteeing global optimal solutions. We also extend our method to incorporate objects' attribute information to improve prediction for unseen/unobserved objects. To the best of our knowledge, this is the first reported method that can make predictions for (1) higher-order relations (2) with multi-relational data (3) with object attribute information and which (4) guarantees global optimal solutions. Using real-world datasets from social web services, we demonstrate that our proposed method is more robust against data sparsity than state-of-the-art methods for higher-order, single/multi-relational data including nonnegative multiple tensor factorization.

## 1. は じ め に

複数の情報源からの関係データが分析のために与えられる問題設定は，Web マイニング [Ma 11, Pan 10, Jiang 12, Hu 13] やバイオインフォマティクス [Acar 13, Acar 12] など様々な分野で一般的なものとなりつつある．例えば Web 上でユーザは他のユーザと “friend” などの関係を築く一方で，ニュース記事を閲覧し，それに関するコメントを投稿したり，情報の再発信を行ったりしている．そしてこれらユーザとリソースの間の異なる “関係” はそれぞれ対象に対する相補的な性質を反映していると期待できるため，複数の情報源からの関係データを統合することにより関連するアプリケーションの質が向上されることが期待されている．実際，Web マイニングの例で言えば，既に協調フィルタリング [Ma 11, Pan 10]，情報抽出 [Gupta 10]，ユーザの行動予測 [Bouchard 13] などのタスクで，複数の情報源からのデータ統合により各タスクにおける特定の性能が向上したことを示す実験結果が報告されている．

複数の情報源からの関係データ統合にあたっては，我々は以下のようなデータの多様性を扱う必要がある．(1) 我々は二項関係だけでなく**多項関係**も扱う必要がある．(2) 我々は関係データの他にオブジェクトの属性情報も活用する必要がある．多項関係は，三つ以上のオブジェクトを巻き込む関係を指す．現実世界においては例えば “花子が太郎から薦められた `http://www.ai-gakkai.or.jp` のサイトを気に入った”などいった複数のオブジェクトの間の相互作用が観測されるため，多項関係を的確に扱う必要が生じる．ここで，“太郎”，“花子”，“`http://www.ai-gakkai.or.jp`”がそれぞれオブジェクトであり，“X が Y から薦められた Z を気に入った”という関係が関係の種類に相当する．

**表 1** ウェブ上の各種多項関係データセットにおける，観測された関係の数（NNZ: Number of NonZero elements）と，オブジェクトの可能な組み合わせ数に対する観測された関係の数の割合（Rel-NNZ: Relative NNZ）.

| データセット | オブジェクト | オブジェクト数 | 関係数 (NNZ) | Rel-NNZ (三項) | Rel-NNZ (二項) |
|---|---|---|---|---|---|
| CiteULike*1 | ユーザ | 12,452 | | | 1.4E-5 |
| | タグ | 312,612 | 5,788,566 | 1.1E-9 | 3.0E-4 |
| (tagging) | 論文 | 1,356,389 | | | 1.5E-3 |
| Delicious*2 | ユーザ | 768 | | | 6.0E-4 |
| | タグ | 8,280 | 33,414 | 7.7E-7 | 6.0E-3 |
| (tagging) | URL | 6,860 | | | 5.0E-3 |
| Twitter*3 | 行為主体ユーザ | 1,144 | | | 1.6E-4 |
| | 発信元ユーザ | 7,935 | 14,221 | 1.4E-7 | 1.1E-3 |
| (retweet) | URL | 11,335 | | | 1.6E-3 |
| Twitter | 行為主体ユーザ | 1,125 | | | 1.2E-4 |
| | 発信元ユーザ | 10,049 | 22,755 | 1.1E-7 | 1.1E-3 |
| (favorite) | URL | 18,244 | | | 2.0E-3 |

(1) の要請は複数テンソルの同時分解の枠組み [Acar 11, Acar 13, Lin 09] によって探求されてきた．しかしながら，それらの多くは非凸最適化問題として定式化されるため，データが疎である場合には局所解による精度悪化が問題となる．特に，多項関係データの解析においては，組み合わせ爆発に伴う疎性とデータの分布に伴う疎性が付随するため，データ過疎への対処は重要な課題となる．以下でこれら二つの疎性について説明する．

**多項関係データに付随する疎性**

多項関係データでは，関係に含まれるオブジェクト数が増えるに従い可能な関係の組み合わせの数は指数的に増える．一方で，観測される関係の数は指数的には増えないため，可能な組み合わせに対して実際に観測されるデータの割合が関係が高次になるほど小さくなる．これが組み合わせ爆発に伴う疎性である．この傾向は表 1 に示すように，本論文中の実験で対象としたようなウェブ上の多項関係データでも確認された．表中では，各データセットについて観測された関係の数（NNZ: Number of NonZero elements）と，オブジェクトの可能な組み合わせの数に対する観測された関係の数の割合（Rel-NNZ）が示されている．ここで Rel-NNZ は右記のように定義される：Rel-NNZ$\equiv \frac{NNZ}{\prod_{k=1}^{K}|N^{(k)}|}$. ここで，$K$ は関係の項数（関係に含まれるオブジェクトの数）であり，$N^{(k)}$ は，$k$ 項目のオブジェクト集合のサイズである．Rel-NNZ（二項）のカラムには，各行の中のオブジェクトカラムの要素を除いた二項関係における Rel-NNZ が示されている．例えば最初の行においてはタグと論文の間の二項関係における Rel-NNZ が示されている．三項関係として Rel-NNZ を計算した場合にはその値は極めて小さく，具体的には最も高くても 0.00008% 以下であることが確認できる．一方，各データセットについて二項の関係にのみ注目して二項関係として Rel-NNZ を計算した場合には，Rel-NNZ は劇的に増加し，高い場合には 0.6% まで上がることが確認できる．これは，三項関係として計算した場合の Rel-NNZ の約 7,500 倍に相当する．

データの分布に伴う疎性とは，二項関係を対象とした複雑ネットワークのデータと同様に，多項関係データにおいてもその分布は例えばべき乗分布などで近似され，多くのオブジェクトは少数の関係にしか関与しない [Cai 11, Yin 13, Zlatic 09] いうものである．このような性質は上記のデータセット*1*2*3でも同様に確認された．少数の関係にしか関与しないオブジェクトはしばしば “コールドノード” などと呼ばれ，コールドノードに対する予測精度はその他のノードと比較して相対的に低くなる．特に推薦タスクにおいてはコールドノードに対する予測精度の向上が重要な問題となっている [Schein 02].

したがって我々は，先述の (1) と (2) の要請に加えもう一つの要請 (3) データ過疎に頑健な多項関係分析のための手法を構築することに応える必要がある．二項関係を対象とした各種データ分析においては，複数の情報源からのデータを組み合わせることでデータ過疎へより良く対処でき，例えば予測精度を向上しうることが報告されている [Takeuchi 13a, Pan 10]．したがって，多項関係予測にあたっても複数の情報源からのデータを統合することにより，データ過疎により良く対処できるようになることが期待できる．実際，近年の研究では，観測データが疎である場合には，多項関係解析においても複数の情報源からのデータを統合することで予測性能などの特定の性能を向上しうることが報告されている [Narita 11, Takeuchi 13b, Acar 13].

本論文では，多項関係分析の様々なタスクの中でも意思決定に直接結びつく予測問題を扱い，複数の情報源からの多項関係データを統合し，かつ理論的大域解を保証する多項関係予測手法を提案する．我々の提案手法は上記の三つの要請を満たす手法である．表 2 に，我々の提案手法と，既存手法の比較をまとめた．我々の提案手法は複数の接続行列の埋め込みによる非線形次元削減手法に相当するため，英訳である Collective Incidence Matrix Embedding の頭文字を取って提案手法を “CIME” と命名している．提案手法は一種類の関係データを用いた多項関係予測のための先行研究 [Nori 12] を複数種類の関係データに対して拡張したものである．一種類の多項関係を接続行列に変換し次元削減を行うことで理論的大域解を保証した先行研究 [Nori 12] に基づき，複数種類の多項関係を接続行列に変換し次元削減を行うことで，本論文では上記三つの要請を満たす手法を実現している．

## 2. 提 案 手 法

### 2·1 問 題 設 定

本論文で用いる記号の概要を 表 3 にまとめた．$K$ 種類のオブジェクト集合，$S^{(1)}, S^{(2)}, \ldots, S^{(K)}$ があり，それぞれの集合が $N^{(k)}$ $(1 \leq k \leq K)$ 個のオブジェクトを有する

*1 http://www.citeulike.org/faq/data.adp
*2 http://www.grouplens.org/datasets/hetrec-2011
*3 http://www.norizm.org/datasets.html

**表 3** 記号の概要

| 記号 | 説明 |
|---|---|
| $K$ | オブジェクトの種類数 |
| $R$ | 関係の種類数 |
| $S^{(k)}$ | $k$ 番目のオブジェクト集合 $(k \in \{1,\ldots,K\})$ |
| $N^{(k)}$ | $k$ 番目のオブジェクト集合に含まれるオブジェクトの総数 $(k \in \{1,\ldots,K\})$ |
| $M^{(r)}$ | $r$ 番目の種類の関係に含まれる関係インスタンスの総数 $(r \in \{1,\ldots,R\})$ |
| $s^{(k,i)} \in S^{(k)}$ | $k$ 番目のオブジェクト集合に含まれる $i$ 番目のオブジェクト $(k \in \{1,\ldots,K\}, i \in \{1,\ldots,N^{(k)}\})$ |
| $I^{(r)}$ | $r$ 番目の種類の関係に関与したオブジェクトの種類のインデックス集合 $(r \in \{1,\ldots,R\})$ |
| $O^{(r)} \subset S^{(r_1)} \times \cdots$ | $r$ 番目の種類の関係について観測された $M^{(r)}$ 個の関係インスタンスの集合 $(r_i \in I^{(r)})$ |
| $o^{(r,j)} \in O^{(r)}$ | $O^{(r)}$ に含まれる $j$ 番目の関係インスタンス $(r \in \{1,\ldots,R\}, j \in \{1,\ldots,M^{(r)}\})$ |
| $\boldsymbol{x}^{(k,i)}$ | オブジェクト $s^{(k,i)}$ の $D^{(k)}$ 次元の属性ベクトル $(k \in \{1,\ldots,K\}, i \in \{1,\ldots,N^{(k)}\})$ |

**表 2** 多項関係予測の実験で用いた様々な手法の比較

| 手法 | 多項関係 | 複数種類の関係 | 属性 | 理論的大域解 |
|---|---|---|---|---|
| CIME（提案手法） | ✓ | ✓ | ✓ | ✓ |
| NLDR [Nori 12] | ✓ | | ✓ | ✓ |
| METAFAC [Lin 09] | ✓ | ✓ | | |
| PARAFAC [Kolda 09] | ✓ | | | |
| Tucker [Kolda 09] | ✓ | | | |

とする．これらのオブジェクトは合計で $R$ 種類の関係に関与し，$r$ 番目の種類の関係について，$M^{(r)}(1 \le r \le R)$ 個の関係インスタンスが観測されているとする．$I^{(r)}$ は，$r$ 番目の関係に関与したオブジェクトの種類のインデックス集合である．$S^{(k)}$ 中の $i$ 番目のオブジェクトを $s^{(k,i)}$ によって示す．例えば，$s^{(1,1)}$ はユーザの花子である．我々はまた，各 $r = 1,2,\ldots,R$ について，$M^{(r)}$ 個の関係インスタンス集合として $O^{(r)} \subset S^{(r_1)} \times S^{(r_2)} \times \cdots$ を得ているとする．ここで $r_i \in I^{(r)}$ である．$O^{(r)}$ 中の $j$ 番目の関係インスタンスを $o^{(r,j)}$ によって示す．観測された各関係インスタンスは，ある特定の種類の関係 ($r$ 番目の種類の関係) が，オブジェクトのある特定の組み合わせについて成立することを示している．例えば，$o^{(r,1)} \in O^{(r)}$ は $Like($花子，`http://www.ai-gakkai.or.jp`，太郎$)$ であり，これは，**花子は太郎が薦めた** *"http://wwww.ai-gakkai.or.jp"* **のサイトを気に入る**ということを示している．

多くの現実的な状況では，各オブジェクトをそれ自身に関する情報と紐付けることができる．例えば，人であれば年齢や性別などの情報を，ウェブページであればドメインやテキストなどの情報を紐付けることができるだろう．したがって，$s^{(k,i)}$ に対して $D^{(k)}$ 次元の属性ベクトル $\boldsymbol{x}^{(k,i)}$ を紐付け，各 $k = 1,2,\ldots,K$ について，まとめて計画行列

$$\boldsymbol{\Phi}^{(k)} \equiv (\boldsymbol{x}^{(k,1)}, \boldsymbol{x}^{(k,2)}, \ldots, \boldsymbol{x}^{(k,N^{(k)})})^{\top},$$

を定義する．以上の前提の下，我々は本論文で扱う問題を以下のような入出力を持つ問題として定義する．

**問題: 複数種類の多項関係予測**

- **入力:**
  - $S^{(1)}, S^{(2)}, \ldots, S^{(K)}$: $K$ 種類のオブジェクト集合
  - $I^{(r)}$: $r$ 番目の種類の関係 $(r = 1,2,\ldots,R)$ に含まれるオブジェクトの種類のインデックス集合
  - $O^{(r)}$ $(\subset S^{(r_1)} \times S^{(r_2)} \times \cdots)$: $r$ 番目の種類の関係について $M^{(r)}$ 個の観測された関係インスタンス集合（$r_i \in I^{(r)}$）
  - $\boldsymbol{\Phi}^{(1)}, \boldsymbol{\Phi}^{(2)}, \ldots, \boldsymbol{\Phi}^{(K)}$: オブジェクトの属性を表現する $K$ 個の計画行列
- **出力:** 各 $r(r = 1,2,\ldots,R)$ について，各オブジェクト $s^{(k,i)} \in S^{(k)}, (k \in I^{(r)}, i \in \{1,...,N^{(k)}\})$ の組み合わせの中で $O^{(r)}$ に含まれない関係の生じやすさ．

ここで，提案手法では計画行列の入力の有無は任意である．すなわち，計画行列が得られる場合はそれを活用した予測が可能だが，計画行列が得られない場合でも関係データのみを用いた予測が可能である．

## 2·2 次元削減を用いた多項関係予測

我々は先行研究 [Nori 12] に基づき，複数種類の多項関係を表現するために以下のような接続行列を構築する．すなわち，$K$ 種類の各オブジェクト集合と $R$ 種類の各関係について，一つの二値行列を構築し，合計で $K \times R$ 個の行列を構築する．これらの行列は，**オブジェクトと関係インスタンスの間の関係**を表現するものであり，グラフ理論の接続行列に相当する．各行列の各要素は，ある特定のオブジェクトがある特定の関係インスタンスに所属するか否かを示す．もし $k$ 番目の種類のオブジェクトが $r$ 番目の種類の関係に関与しない場合，対応する行列は零行列となる．行列 $\boldsymbol{A}^{(k,r)}$ を，$S^{(k)}$ に所属するオブジェクトの，$r$ 番目の種類の関係への所属情報をまとめた $N^{(k)} \times M^{(r)}$ の二値行列とする．$A^{(k,r)}$ の各要素は以下のように定義される．

$$[\boldsymbol{A}^{(k,r)}]_{i,j} \equiv \begin{cases} 1 & (s^{(k,i)} \text{ が } o^{(r,j)} \text{ に参加する場合}) \\ 0 & (\text{そうでない場合}). \end{cases}$$

続いて二部グラフ予測問題 [Yamanishi 09] と似たアイ

デアにより，各オブジェクトとそのオブジェクトが関与した関係インスタンスが近傍に位置するような共通の潜在空間への写像を学習し，各オブジェクトとそのオブジェクトが関与した関係インスタンスを共通の潜在空間に埋め込む．まず最初に簡単のため一次元への埋め込みを考えるが，我々の提案手法は一般に $D$ 次元への埋め込みにも適用可能である．実際，最終的に導出される一般化固有値問題は，$D$ 次元への埋め込みの場合にも成立する．

我々はサイズ $N^{(1)}$ のオブジェクト集合 $S^{(1)}$ をサイズ $N^{(1)}$ のベクトル $\boldsymbol{f}^{(1)}$ として埋め込む．同様に，オブジェクト集合 $S^{(2)},S^{(3)},\dots,S^{(K)}$ をそれぞれサイズ $N^{(2)},N^{(3)},\dots,N^{(K)}$ のベクトル $\boldsymbol{f}^{(2)},\boldsymbol{f}^{(3)},\dots,\boldsymbol{f}^{(K)}$ として埋め込む．同様にして，$r$ 番目の種類の関係をサイズ $M^{(r)}$ のベクトル $\bar{\boldsymbol{f}}^{(r)}$ として共通の潜在空間に埋め込む．各オブジェクトを，関与した関係インスタンスの近傍に埋め込むことを考えると，最小化すべき目的関数を以下のように定義できる[*4]．

$$\begin{aligned} J &= \sum_{k,r,i,j} [\boldsymbol{A}^{(k,r)}]_{i,j} \left( [\boldsymbol{f}^{(k)}]_i - [\bar{\boldsymbol{f}}^{(r)}]_j \right)^2 \\ &= \sum_{k,r} \boldsymbol{f}^{(k)\top} \boldsymbol{D}^{(k,r)} \boldsymbol{f}^{(k)} + \sum_r \bar{\boldsymbol{f}}^{(r)\top} \bar{\boldsymbol{D}}^{(r)} \bar{\boldsymbol{f}}^{(r)} \\ &\quad - 2 \sum_{k,r} \boldsymbol{f}^{(k)\top} \boldsymbol{A}^{(k,r)} \bar{\boldsymbol{f}}^{(r)}. \end{aligned} \tag{1}$$

ここで，$\boldsymbol{D}^{(k,r)}$ は $(i,i)$ 番目の要素が $[\boldsymbol{D}^{(k,r)}]_{i,i} \equiv \sum_j [\boldsymbol{A}^{(k,r)}]_{i,j}$ として定義される対角行列であり，$\bar{\boldsymbol{D}}^{(r)}$ は同様に $[\bar{\boldsymbol{D}}^{(r)}]_{i,i} \equiv \sum_k \sum_j [\boldsymbol{A}^{(k,r)}]_{j,i}$ として定義される．ここで，各 $j$ について $\sum_i [\boldsymbol{A}^{(k,r)}]_{i,j} = 1$ であることを用いた．

この目的関数は $\boldsymbol{f}^{(k)} \equiv \boldsymbol{0}$ かつ $\bar{\boldsymbol{f}}^{(r)} \equiv \boldsymbol{0}$ としてしまうことで最小化できるが，そのような望ましくない解を避けるために下記のスケーリング制約を加える．

$$\sum_{k,r} \boldsymbol{f}^{(k)\top} \boldsymbol{D}^{(k,r)} \boldsymbol{f}^{(k)} = 1. \tag{2}$$

目的関数 $J$ の $\bar{\boldsymbol{f}}^{(r)}$ に関する最小値は以下のように得られる．

$$\bar{\boldsymbol{f}}^{(r)} = \bar{\boldsymbol{D}}^{(r)^{-1}} \sum_k \boldsymbol{A}^{(k,r)\top} \boldsymbol{f}^{(k)}. \tag{3}$$

式 (3) を，式 (1) の正負を逆転させたものに代入し，以下の最大化問題を得る．

$$\sum_{k,\ell,r} \boldsymbol{f}^{(k)\top} \boldsymbol{A}^{(k,r)} \bar{\boldsymbol{D}}^{(r)^{-1}} \boldsymbol{A}^{(\ell,r)\top} \boldsymbol{f}^{(\ell)} - \sum_{k,r} \boldsymbol{f}^{(k)\top} \boldsymbol{D}^{(k,r)} \boldsymbol{f}^{(k)}. \tag{4}$$

ゆえに，

$$\begin{aligned} &- J - \lambda \left( \sum_{k,r} \boldsymbol{f}^{(k)\top} \boldsymbol{D}^{(k,r)} \boldsymbol{f}^{(k)} - 1 \right) \\ &= \sum_{k,\ell,r} \boldsymbol{f}^{(k)\top} \boldsymbol{A}^{(k,r)} \bar{\boldsymbol{D}}^{(r)^{-1}} \boldsymbol{A}^{(\ell,r)\top} \boldsymbol{f}^{(\ell)} \\ &- \sum_{k,r} \boldsymbol{f}^{(k)\top} \boldsymbol{D}^{(k,r)} \boldsymbol{f}^{(k)} \\ &- \lambda \left( \sum_k \sum_r \boldsymbol{f}^{(k)\top} \boldsymbol{D}^{(k,r)} \boldsymbol{f}^{(k)} - 1 \right), \end{aligned} \tag{5}$$

と定義されるラグランジュ関数を最大化することで以下を得る．

$$\sum_{\ell,r} \boldsymbol{A}^{(k,r)} \bar{\boldsymbol{D}}^{(r)^{-1}} \boldsymbol{A}^{(\ell,r)\top} \boldsymbol{f}^{(\ell)} = 2(\lambda+1) \sum_r \boldsymbol{D}^{(k,r)} \boldsymbol{f}^{(k)}. \tag{6}$$

ここで以下の行列を定義する．

$$\boldsymbol{f} \equiv (\boldsymbol{f}^{(1)\top}, \boldsymbol{f}^{(2)\top}, \dots, \boldsymbol{f}^{(K)\top})^\top.$$

$$\bar{\boldsymbol{D}} \equiv \begin{bmatrix} \bar{\boldsymbol{D}}^{(1)} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \bar{\boldsymbol{D}}^{(R)} \end{bmatrix}$$

$$\tilde{\boldsymbol{D}}^{(r)} \equiv \begin{bmatrix} \boldsymbol{D}^{(1,r)} & & \\ & \ddots & \\ & & \boldsymbol{D}^{(K,r)} \end{bmatrix}$$

$$\boldsymbol{A} \equiv \begin{bmatrix} \boldsymbol{A}^{(1,1)} & \dots & \boldsymbol{A}^{(1,R)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{A}^{(K,1)} & \dots & \boldsymbol{A}^{(K,R)} \end{bmatrix}$$

すると，式 (6) は以下の一般化固有値問題として書き直すことができる．

$$\boldsymbol{A} \bar{\boldsymbol{D}}^{-1} \boldsymbol{A}^\top \boldsymbol{f} = \tilde{\lambda} (\sum_r \tilde{\boldsymbol{D}}^{(r)}) \boldsymbol{f}. \tag{7}$$

最大固有値に対応する固有ベクトル $\boldsymbol{f}$ を取得することで，オブジェクトの一次元空間での最適な埋め込み先を得ることができる．$D$ 次元空間での最適な埋め込み先 $\boldsymbol{f}_1$, $\boldsymbol{f}_2,\dots,\boldsymbol{f}_D$ を得るためには，固有値の大きい順に上位 $D$ 個の固有ベクトルを取得すれば良い．

最後に，式 (3) から関係インスタンスの最適な埋め込み先は，関与したオブジェクトの埋め込みベクトルの重

---

*4 関係の種類の影響を調整するためのパラメータを追加することは可能であり，実際この拡張は容易である．その場合に導出される最終的な一般化固有値問題は今回導出される一般化固有値問題と大差ないものとなる．実験では，このスケーリングパラメータの考慮の有無による精度差があまりなかったため，本論文では簡単のためスケーリングパラメータは考慮しない．

心として与えられる．式 (1) で定義されるように，オブジェクトと関係インスタンスの間のユークリッド距離が小さい程，それらのオブジェクトはその関係インスタンスに関与しやすいと考えられる．したがって，$r$ 種類目の関係を考えるとき，$k \in I^{(r)}$ 種類目のオブジェクトのインデックスが $i_k \in \{1,...,N^{(k)}\}$ であるような関係インスタンス $o(r, i_k)$ は，

$$\begin{aligned} &\mathit{diff}(o(r,i_k)) \\ &\equiv \frac{1}{|I^{(r)}|}\sum_{k\in I^{(r)}}\left\|[\boldsymbol{f}^{(k)}]_{i_k} - \frac{1}{|I^{(r)}|}\sum_{k'\in I^{(r)}}[\boldsymbol{f}^{(k')}]_{i_{k'}}\right\|^2 \end{aligned} \quad (8)$$

の値が小さい程生じやすいと考えられる．$|I^{(r)}|$ は集合 $I^{(r)}$ のサイズである．ここでは，異なる種類の関係間で項数が異なる場合に対処するため，$|I^{(r)}|$ で正規化している．

ゆえに，$O^{(r)}$ に含まれないオブジェクトの組み合わせ $o(r, i_k)$ に対して関係の生じやすさを表すスコア $\mathit{score}(o(r,i_k))$ は $\mathit{score}(o(r,i_k)) \equiv -\mathit{diff}(o(r,i_k))$ などとして与えられる．

### 2·3 オブジェクトの属性の活用

続いて，オブジェクトの属性情報 $\{\boldsymbol{\Phi}^{(k)}\}_{k=1}^{K}$ を活用するための拡張を行う．

線形写像：

$$\boldsymbol{f}^{(k)} \equiv \boldsymbol{\Phi}^{(k)}\boldsymbol{w}^{(k)}, \quad (9)$$

を考える．ここで $\boldsymbol{w}^{(k)}$ は $D^{(k)}$ 次元の属性ベクトルを一次元の共通の潜在空間に写像する $D^{(k)}$ 次元のパラメタである．最終的に得られる一般化固有値問題は $D$ 次元への埋め込みにも適用される．

属性を考慮する場合は，各属性とその属性が関与した関係インスタンスが，潜在空間で近傍に位置するような潜在空間への写像を学習する．式 (1) で $\boldsymbol{f}^{(k)}$ を $\boldsymbol{\Phi}^{(k)}\boldsymbol{w}^{(k)}$ に代えて属性を活用しない場合と同様の計算を行うと

$$\begin{aligned} &\sum_{\ell,r}\boldsymbol{\Phi}^{(k)\top}\boldsymbol{A}^{(k,r)}\bar{\boldsymbol{D}}^{(r)^{-1}}\boldsymbol{A}^{(\ell,r)\top}\boldsymbol{\Phi}^{(\ell)}\boldsymbol{w}^{(\ell)} \\ &= 2(\lambda+1)\sum_{r}\boldsymbol{\Phi}^{(k)\top}\boldsymbol{D}^{(k,r)}\boldsymbol{\Phi}^{(k)}\boldsymbol{w}^{(k)}, \end{aligned} \quad (10)$$

を得ることができ，これは下記の一般化固有値問題として表現できる．

$$\boldsymbol{\Phi}^{\top}\boldsymbol{A}\bar{\boldsymbol{D}}^{-1}\boldsymbol{A}^{\top}\boldsymbol{\Phi}\boldsymbol{w} = \tilde{\lambda}\left(\sum_{r}\boldsymbol{\Phi}^{\top}\tilde{\boldsymbol{D}}^{(r)}\boldsymbol{\Phi}\right)\boldsymbol{w}. \quad (11)$$

ここで

$$\boldsymbol{\Phi} \equiv \begin{bmatrix} \boldsymbol{\Phi}^{(1)} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{\Phi}^{(K)} \end{bmatrix}$$

$$\boldsymbol{w} \equiv (\boldsymbol{w}^{(1)\top}, \boldsymbol{w}^{(2)\top}, \ldots, \boldsymbol{w}^{(K)\top})^{\top}.$$

である．

過学習を防ぐために正の値である正則化パラメタ $\sigma > 0$ により正則化項を追加すると，一般化固有値問題，式 (11) は以下のように修正される．

$$\boldsymbol{\Phi}^{\top}\boldsymbol{A}\bar{\boldsymbol{D}}^{-1}\boldsymbol{A}^{\top}\boldsymbol{\Phi}\boldsymbol{w} = \tilde{\lambda}\left(\left(\sum_{r}\boldsymbol{\Phi}^{\top}\tilde{\boldsymbol{D}}^{(r)}\boldsymbol{\Phi}\right) + \sigma\boldsymbol{I}\right)\boldsymbol{w}. \quad (12)$$

### 2·4 計算量について

提案手法のアルゴリズムは，一般化固有値問題を解く部分と，それにより得られた潜在次元表現を用いて予測を行う部分の二つに分けられる．

前者については，属性を考慮した式 (12) で $\boldsymbol{\Phi}$ を単位行列に置き換え，$\sigma = 0$ とおけば式 (7) になるので，まとめて式 (12) について述べると，式 (12) は，適切な行列 $\boldsymbol{M}, \boldsymbol{D}$ を用いて $\boldsymbol{\Phi}^{\top}\boldsymbol{M}\boldsymbol{M}^{\top}\boldsymbol{\Phi}\boldsymbol{w} = \lambda(\boldsymbol{\Phi}^{\top}\boldsymbol{D}\boldsymbol{\Phi} + \sigma\boldsymbol{I})\boldsymbol{w}$ と書き直せる．ここで $\boldsymbol{D}$ は対角行列である．したがって，次元削減部分の計算量は $\boldsymbol{B} = \boldsymbol{\Phi}^{\top}\boldsymbol{M}\boldsymbol{M}^{\top}\boldsymbol{\Phi}, \boldsymbol{C} = \boldsymbol{\Phi}^{\top}\boldsymbol{D}\boldsymbol{\Phi} + \sigma\boldsymbol{I}$ とおいたときの一般化固有値問題 $\boldsymbol{B}\boldsymbol{x} = \lambda\boldsymbol{C}\boldsymbol{x}$ の計算量に相当するので，様々な解法が考えられ，計算量もそれに準じたものとなる．

後者については，$r$ 番目の種類の関係についてある関係インスタンスが与えられたときにスコアを得る計算量は，オブジェクトの属性ベクトルの次元を $D$，潜在次元数を $R$，$r$ 番目の種類の関係における項数を $K$ とすると，式 (9) に基づき属性ベクトルから潜在ベクトルを得る部分で $O(KRD)$，式 (8) の計算量として $O(KR)$ となるので，ある関係インスタンスに対して予測する計算量全体は $O(KRD)$ となる．

## 3. 実　　験

本章では提案手法の有用性を評価するために，ソーシャルメディアから取得した現実のデータセットを用いた実験結果を示す．データとしてソーシャルメディアのデータを選んだのは，ソーシャルメディアにおける関係予測は複数種類の多項関係予測における典型的なタスク例であるためである．しかし，我々の提案手法は他のタスクでも有用であると期待する．

### 3·1 実 験 条 件

**データセット**

[Nori 12] で用いられた *Twitter* のデータセットを用いた[*3]．データの詳細は表 4 にまとめられている．*Twitter* データセットには，二種類の関係：“retweet” (R-1) と “favorite” (R-2) が含まれる．各関係は三種類のオブジェクト：行為主体ユーザ, URL, 発信元ユーザを含む．各タプルは，特定の URL を含む発信元ユーザの発言に対して行為主体ユーザが “retweet” ないし “favorite” の行為を行っ

たことを表している．R-1 の関係を予測する対象の関係として，R-2 の関係を補助情報として用いた．

**比較手法**

提案手法を下記の手法と比較した．手法の比較は，表 2 にまとめた．**METAFAC** は Lin ら [Lin 09] によって提案された非負値テンソル同時分解手法であり，複数テンソルの同時分解手法の state-of-the-art な手法の一つである．我々は Lin らの公開コードを実装として用いた．複数の情報源からの関係データを統合する効果を確認するために，METAFAC においては，R-1 と R-2 の両方を用いた場合（(1) **METAFAC(multi)**）と R-1 のみを用いた場合（(2) **METAFAC(single)**）の二つを評価した．

(2) **PARAFAC** と (3) **Tucker** は多項関係予測に広く使われる典型的なテンソル分解手法である [Kolda 09]．これら二つの手法は一種類の多項関係を対象とした分解手法である．これら二つの手法に関しては，Tensor Toolbox*5 を実装として用い，予測する対象の関係（R-1）のみを用い，R-2 の関係は用いなかった．

**予測設定**

補助情報の関係インスタンス数は固定し，予測する関係種類の関係インスタンス数は一定の条件の下変化させた．具体的には，予測する種類の関係について全関係インスタンスの 20% を訓練データとしてサンプルし，残りの 80% を評価データとして用いた．このサンプリング，評価の一連のプロセスを 5 回繰り返し，t 検定 ($p < 0.05$) を行った．一つのサンプルをデベロップメントデータセットとして全てのパラメタ（潜在次元数）を $\{2^0, 2^1, \ldots, 2^{11}\}$ の範囲で調整した．一方で，補助情報の関係については全関係インスタンスを訓練データとして用いた．予測精度の評価指標としては AUC を採用した．

### 3·2 結 果

本実験では “コールドノード” とは少数の関係にしか関与しないオブジェクトを指す．本実験では全ユーザに対する予測精度とコールドノードに相当するユーザ（コールドユーザ）に対する予測精度を評価した．所謂 80 対 20 の法則により，行動数の多いユーザ上位 2 割が全体の行動の大半を占める傾向にあるため，本実験では行為数の少ないユーザ上位 8 割をコールドユーザーとした．

表 5 に，全ユーザとコールドユーザに対する AUC の平均を示した．t 検定 ($p < 0.05$) で統計的に有意であった手法が太文字で記されている．全ユーザに対する予測精度を評価したときには，我々の手法と比較手法 (1) METAFAC (multi) の間に統計的な有意差は観察されなかったが，コールドユーザに対する予測精度を評価したときには，提案手法は他の全手法を上回る精度を実現した．コールドユーザに対する予測を行うときに提案手法の恩恵が得られたことは特筆すべきである．

*5 `http://www.sandia.gov/~tgkolda/TensorToolbox`

**表 4** 実験で用いたデータセットの詳細

| 関係の種類 | 関係インスタンス数 | オブジェクト | オブジェクト数 |
|---|---|---|---|
| R1 | 14,221 | 行為主体ユーザ | 1,144 |
| | | 発信元ユーザ | 7,935 |
| (retweet) | | URL | 11,335 |
| R2 | 22,755 | 行為主体ユーザ | 1,125 |
| | | 発信元ユーザ | 10,049 |
| (favorite) | | URL | 18,244 |

**表 5** 各手法における予測性能 (AUC) の平均比較．t 検定で統計的に有意 ($p < 0.05$) であった手法が太文字で記されている．

| | 全ユーザ | コールドユーザ |
|---|---|---|
| CIME（提案手法） | 0.873 | **0.851** |
| METAFAC (multi) | 0.878 | 0.780 |
| METAFAC (single) | 0.823 | 0.690 |
| PARAFAC | 0.633 | 0.503 |
| TUCKER | 0.501 | 0.519 |

また，全ユーザに対する予測精度を評価した際には，提案手法との間に有意差が観察されなかった比較手法は METAFAC (multi) のみであった．全ユーザ/コールドユーザに対する予測精度を評価したいずれの場合も提案手法は一種類の多項関係データのみを予測に用いた他の全比較手法：(2) METAFAC (single)，(3) PARAFAC, (4) Tucker を有意に上回る予測精度を実現した．また，比較手法の METAFAC については，全ユーザ/コールドユーザに対する予測精度を評価したいずれの場合も複数種類の関係データを予測に用いた (2) METAFAC (multi) のほうが一種類の関係データのみを予測に用いた (3) METAFAC (single) よりも有意に高い予測精度を実現した．以上から，多項関係予測において複数の情報源からのデータを統合することで予測精度を向上させることができることが示唆される．

## 4. 関 連 研 究

多項関係予測においてはテンソル分解がしばしば用いられており，CP/PARAFAC や Tucker をはじめとする様々な分解モデルやカルバック・ライブラー情報量やユークリッド距離などを元にした様々な目的関数が，様々な最適化アルゴリズムとともに提案されている [Kolda 09]．近年は複数の情報源からの多項関係データを統合する，複数テンソルの同時分解のための様々な手法が提案されている [Lin 09, Takeuchi 13b, Acar 11, Yokota 12]．しかし，多くの既存手法で保証されるのは局所解のみであり，その予測性能が対象アルゴリズムに与える初期値に依存するという問題が存在する．対照的に，本研究が提案する手法は理論的には大域解を保証する定式化となっており，データの疎性に頑健な予測が可能になると期待される．本論文では，一種類の多項関係データを用いた関係予測のための先行研究 [Nori 12] を複数種類の関係に対して拡張することで，複数種類の多項関係のデータを統合し

ながら理論的大域解を保証する手法を実現した．

## 5. おわりに

本論文では，複数の情報源からの多項関係データを統合する多項関係予測のための手法を提案した．理論的大域解を保証する定式化と複数の情報源からの関係データ統合により，提案手法はデータが過疎な状況下で，既存の複数テンソルの同時分解手法や一種類のテンソル分解手法を上回る予測精度を実現した．

## ◇参考文献◇


[Acar 11] Acar, E., Kolda, T. G., and Dunlavy, D. M.: All-at-once Optimization for Coupled Matrix and Tensor Factorizations, in *MLG'11* (2011)

[Acar 12] Acar, E., Gurdeniz, G., Rasmussen, M. A., Rago, D., Dragsted, L. O., and Bro, R.: Coupled Matrix Factorization with Sparse Factors to Identify Potential Biomarkers in Metabolomics, *International Journal of Knowledge Discovery in Bioinformatics*, Vol. 3, No. 3, pp. 22–43 (2012)

[Acar 13] Acar, E., Rasmussen, M. A., Savorani, F., Næs, T., and Bro, R.: Understanding data fusion within the framework of coupled matrix and tensor factorizations, *Chemometrics and Intelligent Laboratory Systems*, Vol. 129, No. 0, pp. 53–63 (2013)

[Bouchard 13] Bouchard, G., Yin, D., and Guo, S.: Convex Collective Matrix Factorization, in *AISTATS '13*, pp. 144–152 (2013)

[Cai 11] Cai, Y., Zhang, M., Luo, D., Ding, C., and Chakravarthy, S.: Low-order tensor decompositions for social tagging recommendation, in *WSDM '11*, pp. 695–704 (2011)

[Csanky 76] Csanky, L.: Fast Parallel Matrix Inversion Algorithms, *SIAM Journal on Computing*, Vol. 5, No. 4, pp. 618–623 (1976)

[Gupta 10] Gupta, S. K., Phung, D., Adams, B., Tran, T., and Venkatesh, S.: Nonnegative Shared Subspace Learning and Its Application to Social Media Retrieval, in *KDD '10*, pp. 1169–1178 (2010)

[Hu 13] Hu, L., Cao, J., Xu, G., Cao, L., Gu, Z., and Zhu, C.: Personalized Recommendation via Cross-domain Triadic Factorization, in *WWW '13*, pp. 595–606 (2013)

[Jiang 12] Jiang, M., Cui, P., Wang, F., Yang, Q., Zhu, W., and Yang, S.: Social Recommendation Across Multiple Relational Domains, in *CIKM '12*, pp. 1422–1431 (2012)

[Kolda 09] Kolda, T. G. and Bader, B. W.: Tensor Decompositions and Applications, *SIAM Review*, Vol. 51, No. 3, pp. 455–500 (2009)

[Lehoucq 96] Lehoucq, R. and Sorensen, D. C.: Deflation Techniques For An Implicitly Re-Started Arnoldi Iteration, *SIAM Journal on Matrix Analysis and Applications*, Vol. 17, No. 14, pp. 789–821 (1996)

[Lin 09] Lin, Y.-R., Sun, J., Castro, P., Konuru, R., Sundaram, H., and Kelliher, A.: MetaFac: Community Discovery via Relational Hypergraph Factorization, in *KDD '09*, pp. 527–536 (2009)

[Ma 11] Ma, H., Zhou, D., Liu, C., Lyu, M. R., and King, I.: Recommender Systems with Social Regularization, in *WSDM '11*, pp. 287–296 (2011)

[Narita 11] Narita, A., Hayashi, K., Tomioka, R., and Kashima, H.: Tensor Factorization Using Auxiliary Information, in *ECML PKDD '11*, pp. 501–516 (2011)

[Nori 12] Nori, N., Bollegala, D., and Kashima, H.: Multinomial Relation Prediction in Social Data: A Dimension Reduction Approach, in *AAAI '12*, pp. 115–121 (2012)

[Pan 87] Pan, V. Y.: Complexity of Parallel Matrix Computations, *Theoretical Computer Science*, Vol. 54, No. 1, pp. 65–85 (1987)

[Pan 10] Pan, W., Xiang, E. W., Liu, N. N., and Yang, Q.: Transfer Learning in Collaborative Filtering for Sparsity Reduction, in *AAAI '10*, pp. 617–624 (2010)

[Schein 02] Schein, A. I., Popescul, A., Ungar, L. H., and Pennock, D. M.: Methods and Metrics for Cold-start Recommendations, in *SIGIR '02*, pp. 253–260 (2002)

[Sorensen 92] Sorensen, D. C.: Implicit application of polynomial filters in a k-step Arnoldi method, *SIAM Journal on Matrix Analysis and Applications*, Vol. 13, No. 1, pp. 357–385 (1992)

[Takeuchi 13a] Takeuchi, K., Kimura, K. I. A., and Sawada, H.: Non-negative Multiple Matrix Factorization, in *IJCAI '13*, pp. 1713–1720 (2013)

[Takeuchi 13b] Takeuchi, K., Tomioka, R., Ishiguro, K., Kimura, A., and Sawada, H.: Non-negative Multiple Tensor Factorization, in *ICDM '13* (2013)

[White 05] White, S. and Smyth, P.: A spectral clustering approach to finding communities in graphs, in *SDM '05*, pp. 274–285 (2005)

[Yamanishi 09] Yamanishi, Y.: Supervised Bipartite Graph Inference, in *NIPS '09*, pp. 1841–1848 (2009)

[Yin 13] Yin, D., Guo, S., Chidlovskii, B., Davison, B. D., Archambeau, C., and Bouchard, G.: Connecting Comments and Tags: Improved Modeling of Social Tagging Systems, in *WSDM '13*, pp. 547–556 (2013)

[Yokota 12] Yokota, T., Cichocki, A., and Yamashita, Y.: Linked PARAFAC/CP Tensor Decomposition and Its Fast Implementation for Multi-block Tensor Analysis, in *ICONIP'12*, pp. 84–91 (2012)

[Zlatic 09] Zlatic, V., Ghoshal, G., and Caldarelli, G.: Hypergraph topological quantities for tagged social networks, *Physical Review E*, Vol. 80, No. 3, p. 036118 (2009)


〔担当委員：鷲尾 隆〕

2014 年 9 月 21 日　受理

## 著者紹介

**則　のぞみ**（学生会員）

2010 年東京大学工学部システム創成学科卒．2012 年東京大学大学院情報理工学系研究科創造情報学専攻修士課程終了．2014 年から京都大学情報学研究科知能情報学専攻博士課程．学術振興会特別研究員（DC1）．統計的機械学習に基づくデータマイニング，特にヘルスケアへの応用に興味を持つ．

**ボレガラ　ダヌシカ**（正会員）

2005 年東京大学工学部電子情報工学科卒．2007 年同大学院情報理工学系研究科修士課程修了．2009 年同研究科博士課程修了．博士 (情報理工学)．同研究科・助教，講師を経て現在は University of Liverpool (Department of Computer Science) の Senior Lecturer (Associate Professor)．自然言語処理に興味を持つ．WWW, ACL, ECAI などの会議を中心に研究成果を発表．

**鹿島　久嗣**（正会員）

1999 年京都大学大学院工学研究科応用システム科学専攻修士課程修了. 2007 年京都大学大学院情報学研究科知能情報学専攻博士課程修了. 博士 (情報学). 1999 年から 2009 年まで日本アイ・ビー・エム株式会社 東京基礎研究所勤務. 2009 年から 2013 年まで東京大学大学院情報理工学系研究科数理情報学専攻・准教授. 2014 年から京都大学情報学研究科知能情報学専攻・教授．機械学習やデータマイニングの研究, 特にグラフやネットワーク構造をもったデータを対象とする予測モデリングに取り組む.