1B2-R-3-8

ソーシャルWebサービスにおけるユーザ行動予測：次元削減アプローチ

Predicting User Actions in Social Web Services: A Dimension Reduction Approach

則　のぞみ[*1]　　ボレガラ　ダヌシカ[*1]　　鹿島　久嗣[*1]
Nozomi Nori　　Danushka Bollegala　　Hisashi Kashima

[*1]東京大学大学院　情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

An important concept in social web services is social actions, such as making connections, communicating with others and adding annotations to web resources. Social actions involve multiple and heterogeneous objects such as users, documents, keywords, and locations, whose high-dimensional property makes it difficult to predict multinomial relations with a limited amount of data. We propose a new multinomial relation prediction method, which is robust to data sparsity. We transform each instance of a multinomial relation into a set of binomial relations between the objects and the multinomial relation of the involved objects. We then apply an extension of a low-dimensional embedding technique to these binomial relations, which results in a generalized eigenvalue problem guaranteeing global optimal solutions. We also incorporate attribute information to address "cold start" problem. Experiments with various datasets demonstrate that the proposed method is more robust against data sparseness as compared to several existing methods.

1. はじめに

ソーシャル Web サービスにおいて、他のユーザとのコミュニケーションや、コンテンツへのアノテーションなどといったユーザの行動は重要な役割を果たす。ユーザの行動は、ユーザ、文書、キーワード、場所など、複数の異種オブジェクトを巻き込んだ関係データとして表現されるため、これらを限定されたデータから予測する際には、観測データの疎性に対して頑強な予測を行うことが課題となる。関係データで生じる観測データの疎性の問題というのは、関係に含まれるオブジェクト数の増加に伴い、可能な関係の組みわせの数が指数的に増えるため、実際に観測される関係の数が、可能な場合の数に対して相対的に小さくなるために生じるものである。実際、Cai [Cai 11] らに報告によると、広く使用されているソーシャルタギングサービスのデータセットでは、可能な組み合わせに対して 0.01%程度の関係しか観測されていない。また、ほとんどのオブジェクトは少数の関係にしか関与せず、その分布はべき乗分布に従っていた。後者の疎性は、推薦タスクにおいて重要な問題と見なされている、「コールドスタート問題」と呼ばれる状況でしばしば遭遇するものである。近年では、多項関係予測を行うにあたって、テンソル分解の手法がしばしば利用されているが、多くの手法は非凸最適化問題として定式化されるため、特に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。本論文では、(1) 多項関係から二項関係の集合への変換と (2) 属性情報の活用という二つのアイデアに基づき、データ過疎に対して頑強な多項関係予測手法を提案する。

2. 提案手法

2.1 問題設定

人やウェブページなど、異なる種類のオブジェクトの間で、ある特定の種類の関係がどれくらい生じやすそうかを、いくつかの観測された関係データを基に推定する問題を考える。例えば、ある人が、別の誰かから勧められたウェブページを気に入るかを予測する状況を考えてみよう。ここでの目標は、「$\texttt{person}_1$ が $\texttt{person}_2$ によって勧められた、URL で指定されるウェブページを気に入る」という関係が、($\texttt{person}_1$, URL, $\texttt{person}_2$) の各組み合わせについて、どれくらい生じやすいかを、(Hanako, ai-gakkai.or.jp, Taro) などの既知の事実を基に予測することである。

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持つとしよう。i 番目のオブジェクト $s^{(k,i)} \in S^{(k)}$ を、$s^{(k,i)}$ のように記載する。例えば、$s^{(1,1)}$ によって Hanako を表す。また、M 個の観測された関係インスタンスから成る集合、$O \subset S^{(1)} \times S^{(2)} \times \cdots \times S^{(K)}$ が得られているとする。各関係インスタンスは、ある特定の関係（例えば先の例では、気に入るという関係）がオブジェクトの特定の組み合わせに対して成立していることを示している。例えば、$o^{(1)} \in O$ は (Hanako, ai-gakkai.or.jp, Taro) などである。

さて、我々の目標は、オブジェクトの各組み合わせの中で、観測された関係インスタンスの集合 O に含まれていないものについて、特定の関係がどれくらい生じやすいかを予測することである。換言すると、$(S^{(1)} \times S^{(2)} \times \cdots \times \ldots S^{(K)}) \setminus O$ に含まれる、オブジェクトの組み合わせについて、関係の生じやすさによってランキングされたリストを求めることが目標である。

多くの現実的な状況では、各オブジェクトは自身に関する何かしらの情報を有する。例えば、人であれば年齢や性別などのデモグラフィックな情報を持つと期待できる。したがって、$s^{(k,i)}$ に対して、$D^{(k)}$ 次元の属性ベクトル $\boldsymbol{x}^{(k,i)}$ を関連付け、各 $k = 1, 2, \ldots, K$ について、まとめて計画行列 Φ

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

としよう。

本論文が提案する多項関係予測の問題は、以下のような入出力を持つ問題として要約できる。

問題：多項関係予測

連絡先: nozomi.nori@gmail.com

- **入力:**
  - $S^{(1)}, S^{(2)}, \ldots, S^{(K)}$: K 種類のオブジェクト集合
  - O ($\subset S^{(1)} \times S^{(2)} \times \cdots \times \ldots S^{(K)}$): 観測された M 個の関係インスタンスから成る集合
  - $\boldsymbol{\Phi}^{(1)}, \boldsymbol{\Phi}^{(2)}, \ldots, \boldsymbol{\Phi}^{(K)}$: オブジェクトの属性を表現する K 個の計画行列
- **出力:** O に含まれない、すなわち、$(S^{(1)} \times S^{(2)} \times \cdots \times \ldots S^{(K)}) \setminus O$ に含まれる、オブジェクトの組み合わせについて、関係の生じやすさによってソートされたランクリスト。

2.2 次元削減を用いた多項関係予測

ここでは、次元削減手法を用いて、一般化固有値問題を一度解くだけで大域解が求まる、新しい多項関係予測手法を提案する。まず、オブジェクトの属性を考慮しない場合から考える。

大域解を保証するための最初のキーアイデアは、多項関係を二項関係に変換することである。K 種類のオブジェクト集合それぞれに対して一つの二値行列を構築し、全部で K 個の行列を得る。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る。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参加しているかを示している。

$\boldsymbol{A}^{(k)}$ を、$S^{(k)}$ に属するオブジェクトの、O に属する関係インスタンスに対する参加情報をまとめた $N^{(k)} \times M$ の二値行列とする。$A^{(k)}$ の各要素は、以下のように定義される。

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

我々の二つ目のキーアイデアは、K 個の行列によって表現された二項関係の低次元への埋め込みである。次元削減を用いた二部グラフ予測の手法 [Yamanishi 09] と似たアイデアを用いて、各オブジェクトと、そのオブジェクトが参加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し、異種のオブジェクトと関係インスタンスを共通の潜在次元に埋め込む。

まず最初に一次元への埋め込みを考えてみよう。サイズ $N^{(1)}$ のオブジェクト集合 $S^{(1)}$ は、長さ $N^{(1)}$ のベクトル $\boldsymbol{f}^{(1)}$ として埋め込まれる。同様にして、各オブジェクト集合 $S^{(2)}, S^{(3)}, \ldots, S^{(K)}$ は、それぞれ $\boldsymbol{f}^{(2)}, \boldsymbol{f}^{(3)}, \ldots, \boldsymbol{f}^{(K)}$ として埋め込まれる。サイズ M の観測された関係インスタンスの集合 O もまた、同じ一次元の潜在空間に、サイズ M のベクトル $\bar{\boldsymbol{f}}$ として埋め込まれる。

もし $s^{(k,n)} \in S^{(k)}$ がある関係インスタンス $o^{(m)} \in O$ に参加するならば、両者の埋め込み先、$[\boldsymbol{f}^{(k)}]_n$ と $[\bar{\boldsymbol{f}}]_m$ を近くすること、すなわち、ユークリッド距離 $([\boldsymbol{f}^{(k)}]_n - [\bar{\boldsymbol{f}}]_m)^2$ を小さくすることを試みよう。このとき、最終的に最小化すべき目的関数は以下のように定義される。

$$\begin{aligned} & J(\{\boldsymbol{f}^{(k)}\}_{k=1}^K, \bar{\boldsymbol{f}}) \qquad (1) \\ &= \sum_k \sum_i \sum_j [\boldsymbol{A}^{(k)}]_{i,j} \left([\boldsymbol{f}^{(k)}]_i - [\bar{\boldsymbol{f}}]_j \right)^2 \\ &= \sum_k \left(\boldsymbol{f}^{(k)\top} \boldsymbol{D}^{(k)} \boldsymbol{f}^{(k)} + \bar{\boldsymbol{f}}^\top \bar{\boldsymbol{f}} - 2 \boldsymbol{f}^{(k)\top} \boldsymbol{A}^{(k)} \bar{\boldsymbol{f}} \right) \end{aligned}$$

ここで、$\boldsymbol{D}^{(k)}$ は、その (i,i) 番目の要素が $[\boldsymbol{D}^{(k)}]_{i,i} \equiv \sum_j [\boldsymbol{A}^{(k)}]_{i,j}$、すなわち、オブジェクト $s^{(k,i)}$ が参加した関係の数として定義される対角行列である。

この目的関数は、$\boldsymbol{f}^{(k)} \equiv \boldsymbol{0}$ かつ $\bar{\boldsymbol{f}} \equiv \boldsymbol{0}$ とすることで容易に最小化できてしまうので、このような望ましくない解を避けるために、以下のスケーリング制約を加える。

$$\sum_{k=1}^K \boldsymbol{f}^{(k)\top} \boldsymbol{D}^{(k)} \boldsymbol{f}^{(k)} = 1 \qquad (2)$$

目的関数 (1) の $\bar{\boldsymbol{f}}$ に関する最小値は以下のように得られる。

$$\bar{\boldsymbol{f}} = \frac{1}{K} \sum_{k=1}^K \boldsymbol{A}^{(k)\top} \boldsymbol{f}^{(k)} \qquad (3)$$

式 (3) を式 (1) の正負を逆転させたものに代入することで、以下の最大化問題を得る。

$$\begin{aligned} & -J(\{\boldsymbol{f}^{(k)}\}_{k=1}^K) \qquad (4) \\ &= \frac{1}{K} \sum_{k,\ell=1}^K \boldsymbol{f}^{(k)\top} \boldsymbol{A}^{(k)} \boldsymbol{A}^{(\ell)\top} \boldsymbol{f}^{(\ell)} - \sum_{k=1}^K \boldsymbol{f}^{(k)\top} \boldsymbol{D}^{(k)} \boldsymbol{f}^{(k)} \end{aligned}$$

ゆえに、

$$\begin{aligned} & L(\{\boldsymbol{f}^{(k)}\}_{k=1}^K, \lambda) = \\ & \quad -J(\{\boldsymbol{f}^{(k)}\}_{k=1}^K) - \lambda \left(\sum_{k=1}^K \boldsymbol{f}^{(k)\top} \boldsymbol{D}^{(k)} \boldsymbol{f}^{(k)} - 1 \right) \end{aligned}$$

として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。

$$\sum_\ell \boldsymbol{A}^{(k)} \boldsymbol{A}^{(\ell)\top} \boldsymbol{f}^{(\ell)} = K(\lambda+1) \boldsymbol{D}^{(k)} \boldsymbol{f}^{(k)}$$

$\tilde{\lambda} \equiv K(\lambda+1)$ とすることで、以下の一般化固有値問題を得ることができる。

$$\boldsymbol{A}\boldsymbol{A}^\top \boldsymbol{f} = \tilde{\lambda} \boldsymbol{D} \boldsymbol{f}$$

ここで、$\boldsymbol{A}$、$\boldsymbol{D}$ と $\boldsymbol{f}$ は以下のように定義される。

$$\boldsymbol{A} \equiv \begin{bmatrix} \boldsymbol{A}^{(1)} \\ \boldsymbol{A}^{(2)} \\ \vdots \\ \boldsymbol{A}^{(K)} \end{bmatrix}, \boldsymbol{D} \equiv \begin{bmatrix} \boldsymbol{D}^{(1)} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{D}^{(K)} \end{bmatrix} \qquad (5)$$

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

最大固有値に対応する固有ベクトル $\boldsymbol{f}$ が、オブジェクトの、一次元空間での最適な埋め込み先である。R 次元の空間における埋め込み先 $\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_R$ を得るためには、固有値の大きい順に上位 R 個の固有ベクトルを取得すれば良い。

最後に、関係の最適な埋め込み先 (3) から示唆されることは、オブジェクトのある組み合わせ $o \equiv (s^{(1,i_1)}, s^{(2,i_2)}, \ldots, s^{(K,i_K)})$ の最適な埋め込み先の r 番目の次元は $\frac{1}{K} \sum_{k=1}^K [\boldsymbol{f}_r^{(k)}]_{i_k}$ で与えられること、および、関係 o の生じやすさは、

$$\sum_{r=1}^R \sum_{k=1}^K \left([\boldsymbol{f}_r^{(k)}]_{i_k} - \frac{1}{K} \sum_{k'=1}^K [\boldsymbol{f}_r^{(k')}]_{i_{k'}} \right)^2$$

に反比例することである。ゆえに、O に含まれない組み合わせのランクリストは、この値をソートすることで得られる。

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 $\{\boldsymbol{\Phi}^{(k)}\}_{k=1}^{K}$ を統合することを考える。

線形写像

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

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潜在空間に写像する $D^{(k)}$ 次元のパラメータである。

属性情報を活用しない場合と同様にして、

$$\sum_{\ell} \boldsymbol{\Phi}^{(k)\top}\boldsymbol{A}^{(k)}\boldsymbol{A}^{(\ell)\top}\boldsymbol{\Phi}^{(\ell)}\boldsymbol{w}^{(\ell)} = K\left(\lambda+1\right)\boldsymbol{\Phi}^{(k)\top}\boldsymbol{D}^{(k)}\boldsymbol{\Phi}^{(k)}\boldsymbol{w}^{(k)}$$

を得ることができ、これは、以下の一般化固有値問題として表現できる。

$$\boldsymbol{\Phi}^{\top}\boldsymbol{A}\boldsymbol{A}^{\top}\boldsymbol{\Phi}\boldsymbol{w} = \tilde{\lambda}\boldsymbol{\Phi}^{\top}\boldsymbol{D}\boldsymbol{\Phi}\boldsymbol{w} \tag{6}$$

ここで、

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

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

である。属性ベクトルの次元が高い場合、次元の呪いと呼ばれる効果により、予測性能がしばしば悪化する。これを避けるため、正の値である正則化パラメータ $\sigma > 0$ により、正則化項を追加するのが一般的である。この場合、我々の一般化固有値問題 (6) は以下のように修正される。

$$\boldsymbol{\Phi}^{\top}\boldsymbol{A}\boldsymbol{A}^{\top}\boldsymbol{\Phi}\boldsymbol{w} = \lambda\left(\boldsymbol{\Phi}^{\top}\boldsymbol{D}\boldsymbol{\Phi} + \sigma\boldsymbol{I}\right)\boldsymbol{w} \tag{7}$$

3. 実験

3.1 データセット

表 1 に、使用した三つのデータセットについて詳細を記載した。最初の二つのデータセットは Twitter というマイクロブログサービスから取得したものであり、ここでは「retweet」と「favoite」という二つの行為を対象とした。各行為は、行為主体ユーザ (subjective user)、発信元ユーザ (mentioned user)、URL という三種類のオブジェクトから構成されるタプルとして表現できる。各タプルにより表現されているのは、発信元ユーザの tweet によって投稿された URL に対して、行為主体ユーザが特定の行為（retweet/favorite）を行ったということである。三つ目のデータセットは delicious というソーシャルタギングサービスから取得したものであり、各行為は、ユーザ (user)、ユーザによって付与されたタグ (tag)、URL から成るタプルで表現される。提案手法では、関係データの他に表 1 に示されるような各オブジェクトの属性も活用した。

3.2 実験設定

実験条件として、導入で述べた、多項関係予測において問題となる二種類の疎な状況、すなわち、関係レベルで疎である状況とオブジェクトレベルで疎である状況を設定した。関係レベルで疎な設定では、全データセットから、観測された関係の一定割合をランダムにサンプリングし、それらを取り除いたものを訓練データとして用いる。残りのデータは評価データとして用いる。オブジェクトレベルで疎な設定では、観測されたオブジェクトの一定割合をランダムにサンプリングし、そのオブジェクトを含む全ての関係を取り除いたものを訓練データとして用いる。残りのデータは評価データとして用いる。それぞれの実験設定で、サンプリング比率を変え、各サンプリング比率についてサンプリング、予測、評価の一連のプロセスを 10 回繰り返した。（以降、この一連のプロセスをスロットと呼ぶ。）評価指標としては AUC を用いた。比較手法としては、CP 分解 (PARAFAC) と、Tucker 分解と呼ばれる二つの標準的なテンソル分解の手法を採用した。また、提案手法、比較手法それぞれについて、各サンプリング比率について、1 スロットを development data として用い、パラメータを AUC の観点からチューニングした。提案手法では、R は $\{16, 32, 64, 128, 256, 512\}$ のつ 6 を候補とした。提案手法で属性を活用する場合には σ を用い、属性を活用しない場合には用いなかった。σ は $\{10^{-2}, 10^{-3}, 10^{-4}\}$ で評価したが、このパラメータの範囲において安定した精度が確認されたため、実際の実験では、10^{-3} で固定した。比較手法においては R は、$\{1, 2, \ldots, 10\}$ までの候補内で十分にチューニングできることを確認したので、これら 10 個を候補とした。

3.3 結果

図 2 が関係レベルで疎な設定での実験結果である。観測された関係の割合を様々に変えたときの AUC の平均を、標準偏差付きで示してある。全データセットにおいて、提案手法がデータ過疎への最も高い頑強性を示していることが確認できる。また、提案手法が見せている相対的に小さい標準偏差は、大域解を保証する定式化によるものと考えられる。図 3 がオブジェクトレベルで疎な設定での実験結果である。属性を活用した提案手法のみ、データ過疎に対して相対的に高い頑強性を示した。この結果は、属性の活用が、コールドスタート問題を解決するにあたって有用であることを示唆している。

4. 関連研究

多項関係予測においてはテンソルを用いた手法がよく使用される。テンソル補完問題等のテンソル分析タスクにおいては、対象となるテンソルに対して低ランク性を仮定することが多く、様々な低ランク分解モデルが、効率的なアルゴリズムと共に提案されてきている [Kolda 09]。しかしながら、多くの既存手法で保証されるのは局所解のみであり、その予測性能は対象アルゴリズムに与える初期値に大きく依存する。対照的に、提案手法では、一般化固有値問題を一度解くだけで大域解を得ることが可能になる。

5. おわりに

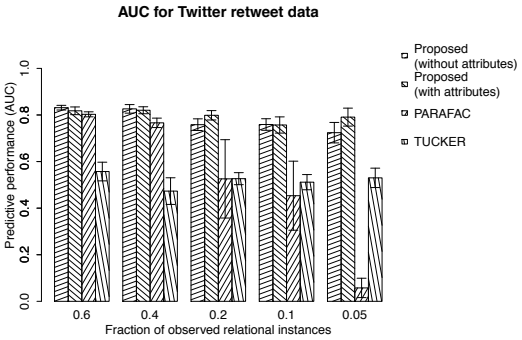
本論文では、異種のオブジェクト間に生じる多項関係を予測するために、関係データとオブジェクトの属性情報の両方を活用する手法を提案した。提案手法は、大域解を保証する定式化とオブジェクトの属性を活用する定式化により、標準的なテンソル分解と比較してデータ過疎への高い頑強性を示した。最後に、ソーシャル Web サービスにおけるユーザ行動予測の重要性や、本実験についての詳細は [Nori 12] を参照されたい。

参考文献

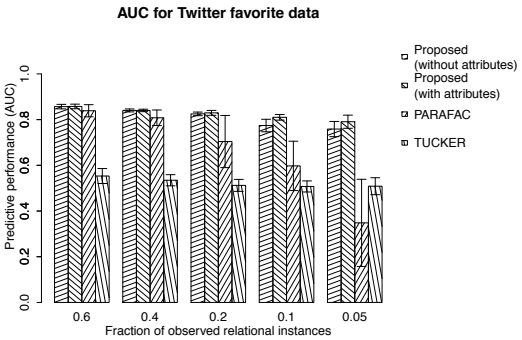
[Cai 11] Cai, Y., Zhang, M., Luo, D., Ding, C., and Chakravarthy, S.: Low-order tensor decompositions for social

表 1: 実験で用いたデータセットの詳細

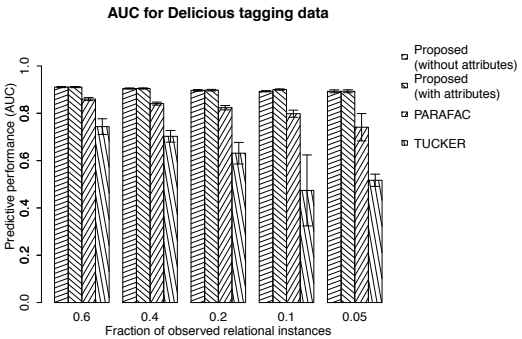
データセット	タプル数	オブジェクト	オブジェクト数	属性	属性数
Twitter	14,221	subjective user	1,144	ユーザの tweet に含まれるキーワード	4,896
		mentioned user	7,935	follow されているユーザ	2,586
(retweet)		URL	11,335	URL と共起した subjective user の tweet に含まれるキーワード	4,757
Twitter	22,755	subjective user	1,125	ユーザの tweet に含まれるキーワード	4,107
		mentioned user	10,049	follow されているユーザ	2,586
(favorite)		URL	18,244	URL と共起した subjective user の tweet に含まれるキーワード	4,107
Delicious	33,414	user	768	friend 関係にあるユーザ	1,098
		tag	8,280	URL と共起したタグ	15,088
(tagging)		URL	6,860	URL と共起したユーザ	1,185



(a) Twitter における Retweet アクション

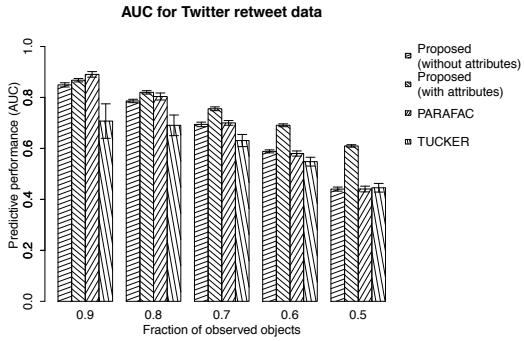


(b) Twitter における Favorite アクション

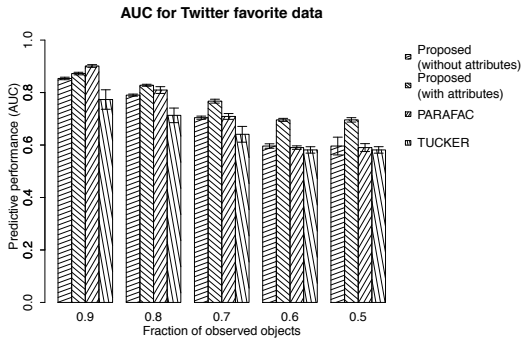


(c) Delicious における Tagging アクション

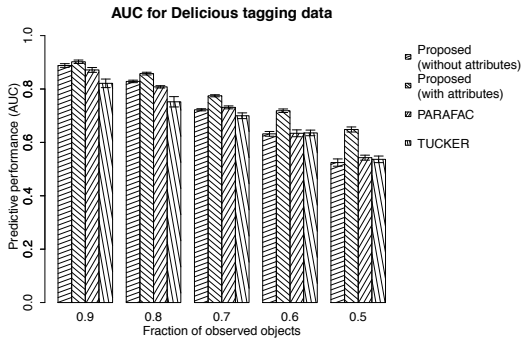
図 1: 関係レベルで疎な状況での予測性能の比較.



(a) Twitter における Retweet アクション



(b) Twitter における Favorite アクション



(c) Delicious における Tagging アクション

図 2: オブジェクトレベルで疎な状況での予測性能の比較.

tagging recommendation, WSDM '11, pp. 695–704 (2011)

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

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

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