

# セマンティクスを用いたテンソル分解手法

## Tensor Factorization That Utilizes Semantics behind Objects

中辻 真  
Makoto Nakatsuji  
NTT サービスエボリューション研究所  
NTT Service Evolution Laboratories  
nakatsuji.makoto@lab.ntt.co.jp

藤原 靖宏  
Yasuhiro Fujiwara  
NTT ソフトウェアイノベーションセンタ  
NTT Software Innovation Center  
fujiwara.yasuhiro@lab.ntt.co.jp

戸田 浩之  
Hiroyuki Toda  
NTT サービスエボリューション研究所  
NTT Service Evolution Laboratories  
toda.hiroyuki@lab.ntt.co.jp

澤田 宏  
Hiroshi Sawada  
(同 上)  
sawada.hiroshi@lab.ntt.co.jp

チェン  
ジン  
Jin Zheng  
Tetherless World Constellation  
Rensselaer Polytechnic Institute  
zhengj3@rpi.edu

ヘンドラー  
ジェームズ  
James Hendler  
(同 上)  
hendler@cs.rpi.edu

**keywords:** tensor factorization, linked open data, recommender systems, link prediction

### Summary

Predicting human activities is important for improving recommender systems or analyzing social relationships among users. Those human activities are usually represented as multi-object relationships (e.g. user's tagging activities for items or user's tweeting activities at some locations). Since multi-object relationships are naturally represented as a tensor, tensor factorization is becoming more important for predicting users' possible activities. However, its prediction accuracy is weak for ambiguous and/or sparsely observed objects. Our solution, Semantic data Representation for Tensor Factorization (SRTF), tackles these problems by incorporating semantics into tensor factorization based on the following ideas: (1) It first links objects to vocabularies/taxonomies and resolves the ambiguity caused by objects that can be used for multiple purposes. (2) It then links objects to composite classes that merge classes in different kinds of vocabularies/taxonomies (e.g. classes in vocabularies for movie genres and those for directors) to avoid low prediction accuracy caused by rough-grained semantics. (3) It finally lifts sparsely observed objects into their classes to solve the sparsity problem for rarely observed objects. To the best of our knowledge, this is the first study that leverages semantics to inject expert knowledge into tensor factorization. Experiments show that SRTF achieves up to 10% higher accuracy (lower RMSE value) than state-of-the-art methods.

### 1. はじめに

人間の行動は一般的に複数のオブジェクトの関係から構成されることが多いため、複数オブジェクト関係を分析することが、将来のユーザ行動を予測するために重要である。複数オブジェクト関係から構成される人間の行動の例としては、コンテンツプロバイダにおいて“ユーザ”が“タグ”を“アイテム”に与えるという行動や、twitterにおいて“ユーザ”が“ある地点”で“ツイート”を記述するという行動など、類挙に暇がない。そのため、複数オブジェクト関係を予測することは、推薦やソーシャルネットワークの分析といった様々なアプリケーションを改善するのに役立つ。例えば、“あるユーザ”がスリラージャンルに属する“映画タイトル”に“ロマンス”というタグ

をつけて評価しており、“別のユーザ”が“同じタイトル”に“カーアクション”というタグをつけて評価していたとする。タグを無視し、ユーザとアイテムの2つのオブジェクト間の関係のみを利用する手法は、上記2ユーザは同じタイトルを評価しているため、類似すると計算する。それに対し、タグを含む3つのオブジェクト関係を利用する手法は、上記2ユーザを、同じアイテムを評価しているという点と、アイテムに対する主張が異なっているという点を考慮し詳細な類似度計算を行う。先行研究 [Nakatsuji and Fujiwara2014] の実験において、こうした主張の違いを協調フィルタリング手法に反映することが推薦精度の向上に繋がることが示されている。

複数オブジェクト関係はテンソルで自然に表現できるため、テンソル分解による将来のユーザの行動予測はますます

重要となっている [Karatzoglou et al.2010] . 今まで知られている提案の中でも, Bayesian Probabilistic Tensor Factorization (BPTF) [Xiong et al.2010] による行動予測は, パラメータ設定が簡単なうえ, 大規模データセットを効率的にサンプリングすることができ, 有望である. しかし, 現状のテンソル分解の枠組みでは, オブジェクトの属する概念(クラス)およびタクソノミ, ポキャブラリで表されるクラス体系として定義されるセマンティクスを活用できないため, 以下で述べる通り意味的な曖昧性があったり観測が希薄なオブジェクト(希薄オブジェクト)を取り扱う際に課題がある.

意味的な曖昧性は, 文書クラスタリングにおける重要な問題の1つである. 例えば, WordNet [Miller1995] や Wikipedia を用いることで, 意味的な曖昧性を解決し文書クラスタリングの性能を向上させたり [Hu et al.2008], 文書の意味的な関連性の計算の性能を向上させる研究 [Gabrilovich and Markovitch2007] がある.

また, 近年のセマンティクスを用いた情報フィルタリングに関する研究では, ユーザの興味対象の背景となるセマンティクスを用いることで高精度なオピニオンマイニング [Saif, He, and Alani2012] や推薦 [Nakatsuji et al.2012, Nakatsuji and Fujiwara2014] を実現している. しかし, 背景となるセマンティクスの粒度が粗い場合に精度が劣化することが報告されている. 例えば図 1-(1)において, 映画作品“Rock”は“Action”ジャンルに属する. しかし, “Action”ジャンルに属する映画作品は非常に多岐にわたって存在しているため, 単純に“Action”ジャンルというクラスが一致しているというだけでユーザの嗜好を測るのは適切ではない. このように, ユーザの嗜好を反映するには, セマンティクスの粒度が粗いことが, 現状のタクソノミ・ポキャブラリを用いる場合に存在する. 近年のセマンティック Web に関する研究 [Parundekar, Knoblock, and Ambite2012] では, Linked Open Data (LOD) [Bizer, Heath, and Berners-Lee2009] 上に存在する複数のポキャブラリやタクソノミにおけるクラスを組合せ複合クラスを作成することで(例えば, geoNames [Jain et al.2010] におけるクラス“populated place”と DBPedia [Bizer et al.2009] におけるクラス“city”を組合せて複合クラスを作成するなど), より詳細なクラス定義を与えようと試みている. 本研究でも, 詳細に定義されたセマンティクスをオブジェクトの背景知識として利用することはテンソル分解による評点予測や消費頻度の予測精度を向上させると考えている.

また, テンソル分解を適用する場合の主要な問題として, 訓練データセットが少ない場合に精度が大きく劣化する希薄問題 (*sparsity problem*) がある [Narita et al.2011]. この問題を解決しテンソル分解の精度を向上するため, [Narita et al.2011, Ermis, Acar, and Cemgil2012, Takeuchi et al.2013, Ermis, Acar, and Cemgil2013, Yilmaz, Cemgil, and Simsekli2011] らは, 複数オブジェクト

関係に加え各オブジェクトに関する補助情報を用いている. しかし, オブジェクトの背景知識であるセマンティクスを効果的に用いるテンソル分解手法は現時点で我々の研究 [Nakatsuji et al.2014] 以外に発表されていない.

本研究で提案する Semantic data Representation for Tensor Factorization (SRTF) は, 以下の3つのアイデアに基づきテンソル分解にセマンティクスを組み込む.(1) オブジェクトと LOD 上のポキャブラリ/タクソノミにおけるインスタンスの類似度を計測し, 最も類似度の高いオブジェクトにインスタンスを関係付けることで曖昧性を解決する. 図 1 は, ユーザ, アイテム, タグからなる3オブジェクト間の関係の例を示している. 図 1-(1)において, ユーザ  $u_m$  が“Rock”という映画アイテム  $v_n$  にタグ“Breathless”を与えていたとする. しかし, “Rock”という映画アイテムは複数存在し, また“Breathless”という単語も複数の意味を持つ. そこで, SRTF はまず  $v_n$  に与えられたメタデータと Freebase\*<sup>1</sup> または DBPedia 上のインスタンスに与えられたプロパティ間の類似度を計算する. そして, アイテム  $v_n$  とクラス“C3: Action”配下のインスタンス“Rock”が最も類似度が高ければ,  $v_n$  をクラス“C3: Action”配下のインスタンス“Rock”に關係付ける. その上で, 図 1 におけるタグ“Breathless”のようなタグ由来の曖昧性も, “C3: Action”配下のインスタンス“Rock”との関連性を計算することで解決をする.

(2) さらに SRTF は, 文献 [Parundekar, Knoblock, and Ambite2012] で提案されている方法を用い, 異なるポキャブラリ内のクラスに属するインスタンスの積集合を計算することで複合クラスを作成する. その上で SRTF は元々のクラスのみではなく複合クラスのインスタンスとオブジェクトを關係付ける. これにより, より詳細なセマンティクスを用いて複数オブジェクト関係を解析することができ, 予測精度の向上に繋がる. 図 1-(2) に複合クラス的具体例を示す. 図において, アイテムはジャンルポキャブラリ配下のクラス“Action”と監督ポキャブラリ配下のクラス“Michael Bay”の両方に含まれるインスタンスへ關係付けられている. SRTF はこの場合, 複合クラス“C4: Action&Michael Bay”を作成し, C3 配下の“Rock”に付け加え C4 配下のインスタンス“Rock”にも  $v_n$  を關係付ける. このように, Action のみでなく, Action&Michael Bay という関係を保持することで, アイテム  $v_n$  に興味を持つユーザは, Action のみでなく, Michael Bay にも興味を持ちうるという詳細な分析を可能とする.

(3) 次に, 観測が希薄なオブジェクトに対して, LOD または WordNet に關係付けされているオブジェクトを所属クラスに置き換え, クラスを用いた複数オブジェクト関係を構築する. SRTF はクラス内の共有知識を用いテンソル分解における希薄オブジェクトへ, オブジェクトのクラスに対し計算される属性値であるセマンティクス由来のバイアスを適用する. 図 1-(3) において, “Rock”ま

\*1 <http://www.freebase.com>

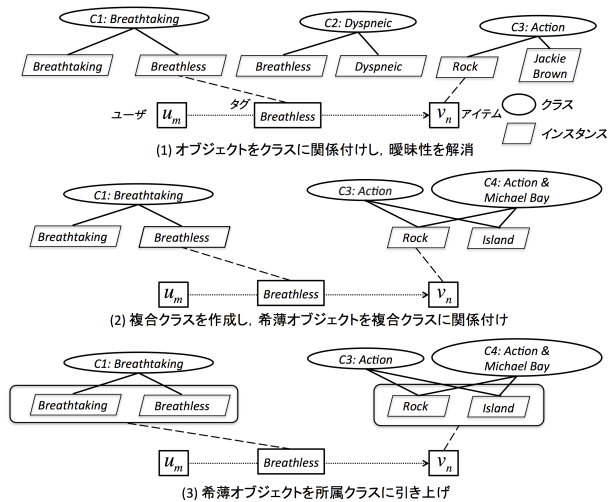


図1 オブジェクトをクラスに關係付けを行う例

たは“Island”に關係付けされているアイテム、もしくは、“Breathhtaking”または“Breathless”に關係付けされているタグが過去にわずかにしか観測されていなかったとする。そうした場合、既存の一般的な分解手法においては、観測が希薄なオブジェクトの影響を無視する傾向があり、結果として精度が劣化する。SRTFはクラスC3や複合クラスC4(またはC1)経由のセマンティックバイアスを、テンソル分解時に希薄オブジェクトに与える。これにより、希薄問題を解決できる。

上述した3つのアイデアをテンソル分解に組み込むにあたり、SRTFはオブジェクトをクラスに關係付けることにより曖昧性解消テンソルを作成し、希薄オブジェクトをクラスに置き換え複数オブジェクト関係を構築した上で、その関係を元のテンソルに付け加えたテンソル(増大テンソル)を作成する。そして、BPTFフレームワーク上において両テンソルを同時に分解する。そして、増大テンソルを分解することで得られるセマンティックバイアスを、曖昧性解消テンソルを分解する際に得られる属性ベクトルに反映させる。これにより、テンソル分解時に発生する希薄問題を解決できる。図2にSRTFの構成を示す。まず多属性からなる行動とボキャブラリ/タクソノミを入力として与え、曖昧性解消およびセマンティックスの粒度詳細化を実施する。そうすることで、セマンティックスに關連付けられたテンソルを構成する。次に、セマンティックスを考慮したテンソル分解を実施することで、希薄性解消を行う。

本論文においては、SRTFをユーザの評点行動の予測と消費頻度の予測に適用し、(1)MovieLensにおける映画タイトルに対する評点/タグデータセット\*2にFreeBase[Bollacker et al.2008]経由のアイテムボキャブラリとWordNet経由のタグタクソノミを組み合わせたもの、(2)Yelpにおけるレストランに対する評点/レビュー

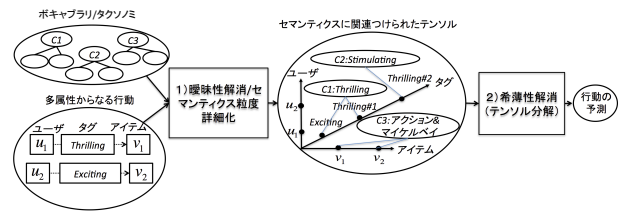


図2 SRTFの構成

データセット\*3にDBpedia[Bizer et al.2009]を用いレビュー内の食品を表すフレーズにタグ付けを行ったもの、(3)Last.fmにおける音楽アーティストに対する聴取頻度\*4にFreeBase[Bollacker et al.2008]経由のアイテムボキャブラリとWordNet経由のタグタクソノミを組み合わせたものの3つのデータセットを用い評価を行った。結果として、提案するSRTFが、テンソルに補助情報を与える研究として現在主流のGeneralize Coupled Tensor Factorization(GCTF)やベースラインであるBPTFよりも最大10%以上の精度向上を果たすことを示した。

なお、本稿における評点行動は、ユーザが決められた評点範囲(1点~5点など)の中でアイテム評価を行う行動をさす。評点分布は一般的にガウス分布に属する傾向が強いとされている[Xiong et al.2010]。本稿で提案するSRTFはガウス分布を事前分布とするため、評点予測に適する。また、本稿でいう消費行動は、ユーザがアイテムを購入・聴取する等、アイテムを消費する行動をさす。消費頻度の分布は一般的にベキ乗分布に属する傾向が強いとされている[Xiong et al.2010]。本稿で提案するSRTFはガウス分布を事前分布とするが、消費行動の予測に適用した場合でも他手法に比べ精度が高いことを実験により確認した。ガウス分布をベキ乗分布に変更することで精度の改善が期待できるが、MCMC処理をベキ乗分布に適用するのは、ガウス分布に適用するよりも複雑であり計算時間を要する[Xiong et al.2010, Sutskever, Tenenbaum, and Salakhutdinov2009]。そのため、本実験結果は有益である。なお、ガウス分布を消費行動の予測に適用するアプローチは他研究でも行われている[Sutskever, Tenenbaum, and Salakhutdinov2009]。

本稿は、[Nakatsuji et al.2014]において提案された手法(SRTF)のMarkov Chain Monte Carlo(MCMC)処理について詳細な説明を与え、さらに検証を深めユーザの商品への評点行動のみでなく、消費頻度の予測にもSRTFが高精度で適用可能であることを示した。結果、多様なアプリケーションにSRTFが適用可能であることを示した。

\*2 <http://www.grouplens.org/node/73> より入手可能

\*3 [http://www.yelp.com/dataset\\_challenge/](http://www.yelp.com/dataset_challenge/)より入手可能

\*4 <http://www.grouplens.org/node/462> より入手可能

## 2. 関連研究

テンソル分解は近年、推薦システム [Karatzoglou et al.2010, Rendle and Schmidt-Thieme2010] やソーシャルネットワーク分析 [Zheng et al.2010] など様々なアプリケーションに適用されるようになってきている。また、確率フレームワークに基づく高速かつパラメータ設定が容易なテンソル分解手法も提案されてきている [Xiong et al.2010]。しかし、実際のアプリケーションにおけるデータはテンソル上で取りうる複数オブジェクト関係の組合せに対し非常に希薄であるため、テンソル分解による予測精度が劣化してしまう [Narita et al.2011]。前章で述べた Generalized Coupled Tensor Factorization (GCTF) [Yilmaz, Cemgil, and Simsekli2011, Ermis, Acar, and Cemgil2012, Ermis, Acar, and Cemgil2013] や、その他のいくつかの研究 [Zheng et al.2010, Acar, Kolda, and Dunlavy2011, Takeuchi et al.2013] は、テンソル分解時に、観測されたテンソルと補助情報を示す行列を同時に分解することで補助情報から得られるバイアスをテンソル分解に反映させ、希薄問題を解決しようとしている。著者らの先行研究 [Nakatsuji, Miyoshi, and Otsuka2006, Nakatsuji, Yoshida, and Ishida2009, Nakatsuji et al.2010, Nakatsuji et al.2011, Nakatsuji et al.2012, Nakatsuji and Fujiwara2014] においても、メモリベースの協調フィルタリングにおける精度を向上させるために、オブジェクトの背景知識を用いているが、推薦等に適用する際に“類似するユーザ数”等のパラメータをデータセット毎にヒューリスティックに用意しないと行けない。一方、タクソノミは現在、ゲノム科学の領域においても機能のアノテーションを伴いゲノムをクラスタリングする際に使われるようになっている [Nakaya et al.2013]。しかし、オブジェクトの背景となるセマンティクスを構成するボキャブラリやタクソノミを活用し、曖昧性解消および希薄問題の解決を行うテンソル分解手法は、我々の提案以外に存在しない。

## 3. 背景技術

本章では本研究で用いる背景技術を説明する。

### 3.1 ボキャブラリとタクソノミ

本稿は Linked Open Data (LOD) で表現されているセマンティクスを、テンソル分解に取り込み、曖昧性解消および希薄問題の解決を行う。そこで、LOD 上に出現するセマンティクスの形式であるボキャブラリおよびタクソノミの紹介を行う。

ボキャブラリとタクソノミは、“シンプルなオントロジ” [McGuinness2003] とも呼ばれ、人間が定義したクラスの集合からなり、グラフ構造 (ボキャブラリ) もしくは木構造 (タクソノミ) からなる階層構造を持つ。本稿は特に DBpedia や Freebase のボキャブラリや WordNet

のタクソノミを用いテンソル分解の精度を向上させる。DBpedia や Freebase は音楽や映画などのアイテムのエントリやそのプロパティを多く持つ。本稿では、アイテムエントリをインスタンスとみなし、アイテムエントリといくつかのプロパティ (例えば、ジャンル、監督など) により指し示されているエントリをクラスとみなし利用する。

WordNet は英語に対する語彙のデータベースであり、文書の自動分析を支援するためのツールとしてよく用いられる。WordNet においては、各単語は 1 つないしは複数の同義語の組 (synset) に分類される。本研究では各 synset をクラスとみなし、synset に分類された単語をインスタンスと考える。

ボキャブラリはインスタンスが複数のクラスを持ちうる一方、タクソノミはインスタンスは 1 つのクラスしか持たない。本稿で提案する手法は、その両方に適用可能であるため、LOD 上のセマンティクスをテンソル分解に取り込むに適切である。

### 3.2 Bayesian Probabilistic Tensor Factorization (BPTF)

まず、テンソル分解で用いる用語の定義を行い、次に BPTF について説明をする。

#### §1 用語の定義

本研究では、ユーザ  $u_m$ 、アイテム  $v_n$ 、タグ  $t_k$  から形成される複数オブジェクト関係を取り扱う。ユーザ数は  $M$ 、アイテム数は  $N$ 、タグ数は  $K$  である。三階テンソル  $\mathcal{R}$  はユーザの集合、アイテムの集合、タグの集合から構成される関係をモデル化し、その  $(m, n, k)$  要素  $r_{m, n, k}$  には、 $m$  番目のユーザが  $n$  番目のアイテムに  $k$  番目のタグを用いて与えた評点が入る。テンソル分解は  $D$  次元の属性ベクトルを各ユーザ、各アイテム、各タグに与え、それぞれを  $\mathbf{u}_m, \mathbf{v}_n, \mathbf{t}_k$  と表記する。結果として、 $\mathcal{R}$  は以下のように近似される。

$$r_{m, n, k} \approx \langle \mathbf{u}_m, \mathbf{v}_n, \mathbf{t}_k \rangle \equiv \sum_{i=1}^D u_{m, i} \cdot v_{n, i} \cdot t_{k, i}$$

ここで、インデックス  $i$  は各ベクトルの  $i$  番目の要素を示す。

#### §2 Bayesian Probabilistic Tensor Factorization (BPTF)

BPTF [Xiong et al.2010] はパラメータに対する事前分布に Gaussian/Wishart を用い評点に対する生成確率モデルを仮定し、Bayesian フレームワークをテンソル分解に適用している。 $\mathbf{u}_m, \mathbf{v}_n, \mathbf{t}_k$  の行列表現をそれぞれ  $\mathbf{U}, \mathbf{V}, \mathbf{T}$  と表記する。ここで  $\mathbf{U}$  のサイズは  $M \times D$  であり、 $\mathbf{V}$  のサイズは  $N \times D$  であり、 $\mathbf{T}$  のサイズは  $K \times D$  である。評点のランダム性を保証するために、BPTF は確率モデル  $\mathcal{R} | \mathbf{U}, \mathbf{V}, \mathbf{T} \sim \prod_{m=1}^M \prod_{n=1}^N \prod_{k=1}^K \mathcal{N}(\langle \mathbf{u}_m, \mathbf{v}_n, \mathbf{t}_k \rangle, \alpha^{-1})$  を用い評点を生成する。この式は  $\mathbf{U}, \mathbf{V}, \mathbf{T}$  が与えられたときの  $\mathcal{R}$  の条件付き分布は、 $(m, n, k)$  要素に対する平均

$\langle \mathbf{u}_m, \mathbf{v}_n, \mathbf{t}_k \rangle$ , 精度  $\alpha$  を持つガウス分布を, 全要素に対し掛け合わせたものから表されることを示す. BPTF の生成プロセスは以下である. ここで,  $\mu_0, \beta_0, \mathbf{W}_0, \nu_0, \tilde{\mathbf{W}}_0, \tilde{\nu}_0$  は BPTF のハイパーパラメータに必要とされるヒューリスティックなパラメータである.

- (1) ガウス分布に対する精度行列  $\Lambda_U, \Lambda_V, \Lambda_T$  を  $\mathcal{W}(\Lambda | \mathbf{W}_0, \nu_0)$  として生成する. ここで  $\mathcal{W}(\Lambda | \mathbf{W}_0, \nu_0)$  は  $D \times D$  のランダム行列  $\Lambda$  に関する分布であり, 自由度  $\nu_0$  および, サイズが  $D \times D$  のスケール行列  $\mathbf{W}_0$  を持つ Wishart 分布 [Bishop2006] であり,  $C$  を定数として以下のように定義される.

$$\mathcal{W}(\Lambda | \mathbf{W}_0, \nu_0) = \frac{|\Lambda|^{(\nu_0 - D - 1)/2}}{C} \exp\left(-\frac{\text{Tr}(\mathbf{W}_0^{-1} \Lambda)}{2}\right).$$

- (2) 平均行列  $\mu_U$  を  $\mu_U \sim \mathcal{N}(\mu_0, (\beta_0 \Lambda_U)^{-1})$  より生成する. 同様にして  $\mu_V, \mu_T$  を  $\mu_V \sim \mathcal{N}(\mu_0, (\beta_0 \Lambda_V)^{-1})$ ,  $\mu_T \sim \mathcal{N}(\mu_0, (\beta_0 \Lambda_T)^{-1})$  から生成する.
- (3)  $\alpha$  を  $\alpha \sim \mathcal{W}(\tilde{\mathbf{W}}_0, \tilde{\nu}_0)$  から生成する.
- (4) 各  $m \in (1 \dots M)$  に対し,  $\mathbf{u}_m$  を  $\mathbf{u}_m \sim \mathcal{N}(\mu_U, \Lambda_U^{-1})$  から生成する.
- (5) 各  $n \in (1 \dots N)$  に対し,  $\mathbf{v}_n$  を  $\mathbf{v}_n \sim \mathcal{N}(\mu_V, \Lambda_V^{-1})$  から生成する.
- (6) 各  $k \in (1 \dots K)$  に対し,  $\mathbf{t}_k$  を  $\mathbf{t}_k \sim \mathcal{N}(\mu_T, \Lambda_T^{-1})$  から生成する.
- (7) 各  $(m, n, k)$  に対し  $r_{m,n,k}$  を  $r_{m,n,k} \sim \mathcal{N}(\langle \mathbf{u}_m, \mathbf{v}_n, \mathbf{t}_k \rangle, \alpha^{-1})$  から生成し観測済みのエントリーとする.

BPTF はハイパーパラメータ  $\alpha, \Theta_U \equiv \{\mu_U, \Lambda_U\}, \Theta_V \equiv \{\mu_V, \Lambda_V\}, \Theta_T \equiv \{\mu_T, \Lambda_T\}$  をランダムな変数とみなし,  $\mathcal{R}$  に対する未観測の評点の予測分布  $\hat{\mathcal{R}}$  を以下のように計算する.

$$p(\hat{\mathcal{R}} | \mathcal{R}) = \int p(\hat{\mathcal{R}} | \mathbf{U}, \mathbf{V}, \mathbf{T}, \alpha) p(\mathbf{U}, \mathbf{V}, \mathbf{T}, \alpha, \Theta_U, \Theta_V, \Theta_T | \mathcal{R}) d\{\mathbf{U}, \mathbf{V}, \mathbf{T}, \alpha, \Theta_U, \Theta_V, \Theta_T\}. \quad (1)$$

BPTF は式(1)を, 事後分布  $p(\mathbf{U}, \mathbf{V}, \mathbf{T}, \alpha, \Theta_U, \Theta_V, \Theta_T | \mathcal{R})$  のうえで,  $p(\hat{\mathcal{R}} | \mathbf{U}, \mathbf{V}, \mathbf{T}, \alpha)$  の期待値とみなし, これを事後分布から獲得されるサンプルの期待値として近似する. 事後分布は直接サンプルを行うには複雑すぎるため, 間接的なサンプリング技術である Markov Chain Monte Carlo (MCMC) を用い, 具体的にはギブスサンプリングに基づき学習を行う.

BPTF における計算複雑性は  $\#nz$  を観測された複数オブジェクト関係の数とすると  $O(\#nz \times D^2 + (M + N + K) \times D^3)$  で与えられる. ここで,  $M, N, K$  は  $D$  よりも十分に大きい. また, 上式においてに第一項が第二項

よりも大きい. そのため, BPTF は一般的なテンソル分解の他の手法よりも計算時間が小さい (例えば, GCTF は [Ermis, Acar, and Cemgil2012] で述べられているように  $O(M \times N \times K \times D)$  の計算量を要求する). さらに BPTF はパラメータ設定が簡素な上, 属性ベクトルを並列に計算することもできる. これらの特性を踏まえ, 本研究では提案するアイデアを BPTF フレームワーク上に実装する.

## 4. 提案手法

本章では提案手法を詳細に説明する. まず 4.1 節で曖昧性解消について述べ, 4.2 節においてセマンティクスを用いたテンソル分解について述べる.

### 4.1 曖昧性解消

まず, アイテムに対する曖昧性解消から説明する.

#### §1 アイテムに対する曖昧性解消

コンテンツプロバイダはアイテムを管理するためにそれぞれに独自かつ簡素なボキャブラリを用いていることが多い. そうした簡素なボキャブラリは粒度の粗いセマンティクスの表現しかできていないため, 1 章の例でも述べたように予測精度をあまり改善しない. そのため, 本提案では, アイテムを Freebase や DBpedia の詳細なボキャブラリに関連付けした上で, Freebase や DBpedia の詳細なボキャブラリを利用する. 例えば, 我々の実験において, MovieLens のアイテムのボキャブラリは元々 19 クラスしか存在しないのに対し, Freebase を用いれば 724 の複合クラスを用意できる. これらの付加された語彙は提案手法の精度に大きく影響を与え, 重要である.

しかし, 簡素なボキャブラリしか持たないアイテムを, 詳細なボキャブラリに関係付けする際, 同じ名前のアイテムが複数存在する場合も多く, 曖昧性の解消を必要とする. ここで SRTF の曖昧性解消処理の手順を説明する.

(i) まずアイテム  $v_n$  のプロパティベクトル  $\mathbf{p}_{v_n}$  を作成する. このベクトルの  $x$  番目の要素は対応するプロパティ  $p_x$  の値に対応する. さらに, Freebase/DBpedia のインスタンス  $e_j$  に対するプロパティベクトル  $\mathbf{p}_{e_j}$  も作成する.  $\mathbf{p}_{e_j}$  は  $\mathbf{p}_{v_n}$  と同じプロパティ要素を持つとする. (ii) SRTF は次にプロパティベクトル  $\mathbf{p}_{v_n}$  と  $\mathbf{p}_{e_j}$  のコサイン類似度を計算し,  $v_n$  と最も類似度の高いインスタンス  $e_j$  にアイテム  $v_n$  を関連付け,  $e_j$  を  $v_{n'}$  とみなす. 同じ名前を持つアイテムは曖昧性解消プロセスの結果, 異なるアイテムインスタンスに関係付けされるため, 曖昧性解消後のアイテムインスタンスの数  $N'$  は, アイテムの数  $N$  とは異なり,  $N \leq N'$  となる. (iii) アイテムをインスタンス  $v_{n'}$  に関連付けした後, SRTF は Freebase/DBpedia のボキャブラリ (例えばジャンルボキャブラリ) を用い  $v_{n'}$  の所属クラスを識別する.

音楽アイテム  $v_n$  が, “名前” と “アルバム名の集合” というプロパティを持つとする. そして, 例えば,  $\mathbf{p}_{v_n}$  が

“ AIR ”, “ Moon Safari ”, “ Talkie Walkie ”, ... , “ The Virgin Suicides ” というベクトルであったとする。Freebase を参照すると、複数の “Air” というアーティストが存在するが、“Moon Safari”, “Talkie Walkie”, “The Virgin Suicides” を出している音楽アーティストは  $e_j$  のみしか存在しない。そのため、提案手法は、 $e_j$  が  $p_{v_n}$  に対し最も類似度が高いと計算し、 $e_j$  に割り当てられている “Electronica”, “Downtempo”, “Space rock”, “Psychedelic pop”, “Progressive rock” というジャンルを  $v_n$  のクラスとして利用する。

本稿では、上述の処理を関数  $f(v_n)$  により表す。 $v_n$  がインスタンス  $v_{n'}$  に曖昧性解消される場合は、 $f(v_n)$  は  $v_{n'}$  を持つクラス集合を返却する。それ以外の場合は、空集合  $\emptyset$  を返す。

## §2 タグレビューに対する曖昧性解消

ユーザはタグを自由にアイテムに与えるため、1章において図を用いて説明したように多くのタグは曖昧性を持つ。そのため、予測精度の向上を行うため、WordNet を用いタグの曖昧性を解消する。なお、本論文は SRTF を “ユーザ” と “アイテム” とユーザがアイテムに与える “レビュー” からなる複数オブジェクト関係にも適用する。この場合、レビューをレビュー内の主題を示す句 (主題タグ) に置き換える。その上で、“ユーザ” と “アイテム” とユーザがアイテムに与える “主題タグ” からなる複数オブジェクト関係を SRTF に入力として与える。これは、主題タグはユーザのアイテムに対する主張を反映しているからである [Yu et al.2011]。本研究では特に主題タグはセマンティクスに基づくマイニングの研究 [Nakatsuji, Yoshida, and Ishida2009] を用い抽出する。上記文献では、特定分野 (例えば、音楽) のレビューを解析し、その分野に用意されたボキャブラリ内のインスタンス (例えば、アーティスト) とレビュー記事内でマッチする句を主題タグとして曖昧性を回避しつつ抽出する。結果として、抽出されたタグはすでにボキャブラリに関係付けされている。

先行研究 [Cantador, Konstas, and Jose2011, Nakatsuji and Fujiwara2014] において、アイテムの内容を表現するタグ (内容タグ) とアイテムに対するユーザの主観を表現するタグ (主観タグ) が予測精度の改善に有効と分析している。そのため、本稿でもまずタグをアイテムの内容を表現するものと、アイテムに対するユーザの主観を表現するものとに分類する。内容タグは一般的に名詞から構成されるため [Cantador, Konstas, and Jose2011], SRTF はタグより名詞から構成される句を抽出する。まず SRTF は接続詞などの不要な単語を除去し、句を品詞の組とみなし、内容タグに特有な品詞の組 ( $[\langle \text{noun} \rangle]$ ,  $[\langle \text{adjective} \rangle \langle \text{noun} \rangle]$ ,  $[\langle \text{determiner} \rangle \langle \text{noun} \rangle]$ ) と比較をすることで、内容タグに分類する。同じく主観タグに関しても、主観タグに特有の品詞の組 ( $[\langle \text{adjective} \rangle]$ ,  $[\langle \text{adjective} \rangle \langle \text{noun} \rangle]$ ,  $[\langle \text{adverb} \rangle]$ ,  $[\langle \text{adverb} \rangle \langle \text{adjective} \rangle]$ ,  $[\langle * \text{pronoun} \rangle \langle \text{adjective} \rangle]$ ) と比較をすることで、主観タグに分類する。なお、

SRTF は Stanford-parser [Klein and Manning2003] を用い “Not good” など、上記のパターンに対する否定形の識別も行う。

次に SRTF は WordNet タクソノミにタグを関係付けする。タグ  $t_k$  はアイテム  $v_n$  の特徴を反映するため、 $t_k$  をタクソノミに関係付けする際、アイテム  $v_n$  に与えられている記述 (解説記事) を分析する。SRTF は VSM [Turney and Pantel2010] に基づく意味的な類似度計測手法 [Zheng et al.2013] を曖昧性解消に以下の手順で適用する。(i) タグ  $t_k$  に現れる単語  $w$  と同名の WordNet インスタンスの記述をまず収集する。同名の WordNet インスタンスは複数存在し得る。具体例としては、“breathless” という単語に対応する WordNet インスタンスの記述は、以下の2つが存在する。

- breathless, dyspneic, dyspnoeic, dyspneal, dyspnoeal (not breathing or able to breathe except with difficulty) “breathless at thought of what I had done”; “breathless from running”; “followed the match with breathless interest”.
- breathless, breathtaking (tending to cause suspension of regular breathing) “a breathless flight”; “breathtaking adventure”.

(ii) 記述  $d_j$  から不要な単語を除去した上で、要素が  $d_j$  における単語、値が対応する単語の観測回数を表すベクトル  $w_j$  を構成する。(iii) 次に、アイテムインスタンス  $v_{n'}$  の記述 (DBpedia/Freebase におけるアイテムに対する解説記事) と  $v_{n'}$  のジャンルの記述 (DBpedia/Freebase におけるジャンルに対する解説記事) を DBpedia/Freebase から収集する。そして、要素が単語であり、値が記述内の単語の出現回数となるベクトル  $i_{n'}$  を構成する。ベクトル  $i_{n'}$  は  $v_{n'}$  とそのジャンルの特徴を示す。(iv) 最後に、 $i_{n'}$  と  $w_j$  のコサイン類似度を計算し、 $t_k$  と最も類似する wordNet インスタンス  $w_j$  に  $t_k$  を関係付けし、 $w_j$  を  $t_{k'}$  とみなす。同じ名前を持つタグは曖昧性解消プロセスの結果、異なるタグインスタンスに関係付けされ得るため、曖昧性解消後のタグインスタンスの数  $K'$  は、タグの数  $K$  と異なり、 $K \leq K'$  となる。

本稿では、上述の処理を関数  $g(v_n, t_k)$  により表す。 $t_k$  がインスタンス  $t_{k'}$  に曖昧性解消される場合は、 $g(v_n, t_k)$  は WordNet インスタンス  $t_{k'}$  を持つクラスを返却する。それ以外の場合は、空集合  $\emptyset$  を返す。

## §3 曖昧性解消テンソルの構築

SRTF は曖昧性解消テンソル  $\mathcal{R}^d$  を以下の手順で作成する。(i) 最初に  $\mathcal{R}$  における各観測エントリ  $r_{m,n,k}$  を取り出す。そして、各観測におけるアイテム  $v_n$  (またはタグ  $t_k$ ) をボキャブラリ/タクソノミ内のインスタンス  $v_{n'}$  ( $t_{k'}$ ) へ関係付けすることにより曖昧性解消を行う。(ii)  $\mathcal{R}$  における観測  $r_{m,n,k}$  を  $\mathcal{R}^d$  内の観測  $r_{m,j,l}^d$  に以下のよう

に置き換える．

$$r_{m,j,l}^d = \begin{cases} r_{m,n,k} & ((f(v_n) = \emptyset)) \\ r_{m,n',k} & ((f(v_n) \neq \emptyset) \cap (g(v_n, t_k) = \emptyset)) \\ r_{m,n',k'} & ((f(v_n) \neq \emptyset) \cap (g(v_n, t_k) \neq \emptyset)) \end{cases}$$

(iii) ステップ (i) から (ii) を  $\mathcal{R}$  内のエントリ  $r_{m,n,k}$  を変えることで繰り返す．結果として，曖昧性解消テンソル  $\mathcal{R}^d$  を獲得できる．

この処理は  $\mathcal{R}$  をセマンティクスへ結びつける．結果として，次に説明するセマンティクスを用いたテンソル分解処理を実施できる．

#### 4.2 セマンティクスを用いたテンソル分解

本処理では，希薄なインスタンスを所属クラスに引き上げ  $\mathcal{R}^d$  に積上げることで増大テンソルを作成する．これにより，テンソル分解時にクラスに共有されている知識を分析することを可能とする．結果として，希薄問題を解決できる．まず，増大テンソルの作成について説明する．

##### §1 増大テンソルの作成

まず，希薄なアイテムインスタンスの集合  $\mathbb{V}_s$  を定義する．これは，全てのアイテムの中で相対的に希薄に観測されているアイテム  $v_s$  の集合である．テンソル  $\mathcal{O}^d$  の要素  $o_{m,n',k'}^d$  は，曖昧性解消テンソル  $\mathcal{R}^d$  におけるユーザ  $u_m$ ，アイテム  $v_{n'}$ ，タグ  $t_{k'}$  により構成される複数オブジェクト関係の観測が存在する場合 1 を要素として持ち，存在しない場合は 0 を要素として持つ．そうすると， $\mathbb{V}_s$  は以下の式で計算される．

$$\mathbb{V}_s = \{v_s : (\sum_{r_{m,s,k'} \in \mathcal{R}^d} o_{m,s,k'}^d / \sum_{r_{m,n',k'} \in \mathcal{R}^d} o_{m,n',k'}^d) < \delta\}. \quad (2)$$

ここで， $\delta$  は  $\mathbb{V}_s$  の中の希薄インスタンスの数を決定するために使われるパラメータである．ロングテール理論 [Anderson2006] に基づき一般的に  $\delta$  は 0.1 から 0.3 に設定できる．なお，希薄インスタンスの所属クラスを  $s_s^v$  とし，希薄インスタンスの所属クラスの数  $S^V = |\bigcup_{v_s} f(v_s)|$  と表記する．SRTF は“ユーザ”，“希薄アイテムインスタンスの所属する複合クラス”，“タグインスタンス”から構成される複数オブジェクト関係を  $\mathcal{R}^d$  に挿入することで増大テンソルを以下のように構築する．

$$r_{m,j,k'}^v = \begin{cases} r_{m,j,k'}^d & (j \leq N') \\ r_{m,s,k'}^d & (N' < j \leq (N' + S^V)) \cap (s_{(j-N')}^v \in f(v_s)) \end{cases} \quad (3)$$

タグに対する意味的な増大テンソル  $\mathcal{R}^t$  も  $\mathcal{R}^v$  と同様にして構築できる．

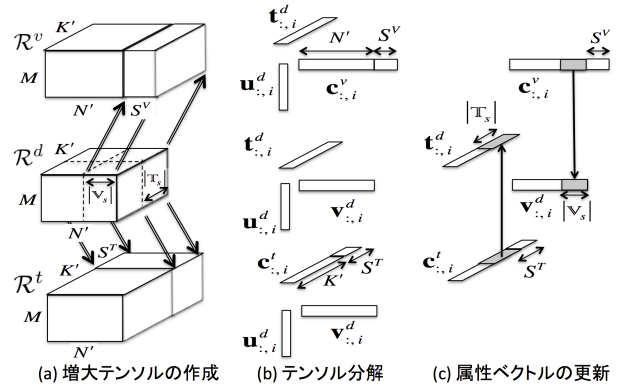


図3 SRTFにおける分解処理の例

図3-(a)において，SRTFは希薄アイテム集合  $\mathbb{V}_s$  のアイテムをそれらの所属クラス（サイズは  $S^V$  である）に持ち上げ，曖昧性解消テンソル  $\mathcal{R}^d$  から増大テンソル  $\mathcal{R}^v$  を作成する．SRTFは同じく希薄タグ集合  $\mathbb{T}_s$  内のタグをそれらの所属クラス（サイズは  $S^T$  である）に持ち上げ，曖昧性解消テンソル  $\mathcal{R}^d$  から増大テンソル  $\mathcal{R}^t$  を作成する．

##### §2 複合クラスを用いたセマンティクスの詳細化

SRTFは，より詳細なセマンティクスをテンソル分解に組込むため，アイテムインスタンスに対し割り当てられている“プロパティとその値”の組合せの積集合として定義される複合クラス [Parundekar, Knoblock, and Ambite2012] を作成する．

プロパティ  $p_x$  により制約を受けているクラスを  $c^x$  と定義する．そして，クラスに所属するインスタンスにより  $X$  種類のプロパティを持つ複合クラスを定義する．複合クラス  $c'$  におけるアイテムインスタンスの集合は  $\{v_j : v_j \in \bigcap_{1 \leq x \leq X} \{v_{n'} \in c^x\}\}$  として計算できる．

そして，関数  $f'(v_n)$  を，もし  $v_n$  が  $v_{n'}$  に関係付けできる場合  $v_{n'}$  を持つ混合クラスの集合を返し，それ以外の場合，空集合  $\emptyset$  を返すものとして定義する．SRTFは式(3)における  $f(v_n)$  の代わりに  $f'(v_n)$  を用い，“ユーザ”，“希薄アイテムインスタンスの所属する複合クラス”，“タグ”から形成される複数オブジェクト関係を  $\mathcal{R}^v$  に追加する．

結果として，SRTFはより粒度の細かいセマンティクスを用いてテンソル分解における複数オブジェクト関係を分析することができる．

##### §3 セマンティクスを用いたテンソル分解

SRTFは曖昧性解消テンソル  $\mathcal{R}^d$  と増大テンソル  $\mathcal{R}^v$ ， $\mathcal{R}^t$  をBPTFフレームワーク上で同時に分解することにより予測値を計算する．図3を用いて説明する．図においてはテンソルを行列  $\mathbf{U}^d$ ， $\mathbf{V}^d$ ， $\mathbf{T}^d$  の行ベクトルに分解している． $\mathbf{U}^d$ ， $\mathbf{V}^d$ ， $\mathbf{T}^d$  の  $i$  番目の行ベクトルは， $\mathbf{u}_{:,i}^d$ ， $\mathbf{v}_{:,i}^d$ ， $\mathbf{t}_{:,i}^d$  ( $1 \leq i \leq D$ ) と表記する．SRTFは，もし  $v_{n'}$  と  $t_{k'}$  が希薄であれば，それらの所属クラスに対する属性ベ

クトル  $\mathbf{c}_j^v$  と  $\mathbf{c}_j^t$  で共有される知識を属性ベクトル  $\mathbf{v}_{n'}$  と  $\mathbf{t}_{k'}$  に与える．同時分解をすることにより，SRTF はセマンティクスを上記共有パラメータ（例えば  $\mathbf{v}_{n'}$  や  $\mathbf{t}_{k'}$ ）を通じ3つのテンソルの分解において共有することができる．結果として，テンソル分解時に分散して存在する希薄オブジェクト間で効果的にセマンティクスを共有することができ，希薄問題を解決することができる．

### i. アプローチ

まず提案する分解手法のアプローチを説明する．

- (A) SRTF はテンソル  $\mathcal{R}^d$ ,  $\mathcal{R}^v$ ,  $\mathcal{R}^t$  を同時に分解する．そして，属性ベクトル  $\mathbf{v}_{n'}^d$  と  $\mathbf{t}_{k'}^d$  を  $\mathcal{R}^d$  を分解することにより計算するとともに，セマンティックバイアスを持つ属性ベクトル  $\mathbf{c}_j^v$  と  $\mathbf{c}_j^t$  を  $\mathcal{R}^v$  と  $\mathcal{R}^t$  を分解することにより計算する．図 3-(b) はテンソル分解により得られた  $i$  番目の行ベクトルを示している．
- (B) SRTF は曖昧性解消テンソル  $\mathcal{R}^d$  を分解することで得られる精度  $\alpha$ ，属性ベクトル  $\mathbf{u}_m^d$ ,  $\mathbf{v}_{n'}^d$ ,  $\mathbf{t}_{k'}^d$ ，そしてそれらのハイパーパラメータを，増大テンソル  $\mathcal{R}^v$  と  $\mathcal{R}^t$  の分解においても共有する．それゆえ，それらの共有されたパラメータを通じ3つのテンソル間でセマンティックバイアスを共有することができる．図 3-(b) において， $\mathbf{u}_{:,i}^d$  と  $\mathbf{t}_{:,i}^d$  は増大テンソル  $\mathcal{R}^v$  を分解する際に共有されている．同様に， $\mathbf{u}_{:,i}^d$  と  $\mathbf{v}_{:,i}^d$  は，増大テンソル  $\mathcal{R}^t$  を分解する際に共有されている．
- (C) SRTF は  $\mathbf{c}_{n'}^v$  由来のセマンティックバイアスを  $\mathbf{v}_{n'}^d$  に入れ籠むことにより，希薄問題を解決する．図 3-(c) においては，各行ベクトル  $\mathbf{c}_{:,i}^v$  が  $N'$  個のアイテムと  $S^V$  個のクラスに対する属性値を持つ． $\mathbf{c}_{:,i}^v$  における属性は希薄アイテムのセマンティクスを共有することで得られたものであるため，希薄アイテムに起因する希薄問題を解決するために有用である．そのため，SRTF はもし  $\mathbf{v}_{n'}^d$  が希薄アイテムインスタンスであれば  $\mathbf{v}_{n'}^d$  を  $\mathbf{c}_{n'}^v$  に置き換える．同様に，SRTF は， $\mathbf{t}_{k'}^d$  が希薄なタグインスタンスであれば， $\mathbf{c}_{k'}^t$  由来のセマンティックバイアスを  $\mathbf{t}_{k'}^d$  に入れ込むことにより希薄問題を解決する．

なお，本研究ではアイテムインスタンスが希薄であるかそうでないかを識別するベクトル  $\mathbf{z}^v$  を導入する． $\mathbf{z}^v$  は長さ  $N'$  のベクトルであり，以下の式の通り，各要素は  $\mathbf{v}_{n'}^d$  が希薄アイテムインスタンスであれば 0，そうでなければ 1 を持つとする．

$$\mathbf{z}_{n'}^v = \begin{cases} 0 & (\mathbf{v}_{n'} \in \mathbb{V}_s) \\ 1 & (\text{otherwise}) \end{cases}$$

同様に，タグインスタンスが希薄であるかそうでないかを識別するベクトル  $\mathbf{z}^t$  を以下の式のように，長さ  $K'$  のベクトルであり各要素は， $\mathbf{t}_{k'}^d$  が希薄タグインスタンスであれば 0，そうでなければ 1 を持つと

する．

$$\mathbf{z}_{k'}^t = \begin{cases} 0 & (t_{k'} \in \mathbb{T}_s) \\ 1 & (\text{otherwise}) \end{cases}$$

これらのベクトルは，式 (2) を使い，生成プロセスの前に一意に決定済みである．

### ii. 生成プロセス

SRTF の生成プロセスは以下の通りである．ここで， $\mu_0$ ,  $\beta_0$ ,  $\mathbf{W}_0$ ,  $\nu_0$ ,  $\tilde{\mathbf{W}}_0$ ,  $\tilde{\nu}_0$  は BPTF のハイパーパラメータに必要とされるヒューリスティックなパラメータである． $\mathbf{z}^v$  と  $\mathbf{z}^t$  は前述した通り，オブジェクトが希薄であるかそうでないかを識別するベクトルである．

- (1)  $\Lambda_U$ ,  $\Lambda_V$ ,  $\Lambda_T$ ,  $\Lambda_{C^v}$ ,  $\Lambda_{C^t}$  をガウス分布に基づく精度行列として  $\mathcal{W}(\Lambda|\mathbf{W}_0, \nu_0)$  より生成する．ここで  $\mathcal{W}(\Lambda|\mathbf{W}_0, \nu_0)$  は  $D \times D$  のランダム行列  $\Lambda$  に関する分布であり，自由度  $\nu_0$  および，サイズが  $D \times D$  のスケール行列  $\mathbf{W}_0$  を持つ Wishart 分布である．
- (2) 平均行列  $\mu_U$  を  $\mu_U \sim \mathcal{N}(\mu_0, (\beta_0 \Lambda_U)^{-1})$  から生成する．同様に， $\mu_V$ ,  $\mu_T$ ,  $\mu_{C^v}$ ,  $\mu_{C^t}$  を  $\mu_V \sim \mathcal{N}(\mu_0, (\beta_0 \Lambda_V)^{-1})$ ,  $\mu_T \sim \mathcal{N}(\mu_0, (\beta_0 \Lambda_T)^{-1})$ ,  $\mu_{C^v} \sim \mathcal{N}(\mu_0, (\beta_0 \Lambda_{C^v})^{-1})$ ,  $\mu_{C^t} \sim \mathcal{N}(\mu_0, (\beta_0 \Lambda_{C^t})^{-1})$  から生成する．
- (3)  $\alpha \sim \mathcal{W}(\tilde{\mathbf{W}}_0, \tilde{\nu}_0)$  を生成する．
- (4) 各  $m \in (1 \dots M)$  に対して， $\mathbf{u}_m$  を以下の式より生成する． $\mathbf{u}_m \sim \mathcal{N}(\mu_U, \Lambda_U^{-1})$ .
- (5) 各  $n' \in (1 \dots N')$  に対して， $\mathbf{v}_{n'}$  を以下の式から生成する． $\mathbf{v}_{n'} \sim \mathcal{N}(\mu_V, \Lambda_V^{-1})$ .
- (6) 各  $k' \in (1 \dots K')$  に対して， $\mathbf{t}_{k'}$  を以下の式から生成する． $\mathbf{t}_{k'} \sim \mathcal{N}(\mu_T, \Lambda_T^{-1})$ .
- (7) 各  $j \in (1, \dots, (N+S^v))$  に対して  $\mathbf{c}_j^v$  を以下の式から生成する． $\mathbf{c}_j^v \sim \mathcal{N}(\mu_{C^v}, \Lambda_{C^v}^{-1})$ .
- (8) 各  $j \in (1, \dots, (K+S^t))$  に対して  $\mathbf{c}_j^t$  を以下の式から生成する． $\mathbf{c}_j^t \sim \mathcal{N}(\mu_{C^t}, \Lambda_{C^t}^{-1})$ .
- (9) 各  $(m, n', k')$  に対し  $r_{m,n',k'}^d$  を以下の式から生成し観測済みのエントリとする．

$$r_{m,n',k'}^d \sim \mathcal{N}\left(\left(\sum_{i=1}^D u_{m,i} \cdot (\mathbf{v}_{n',i}^{z_{n'}^v} \cdot \mathbf{c}_{n',i}^v)^{(1-z_{n'}^v)}\right) \cdot \left(\mathbf{t}_{k',i}^{z_{k'}^t} \cdot \mathbf{c}_{k',i}^t\right)^{(1-z_{k'}^t)}\right), \alpha^{-1}$$

なお，手順 (9) において式  $(\mathbf{v}_{n',i}^{z_{n'}^v} \cdot \mathbf{c}_{n',i}^v)^{(1-z_{n'}^v)}$  は， $\mathbf{v}_{n'}$  が希薄なアイテムインスタンスである場合は，アイテムの属性ベクトルは  $\mathbf{c}_{n'}^v$  のみから決定され，希薄でない場合は  $\mathbf{v}_{n'}$  のみから決定されることを示す．これにより，希薄なアイテムインスタンスに対しては，クラス経由のセマンティックバイアスを属性ベクトルに反映できる．式  $(\mathbf{t}_{k',i}^{z_{k'}^t} \cdot \mathbf{c}_{k',i}^t)^{(1-z_{k'}^t)}$  についても同様の考え方で設計されている．

## iii. MCMC による学習と予測値の計算

SRTF は, BPTF と同様にして MCMC 処理により未観測の評点  $\hat{\mathcal{R}}^d$  に対する予測分布を以下のように計算する .

$$p(\hat{\mathcal{R}}^d | \mathcal{R}^d, \mathcal{R}^v, \mathcal{R}^t, \mathbf{z}^v, \mathbf{z}^t) = \int p(\hat{\mathcal{R}}_{m,n',k'}^d | \mathbf{U}, \mathbf{V}^d, \mathbf{T}^d, \mathbf{C}^v, \mathbf{C}^t, \alpha, \mathbf{z}^v, \mathbf{z}^t) p(\mathbf{U}, \mathbf{V}^d, \mathbf{T}^d, \mathbf{C}^v, \mathbf{C}^t, \alpha, \theta_{\mathbf{U}}, \theta_{\mathbf{V}^d}, \theta_{\mathbf{T}^d}, \theta_{\mathbf{C}^v}, \theta_{\mathbf{C}^t} | \mathcal{R}, \mathcal{R}^v, \mathcal{R}^t, \mathbf{z}^v, \mathbf{z}^t) d\{\mathbf{U}, \mathbf{V}^d, \mathbf{T}^d, \mathbf{C}^v, \mathbf{C}^t, \alpha, \theta_{\mathbf{U}}, \theta_{\mathbf{V}^d}, \theta_{\mathbf{T}^d}, \theta_{\mathbf{C}^v}, \theta_{\mathbf{C}^t}\}.$$

BPTF の詳細な MCMC 処理は [Xiong et al.2010] を参照されたい . MCMC 処理は  $L$  個のサンプルを収集し , 以下の式により未観測の予測値を近似する .

$$p(\hat{\mathcal{R}}^d | \mathcal{R}^d, \mathcal{R}^v, \mathcal{R}^t, \mathbf{z}^v, \mathbf{z}^t) \approx \sum_{l=1}^L p(\hat{\mathcal{R}}^d | \mathbf{U}^d[l], \mathbf{V}^d[l], \mathbf{T}^d[l], \mathbf{C}^v[l], \mathbf{C}^t[l], \alpha[l], \mathbf{z}^v, \mathbf{z}^t) \quad (4)$$

4.2.3 節 i. 項で説明したアプローチを実装するため , SRTF は BPTF のギブスサンプリング手順を改善し , 以下の手順で処理を行う ( 詳細については , 付録も併せて参照されたい ) .

- (1)  $\mathbf{U}^d[1], \mathbf{V}^d[1], \mathbf{T}^d[1], \mathbf{C}^v[1], \mathbf{C}^t[1]$  をガウス分布により初期化する .  $\mathbf{X}$  は属性ベクトル  $\mathbf{x}$  の行列表現である ( 例えば ,  $\mathbf{C}^v$  は  $\mathbf{c}_j^v$  の行列表現である ) (  $\mathbf{C}^v$  と  $\mathbf{C}^t$  は , アプローチ (A) を実装するために必要である . )

次に , ステップ (2) から (6) を  $L$  回繰り返す .

- (2) BPTF と同様にして , ハイパーパラメータを以下のようにサンプルする :

$$\begin{aligned} \circ \alpha[l] &\sim p(\alpha[l] | \mathbf{U}^d[l], \mathbf{V}^d[l], \mathbf{T}^d[l], \mathcal{R}^d) \\ \circ \theta_{\mathbf{U}^d}[l] &\sim p(\theta_{\mathbf{U}^d}[l] | \mathbf{U}^d[l]) \\ \circ \theta_{\mathbf{V}^d}[l] &\sim p(\theta_{\mathbf{V}^d}[l] | \mathbf{V}^d[l]) \\ \circ \theta_{\mathbf{T}^d}[l] &\sim p(\theta_{\mathbf{T}^d}[l] | \mathbf{T}^d[l]) \\ \circ \theta_{\mathbf{C}^v}[l] &\sim p(\theta_{\mathbf{C}^v}[l] | \mathbf{C}^v[l]) \\ \circ \theta_{\mathbf{C}^t}[l] &\sim p(\theta_{\mathbf{C}^t}[l] | \mathbf{C}^t[l]) \end{aligned}$$

ここで ,  $\theta_{\mathbf{X}}$  は  $\{\mu_{\mathbf{X}}, \Lambda_{\mathbf{X}}\}$  を表す .  $\mu_{\mathbf{X}}$  と  $\Lambda_{\mathbf{X}}$  は BPTF と同様にして計算される ( 3 章を参照 ) .

- (3) BPTF と同様にして属性ベクトルをサンプルする .

$$\begin{aligned} \circ \mathbf{u}_m^d[l+1] &\sim p(\mathbf{u}_m^d | \mathbf{V}^d[l], \mathbf{T}^d[l], \alpha[l], \theta_{\mathbf{U}^d}[l], \mathcal{R}^d) \\ \circ \mathbf{v}_{n'}^d[l+1] &\sim p(\mathbf{v}_{n'}^d | \mathbf{U}^d[l+1], \mathbf{T}^d[l], \alpha[l], \theta_{\mathbf{V}^d}[l], \mathcal{R}^d) \\ \circ \mathbf{t}_{k'}^d[l+1] &\sim p(\mathbf{t}_{k'}^d | \mathbf{U}^d[l+1], \mathbf{V}^d[l+1], \alpha[l], \theta_{\mathbf{T}^d}[l], \mathcal{R}^d) \end{aligned}$$

- (4) セマンティックバイアスをかけた属性ベクトル  $\mathbf{c}_{n'}^v[l+1]$  と  $\mathbf{c}_{k'}^t[l+1]$  を  $\alpha[l], \mathbf{U}^d[l+1], \mathbf{V}^d[l+1], \mathbf{T}^d[l+1]$  を用い以下のようにサンプルする .

$$\begin{aligned} \circ \mathbf{c}_{n'}^v[l+1] &\sim p(\mathbf{c}_{n'}^v | \mathbf{U}^d[l+1], \mathbf{T}^d[l+1], \alpha[l], \theta_{\mathbf{C}^v}[l], \mathcal{R}^v) \\ \circ \mathbf{c}_{k'}^t[l+1] &\sim p(\mathbf{c}_{k'}^t | \mathbf{U}^d[l+1], \mathbf{V}^d[l+1], \alpha[l], \theta_{\mathbf{C}^t}[l], \mathcal{R}^t) \end{aligned}$$

(ハイパーパラメータ  $\alpha[l]$  , パラメータ  $\mathbf{U}^d[l+1], \mathbf{V}^d[l+1], \mathbf{T}^d[l+1]$  は , アプローチ (B) を満たすためステップ (3) と (4) において共有されている .)

- (5) 未観測の評点  $\hat{\mathcal{R}}^d[l]$  を ,  $\mathbf{U}^d[l+1], \mathbf{V}^d[l+1], \mathbf{T}^d[l+1], \mathbf{C}^v[l+1], \mathbf{C}^t[l+1], \alpha[l]$  を式 (4) に適用することでサンプルする .

- (6) もし  $v_{n'}$  が希薄なアイテムインスタンスであれば ,  $\mathbf{v}_{n'}^d[l+1]$  を  $\mathbf{c}_{n'}^v[l+1]$  に置き換える . もし希薄なアイテムインスタンスでなければ ,  $\mathbf{v}_{n'}^d[l+1]$  を  $\mathbf{c}_{n'}^v[l+1]$  に置き換えない . 本処理は以下の式で表される .

$$\mathbf{v}_{n'}^d[l+1] = \begin{cases} \mathbf{c}_{n'}^v[l+1] & (z_{n'}^v = 0) \\ \mathbf{v}_{n'}^d[l+1] & (z_{n'}^v = 1) \end{cases}$$

同様に , もし  $t_{k'}$  が希薄なタグインスタンスであれば ,  $\mathbf{t}_{k'}^d[l+1]$  を  $\mathbf{c}_{k'}^t[l+1]$  に置き換える . もし希薄なタグインスタンスでなければ ,  $\mathbf{t}_{k'}^d[l+1]$  を  $\mathbf{c}_{k'}^t[l+1]$  に置き換えない . 本処理は以下の式で表される .

$$\mathbf{t}_{k'}^d[l+1] = \begin{cases} \mathbf{c}_{k'}^t[l+1] & (z_{k'}^t = 0) \\ \mathbf{t}_{k'}^d[l+1] & (z_{k'}^t = 1) \end{cases}$$

なお , 本ステップは , SRTF の生成プロセスの手順 (9)(4.2.3 節 ii. 項参照) に対応する .

我々の MCMC 処理は  $\mathbf{C}^v$  と  $\mathbf{C}^t$  等のパラメータを追加しつつ , ステップ (4) と (6) を BPTF の MCMC 処理に追加するというシンプルなものであるため , 読者は我々のアイデアを自身の研究を進めるために容易に実装することができる .

最後に , 曖昧性が解消される前の元のテンソルにおける未観測の評点  $\hat{r}_{m,n,k}$  に対する予測値を  $v_n$  と  $v_{n'}$  間のリンク関係と ,  $t_k$  と  $t_{k'}$  間のリンク関係を確認することにより ,  $\hat{r}_{m,n',k'}$  から計算をする .

SRTF の計算複雑性はサンプリングの各繰り返しにおいて ,  $O(\#nz \times D^2 + (M+N'+K'+S^V+S^T) \times D^3)$  となる . 3 章で述べた通り , 第一項が第二項に比べ計算複雑性が大きい . 計算複雑性は BPTF の場合とほぼ同程度となる . ただし , パラメータ  $\delta$  はロングテール理論に基づき容易に設定できる . さらに , BPTF フレームワークは Bayesian に基づく処理を行っているため ,  $\mu_0, \beta_0, \mathbf{W}_0, \nu_0, \tilde{\mathbf{W}}_0, \tilde{\nu}_0$  などの各種パラメータを変更しても , 文献 [Xiong et al.2010] では結果があまり変わらないと述べられている .

## 5. 実 験

本実験では , SRTF をユーザの行動予測に適用し評価を行う . 具体的には , ユーザ , アイテム , タグから構成される 3 組に対し与えられている評点 ( または , 消費頻度 ) を用い , 将来起こりうるユーザ , アイテム , タグの 3 組とそれに対応する評点 ( または , 消費頻度 ) を予測するというタスクに対し評価を実施した . なお , 訓練デー

表 1 MovieLens に対するタグクラスの例

$C$	Breathtaking	Historical	Spirited	Unrealistic
$t_k$	Breathtaking	Ancient	Racy	Kafkaesque
	Exciting	Historic	Spirited	Surreal
	Thrilling	Past life	Vibrant	Surreal life

表 2 MovieLens に対する RMSE の平均と分散 ( $\times 10^{-5}$ )

$D$		BPMF	BPTF	SRTF $\alpha$	SRTF $\beta$	SRTF(1.0)	SRTF(0.3)
$D=25$	平均	0.9062	1.0278	0.8842	0.8838	0.8850	0.8807
	分散	8.1712	15.3502	4.7174	5.7223	5.4754	5.9744
$D=50$	平均	0.9043	0.9988	0.8827	0.8818	0.8837	0.8791
	分散	9.1206	39.4048	5.9430	4.6317	5.3072	6.8149
$D=75$	平均	0.9040	0.9863	0.8821	0.8809	0.8829	0.8778
	分散	6.9771	37.2929	5.0868	5.7490	5.5913	6.4982
$D=100$	平均	0.9032	0.9708	0.8823	0.8792	0.8824	<b>0.8774</b>
	分散	7.5844	14.5400	5.3995	5.3136	6.0363	6.1852

タセットにまだ出現していないユーザ、アイテム、タグに対する予測はできないが、訓練データセットに登場したユーザ、アイテム、タグ集合の全ての組合せに対し評点を予測できる。新たに登場したオブジェクトを踏まえた予測はできないが、訓練データセットに膨大な数のオブジェクトがある場合、それを踏まえた予測は十分に実アプリケーションに適用できる問題設定である。

### 5.1 データセット

手法の精度を以下の 3 つのデータセットを用い確認した。

**MovieLens** このデータセットはユーザが映画アイテムに対して与えた評点とタグが含まれている。評点の範囲は 0.5 から 5 である。本データセットに対し、我々は Freebase から集めたジャンル、監督、役者の 3 つのアイテムボキャブラリを用い複合クラスを作成した。また、WordNet から集めたタグタクソノミを用いた。混合クラスを作成するにあたり、各アイテムインスタンスについて、1 つ以下の監督と 2 つ以下の役者のみが割り当てられるという制約をおいた。結果として、724 の複合クラスを持つアイテムボキャブラリと、4,284 のクラスを持つタグタクソノミが得られた。さらに本実験ではクラスに割り当てることがなかったアイテムやタグも実験に利用した。結果として、本データセットは 24,565 の評点と、それに対し 44,595 のタグ (内 33,547 タグはタグクラスを持つ) を持つものとなった。ユーザ、アイテム、タグからなるテンソルのサイズは、 $2,026 \times 5,088 \times 9,160$  となった。表 1 はクラス  $C$  とタグインスタンス  $t_k$  の例を示す。

**Yelp** このデータセットはユーザがレストランに対して与えた評点とレビュー文が含まれている。評点の範囲は 1 から 5 である。実験では Yelp のジャンルをアイテム (レストラン) のボキャブラリとして用いた。アイテムボキャブラリは 179 のクラスからなる。また、本データセットに対しては、ジャンルというプロパティ以外にレストランの内容を表現するのに適切なプロパティがなかったため、複合クラスを作成していない。その上で、[Nakatsuji, Yoshida, and Ishida 2009] らの方法に従い、DBPedia<sup>\*5</sup> の食品に対するボキャブラリを用いレビュー文を解析した。

表 3 Yelp に対する RMSE の平均と分散 ( $\times 10^{-4}$ )

$D$		BPMF	BPTF	SRTF $\alpha$	SRTF $\beta$	SRTF(1.0)	SRTF(0.3)
$D=25$	平均	1.1161	1.2222	-	-	1.2240	1.0858
	分散	9.5985	1.0143	-	-	1.5127	1.4318
$D=50$	平均	1.1154	1.1643	-	-	1.1613	1.0860
	分散	10.1788	0.9400	-	-	1.4273	1.2976
$D=75$	平均	1.1147	1.1620	-	-	1.1627	1.0856
	分散	10.4442	1.2118	-	-	1.6251	1.3544
$D=100$	平均	1.1127	1.1533	-	-	1.1621	<b>1.0851</b>
	分散	10.0000	0.2688	-	-	1.8292	1.3444

表 4 Last.fm に対する RMSE の平均と分散 ( $\times 10^{-5}$ )

$D$		BPMF	BPTF	SRTF $\alpha$	SRTF $\beta$	SRTF(1.0)	SRTF(0.3)
$D=25$	平均	-	1.2579	1.1629	-	1.2203	1.1498
	分散	-	1.7945	1.0921	-	1.1842	0.9823
$D=50$	平均	-	1.2507	1.1621	-	1.2175	<b>1.1402</b>
	分散	-	1.9345	1.0728	-	1.1387	0.9932
$D=75$	平均	-	1.2652	1.1641	-	1.2180	1.1429
	分散	-	1.7492	0.9982	-	1.1212	1.0029
$D=100$	平均	-	1.2578	1.1721	-	1.2182	1.1482
	分散	-	2.2013	1.1923	-	1.2934	1.0392

そのため、本データセットに対しては、曖昧性解消処理は不要である。DBPedia の食品ボキャブラリでは、例えば “Ramen” というインスタンスが “Japanese cousin” というクラスに分類されている。本研究では、食品インスタンスと同一の句がレビュー文に記載されている場合、そのレビュー文は当該食品インスタンスについて評価をしていると考え、食品タグとみなし抽出をした。句のマッチングの際、インスタンスの語長が長いものとのマッチングを優先した。結果、2,038,560 の食品句をレビュー文より抽出した。食品のボキャブラリは 2,316 のクラス配下に 1,957 の食品インスタンスを持つ。レビュー文を持つ評点の数は 158,424 である。その内、33,863 エントリは食品の句を含まなかったため、実験ではダミーの食品タグ ID を与え利用した。そうすることで、データセット中の全てのレストランレビューを利用することができる。ユーザ、アイテム、タグからなるテンソルのサイズは、 $36,472 \times 4,503 \times 1,957$  である。

**Last.fm** このデータセットはユーザが音楽アーティストに対して与えたタグとそのアーティストの聴取回数が含まれている。実験では、他のデータセットと合わせ、ユーザ毎の最大の聴取回数を 5、最小の聴取回数を 1 となるように、聴取回数を線形変換したものを評点として用いた。本データセットに対し、我々は Freebase から集めたジャンルボキャブラリを用い、また、WordNet から集めたタグタクソノミを用いた。本データセットにおいても、ジャンルというプロパティ以外にアーティストの内容を表現するのに適切なプロパティがなかったため、混合クラスを作成していない (なお、[Koenigstein, Dror, and Koren 2011] らが用いたデータセットのように、楽曲をアイテムとするデータセットであれば、アルバム、アーティスト、ジャンルの 3 種類のプロパティから混合クラスを作ることができ、精度向上に効果的と考えられる)。結果として、アイテムタクソノミは 38 クラスとなった。なお、各アイテムは複数のクラスに属する。一方、タグタクソノミは 2,544 クラスを持つ。さらに本実験ではクラスに割り当てることがなかったアイテムやタグも実験に利用した。結果として、本データセットは 20,665 の評点と、それに

\*5 <http://dbpedia.org/ontology/Food>

表 5 MovieLens における  $SRTF(0.3)$  と  $GCTF$  の比較

	BPTF	$SRTF(0.3)$	GCTF
$D=25$	0.9381	0.8832	1.0342
$D=50$	0.9243	0.8791	0.9947
$D=75$	0.9248	<b>0.8788</b>	0.9739
$D=100$	0.9239	0.8809	0.9683

表 6 Yelp における  $SRTF(0.3)$  と  $GCTF$  の比較

	BPTF	$SRTF(0.3)$	GCTF
$D=25$	1.7890	1.2119	1.4672
$D=50$	1.7821	1.1903	1.4192
$D=75$	1.7813	<b>1.1876</b>	1.3827
$D=100$	1.7749	1.1882	1.3291

対し 73,358 のタグ (内 53,964 タグはタグクラスを持つ) を持つものとなった。ユーザ、アイテム、タグからなるテンソルのサイズは、 $1,824 \times 6,854 \times 8,922$  となった。

## 5.2 比較手法

実験では、以下の手法間で比較を行った。

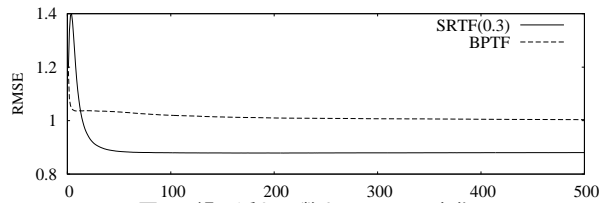
- **BPMF**: [Salakhutdinov and Mnih2008a] らにより提案された Bayesian Probabilistic Matrix Factorization。本手法はタグは用いず、ユーザがアイテムに与えた評点を分析する。個人に特化したタグを推薦することはユーザのアイテムへの理解を深める効果がある [Rendle et al.2009] にもかかわらず、本手法はタグを伴ったアイテム推薦はできない。BPMF は本稿の目指す 3 つ以上のオブジェクト関係の予測ではないため、 $SRTF$  のベースラインとはならないが、参考となる手法という点で比較を行った。
- **BPTF**: [Xiong et al.2010] らにより提案された BPTF。本手法が  $SRTF$  のベースラインとなる。
- **$SRTF\alpha$** : 本手法は曖昧性解消処理において、オブジェクトと同名のインスタンスが複数存在する場合、そのオブジェクトをランダムにその内の 1 つのインスタンスに關係付ける。
- **$SRTF\beta$** : 本手法は複合クラスを作成しない。
- **$SRTF(1.0)$** : 本手法は  $\delta$  を 1.0 に設定し、増大テンソル作成時に全てのオブジェクトをその所属クラスに引き上げる。
- **$SRTF(0.3)$** : 本手法はロングテール理論に基づき  $\delta$  を 0.3 に設定する。
- **GCTF**: 本手法は、希薄問題を解消するために複数のテンソルや行列を同時に分解する、現時点で最も高精度とされる手法 [Ermis, Acar, and Cemgil2012] である。

## 5.3 検証基準とパラメータ設定

BPTF 論文 [Xiong et al.2010] で用いられている検証基準に従い、本稿でも Root Mean Square Error (RMSE) を用いる。RMSE は  $n$  を試験データセット内のエントリの数とした場合、 $\sqrt{(\sum_{i=1}^n (P_i - R_i)^2)/n}$  と計算される。 $P_i$  と  $R_i$  はそれぞれ  $i$  番目のエントリに対する予測値と実際

表 7 Last.fm における  $SRTF(0.3)$  と  $GCTF$  の比較

	BPTF	$SRTF(0.3)$	GCTF
$D=25$	1.1727	1.1147	1.1870
$D=50$	1.1601	<b>1.1112</b>	1.1701
$D=75$	1.1623	1.1118	1.1687
$D=100$	1.1574	1.1124	1.1673

図 4 繰り返し回数と RMSE の変化 ( $D=50$ )。

の値である。RMSE の値が小さいほど誤差は小さく精度が良いといえる。実験ではデータセットを 3 分割し、3 分割交差検定を行った。以降に示す結果は 3 つの実験の平均と分散を与える。[Xiong et al.2010] に従い、パラメータはそれぞれ、 $\mu_0=0$ ,  $\nu_0=D$ ,  $\beta_0=1$ ,  $\mathbf{W}_0=\mathbf{I}$ ,  $\tilde{\nu}_0=1$ ,  $\tilde{W}_0=1$  とした。BPTF のパラメータ設定は、最終的な予測結果への影響が少ないことが報告されているため [Xiong et al.2010]、このように単純な値に設定できる。MCMC における繰り返し回数  $L$  は 500 と設定した。また、 $GCTF$  における目的関数の設定において、本実験の精度向上への寄与が最も大きかったため板倉・斉藤ダイバージェンス<sup>\*6</sup>をコスト関数として設定した。なお、BPTF のギブスサンプリングにおいて、 $\mathcal{R}$  においてタグを積分消去し得られたユーザ、アイテム行列に対して Probabilistic Matrix Factorization (PMF) [Salakhutdinov and Mnih2008b] を計算して得られる *maximum a posteriori* (MAP) の結果を、 $\mathbf{U}[1]$ ,  $\mathbf{V}[1]$  の初期値として用いる工夫を行うことで、MCMC の収束が早くなることが報告されている [Xiong et al.2010] ため、本稿でも同様の設定を行った。

## 5.4 実験結果

定量評価と定性評価の結果を記述する。

### §1 定量評価

まず GCTF を除く各手法の精度 (3 分割交差検定の平均と分散値) を MovieLens, Yelp, Last.fm において比較した結果を表 2, 表 3, 表 4 にそれぞれまとめる。BPMF はタグを伴う予測ができないため単純な精度比較はできないが、BPTF は BPMF よりも予測精度が悪いことがわかる。これは、BPTF はタグもしくはレビュー記事という、付加的な情報を予測問題においてうまく活用できていないことを示している。この原因は、テンソル (ユーザ、アイテム、タグ/レビュー) 表現が行列表現 (ユーザ、アイテム) に比べ、取りうるオブジェクト関係に対する観測値が希薄になるためである。なお、元々の BPTF の

\*6 [http://en.wikipedia.org/wiki/ItakuraSaito\\_distance](http://en.wikipedia.org/wiki/ItakuraSaito_distance)

表 8 MovieLens, Yelp, Last.fm に対する  $BPTF$  と  $SRTF(0.3)$  の予測結果の例の比較

訓練データセット			$BPTF$ と $SRTF(0.3)$ による予測結果				
(レビュー文)とタグ	アイテムジャンル	評点	(レビュー文)とタグ	アイテムジャンル	$BPTF$	$SRTF$	Actual
President History	Air Force One/Thriller&Wolfgang Peterson All the Presidents Men/Historical Fiction	3.5 4.0	disease Historical	Outbreak/Thriller&Wolfgang Peterson Glory/Historical Fiction	2.5 3.0	3.4 3.7	3.5 4.0
Tempura <i>udon</i> was delicious. We enjoyed the <i>foie gras</i> .	A/Japanese C/French	4.0 4.0	<i>Kamen</i> was really good. <i>Grauin</i> is excellent.	B/Hawaiian D/Steakhouses	5.0 2.8	3.8 3.7	4.0 4.0
Hot rockin girl Sun rockabilly	Wanda Jackson/Rockabilly Elvis Presley/Rockabilly	2.7 2.3	Hot rockin girl Hillbilly	KLHM/Rockabilly Eddie Cochran/Rockabilly	2.1 2.2	1.7 1.9	1.6 1.7

論文の実験においては、“ユーザ”，“アイテム”，アイテムを評価した時点を示す“タイムスタンプ”から構成される複数オブジェクト関係を分析している．そして，実験結果は  $BPTF$  は  $BPMF$  よりも予測精度が若干よくなることを示している．この理由については，彼らの実験で用いられているタイムスタンプの数は高々30であり，実験で用いられているテンソルが本実験のテンソルよりも十分密であるためである．そのため付加的な情報（タイムスタンプ）が予測精度改善に効果的に働いていると考えられる．しかし，Web 上に出現してきている複数オブジェクト関係は，本実験で用いている，“ユーザ”，“アイテム”，“タグ/レビュー”から構成される関係だけでなく，twitter 上で観測される“ユーザ”，“話題”，“場所”といった関係など，より大規模なテンソルになることが多く，結果，テンソルが希薄化する機会が多いため  $BPTF$  は現実のアプリケーションにおいては限界があるといえる．なお，Last.fm データセットに対しては， $BPMF$  と比較はしていない．その理由は，ユーザがアイテムを消費した回数の予測と，ユーザがアイテムにタグを付け消費した回数の予測は根本的に別物の予測になり，実験結果の比較をできないからである．一方，評点の場合は，ユーザが恣意的に点数を与えているため，タグがある場合とない場合で参考値として比較をすることは意味があると考えている．つまり，評点はユーザが自身の評価基準に基づき，決められた範囲の評点に従いアイテムに点数をつけている．そのため，評点の予測値というのは，ユーザが具体的に付けた評点との誤差が小さいほど精度が高いことになる．本実験の結果から，提案手法は， $BPMF$  に比較しても精度が良いことがわかった．

また，興味深いことに， $SRTF\alpha$  は単純な曖昧性解消処理を行っているだけであるにもかかわらず， $BPTF$  よりもかなり高精度（RMSE が低い）であることがわかる．この理由は，多くのアイテムオブジェクトは曖昧性が低い（同名のアイテムインスタンスが存在しない）こと，及び，曖昧性解消後のセマンティクスを用いたテンソル分解処理の方が，曖昧性解消時における処理エラーによる精度低減よりも，精度向上に大きく貢献しているためである．この結果は，セマンティクスを用いたテンソル分解というアイデアは例え曖昧性解消処理が単純であっても，良い結果を出すということを示している．このことから，計算時間や処理手順などのシステム制約により曖昧性解消処理が困難なアプリケーションに対しても，本提案は絶大なインパクトを発揮すると考えられる．一方， $SRTF\beta$  は  $SRTF(0.3)$  よりも精度が悪い．この理由

は， $SRTF\beta$  は混合クラスを持たずアイテムボキャブラリにおいて役者や監督のプロパティが使われていないため，テンソルに取り入れるセマンティクスの粒度が粗くなるためである．さらに， $SRTF(1.0)$  は  $SRTF(0.3)$  よりも精度が悪い．これは，セマンティクス由来のバイアスは希薄でないオブジェクトには有用でないことを示している．なお，Yelp における  $SRTF(1.0)$  は MovieLens におけるボキャブラリ/タクソノミよりも詳細でないため，精度が非常に悪い．Yelp における食品のボキャブラリを確認したところ，希薄でないタグとして，“rice”や“bread”などが含まれていることがわかった．しかし，rice や bread は非常に多くのレビューに登場するため，それを持ってユーザの行動を測るのは，ユーザの嗜好を詳細に把握できないため，精度を劣化させる．一方 MovieLens では，映画作品をインスタンスとし，ジャンルと監督・役者を組合せた複合クラスを用い，複合クラスにインスタンスを紐づけている．そのため，映画ボキャブラリ/タクソノミは Yelp におけるボキャブラリ/タクソノミよりも詳細に設計されているといえる．最後に， $SRTF(0.3)$  は，洗練された曖昧性解消処理とテンソル分解処理を持つため，他の手法よりも t 検定に基づき有意に ( $\alpha < 0.05$ ) 精度が良いことがわかった．特に  $SRTF(0.3)$  は MovieLens データセットに対して  $BPTF$  よりも 10%の精度向上を果たしている（ $SRTF(0.3)$  は精度が 0.8774 であるが， $BPTF$  は 0.9708 である）．さらに， $SRTF(0.3)$  は  $BPTF$  もしくは  $BPMF$  の結果の分散値よりも十分値が小さいため，安定した精度で結果を出していることもわかる（Yelp データセットにおいては， $BPTF$  の分散値は  $SRTF(0.3)$  よりも小さいが， $SRTF(0.3)$  は十分高精度な結果を出しているため問題ない）．

次に， $BPTF$ ， $SRTF(0.3)$ ， $GCTF$  の比較を行った． $GCTF$  は  $BPTF$  や  $SRTF$  に比べかなり大きな計算時間を要求する（3章参照）ため，我々は  $GCTF$  を本実験規模のデータセットに適用できない．そのため，ランダムに 200 人のユーザを 5 セット，各データセットより抽出し，各セットに対し実験を行った．表 5，表 6，表 7 はそれぞれ MovieLens，Yelp，Last.fm に対する実験結果を示している．それぞれの値は 5 セットの実験の平均結果を示している．

$SRTF$  は  $GCTF$  よりも十分精度が良い．この理由は， $SRTF$  はユーザ，アイテム，タグから構成されるテンソルに適切な形式（例えば，ユーザ，アイテムクラス，タグ）に変換した上で増大テンソルを構成し，自然な方法論でセマンティクス由来のバイアスを曖昧性解消テンソルに反映しているためである．一方， $GCTF$  は単純に異なる

種類の複数オブジェクト関係（ここでは、ユーザ、アイテム、タグ間の評点関係と、アイテムとアイテムクラスのリンク関係と、タグとタグクラスのリンク関係）を組合せ学習している．さらに、*SRTF* は希薄なオブジェクトを抽出した上でセマンティクス由来のバイアスを与えているのに対し、*GCTF* は全ての複数オブジェクト関係に補助情報（リンク関係）由来のバイアスを与えている．

図 4 は、MCMC 処理における繰り返し回数を変化させた場合の MovieLens データセットにおける精度の変化を示している．この結果より、4.2.3 節でも述べた通り、*SRTF* の精度は  $U$  と  $V$  の初期化に PMF 由来の MAP 値を使っていることもあり、高速に収束していることがわかる．なお、その他のデータセットにおいても同様の結果が得られた．

## §2 定性評価

次に、表 8 において、*BPTF* と *SRTF(0.3)* により出力された結果例を示す．表における列“*BPTF*”、“*SRTF*”および“Actual”は *BPTF* により出力された予測値、*SRTF(0.3)* により出力された予測値、試験データセットにおけるユーザが与えた実際の評価値を示す．MovieLens のアイテム“Outbreak”は学習データセットでは稀にしか観測されていないかった．にも関わらず、*SRTF(0.3)* は、この組合せを高精度に予測できている．また、“Glory”という歴史ジャンルの映画も高精度に予測できている．その理由は、“Wolfgang Petersen”が監督したスリラージャンルの映画と“All the Presidents Men”のような歴史ジャンルの映画は訓練データセットにおいて高く評価される傾向があるという知識を利用できるからである．

Yelp データセットにおいては、レストラン“A”における食品“udon”とレストラン“C”における食品“foie gras”の組合せが学習データセットにおいて高い評点を与えられている．一方、試験データセットにおいては、レストラン“B”における食品“ramen”とレストラン“D”における食品“gratin”の組合せが高く評点を与えられている．これらの食品タグは訓練データセットにおいて希薄にしか観測されていない．*SRTF(0.3)* はこれらの組合せを精度よく予測できている．その理由は、食品タグ“udon”と“ramen”が両方とも食品クラス“Japanese noodles”に入っていること、および、“foie gras”と“gratin”が両方とも食品クラス“French dishes”に入っているという食品ポキャブラリ由来の知識を利用できているからである．さらに、日本食とフレンチの組合せが訓練データに頻出しており、その結果、日本食とフレンチの組合せを高精度に予測できている．なお、レストラン“B”のジャンルは Japanese レストランではない．また、レストラン“D”のジャンルは French レストランではない．このことから、これらの予測はレビュー文の中の食品とその食品の背景知識を食品クラスから取得することによって予測できているわけであり、アイテムの所属クラスから予測できているわけではない．

Last.fm データセットにおいては、アーティスト“Wanda Jackson”におけるタグ“Hot rockin girl”とアーティスト“Elvis Presley”におけるタグ“Sun rockabilly”の組合せが学習データセットにおいて多く観測されている．一方、試験データセットにおいては、アーティスト“Kim Lenz and Her Mjaguars (KLHM)”におけるタグ“Hot rockin girl”とアーティスト“Eddie Cochran”におけるタグ“Hillbilly”の組合せが観測されている．なお、このアーティスト“KLHM”は訓練データセットでは観測されていない．にも関わらず、*SRTF(0.3)* はこれらの組合せを精度よく予測できている．その理由は、訓練データセットにおいて、タグ“Hot rockin girl”と“Rockabilly”ジャンルのアイテムの組合せが多く観測されており、そうしたジャンルポキャブラリ由来の知識を利用できているからである．一方、*BPTF* の予測は評価されるオブジェクトの背景知識であるクラスを利用できていないため、精度が悪い．

## 6. 結論と今後の研究の方向性への議論

本稿は、テンソル分解の精度を向上させるために、オブジェクトの背景にあるセマンティクスをテンソル分解に適用する初めての試みを述べた．テンソル分解においてセマンティクスを効果的に反映させるという試みは、人間の行動をセマンティクスを用いて洗練された機械学習の枠組みで分析するという重要な問題を孕んでおり、人工知能における重要な問題の 1 つである．本稿で提案する *SRTF* は、まずオブジェクトをセマンティクスに関係付けたテンソルを構築し、次に *BPTF* のフレームワーク上にてセマンティクス経由のバイアスをテンソル分解に反映させながら、分解処理を進める．実データセットを用いた実験結果により、*SRTF* は補助情報をテンソル分解に入れ込むアプローチとして現在主流の *GCTF* やベースラインである *BPTF* よりも 10% 以上高い精度を発揮し、多くのアプリケーションを支援できることを示した．

本稿で提案する曖昧性解消処理はテンソル分解の前処理であり、かつ、抽象化処理は比較的単純であるため、本提案は *BPTF* 以外のテンソル分解の枠組みにおいても利用できる．実際、我々は本アイデアを既に *BPTF* とは別の著名なテンソル分解の枠組みである Non-negative Tensor Factorization (NTF) にも適用している．具体的には、文献 [Cemgil2009] で提案されている Variational Non-negative Matrix Factorization (VNMF) を Variational Non-negative Tensor Factorization (VNNTF) に発展させた上で、提案するアイデアを VNNTF に適用し、有用な結果を得ることを確認している．文献 [Cemgil2009] では、*BPTF* のようにガウス分布に基づくモデルではなく、VNMF を、様々な実データセットで観測されるベキ乗分布に比較的効果的に適用できるポアソンモデル [Xiong et al.2010] を採用している．このことから本稿における提案は、様々な分野のアプリケーションに適用でき、大きなインパクトを

持つ。

現在、我々のプロジェクトではSRTFを、例えばユーザ間の将来のコミュニケーション頻度の予測などの、ソーシャルネットワークにおけるリンク予測に適用している。ユーザ間のコミュニケーション頻度はベキ指数に従うため、前段で述べたように、本稿のアイデアをVNTFに適用している。その他の興味深い方向性としては、音楽と映画等ドメインを跨がったユーザ行動の予測である。LODプロジェクトの目的の1つは複数のサービスドメインを跨がるデータのセマンティクスをWeb上で共有することであるため、本研究で提案するLOD上のセマンティクスをテンソル分解に組込むアプローチはクロスドメイン分析に大きな潜在性を持つ。3つ目の興味深い方向性としては、シンプルなボキャブラリやタクソノミではなく、より詳細なセマンティクスを取り扱うように手法を発展させることである。例えば、本研究ではオブジェクトの所属クラスを利用しているが、所属クラスと他クラス間の意味的な関係を効果的に使う手法等が期待される。

#### ◇ 参 考 文 献 ◇

- [Acar, Kolda, and Dunlavy2011] Acar, E., Kolda, T. G., and Dunlavy, D. M. 2011. All-at-once optimization for coupled matrix and tensor factorizations. *CoRR* abs/1105.3422.
- [Anderson2006] Anderson, C. 2006. *The Long Tail: Why the Future of Business Is Selling Less of More*. Hyperion.
- [Bishop2006] Bishop, C. M. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc.
- [Bizer et al.2009] Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., and Hellmann, S. 2009. Dbpedia - a crystallization point for the web of data. *Web Semant.* 7(3):154–165.
- [Bizer, Heath, and Berners-Lee2009] Bizer, C., Heath, T., and Berners-Lee, T. 2009. Linked data - the story so far. *Int. J. Semantic Web Inf. Syst.* 5(3):1–22.
- [Bollacker et al.2008] Bollacker, K., Evans, C., Paritosh, P., Sturge, T., and Taylor, J. 2008. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proc. SIGMOD'08*, 1247–1250.
- [Cantador, Konstas, and Jose2011] Cantador, I., Konstas, I., and Jose, J. M. 2011. Categorising social tags to improve folksonomy-based recommendations. *Web Semant.* 9(1):1–15.
- [Cemgil2009] Cemgil, A. T. 2009. Bayesian inference for nonnegative matrix factorisation models. *Intell. Neuroscience* 2009:4:1–4:17.
- [Ermis, Acar, and Cemgil2012] Ermis, B., Acar, E., and Cemgil, A. T. 2012. Link prediction via generalized coupled tensor factorisation. *CoRR* abs/1208.6231.
- [Ermis, Acar, and Cemgil2013] Ermis, B., Acar, E., and Cemgil, A. T. 2013. Link prediction in heterogeneous data via generalized coupled tensor factorization. *Data Mining and Knowledge Discovery*.
- [Gabrilovich and Markovitch2007] Gabrilovich, E., and Markovitch, S. 2007. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *Proc. IJCAI'07*, 1606–1611.
- [Hu et al.2008] Hu, J., Fang, L., Cao, Y., Zeng, H.-J., Li, H., Yang, Q., and Chen, Z. 2008. Enhancing text clustering by leveraging wikipedia semantics. In *Proc. SIGIR'08*, 179–186.
- [Jain et al.2010] Jain, P., Hitzler, P., Yeh, P. Z., Verma, K., and Sheth, A. P. 2010. A.p.: Linked data is merely more data. In *Proc. AAAI Spring Symposium 'Linked Data Meets Artificial Intelligence'*, 82–86.
- [Karatzoglou et al.2010] Karatzoglou, A., Amatriain, X., Baltrunas, L., and Oliver, N. 2010. Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In *Proc. RecSys'10*, 79–86.
- [Klein and Manning2003] Klein, D., and Manning, C. D. 2003. Accurate unlexicalized parsing. In *Proc. ACL'03*, 423–430.
- [Koenigstein, Dror, and Koren2011] Koenigstein, N., Dror, G., and Koren, Y. 2011. Yahoo! music recommendations: modeling music ratings with temporal dynamics and item taxonomy. In *Proc. RecSys'11*, 165–172.
- [McGuinness2003] McGuinness, D. L. 2003. Ontologies come of age. In *Spinning the Semantic Web*, 171–194. MIT Press.
- [Miller1995] Miller, G. A. 1995. Wordnet: A lexical database for english. *Commun. ACM* 38(11):39–41.
- [Nakatsuji and Fujiwara2014] Nakatsuji, M., and Fujiwara, Y. 2014. Linked taxonomies to capture users' subjective assessments of items to facilitate accurate collaborative filtering. *Artificial Intelligence* 207(0):52–68.
- [Nakatsuji et al.2010] Nakatsuji, M., Fujiwara, Y., Tanaka, A., Uchiyama, T., Fujimura, K., and Ishida, T. 2010. Classical music for rock fans?: novel recommendations for expanding user interests. In *Proc. CIKM'10*, 949–958.
- [Nakatsuji et al.2011] Nakatsuji, M., Fujiwara, Y., Uchiyama, T., and Fujimura, K. 2011. User similarity from linked taxonomies: Subjective assessments of items. In *Proc. IJCAI'11*, 2305–2311.
- [Nakatsuji et al.2012] Nakatsuji, M., Fujiwara, Y., Uchiyama, T., and Toda, H. 2012. Collaborative filtering by analyzing dynamic user interests modeled by taxonomy. In *Proc. ISWC'12*, 361–377.
- [Nakatsuji et al.2014] Nakatsuji, M., Fujiwara, Y., Toda, H., Sawada, H., Zheng, J., and Hendler, J. A. 2014. Semantic data representation for improving tensor factorization. In *Proc. AAAI'14*.
- [Nakatsuji, Miyoshi, and Otsuka2006] Nakatsuji, M., Miyoshi, Y., and Otsuka, Y. 2006. Innovation detection based on user-interest ontology of blog community. In *Proceedings of the 2006 International Semantic Web Conference (ISWC2006)*, 515–528.
- [Nakatsuji, Yoshida, and Ishida2009] Nakatsuji, M., Yoshida, M., and Ishida, T. 2009. Detecting innovative topics based on user-interest ontology. *Web Semant.* 7(2):107–120.
- [Nakaya et al.2013] Nakaya, A., Katayama, T., Itoh, M., Hiranuka, K., Kawashima, S., Moriya, Y., Okuda, S., Tanaka, M., Tokimatsu, T., Yamanishi, Y., Yoshizawa, A. C., Kanehisa, M., and Goto, S. 2013. Kegg oc: a large-scale automatic construction of taxonomy-based ortholog clusters. *Nucleic Acids Research* 41(Database-Issue):353–357.
- [Narita et al.2011] Narita, A., Hayashi, K., Tomioka, R., and Kashima, H. 2011. Tensor factorization using auxiliary information. In *Proc. ECML-PKDD'11*, 501–516.
- [Parundekar, Knoblock, and Ambite2012] Parundekar, R., Knoblock, C. A., and Ambite, J. L. 2012. Discovering concept coverings in ontologies of linked data sources. In *Proc. ISWC'12*, 427–443.
- [Rendle and Schmidt-Thieme2010] Rendle, S., and Schmidt-Thieme, L. 2010. Pairwise interaction tensor factorization for personalized tag recommendation. In *Proc. WSDM'10*, 81–90.
- [Rendle et al.2009] Rendle, S., Marinho, L. B., Nanopoulos, A., and Schmidt-Thieme, L. 2009. Learning optimal ranking with tensor factorization for tag recommendation. In *Proc. KDD'09*, 727–736.
- [Saif, He, and Alani2012] Saif, H., He, Y., and Alani, H. 2012. Semantic sentiment analysis of twitter. In *Proc. ISWC'12*, 508–524.
- [Salakhutdinov and Mnih2008a] Salakhutdinov, R., and Mnih, A. 2008a. Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In *Proc. ICML'08*, volume 25.
- [Salakhutdinov and Mnih2008b] Salakhutdinov, R., and Mnih, A. 2008b. Probabilistic matrix factorization. In *Proc. NIPS'08*, volume 20.
- [Sutskever, Tenenbaum, and Salakhutdinov2009] Sutskever, I., Tenenbaum, J. B., and Salakhutdinov, R. 2009. Modelling relational data using bayesian clustered tensor factorization. In Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., and Culotta, A., eds., *Proc. NIPS'09*. 1821–1828.
- [Takeuchi et al.2013] Takeuchi, K., Tomioka, R., Ishiguro, K., Kimura, A., and Sawada, H. 2013. Non-negative multiple tensor factorization. In *Proc. ICDM'13*, 1199–1204.
- [Turney and Pantel2010] Turney, P. D., and Pantel, P. 2010. From

frequency to meaning: vector space models of semantics. *J. Artif. Int. Res.* 37(1):141–188.

[Xiong et al.2010] Xiong, L., Chen, X., Huang, T.-K., Schneider, J. G., and Carbonell, J. G. 2010. Temporal collaborative filtering with bayesian probabilistic tensor factorization. In *Proc. SDM'10*, 211–222.

[Yilmaz, Cemgil, and Simsekli2011] Yilmaz, Y. K., Cemgil, A.-T., and Simsekli, U. 2011. Generalised coupled tensor factorisation. In *Proc. NIPS'11*, 2151–2159.

[Yu et al.2011] Yu, J., Zha, Z.-J., Wang, M., and Chua, T.-S. 2011. Aspect ranking: identifying important product aspects from online consumer reviews. In *Proc. HLT '11*, 1496–1505.

[Zheng et al.2010] Zheng, V. W., Cao, B., Zheng, Y., Xie, X., and Yang, Q. 2010. Collaborative filtering meets mobile recommendation: A user-centered approach. In *Proc. AAAI'10*, 236–241.

[Zheng et al.2013] Zheng, J. G., Fu, L., Ma, X., and Fox, P. 2013. Sem+: Discover “same as” links among the entities on the web of data. Technical report, Rensselaer Polytechnic Institute, <http://tw.rpi.edu/web/doc/Document?uri=http://tw.rpi.edu/media/2013/10/07/e293/SEM.doc>.

〔担当委員：岡崎 直観〕

2014 年 6 月 27 日 受理

## 7. 付 録

この付録では SRTF の MCMC 処理の中で属性ベクトルとハイパーパラメータをどのようにして学習するのかを詳細に説明する．4.2 節についても併せてご参照願いたい．以下の記号と処理手順上の番号は 4.2 節で使われたものと同じである．SRTF の MCMC 処理は 4.2.3 節 i. 項の“アプローチ”で説明したアプローチに従い BPTF におけるギブスサンプリングを改良したものである．BPTF のサンプリング手順の詳細は BPTF の原論文 [Xiong et al.2010] を参照願いたい．

SRTF は，BPTF と同様に MCMC 処理により，未観測の評点の分布  $\hat{R}^d$  を学習する．中でも具体的にはギブスサンプリングを用い学習を進める．ギブスサンプリング手順は以下の通りである．

(1) まず BPTF と同様に PMF の MAP の結果により， $U^d[1]$ ， $V^d[1]$  を初期化する．さらに， $T^d[1]$ ， $C^v[1]$ ， $C^t[1]$  をガウス分布として初期化する． $C^v$  と  $C^t$  は，それぞれ  $c_j^v$  と  $c_j^t$  の行列表現である．次に，ステップ (2) から (6) を  $L$  回繰り返す，属性ベクトルとそのハイパーパラメータをサンプルする．MCMC 処理における各サンプルは 1 つ前のサンプルにのみ依存する．

(2) 次に，BPTF と同様に  $R^d$ ， $U^d$ ， $V^d$ ， $T^d$  が与えられたときの  $\alpha$  の条件付き分布を以下のようにサンプリングする．

$$p(\alpha | R^d, U^d, V^d, T^d) = \mathcal{W}(\alpha | \tilde{W}_0^*, \nu_0^*),$$

$$\nu_0^* = \tilde{\nu}_0 + \sum_{k'=1}^{K'} \sum_{n'=1}^{N'} \sum_{m=1}^M 1, (\tilde{W}_0^*) = (\tilde{W}_0)^{-1}$$

$$+ \sum_{k'=1}^{K'} \sum_{n'=1}^{N'} \sum_{m=1}^M (r_{m,n',k'}^d - \langle \mathbf{u}_m^d, \mathbf{v}_{n'}^d, \mathbf{t}_{k'}^d \rangle)^2.$$

また BPTF と同様に，ハイパーパラメータ  $\Theta_{U^d}$ ， $\Theta_{V^d}$ ，

$\Theta_{T^d}$  をサンプルする．これらのハイパーパラメータは同様に計算される．例えば， $\Theta_{V^d}$  は  $V^d$  以外のパラメータに依存せず，以下のように計算される．

$$p(\Theta_{V^d} | V^d) = \mathcal{N}(\boldsymbol{\mu}_{V^d} | \boldsymbol{\mu}_0^*, (\beta_0^* \boldsymbol{\Lambda}_{V^d})^{-1}) \mathcal{W}(\boldsymbol{\Lambda}_{V^d} | \mathbf{W}_0^*, \nu_0^*),$$

$$\boldsymbol{\mu}_0^* = \frac{\beta_0 \boldsymbol{\mu}_0 + N' \bar{\mathbf{v}}^d}{\beta_0 + N'}, \beta_0^* = \beta_0 + N', \nu_0^* = \nu_0 + N',$$

$$\mathbf{W}_0^{*-1} = \mathbf{W}_0^{-1} + N' \bar{\mathbf{S}} + \frac{\beta_0 N'}{\beta_0 + N'} (\nu_0 - \bar{\mathbf{v}}^d) (\nu_0 - \bar{\mathbf{v}}^d)^{tr},$$

$$\bar{\mathbf{v}}^d = \frac{1}{N'} \sum_{n'=1}^{N'} \mathbf{v}_{n'}^d,$$

$$\bar{\mathbf{S}} = \frac{1}{N'} \sum_{n'=1}^{N'} (\mathbf{v}_{n'}^d - \bar{\mathbf{v}}^d) (\mathbf{v}_{n'}^d - \bar{\mathbf{v}}^d)^{tr}.$$

さらに我々の MCMC 処理は， $C^v$ ， $\Theta_{C^v}$  をサンプルし， $R^v$  のための評点を生成する． $R^v$  の生成処理は， $V^d$  と  $C^v$  をサンプルする際に， $\alpha$ ， $U^d$ ， $T^d$  を  $R^d$  と共有する以外は，ほぼ  $R^d$  と同様である． $\Theta_{C^v}$  は  $\Theta_{V^d}$  と同様に以下のように計算される．

$$p(\Theta_{C^v} | C^v) = \mathcal{N}(\boldsymbol{\mu}_{C^v} | \boldsymbol{\mu}_0^*, (\beta_0^* \boldsymbol{\Lambda}_{C^v})^{-1}) \mathcal{W}(\boldsymbol{\Lambda}_{C^v} | \mathbf{W}_0^*, \nu_0^*),$$

$$\boldsymbol{\mu}_0^* = \frac{\beta_0 \boldsymbol{\mu}_0 + (N' + S^V) \bar{\mathbf{c}}^v}{\beta_0 + N' + S^V}, \beta_0^* = \beta_0 + N' + S^V, \nu_0^* = \nu_0 + N' + S^V,$$

$$\mathbf{W}_0^{*-1} = \mathbf{W}_0^{-1} + (N' + S^V) \bar{\mathbf{S}} + \frac{\beta_0 (N' + S^V)}{\beta_0 + N' + S^V} (\nu_0 - \bar{\mathbf{c}}^v) (\nu_0 - \bar{\mathbf{c}}^v)^{tr},$$

$$\bar{\mathbf{c}}^v = \frac{1}{N' + S^V} \sum_{j=1}^{N' + S^V} \mathbf{c}_j^v,$$

$$\bar{\mathbf{S}} = \frac{1}{N' + S^V} \sum_{j=1}^{N' + S^V} (\mathbf{c}_j^v - \bar{\mathbf{c}}^v) (\mathbf{c}_j^v - \bar{\mathbf{c}}^v)^{tr}.$$

(3) そして，BPTF と同様にモデルパラメータ  $U^d$ ， $V^d$ ， $T^d$  をサンプルする．例えば， $V^d$  は各アイテムに分解される．ここで， $\mathbf{P}_{m,k'} \equiv \mathbf{u}_m^d \cdot \mathbf{t}_{k'}^d$  とする．各アイテムの属性ベクトルは以下のように並列計算できる．

$$p(\mathbf{v}_{n'}^d | R^d, U^d, T^d, \alpha, \Theta_{V^d}) = \mathcal{N}(\mathbf{v}_{n'}^d | \boldsymbol{\mu}_{n'}^*, (\boldsymbol{\Lambda}_{n'}^*)^{-1}),$$

$$\boldsymbol{\mu}_{n'}^* \equiv (\boldsymbol{\Lambda}_{n'}^*)^{-1} (\boldsymbol{\Lambda}_{V^d} \boldsymbol{\mu}_{V^d} + \alpha \sum_{k'=1}^{K'} \sum_{m=1}^M r_{m,n',k'}^d \mathbf{P}_{m,k'}),$$

$$\boldsymbol{\Lambda}_{n'}^* = \boldsymbol{\Lambda}_{V^d} + \alpha \sum_{k'=1}^{K'} \sum_{m=1}^M o_{m,n',k'}^d \mathbf{P}_{m,k'} \mathbf{P}_{m,k'}^{tr}.$$

(4) 次に，モデルパラメータ  $C^v$  と  $C^t$  をサンプルする．例えば  $C^v$  の条件付き分布は以下のように各アイテムと各アイテムクラスに分解される．各属性ベクトルはステップ (3) と同様に以下のように並列計算できる．

$$p(\mathbf{c}_j^v | R^v, U^d, T^d, \alpha, \Theta_{C^v}) = \mathcal{N}(\mathbf{c}_j^v | \boldsymbol{\mu}_j^*, (\boldsymbol{\Lambda}_j^*)^{-1}),$$

$$\boldsymbol{\mu}_j^* \equiv (\boldsymbol{\Lambda}_j^*)^{-1} (\boldsymbol{\Lambda}_{C^v} \boldsymbol{\mu}_{C^v} + \alpha \sum_{k'=1}^{K'} \sum_{m=1}^M r_{m,j,k'}^v \mathbf{P}_{m,k'}),$$

$$\boldsymbol{\Lambda}_j^* = \boldsymbol{\Lambda}_{C^v} + \alpha \sum_{k'=1}^{K'} \sum_{m=1}^M o_{m,j,k'}^v \mathbf{P}_{m,k'} \mathbf{P}_{m,k'}^{tr}.$$

ここで，テンソル  $\mathcal{O}^v$  の要素  $o_{m,j,k'}^v$  は  $R^v$  においてユーザ  $u_m$  と  $c_j^v$  に対応する  $j$  番目のアイテム/クラス，タグ  $t_{k'}$  が存在すれば 1，存在しなければ 0 を与える．ステップ (3) と (4) は  $\alpha$  と属性ベクトル  $\mathbf{u}_m^d$ ， $\mathbf{t}_{k'}^d$  ( $\mathbf{P}_{m,k'}$  の計算で使われている) を共有している．これは，SRTF はテンソル  $R^d$  と  $R^v$  の分解において上記パラメータを共有し

ていることを意味する。故に、これらのパラメータを通して、上記テンソル間でセマンティクス由来のバイアスが共有されることになる。条件付き分布  $C^l$  も同様にして計算される。

(5) 各繰り返しにおいて、未観測の評点  $\hat{r}^d[l]$  を、 $U^d[l+1]$ ,  $V^d[l+1]$ ,  $T^d[l+1]$ ,  $C^v[l+1]$ ,  $C^t[l+1]$ ,  $\alpha[l]$  を式(4)に適用することでサンプルする。

(6) もし  $v_{n'}$  が希薄なアイテムインスタンスであれば、 $v_{n'}^d$  を  $c_{n'}^v$  に置き換える。もし希薄なアイテムインスタンスでなければ、 $v_{n'}^d$  を  $c_{n'}^v$  に置き換えない。本処理は以下の式で表される。

$$v_{n'}^d = \begin{cases} c_{n'}^v & (z_{n'}^v = 0) \\ v_{n'}^d & (z_{n'}^v = 1) \end{cases}$$

同様に、もし  $t_{k'}$  が希薄なタグインスタンスであれば、 $t_{k'}^d$  を  $c_{k'}^t$  に置き換える。もし希薄なタグインスタンスでなければ、 $t_{k'}^d$  を  $c_{k'}^t$  に置き換えない。本処理は以下の式で表される。

$$t_{k'}^d = \begin{cases} c_{k'}^t & (z_{k'}^t = 0) \\ t_{k'}^d & (z_{k'}^t = 1) \end{cases}$$

なお、本ステップは、SRTFの生成プロセスの手順(9)(4.2.3節項参照)に対応する。

SRTFは最終的に  $v_n$  と  $v_{n'}$  間、および、 $t_k$  と  $t_{k'}$  間のリンク関係を確認することにより、 $\hat{r}_{m,n',k'}^d$  から元のテンソル  $\hat{r}_{m,n,k}$  における未観測の評点に対する予測分布を計算する。

このようにして、SRTFはMCMC処理の各繰り返しにおいてセマンティクスに基づくバイアスを希薄オブジェクトの属性ベクトルに効果的に反映する。これにより、希薄問題を解決できる。

——— 著 者 紹 介 ———



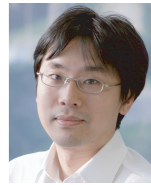
中辻 真(正会員)

2001年京都大学工学部数理工学卒業。2003年同大学大学院情報学研究所システム科学専攻修士課程修了。2010年同研究所社会情報学専攻博士課程修了。情報学博士。2003年日本電信電話株式会社入社。2013年レンセラー工科大学研究員。現在、日本電信電話株式会社サービスエボリューション研究所勤務。データマイニング、都市情報学、セマンティック Web に関する研究に従事。電子情報通信学会、日本データベース学会各会員。



藤原 靖宏(正会員)

2003年早稲田大学大学院理工学研究科電気工学専攻修士課程修了。同年、日本電信電話株式会社入社。2011年東京大学大学院情報理工学系研究科電子情報学専攻博士課程修了。2014年ニューヨーク大学客員研究員。現在、日本電信電話株式会社 NTT ソフトウェアイノベーションセンター特別研究員。博士(情報理工学)。グラフマイニングの研究開発に従事。第27回テレコムシステム技術賞、第9回上林奨励賞などを受賞。情報処理学会、電子情報通信学会、日本データベース学会各会員。



戸田 浩之(正会員)

1997年名古屋大学工学部材料プロセス工学科卒業。1999年同大学院工学研究科博士課程前期課程修了。同年、日本電信電話株式会社入社。2007年筑波大学大学院システム情報工学研究科コンピュータサイエンス専攻博士後期課程修了。博士(工学)。現在 NTT サービスエボリューション研究所所属。情報検索、情報抽出、Web マイニングの研究に従事。電子情報通信学会、情報処理学会、日本データベース学会、ACM 各会員。



澤田 宏(正会員)

1991年京都大学工学部情報工学科卒業。1993年同大学院工学研究科修士課程修了。同年、日本電信電話株式会社(NTT)入社。以来、同社コミュニケーション科学基礎研究所にて、VLSI 向け CAD および計算機アーキテクチャの研究に従事。2000年より、信号処理、特にブラインド音源分離の研究に従事。2009年より、知能創環境研究グループグループリーダー。2013年より、同社サービスエボリューション研究所、現在に至る。2001年京都大学博士(情報学)。電子情報通信学会、日本音響学会、IEEE 各会員。



チェン ジン

Jin Guang Zheng received the B.S. degree and Ph.D degree in computer science from Rensselaer Polytechnic Institute, Troy, NY, USA, in 2009 and 2014 respectively. Since 2015, he has been a Research Associate with the Institute of Data Exploration and Application, the Tetherless World Constellation, and the Blender Lab for Natural Language Processing in Rensselaer Polytechnic Institute. His research interests are Semantic eScience, Semantic Similarity Computation, and Automatic Knowledge Base Construction using Semantic Web technologies and Natural Language Processing technologies. Recently, he has contributed to the NSF Global Change Information System for Semantic eScience project, the NIST Entity Discovery and Linking Competition for Knowledge Base Population project, and the Disney NLP project.



ヘンドラー ジェームズ

James Hendler is the Director of the Institute for Data Exploration and Applications and the Tetherless World Professor of Computer, Web and Cognitive Sciences at RPI. Hendler has authored over 350 books, technical papers and articles in the areas of Semantic Web, artificial intelligence, agent-based computing and high performance processing. One of the originators of the " Semantic Web, " Hendler was the recipient of a 1995 Fulbright Foundation Fellowship, is a former member of the US Air Force Science Advisory Board, and is a Fellow of the American Association for Artificial Intelligence, the British Computer Society, the IEEE and the AAAS. He is also the former Chief Scientist of the Information Systems Office at the US Defense Advanced Research Projects Agency (DARPA) and was awarded a US Air Force Exceptional Civilian Service Medal in 2002. He is also the first computer scientist to serve on the Board of Reviewing editors for Science. In 2010, Hendler was named one of the 20 most innovative professors in America by Playboy magazine and was selected as an " Internet Web Expert " by the US government. In 2012, he was one of the inaugural recipients of the Strata Conference " Big Data " awards for his work on large-scale open government data, and he is a columnist and associate editor of the Big Data journal. In 2013, he was appointed by the governor as the Open Data Advisor to New York State.