Personalizing Tags: A Folksonomy-like Approach for Recommending Movies

Alan Said Benjamin Kille Ernesto W. De Luca Sahin Albayrak
TU Berlin / DAI Lab
{alan.said, benjamin.kille, ernesto.deluca, sahin.albayrak}@dai-lab.de

ABSTRACT
Movie recommender systems attempt to find movies which are of interest for their users. However, as new movies are added, and new users join movie recommendation services, the problem of recommending suitable items becomes increasingly harder. In this paper, we present a simple way of using a priori movie data in order to improve the accuracy of collaborative filtering recommender systems. The approach decreases the sparsity of the rating matrix by inferring personal ratings on tags assigned to movies. The new tag ratings are used to find which movies to recommend. Experiments performed on data from the movie recommendation community Moviepilot show a positive effect on the quality of recommended items.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval - Retrieval models

General Terms
Algorithms, Design, Experimentation, Human Factors

Keywords
Recommender System, Folksonomy, Movie recommendation

1. INTRODUCTION
Movie recommendation websites such as Netflix[1] and Movielens[2] have become popular among the users of the Web. Commonly, movie recommendation systems will employ Collaborative Filtering (CF)[3] to find movies of interest to their users.

CF calculates the relevance of an item for a user based on other users' rating information on items co-rated by the user and others. CF approaches are commonly categorized as either model-based or memory-based [3]. In this work we focus on the latter, which creates item predictions for a user by finding users similar to that user (in terms of co-rated items) in a neighborhood, and using information from the neighborhood, predicts items not rated by the user.

As new movies and new users are added to the systems, the user-item matrix which represents the opinions of users on movies, becomes increasingly sparse making accurate recommendations more and more difficult [2].

In this paper, we present a simple extension to the regular CF approach utilizing information perhaps most associated with the concept of folksonomies, i.e. tags. The approach uses tags in order to decrease the sparsity of the user-item matrix to make collaborative filtering-based recommendations more accurate.

In the context of item recommendation in folksonomies, tags have been put to use with good results [13], however most movie recommendation websites either lack this functionality, or do not release their data for research purposes (Movielens being a rather well-know exception).

Our approach makes use of tags for the purpose of recommending movies, the principle is however applicable to any information related to movies, e.g. genres, actors, etc. For this purpose we have used a dataset provided from the German movie recommendation community Moviepilot[4] which contains, users, movies, ratings, and categorical tags assigned to each movie. Tags are assigned by a group of moderators, users can then rate how well the tags fit each movie.

Using this data, our approach creates personalized tag ratings based on the ratings given by users to tagged movies. One could compare this to a narrow folksonomy, except for the fact that tags are connected *to* users through ratings and not to movies *by* users.

We show that when using selected tag categories, the overall quality of the recommendations can be improved considerably.

2. PERSONALIZING A NARROW FOLKSONOMY
The term folksonomy was coined by Vander Wal in 2004 [12] and is defined as *"the result of personal free tagging of information and objects ...for one's own retrieval"*. Tags offer a short *content-related* description of items to which they are assigned. Traditionally, folksonomies are classified as either *broad* or *narrow*. Vander Wal states that *"[a] broad*

[1]http://www.netflix.com
[2]http://www.movielens.org
[3]http://movielens.umn.edu/html/tour/index.html
[4]`http://www.moviepilot.de`

HetRec '11, October 27, 2011, Chicago, IL, USA

folksonomy has many people tagging the same object and every person can tag the object with their own tags", and continues by stating *"[a] narrow folksonomy is done by one or a few people providing tags that the person uses to get back to that information"* [11]. Delicious[5] is an example of a broad folksonomy, whereas Flickr[6] exemplifies a narrow one.

In order to create a personalized relation between users and tags, the per-user average rating for each movie tagged with a certain tag was calculated. This created a structure similar to a broad folksonomy but since each user only has one rating per tag, the structure remains different from a traditional broad folksonomy. The Motivation for this broadening was to decrease sparsity of the rating matrix. We quantify a personalized relation between users u and tags t by calculating an average rating of tag t by user u over all movies containing tag t and rated by user u. We can summarize all user-tag average ratings in a matrix $W = (w_{ij})$, where w_{ij} denotes the average rating of tag j by user i.

To specify the user-tag matrix W more precisely, we suppose that

$$\begin{aligned} U &= \{u_1, \ldots, u_f\} \\ M &= \{m_1, \ldots, m_g\} \\ T &= \{t_1, \ldots, t_h\} \end{aligned}$$

is the set of users, movies, and tags, respectively. By $R = (r_{ij})$ we denote the $(f \times g)$-matrix with elements r_{ij} from some discrete interval $[\min_r, \max_r]$ of non-negative, real values. We assume that elements of R corresponding to missing ratings have value $< \min_r$. In addition, the binary $(h \times g)$-matrix $V = (v_{ij})$ associates tags with movies by

$$v_{ij} = \begin{cases} 1 & : \text{ if tag } i \text{ is present in movie } j \\ 0 & : \text{ otherwise} \end{cases}$$

For each user u_a, we define a tag-movie rating $(h \times g)$-matrix $V^a = (v^a_{ij})$ of the form

$$v^a_{ij} = \begin{cases} r_{aj} \cdot v_{ij} & : \text{ if user } a \text{ rated movie } j \\ 0 & : \text{ otherwise} \end{cases}$$

The matrix V^a has the same structure as the binary tag-movie matrix, where an element v^a_{ij} with value one is replaced by rating r_{aj} of movie j given by user a. Then the user-tag average rating matrix is a $(f \times h)$-matrix $W = (w_{ij})$ with

$$w_{ij} = \frac{s_{ij}}{r_i},$$

where r_i denotes the number of ratings of movies given by user i that contain tag j and s_{ij} is the sum of the elements of the j-th row of matrix V^i.

We construct an extended matrix $\hat{R} = (R, W)^t$ by stacking the matrices R and W as shown in Fig. 1. The extended matrix $\hat{R}$ has lower sparsity than the original rating matrix R.

3. EXPERIMINETS

In our experiments, we assessed the performance of the proposed approach and investigated the effects of enhancing a neighborhood-based recommender with folksonomical data.

[5] http://www.delicious.com

[6] http://www.flickr.com

$$\left[\begin{array}{cccc}
\multicolumn{4}{c}{\overbrace{\hphantom{r_{1,1}\quad r_{1,2}\quad \cdots\quad r_{1,n}}}^{users}} \\
r_{1,1} & r_{1,2} & \cdots & r_{1,n} \\
r_{2,1} & r_{2,2} & \ddots & r_{2,n} \\
\vdots & & \ddots & r_{m-1,n} \\
r_{m,1} & & \cdots & r_{m,n} \\ \hdashline
w_{1,1} & w_{1,2} & \cdots & w_{1,n} \\
w_{2,1} & w_{2,2} & \ddots & w_{2,n} \\
\vdots & & \ddots & w_{k-1,n} \\
w_{k,1} & & \cdots & w_{k,n}
\end{array}\right]
\begin{array}{l} \left.\vphantom{\begin{array}{c}r\\r\\r\\r\end{array}}\right\} movies \\ \left.\vphantom{\begin{array}{c}w\\w\\w\\w\end{array}}\right\} tags \end{array}$$

Figure 1: The extended matrix after R and W are joined, used as the input to our system.

3.1 Data

The dataset used contains 840 users, 15,613 movies, 333,061 movie ratings, and 6,580 tags assigned movies. All ratings were performed during the same 6 month period.

Tags are divided into five categories, each category can be seen as a genre, however since each category contains several tags, the representation is much more fine grained than that of genres, e.g., the *emotion* category represents the emotion of the movie, such as happy, sad, exciting, touching, etc. The *intended audience* tag tells who the movie is intended for, e.g. family, teens, etc. Similarly the *place* and *time* categories describe where and when the movie take place, e.g. Chicago, North Pole, and today and the future, respectively. The number of tags and the percentage of ratings containing tags from each category are shown in Table 1.

Users in the dataset have rated at least 30 movies during the 6 month period considered. This filtering was performed in order to minimize the effects related to the cold start problem as it is not the main focus of this paper.

tag category	# of elements	% rating coverage
Emotion	16	61.85
Intended Audience	12	35.50
Place	763	75.39
Plot	5,565	90.00
Time	224	64.02

Table 1: Tag categories with number of elements, and percentage of all ratings covered.

3.2 Experimental Setup

The dataset was split into fifty randomly selected training- and test datasets, where each test dataset contained 20% of the 840 users. 20% of the test users ratings, with a value above the users' average were selected into the testset, the remaining ratings (tag and movie) were used for training.

For each tag category as well as for the set of all tags, we constructed an extended matrix $\hat{R} = (R, W)^t$ as described in Section 2. The recommendation algorithm used for our experiments was a standard k-NN collaborative filtering-based recommender using Pearson Correlation as the similarity metric. The neighborhood size was set to 50 and the algorithm was issued once for movies only (the R matrix, which was used as the baseline), once for a combination of movies and each tag category and finally once for the full dataset

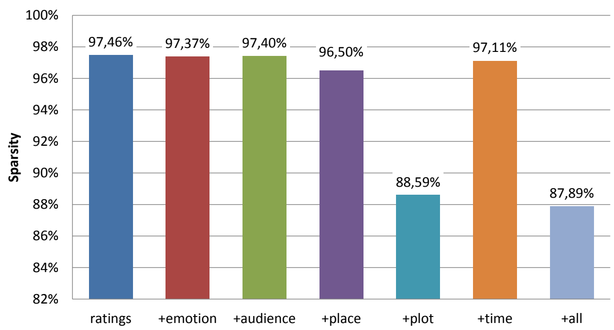

Figure 2: Sparsity of rating matrices.

combined into one (the R and W matrices combined), for each of the fifty training and testing datasets. Additionally, a constraint to only recommend movies, not tags, was introduced.

After each run, Mean Average Precision (MAP), and Precision@N for N={5, 10, 50} values were calculated and finally averaged over all 50 runs for each of the approaches.

3.3 Results and Discussion

Fig. 3 and Table 2 summarize the results of the experiments. The resulting MAP and P@N values confirm the assumption that adding personal tag rating to the users' rating profiles is beneficial in terms of recommendation quality. All tag categories perform well, with the lowest improvement being just above 50% when including time tags, to an exceptional 2850% improvement when using all tags. The single tag category which improves results most is by far the plot category which improves MAP values by 2500%.

When taking the level of sparsity for each of the tag categories, it seems natural that the plot category performs best. It should however also be noted that the plot category has the largest coverage of ratings at 90% (see Table 1), which might serve as an explanation to why the results are improved by several orders of magnitude. Another observation is that the place category provides an improvement comparable to the emotion tags, even though plot outranks emotion in number of tags and coverage percentage.

Plot tags categorize movies based on what happens. With 5,565 different tags a broad variety of incidents and story lines can be represented. This allows for a more fine-grained representation of users' tastes and interests for specific plots.

Emotion tags mirror a movie's emotion, and even though there are relatively few tags, MAP and P@N values were improved by over 200% compared to the baseline. We assume that these "pseudo-movies" (i.e. the tags in the joint matrix) representing a specific emotion enable clustering of users into neighborhoods of certain personal categories, i.e. people who enjoy specific types of movies e.g. "funny" or "exciting" films.

Intended audience tags can be seen analogously to emotion tags. Instead of reflecting an emotion, they represent a clustering of movies according to major audiences. Movies dedicated to minors are then recommended with higher probability to people who already rated similar movies.

Tags contained in the place and time categories improve recommendation results as well, however, given the coverage and amount of tags in those categories, we found it surprising that improvement was not higher. Place being the category which, in terms of coverage and number of tags, mostly resembles the plot category, gives an improvement of "only" 296% in terms of MAP (which in comparison to the 2500% improvement of plot is low). We believe the reason for this is the fact that time and place, even though always present in movies, do not really capture anything related to the actual content of the movie.

Summing up, the rated tags can be used for a more fine grained recommendation than that of only movie ratings, especially since the approach enables the finding of users who have broader similarities (emotions, plots) without actually having seen the same movies.

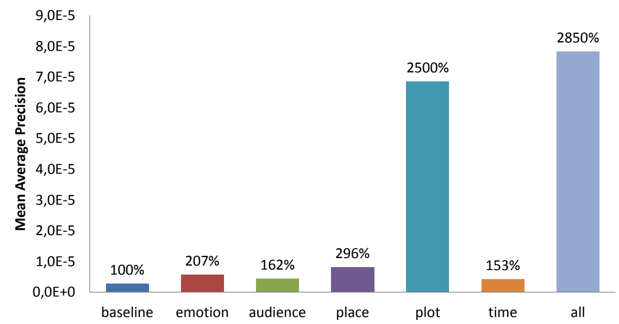

Figure 3: MAP values for the baseline and tag-based approaches. Numbers above columns show percentage improvement towards baseline.

Setting	P@5	P@10	P@50
ratings	$0.981E-2$	$1.409E-2$	$2.719E-2$
+emotion	$1.079E-2$	$1.465E-2$	$2.713E-2$
+audience	$1.171E-2$	$1.503E-2$	$2.804E-2$
+place	$1.628E-2$	$1.831E-2$	$2.929E-2$
+plot	$6.764E-2$	$6.448E-2$	$6.461E-2$
+time	$0.804E-2$	$1.222E-2$	$2.367E-2$
+all	$6.992E-2$	$6.700E-2$	$6.866E-2$

Table 2: Precision@N values for N={5, 10, 50}

4. RELATED WORK

Since the early '90s recommender systems research [9] has been an active research topic in information retrieval and machine learning.

In [1] three approaches in recommender systems are identified: content-based, collaborative filtering-based, and hybrid incorporating both. While content-based recommendations are generated by calculating similarities between item content, collaborative recommendations are based on similarities between users.

In order to improve recommendation quality, more information about either users or items can be used. These approaches are usually clustered into two sets, one using contextual information [6], and one using content for either enriching the items [5] or enriching the profiles of users [4, 10].

Online use of tags has evolved to collaborative tagging (social tagging), which is the process where users assign tags to a set of resources with the purpose of sharing, discovering, and recovering them [11]. Folksonomies refer to the classification systems that emerge from collaborative tagging. Tags consist of free keywords given by users and can be considered as user-generated taxonomies [7]. There are two ways of analyzing and discovering this information; Users can pose queries with specific search criteria, which are compared to the contents of the searched items (retrieval) or through a more automatic information discovery models (recommendation).

The aspects in which our approach differs from the ones presented above are several. The dataset used in our experiments is not a folksonomy per se, rather a collaborative rating dataset, with tags. The tags are not assigned by users, they are assigned by a group of experts (the moderators of the website). However users can rate the accuracy of the tags that have been assigned. To our knowledge, this type of personalization of a narrow folksonomy-like structure has not been investigated earlier.

Furthermore, most approaches are evaluated on datasets from the movie rating community Movielens which is a community built for the purpose of gathering data for research purposes [8] and might not reflect the activities of users in a commercial community. Our approach is evaluated on data from a commercial website centered around movie recommendation, thus having a different incentive in how it interacts with it's users.

5. CONCLUSIONS AND FUTURE WORK

In this paper we presented an approach for enhancing neighborhood-based CF with folksonomic data by extending the rating matrix to include personalized tag ratings.

Results on the Moviepilot datasets show that categorical tags, even though very simplistic in nature, do seem to be able to increase the performance of recommender systems, when put to use correctly. In our experiments we have shown that some tag categories, such as the plot category, perform much better than baseline in terms of MAP and P@N values. Tags from other categories, such as emotion or intended audience, achieved the best relative improvements.

Compared to other recommendation approaches, our tag-based extension of the standard rating matrix improves recommendation quality considerably, without necessarily adding to the overall complexity of the algorithm.

However, as the approach requires a set of tags (or other content-related) information, it is not a suitable on purely collaborative approaches, where factorization models might be the better choice.

One of the research topics we will be focusing on in the future is whether a combined tag- and time-based approach would result in better predictions of items. Additionally, we aim to explore how the proposed approach could be used for limiting the effects of cold start. Another question we intend to address is whether tagging performed by experts provides superior results to tagging performed by users. Our dataset only contains one type of tags. However, the Movielens dataset[7] for instance, contains tags generated by regular users.

The work presented in this paper describes a model for transforming non-personalized content data assigned to items, into a personalized, rated representation of the same content in order to minimize sparsity and improve collaborative filtering-based recommendations.

[7] http://www.grouplens.org/node/73

6. ACKNOWLEDGMENTS

The authors would like to express their gratitude to the Moviepilot team who contributed to this work with dataset, relevant insights and support.

The work in this paper was conducted in the scope of the KMulE project which was sponsored by the German Federal Ministry of Economics and Technology (BMWi).

7. REFERENCES

[1] Gediminas Adomavicius and Alexander Tuzhilin, 'Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions', *IEEE Trans. on Knowledge and Data Engineering*, **17**(6), (06/2005).

[2] Jesus Bobadilla and Francisco Serradilla, 'The effect of sparsity on collaborative filtering metrics', in *Proc. of ADC'09*. ACS, (2009).

[3] J S Breese, D Heckerman, and C Kadie, *Empirical analysis of predictive algorithms for collaborative filtering*, volume 461, San Francisco, CA, 1998.

[4] Iván Cantador, Martin Szomszor, Harith Alani, Miriam Fernández, and Pablo Castells, 'Enriching ontological user profiles with tagging history for multi-domain recommendations', in *CISWeb*, (2008).

[5] Claudiu S. Firan, Wolfgang Nejdl, and Raluca Paiu, 'The benefit of using tag-based profiles', *LA-Web'07*, (2007).

[6] Alexandros Karatzoglou, Xavier Amatriain, Linas Baltrunas, and Nuria Oliver, 'Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering', in *Proc. of RecSys'10*. ACM, (2010).

[7] A. Nanopoulos, 'Item recommendation in collaborative tagging systems', *IEEE Trans. on Systems, Man and Cybernetics, Part A: Systems and Humans*, **41**(4), (7/2011).

[8] Julio Ojeda-Zapata, 'New site personalizes movie reviews', *St. Paul Pioneer Press*, (9/15/1997).

[9] P. Resnick, N. Iacovou, M. Sushak, P. Bergstrom, and J. Riedl, 'Grouplens: An open architecture for collaborative filtering of netnews', in *Proc. of CSCW*. ACM, (1994).

[10] Alan Said, Ernesto W. De Luca, and Sahin Albayrak, 'Using social and pseudo social networks to improve recommendation', in *Proc. of ITWP'11*, (2011).

[11] Thomas Vander Wal. Explaining and Showing Broad and Narrow Folksonomies :: Personal InfoCloud. http://www.personalinfocloud.com/2005/02/explaining_and_.html, retrieved 07/2011, 2005.

[12] Thomas Vander Wal. Folksonomy :: vanderwal.net. http://www.vanderwal.net/folksonomy.html, retrieved 07/2011, 2007.

[13] Robert Wetzker, Winfried Umbrath, and Alan Said, 'A hybrid approach to item recommendation in folksonomies', in *Proc. of ESAIR 2009*. ACM, (2009).