An Efficient Tagging Data Interpretation and Representation Scheme for Item Recommendation

Noor Ifada[1,2], Richi Nayak[1]
[1]School of Electrical Engineering and Computer Science
Queensland University of Technology, Australia
[2]Informatics Department
University of Trunojoyo Madura, Indonesia
`noor.ifada@{qut.edu.au, if.trunojoyo.ac.id}, r.nayak@qut.edu.au`

Abstract

A tag-based item recommendation method generates an ordered list of items, likely interesting to a particular user, using the users past tagging behaviour. However, the users tagging behaviour varies in different tagging systems. A potential problem in generating quality recommendation is how to build user profiles, that interprets user behaviour to be effectively used, in recommendation models. Generally, the recommendation methods are made to work with specific types of user profiles, and may not work well with different datasets. In this paper, we investigate several tagging data interpretation and representation schemes that can lead to building an effective user profile. We discuss the various benefits a scheme brings to a recommendation method by highlighting the representative features of user tagging behaviours on a specific dataset. Empirical analysis shows that each interpretation scheme forms a distinct data representation which eventually affects the recommendation result. Results on various datasets show that an interpretation scheme should be selected based on the dominant usage in the tagging data (i.e. either higher amount of tags or higher amount of items present). The usage represents the characteristic of user tagging behaviour in the system. The results also demonstrate how the scheme is able to address the cold-start user problem.

Keywords: tagging, data interpretation, data representation, item recommendation, user profile, cold-start user

1 Introduction

Learning users past tagging behaviour is essential to generate quality recommendations (Ifada and Nayak, 2014; Rendle et al., 2009). In social tagging systems, the user tagging behaviour can be inferred from the ternary relation between users, items, and tags. Users annotate items of their interest by using freely-defined tags. The task of tag-based item recommendation system is to predict an ordered list of items to a user based on users tagging behaviour. In real practice, users can use different tags to annotate the same item as well as the same tag can be used for annotating different items. This approach of freely using tags emphasizes on implementing a personalization approach in the recommendation system by using user past tagging behaviour to build the user profile.

A typical tag-based recommendation system customarily interprets the observed data as positive entries. Observed data is the state which users have expressed their interest to items by annotating those items using tags, as illustrated in Figure 1. On the contrary, how should the non-observed data be interpreted, it remains disputed. Non-observed data is a mixture of the following two states: (1) user is not interested with the items – negative entries, and (2) user might be interested to the items in the future – null values. The selection of interpretation scheme is crucial at this stage as the user profiles will be defined based on this representation.

Tensor modelling is a natural approach to represent and analyse the latent relationships inherent in a three-dimensional tagging data model (Ifada and Nayak, 2014; Symeonidis, Nanopoulos and Manolopoulos, 2010). A tensor model is commonly built by implementing the *boolean* scheme to interpret the data (Symeonidis, Nanopoulos and Manolopoulos, 2008). Rendle et al. (2009) have identified drawbacks of using the *boolean* scheme and proposed a *set-based* scheme which has shown more accurate interpretation of the tagging data. However, this *set-based* scheme is customized to the recommendation method proposed. It does not detail how various kinds of *set-based* schemes can be selected according to data characteristics for a recommendation method.

The *set-based* scheme creates an object pairwise ranking representation between the positive entries interpreted from the observed tagging data and the negative entries interpreted from the non-observed data, by two means: (1) user-item set, or (2) user-tag set. These set data represents users tagging behaviours in the system. The user-item set expresses that a user can use multiple tags to annotate an item. From this point of view, the pairwise ranking representation is created by using tags as the pairwise ranking objects. Accordingly, for each user-item set, the pairwise ranking is generated to represent that the user is more favourable to use tags of positive entries than those of negative entries to annotate an item. Alternatively, the user-tag set indicates that a user can use the same tag to annotate different items and, consequently, the items become the pairwise ranking objects for the data representation. For each user-tag set, the pairwise ranking is generated to represent that the user is more favourable to annotate items of positive entries than those of negative entries using a tag.

The set selection is influenced by the underlying recommendation task, i.e. using the user-item set representation for tag recommendation and the user-tag set for item recommendation. However, in practice, not all recommendation methods can perform their best when they are implemented on different datasets (Gemmell et al., 2011; Ifada and Nayak, 2014; Rendle et al., 2009). Given that users tagging behaviours are captured and represented differently in each social tagging system, the data set to be used in a recommendation system should be selected based on the feature set that defines the representation of user tagging behaviour in that system, i.e. the dominant set observed from the tagging data. The user-item set is more dominant than the user-tag set when the users prefer to use less number of tags in annotating items. On the contrary, the user-tag set is considered more dominant than the user-item set when the users prefer to use more number of tags in annotating items.

The consequence of restricting the selection of interpretation scheme to the task of recommendation is that the method could possibly outperform other methods when implemented to a certain tagging data, however, it yields poor results when it is implemented on a different tagging data (Ifada and Nayak, 2014; Rendle and Schmidt-Thieme, 2010). Another limitation is that a recommendation method could not possibly be implemented with an appropriate scheme, i.e., generating item recommendation by implementing the user-tag set scheme while the dominant feature of the tagging data is user-item set (Gemmell et al., 2011), or generating tag recommendation by implementing the user-item set scheme when the dominant feature is user-tag set (Rendle et al., 2009).

This paper investigates and compares several tagging data interpretation and representation schemes that can affect the performance of tag-based item recommendation systems. These schemes vary in manners how they highlight the representative features of the user tagging behaviours on different systems. Six tagging data interpretations and representations, which are constructed using the *set-based* ranking scheme, are proposed. We analyse how each interpretation scheme forms a distinct data representation which eventually affects the recommendation results.

To test and evaluate the concept, we needed to implement the proposed representations on a recommendation method which employs a *set-based* scheme for building and learning the tensor model. We found no item recommendation method that can be used in experiments. Pairwise Interaction Tensor Factorization (PITF) (Rendle and Schmidt-Thieme, 2010), a tensor based tag recommendation method, uses the *set-base*d scheme and has reported high accuracy and scalability performance. We adopted this tag recommendation method to recommend item and named it as Pairwise Interaction Tensor Factorization for Item Recommendation (PITF-I). The tag and item recommendations are two distinct tasks. Tag recommendations are generated with two specified dimensions, i.e. user and item, while the item recommendations are made with only users identity specified and, therefore, the item ranking scores must be calculated for the whole available tags before being sorted as a list of top-*N* recommendation. In this paper, we customize the PITF method so that it is applicable for both user-item and user-tag set schemes. We then show how to select the best set of interpretation scheme to be used for different datasets, given that users tagging behaviours in the different tagging systems are different from one another.

The contribution of this paper is as follows: (1) introducing comprehensive data interpretation schemes to generate user profile representation on tensor models for tag-based item recommendation, (2) proposing a process of selecting the best interpretation scheme to be used on a certain dataset, (3) adapting a pairwise ranking tensor factorization method for implementing various interpretation schemes for tag-based item recommendation, (4) establishing that an efficient user profile presentation is more important than just simply trying to get more dense data representation to generate quality recommendations, and (5) finally, showing how an efficient interpretation scheme is able to address the cold-start user problem.

The remainder of this paper is organized as follows. Section 2 details the related works. Section 3 describes briefly about the basics in the tag-based item recommendation. Section 4 explains the proposed interpretation and representation schemes, a brief description about the method used, and the process of selecting the best interpretation scheme to be used. Section 5 presents the experimental results based on real world datasets. Section 6 concludes the paper.

2 Related Work

Recommendation is a well-established research area (Adomavicius and Tuzhilin, 2005; Zhang, Zhou and Zhang, 2011). In the last few years, tensor modelling (Kolda and Bader, 2009), a well-known approach to represent and analyse latent relationships inherent in multi-dimensions data, is adapted in recommendation systems. The tensor modelling based recommendation methods have shown improved results over the matrix based methods (Rafailidis and Daras, 2013; Symeonidis, Nanopoulos and Manolopoulos, 2010). Existing tensor methods solely interpret the tagging data using *boolean* scheme to build the model and then directly use the tensor reconstruction results from the factorized tensor to generate the recommendations. The *boolean* scheme simply interprets the positive observed tagging data as 1 and the non-observed ones as 0. This scheme fits both the negative and null values as 0 which makes it difficult to predict the ranking list in the future (Rendle et al., 2009).

In other words, by using the *boolean* scheme, these approaches ignore the user's past tagging activities that have been found most influencive in forming user likelihood for matrix-based recommended methods (Kim et al., 2010). A recent work (Ifada and Nayak, 2014) solves this problem by ranking the reconstructed tensor results utilising the past collaborative data and make the final recommendations. A pairwise tensor factorization model (Rendle and Schmidt-Thieme, 2010), for recommending tags (unlike our paper that recommends items), has also been proposed to solve the recommendation ranking problem. This method uses the *set-based* ranking interpretation scheme when building the tensor.

The *set-based* ranking scheme distinguishes the interpretation between observed and non-observed data. The representation creates pairwise ranking objects between the positive entries interpreted from the observed data and the negative entries interpreted from the non-observed data. The rest of other entries are left as null values. Within each set, the positive entries are assigned higher values than the negative ones (instead of assigning a fixed numeric values to both entries). This scheme interprets the null values as rankings that can be predicted in the future, unlike the *boolean* scheme that over fits the null values using the negative examples as 0. The model based on the *boolean* scheme tries to learn and predict a 0 for each of the negative and null case (Rendle et al., 2009). By using the *set-based* ranking scheme, entries derived by the factorised tensor model can directly be used for generating recommendations. Selecting the appropriate interpretation scheme for tensor modelling is crucial as the tensor model has to learn the total interaction between users, items, and tags represented by the data using this scheme, as well as, the model has to expose the latent relationship among those dimensions to be used for generating the recommendations.

In this paper, we investigate the process of selecting an interpretation scheme for tagging data to improve the performance of tag-based item recommendation systems on various social tagging systems that employ a system specific method to collect user tagging behaviour. As users tagging behaviours in different tagging systems are different from one another, recommendation methods are usually not able to generalize their outperformance if they are implemented on different datasets. Therefore, though the methods are able to show that they outperform their benchmark methods on a dataset, yet they do not show the similar performance when applied on another dataset (Ifada and Nayak, 2014; Rendle and Schmidt-Thieme, 2010).

We demonstrate how different interpretation scheme forms different data representation which eventually affects the recommendation results. Moreover we modify the pairwise tensor factorization method (Rendle and Schmidt-Thieme, 2010), designed for tag recommendation, to generate an ordered list of item recommendations using the six proposed tagging data interpretation and representation schemes.

To the best of our knowledge, this is the first paper that studies the data interpretation schemes for tensor models for generating tag-based item recommendations, in detail.

3 Tag-based Item Recommendation

The task of tag-based item recommendation is to generate an ordered list of items that might be of interest to a user using the collaborative tags. The list of recommended items can be learned from user past tagging behaviour inferred from observed and non-observed tagging data.

3.1 Tagging Data

Let $U = \{u_1, u_2, u_3, \dots, u_{|U|}\}$ be the list of all users, $I = \{i_1, i_2, i_3, \dots, i_{|I|}\}$ be the list of all items, and $T = \{t_1, t_2, t_3, \dots, t_{|T|}\}$ be the list of all tags. Tagging data forms a ternary relationship between users, items, and tags. The observed tagging data can be denoted as $A \subseteq U \times I \times T$, where a vector of $(u, i, t) \in A$ represents the tagging activity of user u who has tagged item i using tag t. The user-vocabulary V_u denotes the list of distinct tags that have been used by user u to annotate any items:

$$V_u = \{t|(u,*,t) \in A\}$$

Whereas the user-collection C_u denotes the list of distinct items that have been tagged by user u using any tags:

$$C_u = \{i|(u,i,*) \in A\}$$

The tagging data can be naturally modelled as a three-dimensional tensor of $\mathcal{Y}^{U\times I\times T}$. Figure 1 illustrates a tensor model representing a toy example of the observed tagging data, $\mathcal{Y}^{3\times 4\times 5}$ where $U = \{u_1, u_2, u_3\}$, $I = \{i_1, i_2, i_3, i_4\}$, and $T = \{t_1, t_2, t_3, t_4, t_5\}$. Each slice of the tensor represents a user matrix which contains the user tag usage for an item.

For generating tag-based item recommendations, the ranking representation of tagging data can be inferred from either using the user-item or user-tag sets. The user-item (u, i) sets are all distinct user-item combinations in A as a user can annotate an item with multiple tags. The user-tag (u, t) sets are all distinct user-tag combinations in A since a user can use the same tag to annotate multiple items.

3.2 Interpretation Scheme

The *boolean* scheme interprets the positive observed tagging data as 1 and denotes any other data as 0. Consequently, the user profile is only made from those positive entries exist in A. This scheme unfortunately overfits the negative entries and the null values as the same 0 value (Rendle et al., 2009).

Figure 2 shows an example of the scheme and its representation to build the user profile generated from the tagging data in Figure 1. For ease of illustration, it is only showing entries for User 1. Figure 1 shows that User 1 (u_1) has revealed his interest to i_2 and i_3 by annotating them using tags $\{t_1, t_3\}$ and $\{t_1, t_4\}$, respectively.

In contrast to the *boolean* scheme, the *set-based* ranking interpretation scheme distinguishes the positive, negative, and null values. It creates a pairwise classification ranking from the positive entries interpreted from the observed data and the negative entries interpreted from the non-observed data, on each user-item or user-tag set. In this case, the missing or null values are interpreted as the entries to be predicted for generating recommendations.

For representing the ranking within each set, the positive entries are simply given higher values than the negative ones. This indicates that the user favours the positive entries more than the negative entries (Rendle et al., 2009). The ranking order can be learned from A by creating: (1) tag-pairwise ranking on each distinct user-item set (u, i, t_P, t_N); or (2) item-pairwise ranking on each distinct user-tag set (u, t, i_P, i_N).

4 Proposed Tagging Data Interpretation and Its Representation

Given that users tagging behaviours in the different tagging systems are different from one another, the big question is how to best interpret the users tagging data as the user profile will be defined based on this representation. This sections details the six proposed tagging data interpretations and representations schemes, the recommendation method, as well as the process of selecting the best set of interpretation scheme to be used. The map of the interpretation schemes is represented in Figure 3 and the detailed examples are described in Figure 4.

4.1 Tag-Pairwise Ranking on User-Item Set

The tag-pairwise ranking on user-item (u, i) set is the ranking between the tags of positive entries (t_P) and the tags of negative entries (t_N), inferred from each $(u, i) \in A$. The tags of positive entries can easily be derived from the observed data.

Let us consider the toy example as shown in Figure 1. For User 1 (u_1), the tags of positive entries generated from the (u_1, i_1), (u_1, i_2), (u_1, i_3), and (u_1, i_4) sets are: $t_{P(u_1,i_1)} = \emptyset$, $t_{P(u_1,i_2)} = \{t_1, t_3\}$, $t_{P(u_1,i_3)} = \{t_1, t_4\}$, and $t_{P(u_1,i_4)} = \emptyset$, respectively. However, finding the tag values for the non-observed or negative entries is difficult.

We propose to infer and represent these tag values using the following schemes: (a) *all-tag*, (b) *user-vocabulary*, or (c) *non-user-vocabulary*. The ranking function can be formulated as $f(u, i, t_P, t_N) \rightarrow \mathbb{R}$.

4.1.1 All-Tag Pairwise Ranking

The *all-tag* ranking scheme interprets a negative entry (t_N) as follows. For each $(u, i) \in A$, user u is less favourable to annotate item i using any tags other than those appearing in positive entries $(T \backslash t_P)$ (Rendle and Schmidt-Thieme, 2010). The representation of this can be formulated as:

$$D = \{(u, i, t_P, t_N): (u, i, t_P) \in A \wedge (u, i, t_N) \notin A \wedge t_N \in T \backslash t_P\}$$

In Figure 4, the positive entries show that u_1 has revealed his interest for i_2 by tagging the item using tags $\{t_1, t_3\}$. Given $T = \{t_1, t_2, t_3, t_4, t_5\}$, $t_{P(u_1,i_2)} = \{t_1, t_3\}$, and $T \backslash t_{P(u_1,i_2)} = \{t_2, t_4, t_5\}$, this tagging data is interpreted as u_1 favours $\{t_1, t_3\}$ more than $\{t_2, t_4, t_5\}$ to annotate i_2. The representation can then be generated from the pairwise ranking of $t_P = \{t_1, t_3\}$ and $t_N = \{t_2, t_4, t_5\}$ on (u_1, i_2) set.

4.1.2 User-Vocabulary Pairwise Ranking

The *user-vocabulary* ranking scheme interprets a negative entry (t_N) as follows. For each $(u, i) \in A$, user u is less favourable to annotate item i using any tags of user-vocabulary (V_u) than those appearing in positive entries $(V_u \backslash t_P)$. The representation of this can be formulated as:

$$D = \{(u, i, t_P, t_N): (u, i, t_P) \in A \wedge ((u, i, t_N) \notin A \wedge t_N \in V_u \backslash t_P)\}$$

In Figure 4, the positive entries show that u_1 has revealed his interest for i_2 using tags $\{t_1, t_3\}$ and for i_3 using tags $\{t_1, t_4\}$. Knowing $V_{u_1} = \{t_1, t_3, t_4\}$, $t_{P(u_1,i_2)} = \{t_1, t_3\}$, and $V_{u_1} \backslash t_{P(u_1,i_2)} = \{t_4\}$, this tagging data is interpreted as u_1 favours $\{t_1, t_3\}$ more than $\{t_4\}$ to annotate i_2. The representation can then be generated from the pairwise ranking of $t_P = \{t_1, t_3\}$ and $t_N = \{t_4\}$ on (u_1, i_2) set.

4.1.3 Non-User-Vocabulary Pairwise Ranking

The *non-user-vocabulary* ranking scheme interprets a negative entry (t_N) as follows. For each $(u, i) \in A$, user u is less favourable to annotate item i using other tags that have not been used in any other items $(T \backslash V_u)$. The representation of this can be formulated as:

$$D = \{(u, i, t_P, t_N): (u, i, t_P) \in A \wedge ((u, i, t_N) \notin A \wedge t_N \in T \backslash V_u)\}$$

In Figure 4, the positive entries show that u_1 has revealed his interest for i_2 using tags $\{t_1, t_3\}$ and for i_3 using tags $\{t_1, t_4\}$. Given $T = \{t_1, t_2, t_3, t_4, t_5\}$, $V_{u_1} = \{t_1, t_3, t_4\}$ and $T \backslash V_{u_1} = \{t_2, t_5\}$, this tagging data is interpreted as u_1 favours $\{t_1, t_3\}$ more than $\{t_2, t_5\}$ to annotate i_2. The representation can then be generated from the pairwise ranking of $t_P = \{t_1, t_3\}$ and $t_N = \{t_2, t_5\}$ on (u_1, i_2) set.

4.2 Item-Pairwise Ranking on User-Tag Set

The item-pairwise ranking on user-tag (u, t) set is the ranking between the items of positive entries (i_P) with the items of negative entries (i_N) inferred from each $(u, t) \in A$. The items of positive entries can easily be derived from the observed data.

Let us consider the toy example as shown in Figure 1. For User 1 (u_1), the items of positive entries generated from the (u_1, t_1), (u_1, t_2), (u_1, t_3), (u_1, t_4), and (u_1, t_5) sets are: $i_{P(u_1,t_1)} = \{i_2, i_3\}$, $i_{P(u_1,t_2)} = \emptyset$, $i_{P(u_1,t_3)} = \{i_2\}$, $i_{P(u_1,i_4)} = \{i_3\}$, and $t_{P(u_1,t_5)} = \emptyset$, respectively. However, finding the item values for the non-observed or negative entries is difficult.

The item values are inferred and represented using the following schemes: (a) *all-item*, (b) *user-collection*, or (c) *non-user-collection*. The ranking function can be formulated as $f(u, t, i_P, i_N) \rightarrow \mathbb{R}$.

4.2.1 All-Item Pairwise Ranking

The *all-item* ranking scheme interprets a negative entry (i_N) as follows. For each $(u, t) \in A$, user u is less favourable to use tag t to annotate any items other than those appearing in positive entries $(I \backslash i_P)$ (Gemmell et al., 2011). The representation of this interpretation can be formulated as:

$$D = \{(u, t, i_P, i_N): (u, t, i_P) \in A \wedge (u, t, i_N) \notin A \wedge i_N \in I \backslash i_P\}$$

In Figure 4, the positive entries show that u_1 has used t_1 to reveal his interest for items $\{i_2, i_3\}$. Given $I = \{i_1, i_2, i_3, i_4\}$ and $i_{P(u_1,t_1)} = \{i_2, i_3\}$ so that $I \backslash i_{P(u_1,t_1)} = \{i_1, i_4\}$, this tagging data is interpreted as u_1 favours $\{i_2, i_3\}$ more than $\{i_1, i_4\}$ to be annotated using t_1. The representation can then be generated from the pairwise ranking of $i_P = \{i_2, i_3\}$ and $i_N = \{i_1, i_4\}$ on (u_1, t_1) set.

4.2.2 User-Collection Pairwise Ranking

The *user-collection* ranking scheme interprets a negative entry (i_N) as follows. For each $(u, t) \in A$, user u is less favourable to use tag t to annotate any items of user-collection (C_u) than those appearing in positive entries ($C_u \backslash i_P$). The representation of this interpretation can be formulated as:

$$D = \{(u, i_P, i_N, t) : (u, i_P, t) \in A \wedge ((u, i_N, t) \notin A \wedge i_N \in C_u \backslash i_P)\}$$

In Figure 4, the positive entries show that u_1 has used t_1 to reveal his interest for items $\{i_2, i_3\}$, t_3 for $\{i_2\}$, and t_4 for $\{i_3\}$. Knowing $C_{u_1} = \{i_2, i_3\}$ and $i_{P(u_1,t_1)} = \{i_2, i_3\}$ so that $C_{u_1} \backslash i_{P(u_1,t_1)} = \emptyset$, this tagging data is interpreted as u_1 has no other favours than $\{i_2, i_3\}$ to be annotated using t_1. The representation then cannot be generated on (u_1, t_1) set. On the other hand, a representation can be generated on (u_1, t_3) set as u_1 has only used t_3 to annotate $\{i_2\}$. The tagging data is interpreted as u_1 favours $\{i_2\}$ more than $\{i_3\}$ to be annotated using t_3. The representation of this is the pairwise ranking of $i_P = \{i_2\}$ and $i_N = \{i_3\}$.

4.2.3 Non-User-Collection Pairwise Ranking

The *non-user-collection* ranking scheme interprets a negative entry (i_N) as follows. For each $(u, t) \in A$, user u is less favourable to use tag t to annotate other items that have not been tagged by u with other tags ($I \backslash C_u$). The representation of this can be formulated as:

$$D = \{(u, i_P, i_N, t) : (u, i_P, t) \in A \wedge ((u, i_N, t) \notin A \wedge i_N \in I \backslash C_u)\}$$

In Figure 4, the positive entries show that u_1 has used t_1 to reveal his interest for items $\{i_2, i_3\}$, t_3 for $\{i_2\}$, and t_4 for $\{i_3\}$. Given $I = \{i_1, i_2, i_3, i_4\}$ and $C_{u_1} = \{i_2, i_3\}$ so that $I \backslash C_{u_1} = \{i_1, i_4\}$, this tagging data is interpreted as u_1 favours $\{i_2, i_3\}$ more than $\{i_1, i_4\}$ to be annotated using t_1. The representation can then be generated from the pairwise ranking of $i_P = \{i_2, i_3\}$ and $i_N = \{i_1, i_4\}$ on (u_1, t_1) set.

4.3 The Pairwise Ranking Method for Item Recommendation

The Pairwise Interaction Tensor Factorization (PITF) (Rendle and Schmidt-Thieme, 2010) is a well-known tag recommendation method. In this paper, we utilise this method for the task of item recommendation. The adaptation is necessarily as the task of recommending tags differs from the task of recommending items.

For tag recommendation, predictions are generated for each predefined user and item combination, i.e. the recommendation system predicts tags for an item to a user. However, for item recommendation, the recommendation system predicts items based on the user information only. Consequently, a method must calculate the item ranking score from the whole available tags before deciding which items are in the top-N recommendation list for the user. The original method works only for a (u, i) set scheme, however, our proposed method, called as PITF-I, is able to generate the recommendations by implementing the two sets: (u, i) and (u, t) schemes.

Using the (u, i) set interpretation scheme, PITF-I method represents the ranking of tagging data as (u, i, t_P, t_N), where (u, i, t_P) is a triple of positive entry and (u, i, t_N) is a triple of negative entry. It then creates a tensor factorization model which employs an iterative gradient descent algorithm for optimizing the ranking function so that the positive entries are assigned with higher values than the negative entries. This ensures the notion that the user favours the positive entries more than the negative ones. The model is formulated as:

$$\hat{\mathcal{Y}} \approx \hat{U}_k \cdot \hat{T}_k^U + \hat{I}_k \cdot \hat{T}_k^I \tag{1}$$

where $\hat{U}_k$ is the user factor matrix, $\hat{I}_k$ is the item factor matrix, $\hat{T}_k^U$ is the tag factor matrix with respect to users, $\hat{T}_k^I$ is the tag factor matrix with respect to items, k is the size of factors, and $\hat{\mathcal{Y}}$ is the reconstructed personalized tag-ranking tensor. The element-wise relevance recommendation ranking score is calculated as follows:

$$\hat{y}_{u,i,t} \approx \sum_{j=1}^{k} \hat{u}_{u,j} \cdot \hat{t}_{t,j}^U + \sum_{j=1}^{k} \hat{\imath}_{i,j} \cdot \hat{t}_{t,j}^I \tag{2}$$

Using the (u, t) set interpretation, PITF-I exchanges the roles of items and tags with respect to each other. The data representation is the ranking of tagging data as (u, t, i_P, i_N), where (u, t, i_P) is a triple of positive entry and (u, t, i_N) is a triple of negative entry. Consequently, the model formulation becomes:

$$\hat{\mathcal{Y}} \approx \hat{U}_k \cdot \hat{I}_k^U + \hat{T}_k \cdot \hat{I}_k^T \tag{3}$$

where $\hat{U}_k$ is the user factor matrix, $\hat{T}_k$ is the tag factor matrix, $\hat{I}_k^U$ is the item factor matrix with respect to users, $\hat{I}_k^T$ is the item factor matrix with respect to tags, k is the size of factors, and $\hat{\mathcal{Y}}$ is the new tensor. The relevance recommendation ranking score is calculated as:

$$\hat{y}_{u,t,i} \approx \sum_{j=1}^{k} \hat{u}_{u,j} \cdot \hat{\imath}_{i,j}^U + \sum_{j=1}^{k} \hat{t}_{i,j} \cdot \hat{\imath}_{i,j}^T \tag{4}$$

4.4 Selecting the Set

Selecting a set for user profile representation cannot trivially be restricted based upon the recommendation task, i.e. based on the predefined notion of using the (u, i) set scheme for tag recommendation and using the (u, t) set scheme for item recommendation. This limitation has shown to cause the recommendation methods to perform at a varied level when they are implemented on various datasets (Gemmell et al., 2011; Ifada and Nayak, 2014; Rendle et al., 2009). A recommendation method performs best to its capacity when it is applied on the social tagging system for which it was built for. The performance degrades when it is applied to another social tagging system in which the user tagging behaviour varies. For example, the user tagging behaviour on the Delicious website (http://delicious.com) differs from the user behaviour on the LastFM website (http://www.last.fm/).

The difference can easily be perceived from the statistic of tagging data listed in Section 5.1. As the number of unique tags is much larger than the number of unique items in the Delicious data, the average number of unique tags used by each user to annotate an item overrides the number of unique items to be annotated using a tag. This analysis suggests that the (u, t) set in Delicious is more dominant than the (u, i) set. On the contrary, the number of unique tags is much less than the number of unique items in the LastFM data. This shows that the (u,t) set is less dominant than the (u,i) set as the average number of unique tags used by each user to annotate an item is less than the number of unique items to be annotated using a tag. The comparison between these two datasets (including other two datasets) is illustrated in Figure 5.

The characteristic of user tagging behaviour can be assessed by comparing the number of (u,i) and (u,t) sets in the data. It can be said that the dominant feature set (reflected by the larger distribution) is representative of the user tagging behaviour in that system and, therefore it can determine the interpretation scheme for building user profiles. The (u,i) set is more dominant than the (u,t) set when the users tend to use less number of tags in annotating items. The (u,t) set is more dominant than the (u,i) set when the users tend to use more number of tags in annotating items. We conjecture that, for Delicious data, the best performance of recommendation methods can be obtained when the (u,t) set interpretation scheme is implemented to build the user profile model. However, the (u,i) set interpretation scheme will become the best choice for modelling the user profile of LastFM data.

5 Empirical Analysis

Experiments are conducted to investigate the tagging data interpretation and representation schemes that can improve the performance of tag-based item recommendation systems by highlighting the representative features of user tagging behaviours on different datasets. We implemented six tagging data interpretations which resulted in different representations. The representation of each set is generated as pairwise ranking between the positive entries generated from the observed tagging data and the negative entries generated from the non-observed data using the following schemes: (a) *all-tag*, (b) *user-vocabulary*, (c) *non-user-vocabulary*, (d) *all-item*, (e) *user-collection*, and (f) *non-user-collection*. The first three schemes are based on the (u,i) set interpretation scheme, while the last three are based on the (u,t) set interpretation scheme.

5.1 Dataset

The offline experiments use several real-world tagging datasets to implement the proposed user profile representations for generating item recommendation. Adapting the standard practice of eliminating noise and decreasing the data sparsity (Nanopoulos, 2011; Rafailidis and Daras, 2013; Symeonidis, Nanopoulos and Manolopoulos, 2010), the datasets are refined by using the p-core technique (Batagelj and Zaveršnik, 2002), i.e. selecting users, items, and tags that have occurred in at least p number posts. Post is the set of distinct user-item combinations in the observed tagging data. In general we choose $p = 10$ to refine the dataset as this core value has shown a stable recommendation performance in our previous work after a systematic and extensive experiments (Ifada and Nayak, 2014). However, we use $p = 5$ instead for CiteULike and MovieLens datasets as they do not contain a 10-core.

The details of four tagging datasets used in this paper are:

Delicious (http://delicious.com/)**.** It is a website that facilitates its users to save, organize and discover interesting links on the web. The Delicious dataset is generated with 50,991 observed tagging data resulted from 10–core refinement, and consists of 2,009 users, 1,485 items and 2,589 tags.

LastFM (http://www.last.fm/)**.** It is a website that gives user personalized recommendations based on the music the user listens to. The LastFM dataset is generated with 99,211 observed tagging data resulted from 10-core refinement, and consists of 867 users, 1,715 items and 1,423 tags.

CiteULike (http://www.citeulike.org/)**.** It is a website that provides a service for managing and discovering scholarly references. The CiteULike dataset is generated with 59,832 observed tagging data resulted from 5-core refinement, and consists of 2,536 users, 3,091 items and 6,949 tags.

MovieLens (http://movielens.org)**.** It is a website which provides a personalized movie recommendation. The MovieLens dataset is generated with 25,103 observed tagging data resulted from 5-core refinement, and consists of 571 users, 1,684 items and 1,559 tags.

Figure 5 shows the set-based statistics to observe the characteristic of user tagging behaviour on each dataset. This information is used for selecting a profile presentation, i.e. based on the dominant number of sets.

5.2 Evaluation Criteria

To evaluate the quality of recommendation, we implemented the 3-fold cross-validation and we divided the dataset randomly into a training set D_{train} (80%) and a test set D_{test} (20%) based on the number of posts data. D_{train} and D_{test} do not overlap in posts, i.e., there exist no triplets for a user-item combination in the training set if a triplet $(u,i,*)$ is present in the test set. The recommendation task is to predict and rank the Top-N items for the users present in D_{test}.

The performance is measured using F1-Score. F1-Score is a harmonic mean of overall precision and recall. Precision is the ratio of number of relevant items (all items in the post by the user in D_{test}) in the Top-N list to the total number of Top-N recommended items. Recall is the ratio of the number of relevant items in the Top-N list to the total number of relevant items.

$$Precision(D_{test}, N) = avg_{(u,i)\in D_{test}} \frac{|Test_u \cap TopN_u|}{|TopN_u|} \quad (1)$$

$$Recall(D_{test}, N) = avg_{(u,i)\in D_{test}} \frac{|Test_u \cap TopN_u|}{|Test_u|} \quad (2)$$

$$F1(D_{test}, N) = \frac{2\cdot Precision(D_{test},N)\cdot Recall(D_{test},N)}{Precision(D_{test},N)+Recall(D_{test},N)} \quad (3)$$

Where $Test_u$ is the set of items tagged by target user in the D_{test} and $TopN_u$ is the Top-N list of items recommended to user from the reconstructed tensor $\hat{\mathcal{Y}}$.

5.3 Result and Discussion

5.3.1 Effect of Interpretation Set Selection on Recommendation Accuracy

The characteristic of users tagging behaviours of each dataset can be identified from the set-based statistic, as illustrated in Figure 5. The statistic shows that users of Delicious dataset have the same tagging behaviour as those of CiteULike, i.e. the (u,i) set is less dominant than the (u,t) sets. Conversely, the users of LastFM and MovieLens have similar tagging behaviour, i.e. the (u,i) set is more dominant than the (u,t) sets. We conjecture that the dominant set on each dataset defines which set should be used for interpreting the tagging data. This means that Delicious and CiteULike datasets should be best interpreted using the (u,t) set interpretation scheme, while LastFM and MovieLens datasets should be best interpreted using the (u,i) set scheme.

Figure 6 displays the F1-Score comparison on Top-N lists for the recommendation accuracy on each dataset which ascertains our claim that both Delicious and CiteULike achieve their best recommendation performance when their tagging data is interpreted using the *non-user-collection* of (u,t) set pairwise ranking interpretation scheme. Similarly, LastFM and MovieLens perform best when their tagging data is interpreted using the *non-user-vocabulary* of (u,i) set pairwise ranking interpretation scheme.

These results also verify that, for generating the best pairwise ranking representation, the negative entries of non-observed data should be interpreted from: (1) the tags that have not been used in any other items by user u for annotating item i on each (u,i) set, and (2) the items that that have not been tagged with any other tags by user u using tag t on each (u,t) set. Figure 6 also illustrates that the *all-tag* scheme which uses any tags other than those appearing in positive entries to annotate item i (Rendle and Schmidt-Thieme, 2010) or the *all-item* scheme which uses any items other than those appearing in positive entries using tag t (Gemmell et al., 2011) cannot interpret non-observed data as negative entries properly and results in inferior recommendation performance.

On the other hand, though it sounds reasonable that, for the (u,i) set scheme, i.e. using the *user-vocabulary* interpretation, the negative entries should be interpreted from the tags of user-vocabulary other than those appearing in positive entries to annotate item i, however our experiments show that this interpretation severely impacts the recommendation performance for all datasets. Likewise, the items of user-vocabulary other than those appearing in positive entries using tag t should not be interpreted as the negative entries for the (u,t) set scheme, i.e. using the *user-collection* interpretation.

5.3.2 Data Representation Density

We examine the data representation density resulting from the six proposed interpretation schemes and compare with that of the *boolean* scheme as listed in Table 1. This observation is not merely to show that the *set-based* interpretation scheme is able to generate more dense data representation than the *boolean* scheme as this conclusion has been preliminary discovered (Rendle et al., 2009). Our focus is mainly to explore the representation density resulted from various *set-based* interpretation schemes, particularly comparison of the *non-user-vocabulary* and *non-user-collection* schemes to their counterparts, i.e. *all-tag* and *all-item*, respectively. Accordingly, when we highlight the superiority of the *non-user-vocabulary* scheme over *all-tag* scheme, it means that we can also state the same thing about that of the *non-user-collection* scheme over the *all-item* scheme.

Results in Table 1 show that the data representation densities of *user-vocabulary* and *user-collection* schemes are significantly lower than the other *set-based* schemes. It indicates that these two schemes, in comparison to others, are not able to interpret the tagging data efficiently and therefore the generated data representations are not able to include all relationships properly.

Table 1 shows that the data representation density of *non-user-vocabulary* is less than that of *all-tag* scheme. Yet, as established in Section 5.3.1, the former scheme outperforms the later. This fact clarifies that the *all-tag* scheme includes some relationships that are not meant to be. On each (u,i) set, the scheme uses all tags other than those appearing in positive entries to annotate item i as negative entries for the pairwise ranking representation. This interpretation is incorrect as some of those tags have actually been used by user u to annotate other items which means that those used tags should not be interpreted as negative entries. It confirms that the interpretation scheme strongly determines the recommendation performance. To generate quality recommendations, an efficient user profile presentation is more important, instead of just simply trying to get more dense data representation as other researchers had done previously (Cui et al., 2011; Leginus, Dolog and Žemaitis, 2012; Rafailidis and Daras, 2013). These methods practically implement the *boolean* scheme to interpret the tagging data and then applied clustering techniques for reducing the tag dimension to represent the semantically similar tags. These approaches are able to generate more dense data, yet they still interpret the tagging data improperly.

5.3.3 Cold-start User Problem

We carry out further experiments to examine the impact of an interpretation scheme to recommendation performance in addressing the cold-start user problem. We identify the cold-start user problem as a situation in which a user has annotated a single item only with limited number of tags. Due to limited usage data, we cannot infer the user preferences on the system.

We compare the recommendation performance on MovieLens dataset by implementing the (u,i) set interpretation scheme as it is the best choice of scheme to model the user profile. Since our main focus is on how the interpretation scheme is able to address the cold-start user problem, we use the recall metric to demonstrate the coverage of recommendations. Figure 7 shows the recall of recommendations generated using the *non-user-vocabulary* scheme outperforms the results of *all-tag*. As expected, the results of the *user-vocabulary* scheme have shown the worst performance. This fact confirms that the approach of interpreting negative entries from tags that have not been used by user u in any other items is resulting quality recommendations for both the active and the cold-start users. We can also state the same thing for the (u, t) set interpretation scheme, i.e. the negative entries are best interpreted from items that have not been tagged by user u using any other tags.

6 Conclusion

In this paper, we proposed six *set-based* tagging data interpretation schemes and representations to investigate an efficient scheme leading to build effective user profiles for generating item recommendations. For each interpretation scheme, a data representation is produced, where for each set, pairwise ranking is generated between the positive entries of observed tagging data and the negative entries of non-observed data. We have shown that the set to be used as the interpretation scheme must be selected based on the dominant number of set observed from the tagging data as it represents the unique characteristic of user tagging behaviour in the system. We implemented the PITF-I method, using a pairwise ranking representation between the positive entries and the negative entries. The PITF-I method represents the ranking of tagging data as (u, i, t_P, t_N) when the dataset has a dominant number of (u, i) sets. Alternatively, the method represents the ranking of tagging data as (u, t, i_P, i_N) when the dataset has a dominant number of (u, t) sets.

The proposed representations are extensively evaluated on four datasets which exhibit different tagging behaviour characteristics. Empirical analysis shows that the improper interpretation, of negative entries to be ranked pairwise with the positive entries, results in inferior recommendation performance. The best scheme for pairwise ranking representation should generate the negative entries interpreted from either the tags or the items that have not been used by user u in any other items or tags, i.e. *non-user-vocabulary* or *non-user-collection* schemes, respectively. We also show how this scheme is able to address the cold-start user problem. In the future, we are planning to investigate the tagging data interpretation and representation schemes for a list-wise ranking tensor factorization method.

7 Acknowledgement

This work is supported by the Directorate General of Higher Education (DGHE) Indonesia. Computational resources and services are provided by the HPC and Research Support Group, Queensland University of Technology, Brisbane, Australia.

8 Reference

Adomavicius, G., and Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering, 17*(6), 734-749.

Batagelj, V., and Zaveršnik, M. (2002). Generalized Cores. *arXiv preprint cs/0202039*,

Cui, J., Liu, H., He, J., Li, P., Du, X., and Wang, P. (2011). TagClus: a random walk-based method for tag clustering. *Knowledge and information systems, 27*(2), 193-225.

Gemmell, J., Schimoler, T., Mobasher, B., and Burke, R. (2011). Tag-based Resource Recommendation in Social Annotation Applications. In *User Modeling, Adaption and Personalization*. 111-122, Springer.

Ifada, N., and Nayak, R. (2014). A Two-Stage Item Recommendation Method Using Probabilistic Ranking with Reconstructed Tensor Model. In *User Modeling, Adaptation, and Personalization*. 98-110, Springer.

Kim, H.-N., Ji, A.-T., Ha, I., and Jo, G.-S. (2010). Collaborative Filtering based on Collaborative Tagging for Enhancing the Quality of Recommendation. *Electronic Commerce Research and Applications, 9*(1), 73-83.

Kolda, T., and Bader, B. (2009). Tensor Decompositions and Applications. *SIAM Review, 51*(3), 455-500.

Leginus, M., Dolog, P., and Žemaitis, V. (2012). Improving Tensor Based Recommenders with Clustering. In *User Modeling, Adaptation, and Personalization*. 151-163, Springer Berlin Heidelberg.

Nanopoulos, A. (2011). Item Recommendation in Collaborative Tagging Systems. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 41*(4), 760-771.

Rafailidis, D., and Daras, P. (2013). The TFC Model: Tensor Factorization and Tag Clustering for Item Recommendation in Social Tagging Systems. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 43*(3), 673-688.

Rendle, S., Balby Marinho, L., Nanopoulos, A., and Schmidt-Thieme, L. (2009). Learning optimal ranking with tensor factorization for tag recommendation. *Proc. The 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Paris, France, 727-736. ACM.

Rendle, S., and Schmidt-Thieme, L. (2010). Pairwise Interaction Tensor Factorization for Personalized Tag Recommendation. *Proc. The 3rd ACM International Conference on Web Search and Data Mining*, New York, USA, 81-90. ACM.

Symeonidis, P., Nanopoulos, A., and Manolopoulos, Y. (2008). Tag Recommendations based on Tensor Dimensionality Reduction. *Proc. The 2008 ACM Conference on Recommender Systems*, Lausanne, Switzerland, 43-50. ACM.

Symeonidis, P., Nanopoulos, A., and Manolopoulos, Y. (2010). A Unified Framework for Providing Recommendations in Social Tagging Systems Based on Ternary Semantic Analysis. *IEEE Transactions on Knowledge and Data Engineering, 22*(2), 179-192.

Zhang, Z.-K., Zhou, T., and Zhang, Y.-C. (2011). Tag-Aware Recommender Systems: A State-of-the-Art Survey. *Journal of Computer Science and Technology, 26*(5), 767-777.

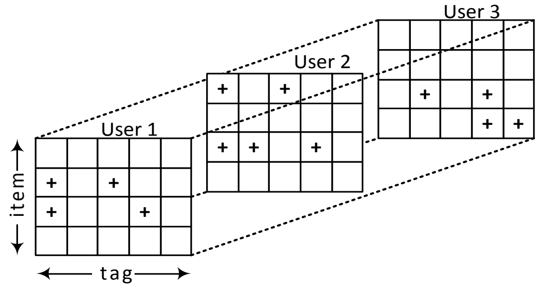

Figure 1: Toy example of observed tagging data with $U = \{u_1, u_2, u_3\}$, $I = \{i_1, i_2, i_3, i_4\}$, and $T = \{t_1, t_2, t_3, t_4, t_5\}$

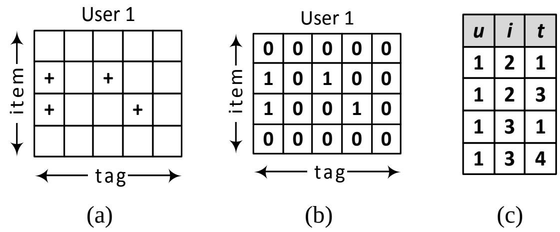

Figure 2: The *boolean* interpretation scheme and its representation for User 1 (u_1). (a) Observed entries, (b) Data interpretation, (c) Data representation generated only from the positive entries.

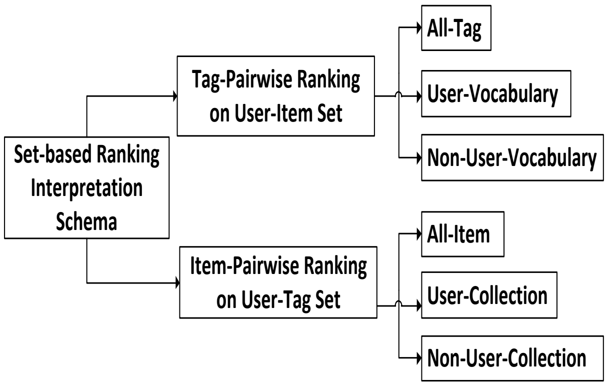

Figure 3: *Set-based* ranking interpretation schemes

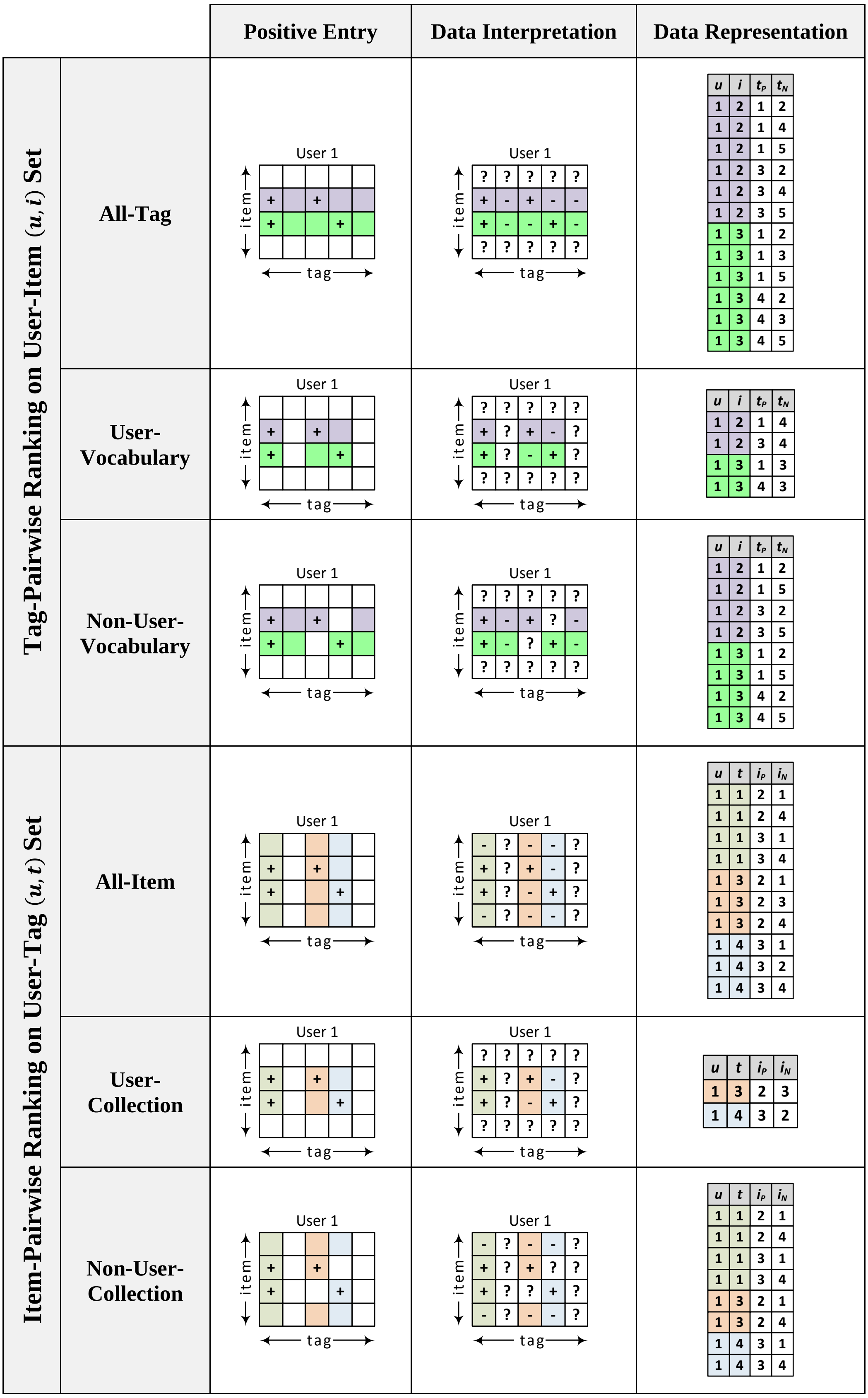

Figure 4: Examples showing the *set-based* ranking interpretation scheme and its various representation for u_1. The *set-based* ranking scheme uses the pairwise ranking of positive entries (interpreted from the observed data) and negative entries (interpreted from the non-observed data) for generating data representation. Other entries, that are to be predicted as recommendations, are noted as missing values (denoted as "?"). For the user-item set, the representation is generated by using tags as the pairwise ranking object (u, i, t_P, t_N) while the user-tag set generates the representation by using items as the pairwise ranking object (u, t, i_P, i_N).

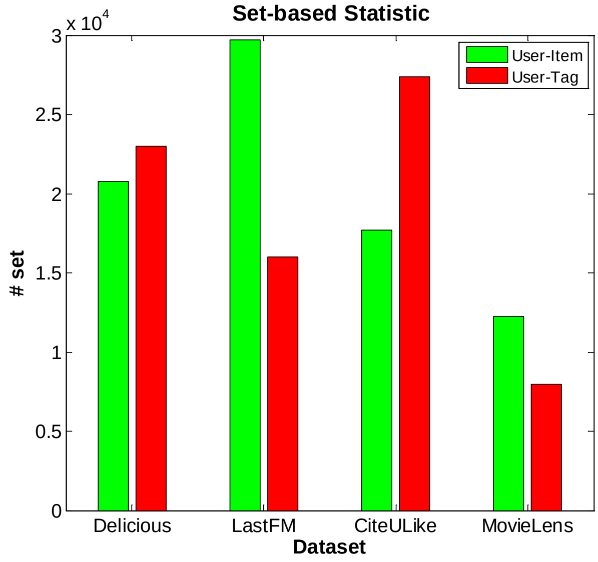

Figure 5: Set-based Statistic on Each Dataset

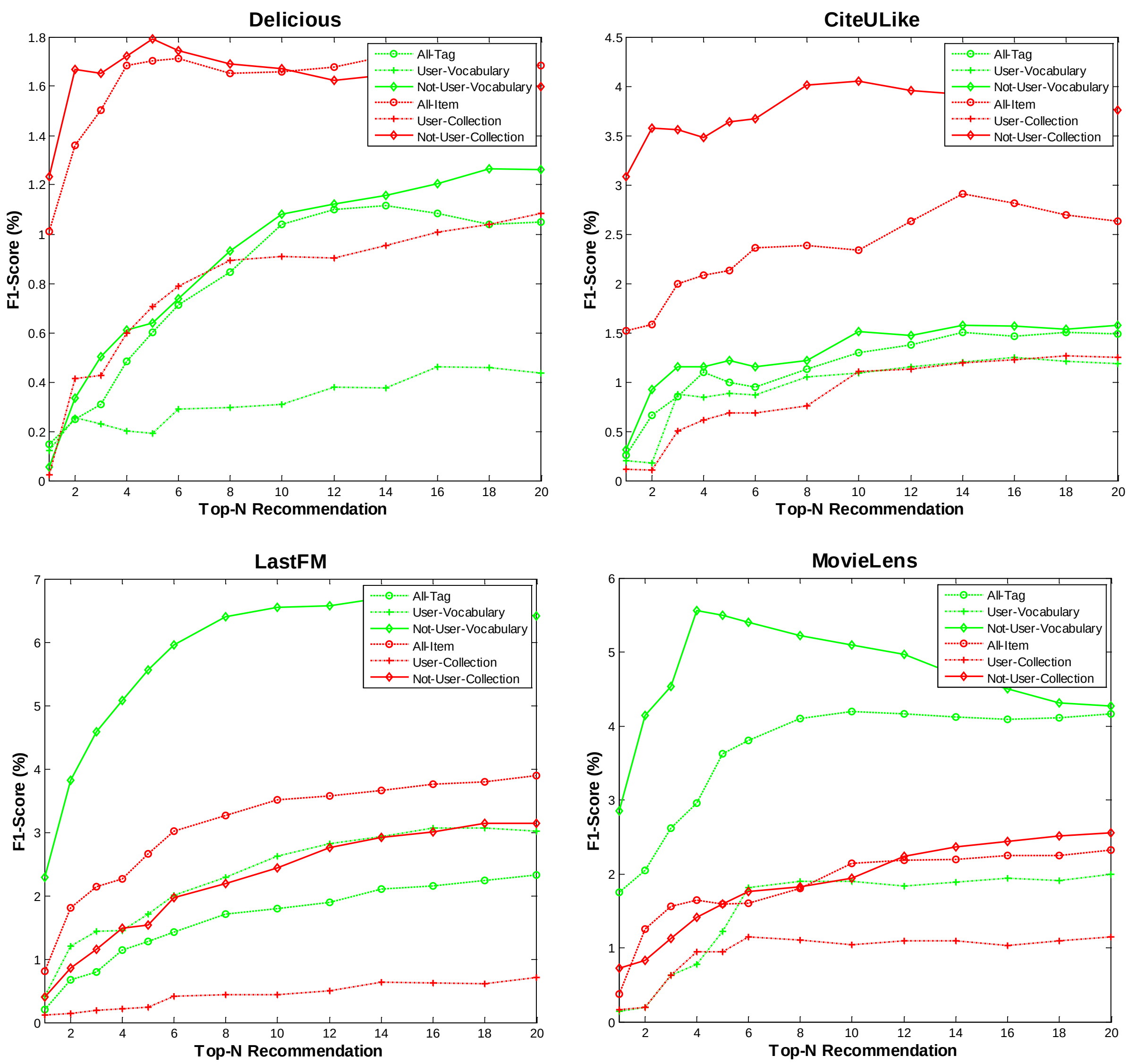

Figure 6: F1-Score comparison on Top-*N* lists for the recommendation accuracy resulting from the six proposed *set-based* interpretation schemes on various datasets

INTERPRETATION SCHEME \ DATASET		Delicious	LastFM	CiteULike	MovieLens
boolean		0.0006%	0.0042%	0.0001%	0.0015%
Tag-Pairwise Ranking on User-Item (u,i) Set	**All-Tag**	1.4980%	6.0184%	0.6902%	2.2693%
	User-Vocabulary	0.0273%	0.1262%	0.0099%	0.1626%
	Non-User-Vocabulary	***1.4707%***	***5.8922%***	***0.6803%***	***2.1067%***
Item-Pairwise Ranking on User-Tag (u,t) Set	**All-Item**	0.8540%	7.2087%	0.3064%	2.4224%
	User-Collection	0.0241%	0.3641%	0.0069%	0.2489%
	Non-User-Collection	***0.8299%***	***6.8447%***	***0.2996%***	***2.1735%***

Table 1: The Comparison of data representation density resulting from the *boolean* and *set-based* ranking interpretation schemes on various datasets

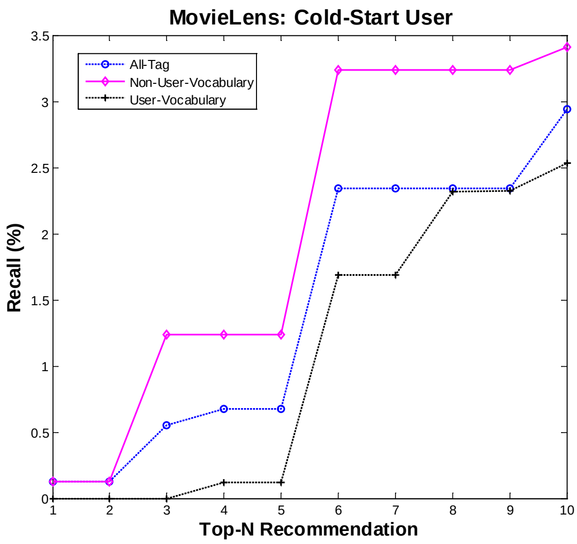

Figure 7: Cold-Start User Recall Recommendation on *all-tag*, *non-user-vocabulary*, and *user-vocabulary*