Multinomial Relation Prediction in Social Data: A Dimension Reduction Approach

Nozomi Nori nozomi.nori@gmail.com
Danushka Bollegala danushka@iba.t.u-tokyo.ac.jp
Hisashi Kashima kashima@mist.i.u-tokyo.ac.jp
Graduate School of Information Science and Technology, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

Abstract

The recent popularization of social web services has made them one of the primary uses of the World Wide Web. An important concept in social web services is social actions such as making connections and communicating with others and adding annotations to web resources. Predicting social actions would improve many fundamental web applications, such as recommendations and web searches.

One remarkable characteristic of social actions is that they involve multiple and heterogeneous objects such as users, documents, keywords, and locations. However, the high-dimensional property of such multinomial relations poses one fundamental challenge, that is, predicting multinomial relations with only a limited amount of data.

In this paper, 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 as side information to address the "cold start" problem in multinomial relation prediction.

Experiments with various real-world social web service datasets demonstrate that the proposed method is more robust against data sparseness as compared to several existing methods, which can only find sub-optimal solutions.

Introduction

Rise of social data – Multinomial relations facilitate information flow on the contemporary web

The recent rapid popularization of social web services such as Twitter, Facebook, Tumblr, and Google+ has made them one of the major uses of the World Wide Web, along with web searches. One of the important concepts in social web services is *social actions*. Annotation is one of the typical social actions in social web services. Users can annotate various web resources, such as webpages, photos, and scientific literature, with keywords, and then share them with other users through social web services such as del.icio.us, flickr, and CiteULike. Users also forward resources to their friends or repeat information, for example, by "retweet"ing on Twitter and "reblogging" articles on Tumblr. Such actions reflect user preferences (Nori, Bollegala, and Ishizuka 2011a), perspectives (Sigurbjörnsson and van Zwol 2008), trust (Matsuo and Yamamoto 2009), and so on. An action by a user, such as informing his or her social network of some news, could trigger others' actions, such as commenting on it or forwarding it to others through their social networks. Such chains of actions facilitate information flow in the users' social networks. Understanding the context of users' actions, that is, how they are related to other users, their actions, and available resources, is the clue to understanding various tasks related to social web services (Lin et al. 2009).

Social actions also provide useful information to improve many web applications. For example, user annotations enhance (personalized) searches (Bao et al. 2007; Heymann, Koutrika, and Garcia-Molina 2008; Xu et al. 2008), inference of social relations (Schifanella et al. 2010), and the discovery of emerging ontologies (Mika 2005). Moreover, if we can *predict* such social actions, it would further extend their applicability. For example, a tag recommendation is one of the typical prediction tools for social actions, which suggests personalized tags with which each user can annotate each web resource. Tag recommendation automatically enriches resource information (Sigurbjörnsson and van Zwol 2008; Song et al. 2008) and improves the quality of information retrieval (Guan et al. 2009). Naveed et al. (2011) also reported that the accurate prediction of "retweet" actions improved search quality on Twitter.

In contrast with hyperlinks on the ordinary World Wide Web that connect two resources, social actions often involve more than two objects, such as multiple users, documents, keywords, and locations. This introduces complications and makes prediction tasks difficult.

A challenge in multinomial relation prediction: data sparsity

We address the *multinomial relation prediction* problem, which aims to predict relations involving multiple heterogeneous objects. Despite the increasing importance of multinomial relation prediction, a fundamental challenge faced in addressing this problem is prediction with sparse observations. Since the number of possible combinations of objects increases exponentially with respect to the number of objects involved in the relations, the number of observed relations is much smaller than the number of possible combinations in many social web applications. In fact, Cai et al. (2011) reported that the widely used datasets from social tagging services such as Last.fm, MovieLens, and del.icio.us are quite sparse, that is, less than 0.01% of the possible combinations of users, URLs, and tags are observed. They also pointed out that most objects are involved in only a small number of relations and that their numbers follow power-low distributions. In particular, we often encounter the latter sparseness in "cold start" situations, which is regarded as an important problem in recommender systems research (Schein et al. 2002).

Therefore, we need a precise relation prediction method that is robust to data sparseness. Recently, tensor decomposition methods (Kolda and Bader 2009) have often been used for relation prediction. However, they are usually formulated as non-convex optimization problems that suffer from local optimal solutions (especially when observations are sparse).

Proposed solution: binomial reduction, dimension reduction, and use of side information

We propose a new relation prediction method that is robust to data sparseness. Our proposed method is based on two ideas: (1) reduction from multinomial relations to sets of binomial relations, and low-dimensional embedding of the binomial relations for guaranteeing global optimal solutions, and (2) the use of attribute information of objects to cope with highly sparse data situations.

First, we transform each observed multinomial relation of K objects into a set of K binomial relations between the objects and the instance of the multinomial relation involving the objects. Figure 1 shows the transformation of a multinomial relation instance (`Alice`, `aaa.com`, `Bob`) into three binomial relations between the multinomial relation instance and the three objects (that are, `Alice`, `aaa.com`, and `Bob`). This transformation corresponds to the incidence matrix representation of a hypergraph (Voloshin 2009).

Next, we apply a nonlinear dimensionality reduction technique (Belkin and Niyogi 2003) to the binomial relations to embed the heterogeneous objects into a common latent space, so that each object and its participating relations are placed in close proximity in the latent space. The resultant optimization problem is formulated as a generalized eigenvalue problem that guarantees global optimal solutions. This results in robustness against data sparseness.

In addition, we exploit various attributes of the objects to tackle the "cold start" problem in the multinomial relation prediction problem. For example, a user might be represented by attributes such as age, gender, and occupation, whereas a URL might be associated with its domain and filetype. These attributes are helpful in situations where most of the objects are involved in only a small number of relations.

Finally, we empirically demonstrate the robustness of our proposed method using three real-world data sets, namely, the "retweet" and "favorite" actions on Twitter and tags for web pages on del.icio.us. The proposed method outperforms standard tensor decomposition methods in prediction accuracy in sparse data situations, such as (1) when only a small number of relations are available in the training phase or (2) when there exist a large number of new objects in the prediction phase.

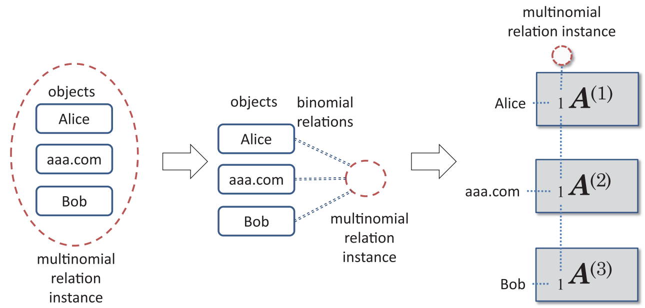


Figure 1: Transformation from a multinomial relation involving three ($K = 3$) different objects into a set of three binomial relations, which is further represented as elements of incidence matrices.

Relation prediction problem

We consider the problem of estimating the likelihood that a certain kind of relationship occurs among different kinds of objects, such as people and webpages, given some observed relations. For example, let us assume that we want to predict the preference of people to webpages recommended by someone. The goal is to predict the likelihood that the relation "$\texttt{person}_1$ likes the webpage at `URL` recommended by $\texttt{person}_2$" holds for each combination of ($\texttt{person}_1$, `URL`, $\texttt{person}_2$), given several known facts, such as that the relation holds for (`Alice`, `aaa.com`, `Bob`).

Let us assume we have K sets of objects $S^{(1)}, S^{(2)}, \dots, S^{(K)}$, each of which contains $N^{(k)}$ $(1 \leq k \leq K)$ objects. In the previous example, we can take $S^{(1)}$ and $S^{(3)}$ as people[1] and $S^{(2)}$ as a set of URLs. We denote the i-th object $s^{(k,i)} \in S^{(k)}$ by $s^{(k,i)}$. For example, $s^{(1,1)}$ can be `Alice`. We are also given $O \subset S^{(1)} \times S^{(2)} \times \cdots \times S^{(K)}$, which is a set of M observed relation instances. Each observed relation instance indicates that a certain relation (such as the `like` relation mentioned above) holds for a particular combination of objects. For example, $o^{(1)} \in O$ can be (`Alice`, `aaa.com`, `Bob`). Now, our goal is to predict the likelihood that the relation holds for each combination of objects not included in O. In other words, we require a ranked list of object combinations included in $(S^{(1)} \times S^{(2)} \times \cdots \times \dots S^{(K)}) \setminus O$.

In many realistic situations, each object is associated with information about itself. For example, each person has his or her demographic information and each webpage has its

[1] We consider $\texttt{person}_1$ and $\texttt{person}_2$ belong to different object sets, because they have different role in the relation. We can consider they belong to a same set when we model symmetric relations such as friendship relation in social networks.

content. Hence, we associate $s^{(k,i)}$ with a $D^{(k)}$-dimensional attribute vector $x^{(k,i)}$, and summarize these as a design matrix given by

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

for each $k = 1, 2, \ldots, K$.

The multinomial relation prediction problem that we focus on in this paper is summarized as follows.

Problem: Multinomial relation prediction

- **INPUT:**
  - $S^{(1)}, S^{(2)}, \ldots, S^{(K)}$: K sets of objects
  - O ($\subset S^{(1)} \times S^{(2)} \times \cdots \times \ldots S^{(K)}$): a set of M observed relation instances
  - $\Phi^{(1)}, \Phi^{(2)}, \ldots, \Phi^{(K)}$: K design matrices representing object attributes
- **OUTPUT:** A ranked list of object combinations not included in O, i.e., included in $(S^{(1)} \times S^{(2)} \times \cdots \times \ldots S^{(K)}) \setminus O$, sorted according to the likelihood of the relation

Proposed solution

One of the standard approaches to analyzing multinomial relation is to use tensor decomposition (Kolda and Bader 2009). However, most of the existing methods for tensor decomposition iteratively apply eigen-decomposition or gradient-based optimization, which do not guarantee global optimal solutions.

In this section, we propose a new multinomial relation prediction method using a dimension reduction technique, which can obtain global optimal solutions by solving a generalized eigenvalue problem only once. We then extend the proposed method to handle the attribute information of objects.

Multinomial relation prediction using dimension reduction

We start with the case where there is no attribute information. Our first key idea to guarantee global optimal solutions is to reduce multinomial relations to binomial relations. We create one binary matrix for each of the K types of objects to obtain K matrices in total. The matrices represent *relations between the objects and the relation instances*. Each element of the matrix indicates whether a particular object belongs to a particular relation instance. Let $A^{(k)}$ be an $N^{(k)} \times M$ binary matrix summarizing the participation of the objects in $S^{(k)}$ in the relation instances in O. Each element in $A^{(k)}$ is defined as

$$[A^{(k)}]_{n,m} \equiv \begin{cases} 1 & (\text{if } s^{(k,n)} \in S^{(k)} \text{ participates in } o^{(m)} \in O) \\ 0 & (\text{otherwise}). \end{cases}$$

Figure 1 depicts an example of this transformation.

Our second key idea is the low-dimensional embedding of the binomial relations represented by the K matrices. Using a similar idea to that in bipartite graph prediction using dimension reduction (Yamanishi 2009), we embed both the objects and relation instances into a common low-dimensional latent space to ensure close proximity between each object and its participating relations.

Let us first consider one-dimensional embedding. The objects $S^{(1)}$ of size $N^{(1)}$ are embedded as a vector $f^{(1)}$ of length $N^{(1)}$. Similarly, the objects $S^{(2)}, S^{(3)}, \ldots, S^{(K)}$ are embedded as $f^{(2)}, f^{(3)}, \ldots, f^{(K)}$, respectively. The observed relations O (of size M) are also embedded in the same one-dimensional latent space as $\bar{f}$ (of length M).

If an object $s^{(k,n)} \in S^{(k)}$ participates in a relation instance $o^{(m)} \in O$, we attempt to bring their embeddings $[f^{(k)}]_n$ and $[\bar{f}]_m$ close to each other, i.e., to make the Euclidean distance $([f^{(k)}]_n - [\bar{f}]_m)^2$ small. Therefore, the total objective function to be minimize is defined as

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

where $D^{(k)}$ is a diagonal matrix, whose (i,i)-th element is defined as $[D^{(k)}]_{i,i} \equiv \sum_j [A^{(k)}]_{i,j}$, which is the number of relation instances that the object $s^{(k,i)}$ participates in.

Since this objective function can easily be minimized by taking $f^{(k)} \equiv 0$ and $\bar{f} \equiv 0$, we impose the following additional scaling constraints to avoid the undesired solution.

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

The minimum of the objective function (1) with respect to $\bar{f}$ is obtained as

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

Plugging Eq. (3) into the negative of Eq. (1) results in the maximization of

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

Therefore, by maximizing the Lagrangian defined as

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

we obtain

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

By taking $\tilde{\lambda} \equiv K(\lambda + 1)$, a generalized eigenvalue problem can be obtained:

$$AA^\top f = \tilde{\lambda} D f,$$

where A, D, and f are defined as

$$A \equiv \begin{bmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(K)} \end{bmatrix}$$

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

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

The generalized eigenvector (GEV) f, corresponding to the largest generalized eigenvalue (GE), is the optimal one-dimensional embedding of the objects. To obtain R-dimensional embeddings, $f_1, f_2, \ldots, f_R$, we select the top-R GEs and GEVs.

Finally, the optimal embeddings of the relations (3) implies that the r-th dimension of the embedding of a combination of objects $o \equiv (s^{(1,i_1)}, s^{(2,i_2)}, \ldots, s^{(K,i_K)})$ is given as $\frac{1}{K}\sum_{k=1}^{K}[f_r^{(k)}]_{i_k}$, and the likelihood of a relation existing for o should be inversely proportional to

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

Therefore, a ranked list of combinations not included in O is obtained by sorting the scores.

It is notable that the optimization problem is reduced to the GE problem. Since GE problems are solved exactly despite its non-convex objective functions, we can obtain global optimal solutions. This is in contrast with the other tensor decomposition methods, which provide only locally optimal solutions. In addition, our method requires solving the GE problem only once, while most of the existing methods (Kolda and Bader 2009) require multiple calls to an eigensolver (and optimal solutions are not guaranteed).

Incorporating attribute information

Next, we consider incorporating the attribute information $\{\Phi^{(k)}\}_{k=1}^{K}$ into relation prediction. This is particularly important because using only the observed relations is insufficient for accurately predicting relations, especially in cases where observations are sparse. For example, some objects participate in none or only a few relations. In addition, we must sometimes make predictions about new objects not included in the original object sets.

Let us consider the linear projection model:

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

where $w^{(k)}$ is a $D^{(k)}$-dimensional parameter that projects a $D^{(k)}$-dimensional attribute vector to one-dimensional latent space. Similar to the case with only relation information, we obtain

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

which is summarized as the following GE problem:

$$\Phi^\top A A^\top \Phi w = \tilde{\lambda} \Phi^\top D \Phi w, \tag{5}$$

where

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

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

When the dimensionality of the attribute vectors is large, the predictive performance sometimes suffers because of the over-fitting of the data owing to the "curse of dimensionality" effect. To avoid this, it is common to add small regularization terms with a positive regularization parameter $\sigma > 0$. In our case, the GE problem (5) is modified to

$$\Phi^\top A A^\top \Phi w = \lambda\left(\Phi^\top D \Phi + \sigma I\right) w. \tag{6}$$

Experiments

Here, we show experimental results on multinomial relation prediction by using several datasets from social web services. Overall, the results demonstrate that our proposed method is quite robust against data sparsity.

Experimental Settings

Datasets. We used three datasets obtained from the two social web services summarized in Table 1, which details their relation types, numbers of observed relation instances, types of objects involved in the relations, and available object attributes and their dimensionality.

The first two datasets[2] were collected from the Twitter microblogging service: one for the prediction of "retweet" actions, and the other, for "favorite" actions. On Twitter, users can post short texts called tweets. "Retweet" is a function used to re-post other users' tweets, and users can also mark tweets as their "favorite". We used the "retweet" and "favorite" actions in our experiments. Each action consists of three objects: the user, the URL, and the original user, where the subject of the action is the user and the target is the URL posted by the original user. Starting from a specific user, we identified other users whose distances from the initial user were less than three in his or her social network. We identified the actions of users with time stamps between August 1 and 30, 2010. More details about crawling condition is described in our previous work (Nori, Bollegala, and Ishizuka 2011b; 2011a).

The other dataset, named "Delicious", deals with tagging actions carried out with the del.icio.us social tagging service, where a tagging action is represented as a tuple of a user, a tag, and a URL. This dataset is called "hetrec2011-delicious-2k"[3]. We extracted tagging actions with time stamps of August, 2010.

[2] Available from http://nozomi.shi-ba.org/datasets.html“ #TwitterActions2011.

[3] Available from http://www.grouplens.org/node/462“ #attachments.

In addition to the relation data, our method can deal with object attributes such as keywords that users have used in their tweets on Twitter. Table 1 summarizes the attributes we created for each object type. Except for "friends in contact" on Delicious, we extracted the top five attributes for each object. When constructing feature vectors, we used TF-IDF and scaled the ranges to fit into $[0.0, 1.0]$ for all the attributes.

Two data sparsity assumptions: relation-wise sparsity and object-wise sparsity. The main challenge we focus on in this paper is multinomial relation prediction in data-sparse situations, where only a small number of relation instances are available among numerous potential candidates. In our experiments, we consider two kinds of sparsity assumptions: one is *relation-wise sparsity* and the other is *object-wise sparsity*.

- Relation-wise sparsity assumes that some relation instances are missing at random.
- Object-wise sparsity assumes that all relation instances involving particular objects are completely missing.

The latter situation is known as the "cold start" problem in recommender systems. When a new user first joins an online shopping site, very little (or no) information about the user's actions is available. Hence it becomes quite difficult to make predictions for the user.

For the relation-wise sparsity setting, observed relations were randomly sampled from the entire dataset for model estimation, and the remaining data were used for performance evaluation. For the object-wise sparsity setting, we randomly sampled some objects, and all relations that did not involve the sampled objects were used as an evaluation dataset, with the remaining relation instances used for model estimation. In both the situations, we varied the sampling ratio, and we repeated the experimental procedure of sampling, prediction, and evaluation 10 times for each sampling ratio.

As an evaluation metric for predictive performance, we used the AUC (area under the ROC curve), which is widely used because it does not depend on the decision threshold.

Competing methods. We compared our proposed method with two standard tensor decomposition methods widely used for high-order relation analysis (Kolda and Bader 2009). One is PARAFAC/CANDECOMP which decomposes a tensor as a sum of rank-one tensors. The other is Tucker decomposition which decomposes a tensor into a core tensor and several factor matrices. We used *Tensor Toolbox*[4] as their implementations.

Parameter Settings. We optimized the hyperparameters in terms of the AUC by using a development dataset for each sampling ratio. For our method, we tuned R among 6 candidates, $\{16, 32, 64, 128, 256, 512\}$. For the proposed method with attributes, we tuned σ among $\{10^{-2}, 10^{-3}, 10^{-4}\}$, but the results were quite stable against this parameter, so we fixed σ as 10^{-3} in the experiments. For the competing methods, we tuned R among 10 candidates, $\{1, 2, \ldots, 10\}$.

[4]Available from http://csmr.ca.sandia.gov/˜tgkolda/TensorToolbox/.

Results

Our results show that the proposed methods (with and without attributes) are quite robust to data sparseness in relation-wise sparse situations and that the use of attributes is particularly effective in the object-wise sparse cases.

Relation-wise sparse situations. Figure 2 shows the averaged AUCs with standard deviations for the three datasets in relation-wise sparse situations with various observation ratios. For all the datasets, our proposed method showed the highest robustness against data sparseness. The small standard deviation of the proposed method shows the stability of its formulation guaranteeing globally optimal solutions.

Object-wise sparse situations. Figure 3 shows the results in object-wise sparse situations. Since we have no relational information for the test objects, the proposed method without attribute information performed poorly, while the proposed method with attribute information maintained relatively high performance, even with low observation ratios. The results show that the use of object information can mitigate the "cold start" problem in multinomial relation prediction.

Related Work

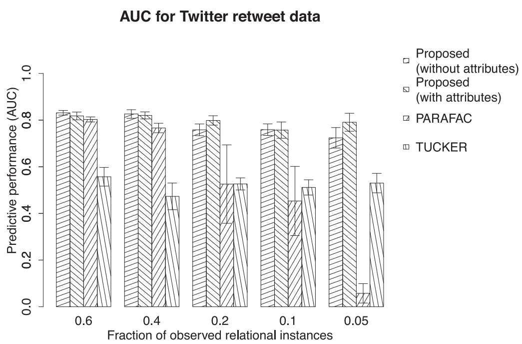
Tensor analysis is a popular approach for dealing with multinomial relations. In various tensor analysis tasks, including tensor completion, target tensors are often assumed to be of low rank, and various low-rank tensor decomposition models with efficient algorithms (Kolda and Bader 2009) have been proposed. In web mining research, Symeonidis et al. (2008) and Rendle et al. (2009; 2010) proposed tag recommendation methods using tensor decomposition. However, most of the existing methods guarantee only local optimal solutions and their quality depends highly on their initial values. On the other hand, our proposed method can obtain global optimal solutions by solving a generalized eigenvalue problem only once. This property of the proposed method, added to the incorporation of the attribute information of objects, makes it more robust and stable against data sparsity.

Conclusion

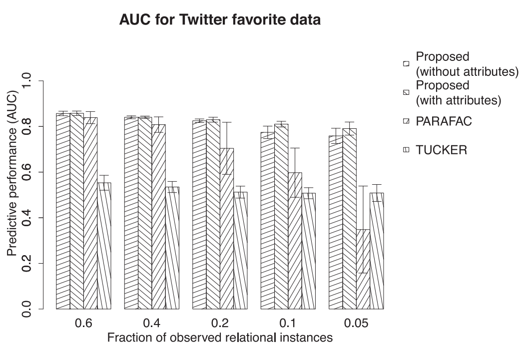
In this paper, we proposed a method to predict multinomial relations among heterogeneous objects using both relation information and object attribute information. In contrast with the existing relation prediction methods based on tensor decomposition, the proposed method is highly robust to observation sparsity because it deals with object attributes as well as relation information and because its formulation guarantees globally optimal solutions.

Table 1: Summary of the datasets used in the experiments.

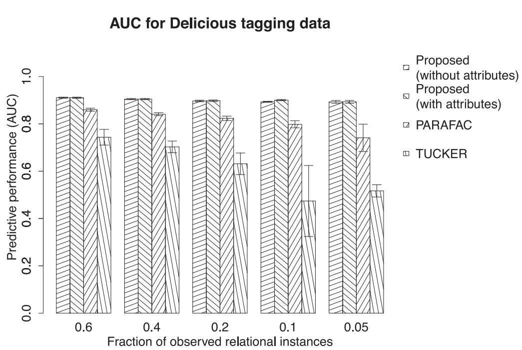
Dataset	# of observations	Objects	# of objects	Attributes	# of attributes
Twitter	14,221	subjective user	1,144	keywords extracted from the user's tweets	4,896
		mentioned user	7,935	followers of the user	2,586
(retweet)		URL	11,335	keywords extracted from subjective users co-occurring with the URL	4,757
Twitter	22,755	subjective user	1,125	keywords extracted from the user's tweets	4,107
		mentioned user	10,049	followers of the user	2,586
(favorite)		URL	18,244	keywords extracted from subjective users co-occurring with the URL	4,107
Delicious	33,414	user	768	friends in contact	1,098
		tag	8,280	URLs co-occurring with the tag	15,088
(tagging)		URL	6,860	users co-occurring with the URL	1,185



(a) "Retweet" actions on Twitter (microblogging)

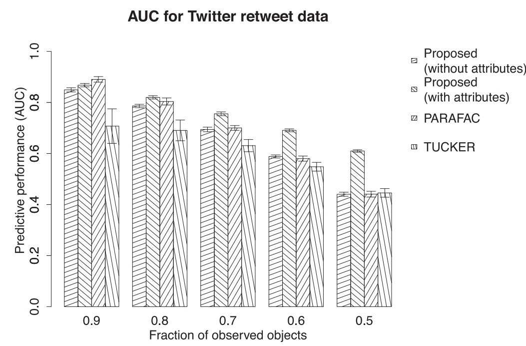


(b) "Favorite" actions on Twitter (microblogging)

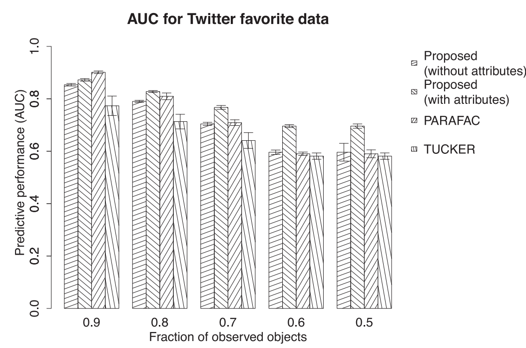


(c) Tagging actions on Delicious (social tagging)

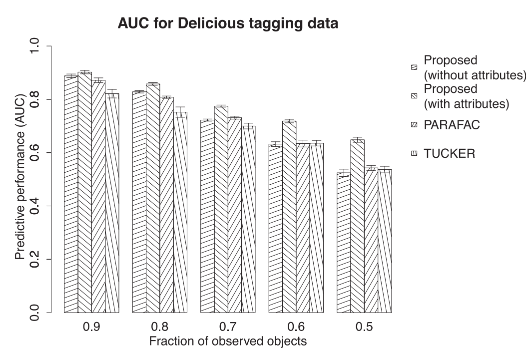
Figure 2: Comparison of predictive performance in *relation-wise sparse* situations. The proposed method showed high robustness against data sparseness.



(a) "Retweet" actions on Twitter (microblogging)



(b) "Favorite" actions on Twitter (microblogging)



(c) Tagging actions on Delicious (social tagging)

Figure 3: Comparison of predictive performance in *object-wise sparse* situations. The proposed method using attribute information maintains relatively high performance even with many missing objects.

References

Bao, S.; Xue, G.; Wu, X.; Yu, Y.; Fei, B.; and Su, Z. 2007. Optimizing web search using social annotations. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, 501–510.

Belkin, M., and Niyogi, P. 2003. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* 15(6):1373–1396.

Cai, Y.; Zhang, M.; Luo, D.; Ding, C.; and Chakravarthy, S. 2011. Low-order tensor decompositions for social tagging recommendation. In *Proceedings of the fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, 695–704.

Guan, Z.; Bu, J.; Mei, Q.; Chen, C.; and Wang, C. 2009. Personalized tag recommendation using graph-based ranking on multi-type interrelated objects. In *Proceedings of the 32st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '09, 540–547.

Heymann, P.; Koutrika, G.; and Garcia-Molina, H. 2008. Can social bookmarking improve web search? In *Proceedings of the first International Conference on Web Search and Web Data Mining*, WSDM '08, 195–206.

Kolda, T. G., and Bader, B. W. 2009. Tensor decompositions and applications. *SIAM Review* 51(3):455–500.

Lin, Y.-R.; Sun, J.; Castro, P.; Konuru, R.; Sundaram, H.; and Kelliher, A. 2009. Metafac: community discovery via relational hypergraph factorization. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, 527–536.

Matsuo, Y., and Yamamoto, H. 2009. Community gravity: Measuring bidirectional effects by trust and rating on online social networks. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, 751–760.

Mika, P. 2005. Ontologies are us: A unified model of social networks and semantics. In *Proceedings of the fourth International Semantic Web Conference*, volume 3729 of *ISWC '05*, 522–536. Springer.

Naveed, N.; Gottron, T.; Kunegis, J.; and Alhadi, A. C. 2011. Searching microblogs: Coping with sparsity and document quality. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11.

Nori, N.; Bollegala, D.; and Ishizuka, M. 2011a. Exploiting user interest on social media for aggregating diverse data and predicting interest. In *Proceedings of the fifth International Conference on Weblogs and Social Media*, ICWSM '11, 241–248.

Nori, N.; Bollegala, D.; and Ishizuka, M. 2011b. Interest prediction on multinomial, time-evolving social graphs. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, IJCAI '11, 2507–2512.

Rendle, S., and Schmidt-Thieme, L. 2010. Pairwise interaction tensor factorization for personalized tag recommendation. In *Proceedings of the third ACM International Conference on Web Search and Data Mining*, WSDM '10, 81–90.

Rendle, S.; Balby Marinho, L.; Nanopoulos, A.; and Schmidt-Thieme, L. 2009. Learning optimal ranking with tensor factorization for tag recommendation. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, 727–736.

Schein, A. I.; Popescul, A.; Popescul, R.; Ungar, L. H.; and Pennock, D. M. 2002. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '02, 253–260.

Schifanella, R.; Barrat, A.; Cattuto, C.; Markines, B.; and Menczer, F. 2010. Folks in folksonomies: social link prediction from shared metadata. In *Proceedings of the third ACM International Conference on Web Search and Data Mining*, WSDM '10, 271–280.

Sigurbjörnsson, B., and van Zwol, R. 2008. Flickr tag recommendation based on collective knowledge. In *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, 327–336.

Song, Y.; Zhuang, Z.; Li, H.; Zhao, Q.; Li, J.; Lee, W.-C.; and Giles, C. L. 2008. Real-time automatic tag recommendation. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, 515–522.

Symeonidis, P.; Nanopoulos, A.; and Manolopoulos, Y. 2008. Tag recommendations based on tensor dimensionality reduction. In *Proceedings of the second ACM Conference on Recommender Systems*, RecSys '08, 43–50.

Voloshin, V. I. 2009. *Introduction to Graph and Hypergraph Theory*. Nova Kroshka Books.

Xu, S.; Bao, S.; Fei, B.; Su, Z.; and Yu, Y. 2008. Exploring folksonomy for personalized search. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, 155–162.

Yamanishi, Y. 2009. Supervised bipartite graph inference. In *Proceedings of the 22nd Annual Conference on Neural Information Processing Systems*, NIPS '09, 1841–1848.