Towards Understanding Latent Factors and User Profiles by Enhancing Matrix Factorization with Tags

Tim Donkers, Benedikt Loepp, Jürgen Ziegler
University of Duisburg-Essen
Duisburg, Germany
{tim.donkers, benedikt.loepp, juergen.ziegler}@uni-due.de

ABSTRACT

With the interactive recommending approach we have recently proposed, users are given more control over model-based Collaborative Filtering while the results are perceived as more transparent. Integrating the latent factors derived by Matrix Factorization with tags users provided for the items has, however, even more advantages. In this paper, we show how general understanding of the abstract factor space, and of user and item positions inside it, can benefit from the semantics introduced by considering additional information. Moreover, our approach allows us to explain the user's (former latent) preference profile by means of tags.

CCS Concepts

•**Information systems** → **Recommender systems;** *Personalization; Search interfaces;*

1. INTRODUCTION

Complementing *Matrix Factorization* (MF) with further data, e.g. implicit feedback, temporal information or predefined metadata, has widely been accepted to increase algorithmic accuracy [4]. In line with others, we have shown that this is also true when user-provided tags are taken into account [1]. As part of a comprehensive user study of tag-enhanced *Recommender Systems* (RS), we could confirm that the recommendation quality perceived by users benefits as well [2]. However, latent factor models derived by MF have rarely been exploited for purposes other than improving effectiveness or performance. Few exceptions rely on the factor space to elicit user preferences in a choice-based manner [5, 3] or to visualize e.g. item characteristics [6]. Although considered particularly difficult from a system-perspective [4], in [8], a first step towards relating the factors to an intelligible meaning has been taken. Still, latent factors are overall hard to explain. It is also a more general problem that users often lack a deeper understanding in model-based *Collaborative Filtering* (CF) systems.

By enhancing standard SVD-like MF with tags users provided for the items, we however not only proposed novel interaction possibilities, especially for cold-start situations, but have also shown that introducing the easy-to-understand semantics of tags to a latent factor model positively affects perceived recommendation transparency [2]. In this paper, we demonstrate that our approach has even more advantages. It allows us to get a general understanding of the factor space and how users and items are positioned inside it. Furthermore, users can be presented with textual tags explicitly explaining their preference profile they have expressed implicitly in the (intransparent) latent factor space.

RecSys 2016 Poster Proceedings, Sept 15-19, 2016, Boston, MA, USA.

2. UNDERSTANDING LATENT FACTORS AND EXPLAINING USER PROFILES

With the steps described in [1], we have integrated item-specific tag relevance information into a MF algorithm in order to derive corresponding user-tag relevance scores as well as tag-factor relations. Since the matrix holding these general relations, $\mathbf{\Theta\Theta}^{\mathrm{T}}$, is a square diagonalizable matrix, we can use eigendecompositon to represent it in terms of eigenvalues and eigenvectors:

$$\mathbf{R} \approx {}^{\mathbf{u}}\mathbf{A}\mathbf{\Theta}\mathbf{\Theta}^{\mathrm{T}}\,{}^{\mathbf{i}}\mathbf{A}^{\mathrm{T}} = {}^{\mathbf{u}}\mathbf{A}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}\,{}^{\mathbf{i}}\mathbf{A}^{\mathrm{T}} \quad (1)$$

The diagonal matrix $\mathbf{\Lambda}$ contains the eigenvalues of $\mathbf{\Theta\Theta}^{\mathrm{T}}$ in non-increasing order. The eigenvectors in $\mathbf{V}$ hold the importance of every tag with respect to a certain direction. Since $\mathbf{\Theta\Theta}^{\mathrm{T}}$ is symmetric, eigenvectors are chosen orthogonal to each other. Latent factors are thus incorporated into the tag information space by stretching it along the directions of the eigenvectors to the amount of corresponding eigenvalues.

This way considering tag information in model-based CF offers us the opportunity to access the previously abstract user-factor and item-factor vectors in a much more comprehensible way. As the tag concept is easily understood, this allows us to both explain a user's vector, i.e. his or her preference profile, and to let users actively adjust it. We have described the latter in our previous work [1, 2]. Now, we will concentrate on how enhancing MF with additional information allows to gain a better understanding of users, items and the latent factor space itself, and in particular to express a latent preference profile through explicit tags.

Understanding the Factor Space: Applying eigendecomposition as described above provides us information on the importance of each dimension of the factor space and their relationship to the tags. By looking at the most positively/negatively related tags, this gives us a more general understanding of what is expressed by factors derived automatically by MF. For instance, using the configuration from [1, 2] with MovieLens 10M and Tag Genome dataset, Table 1 shows that the first latent factor is best described by tags such as "twist ending" or "psychology", the second by "time travel" or "comedy". Consequently, we found movies such as "Seven" or "Back to the Future" having highest values for the respective factors in the actual item-factor matrix ${}^{\mathbf{i}}\mathbf{A}\mathbf{V}\mathbf{\Lambda}^{1/2}$.

Table 1: Most important factors, their eigenvalues and most positively or negatively related tags, as well as some representative movies for each factor.

Factor	Positively related		Negatively related
1 (0.68)	twist ending (0.50), psychology (0.38), classic (0.34)	···	romance (-0.40), quirky (-0.30), sci-fi (-0.23)
	Seven, The Machinist, ...		
2 (0.45)	time travel (0.36), dystopia (0.31), comedy (0.27)	···	vis. appealing (-0.54), stylized (-0.35), romance (-0.23)
	Bill & Ted's Excellent Advent., Back to the Future, ...		
3 (0.32)	sci-fi (0.43), vis. appealing (0.24), twist ending (0.24)	···	dark (-0.53), surreal (-0.38), violence (-0.19)
	Planet of the Apes, When Worlds Collide, ...		
4 (0.25)	psychology (0.36), dystopia (0.30), romance (0.28)	···	classic (-0.32), vis. appealing (-0.29), based on a book (-0.27)
	A Life Less Ordinary, Splendor, ...		
5 (0.14)	dystopia (0.29), violence (0.23), sci-fi (0.20)	···	psychology (-0.41), time travel (-0.41), based on a book (-0.37)
	Nausicaä, Robocop, ...		

The regression-constrained formulation also allows us to gain insights on how users and items are positioned inside the information space. For items, Figure 1 illustrates this in an example with two tags, which are then used to enhance a two-factorical MF for demonstration purposes. The original tag-related item positions are represented by row vectors of ${}^{\mathbf{i}}\mathbf{A}$. In the left plot, movies are shown accordingly with respect to tag relevances. The influence of latent information can then be examined by considering the item-factor matrix ${}^{\mathbf{i}}\mathbf{A}\mathbf{V}\mathbf{\Lambda}^{1/2}$ which is used eventually for generating recommendations by MF, i.e. usually by calculating dot products. Consequently, the right plot in Figure 1 shows how items are arranged with respect to the resulting factors. Similarities between movies in terms of tag relevance can still be found when considering the latent information, especially in this case where we used the same amount of tags and factors.

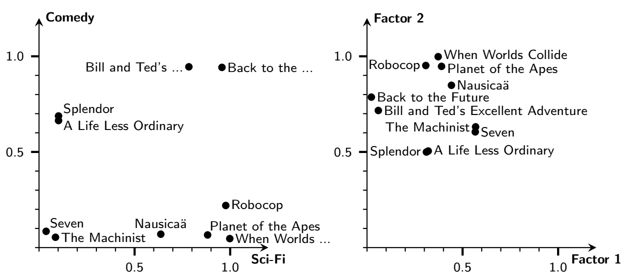

Figure 1: Normalized movie positions with respect to tag relevance (left) and latent factors (right).

Explaining User Profiles: By enhancing MF, preference profiles now comprise information related to both tags and latent factors. Thus, we can utilize ${}^{\mathbf{u}}\mathbf{A}$ to explain the user's profile by means of those tags considered most important to him or her. In an extension of our web-based movie RS, a dialog comprising a tag cloud enables users to inspect their existing profile by means of such tags (Figure 2). Individual preferences expressed implicitly with respect to the factor space, e.g. by rating items, can thus be presented more explicitly. When ${}^{\mathbf{u}}\mathbf{A}$ is derived according to our approach [1], this is independent of the tags users actually have assigned. We can thereby present the tag cloud even in the common case where the current user never tagged any items.

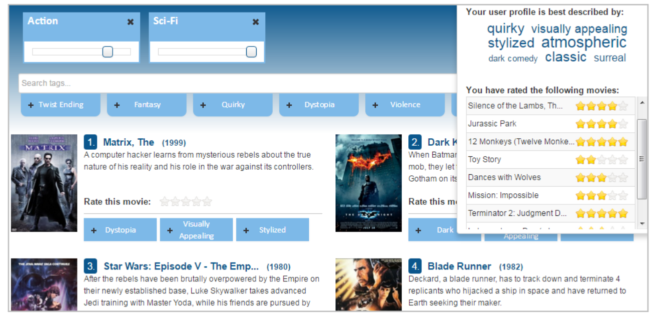

Figure 2: Screenshot of our RS: The user's existing preference profile is explained by a tag cloud.

3. CONCLUSIONS AND OUTLOOK

Enhancing MF with tags improves accuracy [1] as well as subjective perception of RS [2]. In this paper, we have proposed that such additional data may also be leveraged for the purpose of explaining latent factors. In particular, our approach seems promising to present users with an explicit description of their—in model-based CF usually intransparent—preference profile. However, it will be subject of our future work to examine this contribution in more detail, for instance, by determining the information theoretic value as in [7] or by conducting user studies to compare with other approaches that explain recommendations using tags, e.g. [9]. We also aim to investigate how latent factor models can be better semantically enriched, e.g. by exploiting further data. Finally, there is also room left for improvement especially regarding visual explanations.

4. REFERENCES

[1] T. Donkers, B. Loepp, and J. Ziegler. Merging latent factors and tags to increase interactive control of recommendations. In *Poster Proc. RecSys '15*, 2015.

[2] T. Donkers, B. Loepp, and J. Ziegler. Tag-enhanced collaborative filtering for increasing transparency and interactive control. In *Proc. UMAP '16*, pages 169–173. ACM, 2016.

[3] M. P. Graus and M. C. Willemsen. Improving the user experience during cold start through choice-based preference elicitation. In *Proc. RecSys '15*, pages 273–276. ACM, 2015.

[4] Y. Koren, R. M. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8):30–37, 2009.

[5] B. Loepp, T. Hussein, and J. Ziegler. Choice-based preference elicitation for collaborative filtering recommender systems. In *Proc. CHI '14*, pages 3085–3094. ACM, 2014.

[6] B. Németh, G. Takács, I. Pilászy, and D. Tikk. Visualization of movie features in collaborative filtering. In *Proc. SoMeT '13*, pages 229–233, 2013.

[7] T. T. Nguyen and J. Riedl. Predicting users' preference from tag relevance. In *Proc. UMAP '10*, pages 274–280. Springer, 2013.

[8] M. Rossetti, F. Stella, and M. Zanker. Towards explaining latent factors with topic models in collaborative recommender systems. In *Proc. DEXA '13*, pages 162–167, 2013.

[9] J. Vig, S. Sen, and J. Riedl. Tagsplanations: Explaining recommendations using tags. In *Proc. IUI '09*, pages 47–56. ACM, 2009.