The semantics of movie metadata: enhancing user profiling for hybrid recommendation

Márcio Soares[1], Paula Viana[1,2]

[1] INESC TEC, Campus da FEUP Rua Dr. Roberto Frias,
Porto, 4200 - 465, Portugal
[2] Polytechnic of Porto, School of Engineering, Rua Dr. António Bernardino de Almeida,
Porto, 4249 - 015, Portugal
{mmsoares, paula.viana}@inesctec.pt

Abstract. In movie/TV collaborative recommendation approaches, ratings users gave to already visited content are often used as the only input to build profiles. However, users might have rated equally the same movie but due to different reasons: either because of its genre, the crew or the director. In such cases, this rating is insufficient to represent in detail users' preferences and it is wrong to conclude that they share similar tastes. The work presented in this paper tries to solve this ambiguity by exploiting hidden semantics in metadata elements. The influence of each of the standard description elements (actors, directors and genre) in representing user's preferences is analyzed. Simulations were conducted using Movielens and Netflix datasets and different evaluation metrics were considered. The results demonstrate that the implemented approach yields significant advantages both in terms of improving performance, as well as in dealing with common limitations of standard collaborative algorithm.

Keywords: user profiling; hybrid recommendation; movie metadata; semantic knowledge.

1 Introduction

Automated recommendations have become a pervasive part of the daily user experience on the TV. Today, many major TV providers and media streaming platforms, use a part of the user interface to display recommendations to their users.

Standard recommendation mechanisms are usually based on collaborative filtering (CF) [1], content-based filtering (CB) [2, 3] or on a combination of these two methods – the hybrid approaches [4] that try to overcome limitation from the two previous solutions.

Most of existing recommender systems (RS) are based on a single numerical rating that represents the user's opinion about an item. However, in single rating approaches many users may have decided to rate an item with the same scores, but due to different reasons. For instance, recommending vacation packages, restaurants or hotels may require more than a single rating to take into consideration different aspects like breakfast, view, localization, etc. [5, 6]. Likewise, in a movie or TV

program scope, preferences may be driven by different aspects, such as the actors, the directors or the genre of the program. In such scenarios, a single rating approach completely ignores the semantics that can be extracted from the information contained in the metadata content. The key to more effective personalization services requires a system able to understand not only what people like, but why they like it. In other words, the ability of creating a more effective preference representation schema, will potentially lead to the design of a recommendation algorithms with increased performance [7]. To go beyond and overcome the common limitations of the use of preferences expressed only in form of ratings, a research trend which can exploit both user preferences and semantic contents, has been emerging [8, 9].

This paper presents a content-collaborative hybrid approach that explores different movie and TV metadata elements by assuming that the semantics of each element should be used to help creating better user profiles that enable relating users according to their metadata preferences. For that, user ratings are correlated with metadata, bringing to surface the real reasons that drove user rating behaviour. In addition, a deep analysis on the impact of using individual metadata categories or a combination of a set of elements in representing user's preferences is also done.

Simulations have been run using Netflix and Movielens datasets and different sparsity conditions were taken into account. Finally, in order to evaluate the performance of the proposed recommendation approach, besides Precision, a set of other emerging metrics (Diversity, Novelty and Serendipity), considered relevant in literature, and rarely explored in the semantic knowledge domain, were applied.

The remaining part of paper is organized as follows: Section 2 describes the related work. Section 3 provides an overview on the methodology used to create semantic enhanced user profiles and describes the hybrid recommendation approach implemented. Section 4 presents the proposed experimental setup used to evaluate the solution and results are presented and discussed in detail in Section 5. Finally, in Section 6 some conclusions are taken and future work is highlighted.

2   Related work

One of the main drivers for building successful RS is the availability of accurate user profiles. In order to explore movie semantic features to more accurately find similarities between users, some work has been proposed [10–14]. In [13] a recommender algorithm that is based on a factorized matrix composed of user preferences associated to the movies' genres is proposed. By using a profile enrichment approach together with collaborative methods, the author demonstrates an increase in the quality of the recommendations. The work in [10] presents an approach to automatically identify Communities of Interest (CoI) from ontology-based user profiles. Taking into account the semantics preferences of several users, common topics of interest are found by using a clustering algorithm. Thus, users who share interests on a specific concept cluster are connected and linked in a community that can be further exploited by collaborative filtering techniques. Another cluster approach is presented in [11] where authors introduce a method using a clustering algorithm to combine content-based and collaborative filters. Users' profiles are first

grouped into clusters and these clustered are then used to create a new user-item matrix for recommendations. Finally, predictions are calculated by using the classic collaborative algorithm based on the new user-item matrix.

To explore the real reasons of users' rating behaviour, a feature-weighted user profile model is proposed in [14]. In addition, a new top N generation list algorithm based on features' frequency is also presented. A final experimental comparison of the proposed method against the state-of-art CF, CB and a hybrid algorithm shows significant improvements of this approach.

In order to explore user generated metadata (i.e. social tags), [12] proposes a new recommendation method that exploits social tags to annotate multimedia items. Tag information is used to analyse user's preferences and make collaborative recommendations. Because social tags can measure user preferences from different semantic dimensions, conducted experiments prove that it surpasses other methods.

The work presented in this paper extends previously published work by deeply analysing the impact of movie metadata in enhancing user profiles from different points of view. For validation purpose, we conducted simulations using two distinct datasets - Movielens and Netflix. This allows results' generalization, by confirming the achievements in independent samples, which was not provided in previously related work. In addition, the impact of slight variation on the sparsity of the datasets is also evaluated. Most of the published work evaluates the performance of the system using standard accuracy metrics like the Mean Average Error (MAE) or the Precision. The appropriateness of such a metric for evaluating the quality of the top-N recommendations has been questioned by several authors [15, 16]. In order to improve the quality of the evaluation, emerging novel metrics (Diversity, Novelty and Serendipity), considered relevant in literature, were used to validate our approach.

3   A semantic approach for movie recommendation

3.1   Metadata preferences profile

In the movie domain, preferences and tastes of the users may be guided by the genre of the content or by the film crew. Information on the rating and on the metadata associated to the content can be analysed and used to distinguish users and to create a profile that represents the level of interest that a user has for each of the existing metadata element. For example, if a user rated 5 all comedies that he saw while for romances he decided for a 2, the system should infer that the user prefers comedy movies, recommending him comedies instead of romances.

The level of preference for a given metadata can be calculated according to Eq. 1, where r_i is the rating assigned by the user to a movie, $n_{r_{movies\ rated}}$ represents all the movies that contain the metadata j with the rating r_i assigned and $n_{r_{watched\ movies}}$ represents the number of movies watched by the user and that contain the evaluated metadata. r_{min} and r_{max} correspond, respectively, to the minimum and maximum rating that a user can assign to an item.

The final preference profile for a selected metadata is represented as a vector. The vector size is the number of concepts that the metadata consists of (p.e. $p_{ugenre}=M_{action},M_{romance},\ldots,M_{terror}$).

$$M_j = \frac{\sum_{i=r_{min}}^{r_{max}}\left(r_i \times n_{r_{movies\ rated}}\right)}{n_{r_{watched\ movies}}} \quad . \tag{1}$$

3.2 Hybrid recommender approach

In a typical collaborative approach, the pattern of numerical ratings for individual users is used to find the similarity between them. In contrast, in a collaborative-content approach, the content-based profile for each user is exploited to detect similarities among users. In this approach, profiles are essentially vectors of terms and weights that can be compared across users to compute predictions. These users' weight vectors are a compressed representation of a user's interest, and the collaborative mechanism that follows can operate on this dense information representation more easily than on raw rating data. In our work, we focus on a sequential combination of content-based filtering and collaborative filtering where, initially, a content-based algorithm is applied to find users who share similar interests, and then, a collaborative algorithm is applied to make predictions. This methodology uses a prediction scheme similar to the standard collaborative filtering while maintaining the content-based profiles for each user. The outline of our approach includes the following steps:

1. Build a content-based user profile considering movie metadata and ratings given by users.
2. Find user's neighbourhoods by calculating the similarity between each user using the Pearson correlation coefficient. To find similar users, movie metadata elements (genre, actors and directors) are used individually and by combining them. The user can have one to three metadata preference profiles associated with each of metadata categories.
3. Implement a standard user-based collaborative filtering algorithm, as presented in [17], considering neighbours having the most similar metadata preference profile, as discussed above.

4 Experiments setup

4.1 Evaluation metrics

Several accuracy metrics have been proposed to evaluate the performance of recommenders [16, 18]. In recent years, the academic community has also discussed the use of novel metrics which may influence users' satisfaction [19, 20]. The results of our work were evaluated using both traditional and novel metrics.

Precision can be calculated as the ratio of relevant items in the total number of items selected as presented in Eq. 2. In an n-ary classification systems, an appropriate

threshold may be used to classify as relevant or irrelevant each item. For a rating scale ranging from 0 to 5, classifications above 4 could be considered as a "like" (relevant) and, below that, as a "dislike" (irrelevant).

$$precision = \frac{N_{RelevantSelected}}{N_{Selected}} \; . \tag{2}$$

Diversity has been considered as one important aspect towards recommendation quality [20, 21] as it helps users to discover new interests that they might not have discovered by themselves, enhancing then user's experience. It is frequently defined as the opposite of similarity [22]. In our work, we followed the definition proposed by [23] where diversity of a set of items, $i_1, \dots i_n$, is defined as the average dissimilarity between all pairs of items in the result set as presented in Eq. 3.

$$diversity(i_1, \dots, i_n) = \frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \left(1 - sim(i_i, i_j)\right)}{\frac{k(k-1)}{2}} \; . \tag{3}$$

The similarity, *sim*, between movies was calculated considering information on the genre, actors and directors of each movie. The final value is the average similarity to the aforementioned metadata. For the similarity based on the genre, the cosine distance was used while for the similarity based on the actors and directors the Inverse Rank Measure [22, 24] was adopted.

The ability to find surprising TV contents may also contribute to user satisfaction. In order to measure the ability of our algorithm to recommend novel items in a top-N list, the novelty metric proposed in [25], named Expected Popularity Complement(EPC), was applied:

$$EPC = \frac{\sum_{u \in U} \sum_{r=1}^{N} \frac{rel(u,i_r) * (1 - pop(i_r))}{\log_2(r+1)}}{\sum_{u \in U} \sum_{r=1}^{N} \frac{rel(u,i_j)}{\log_2(r+1)}} \; . \tag{4}$$

where i_r represents the item that is at the ranking position r of the current recommendation list with size N; $rel(u,i_r)$ is a binary value (0 or 1) since for this metric only relevant items are considered; $pop(i_r)$ - the popularity - is calculated as the ratio between the number of items that have been rated so far, *Rat(i)*, and the number of ratings of the most rated item in the item set I as presented in (5). Additionally, the items are weighted according to their position r in the recommendation list by using a logarithmic discount.

$$pop(i) = \frac{|Rat(i)|}{\max_{i \in I}|Rat(i)|} \; . \tag{5}$$

Another concept widely acknowledged as a key aspect in RS quality is serendipity (the capacity to surprise the user by suggesting fortuitous and expected content) [26]. In order to measure the performance of our method in suggesting serendipitous items in a top-N recommendation list, the approach presented in [27, 28], which captures two aspects of serendipity (unexpectedness and usefulness), was adopted:

$$serendipity(u) = \frac{|UNEXP(u) \cap USEFUL(u)|}{N}. \tag{6}$$

where *UNEXP(u)* represents an unexpected set of recommendations for user *u* and *USEFUL(u)* is the useful (relevant) items for user *u* which, for this work, corresponds

to items rated by that user above a defined threshold and *N* represents the size of the recommendation set *RS(u)*.

4.2 Datasets partitioning

Tests were conducted using the Netflix and Movielens10M datasets, two well-known datasets in the movies domain. Since these datasets do only have information concerning ratings that users gave to items, available APIs were used to extract the required metadata from existing services and to enhance the datasets.

Given the datasets contain a large set of ratings (a few millions) a dataset split and resizing was made to reduce computational costs. Following the approach presented in [18] the top 3000 users, that is, the ones that contributed with more ratings, were selected. These users were further split into 3 groups, according to the percentage of ratings. As shown in Table 1, the sparsity of the defined datasets is still notably large.

Table 1: Sub-datasets constructed for the experiments (based on the number of ratings)

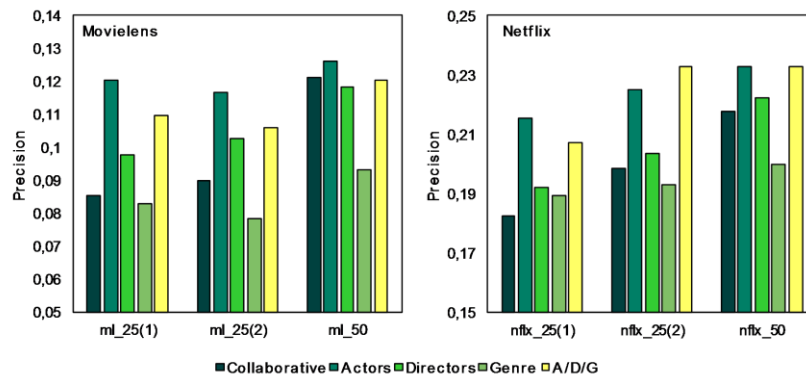| Datasets					
Movielens			Netflix		
Name	Ratings percentage	Sparsity	Name	Ratings percentage	Sparsity
ml_25(1)	1114 users, ~25% of all the ratings	94%	nflx_25(1)	972 users, ~25% of all the ratings	89%
ml_25(2)	872 users, ~25% of all the ratings	93%	nflx_25(2)	874 users, ~25% of all the ratings	88%
ml_50	1025 users, ~50% of all the ratings	87%	nflx_50	1197 users, ~50% of all the ratings	82%

4.3 Testing methodology

For testing our approach, the user-based collaborative filtering algorithm from the Apache Mahout framework [29] was used. External modules to allow movie metadata to be integrated into the framework were developed. Simulations include:

1. Comparing the performance of collaborative single rating and hybrid approaches considering each of the metadata elements individually, as well as all the possible combinations: the three elements together (A/D/G); actors and directors (A/D); actors and genres (A/G); genres and directors (G/D). The Pearson correlation was used to compute users' similarity. A neighbourhood size of 5 was considered;

2. Analysing performance using different evaluation metrics (Precision, Diversity, Novelty and Serendipity). Items rated 4 or up were defined as relevant and a top-N methodology with N defined as 10 was used.

5 Results

The first conclusion that can be drawn (Fig. 1) is that our approach increases the precision when compared to the standard collaborative algorithm that computes the similarity between users using single numerical ratings. These results were validated using both datasets. The users' metadata profiles that enable the best performance result from the aggregation of information on the genre, actors and directors (A/D/G) and on the one that used just the information on the actors. Given that the impact on using a more complete set of metadata (A/D/G) does not contribute to increase significantly the performance and that computational costs are significantly higher when compared to using just one metadata element, an enhanced profile based on the actors can be selected as the best approach. As shown also in Fig. 1, the different sparsity levels of the datasets (small number of ratings or large number of new items) does not affect the results.

Fig. 1. Precision for a) Movielens dataset; b) Netflix dataset

Fig. 2 shows the results when analysing the diversity of the items recommended for two sample datasets. The simulations show a slight improvement of our hybrid approach when compared to the standard collaborative filtering algorithm. This conclusion is independent of the dataset and of the dataset sparsity. Despite none of the metadata elements outperforms another, the combination of all the metadata information presented fairly consistent results for all datasets samples.

Novelty and serendipity results are depicted in Fig. 3. It is highly noticeable that performance patterns are very similar for these two indicators. Again, performance was enhanced when compared to the simple approach that uses single ratings. No metadata that consistently stands out can be identified. However, once again, the profile that combines all metadata seems to be the approach that demonstrates the overall consistency and best results.
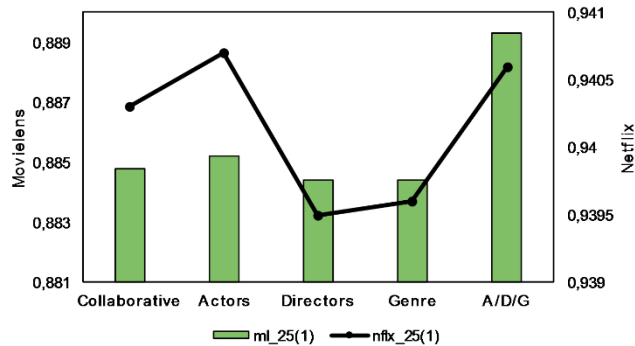
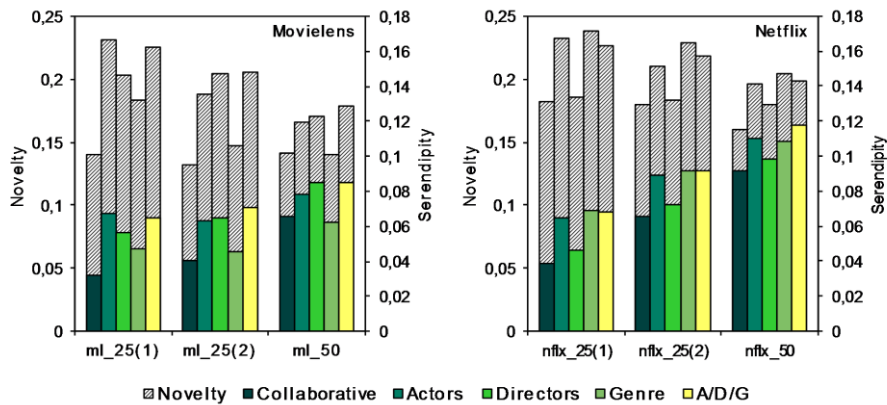**Fig. 2.** Diversity for Movielens and Netflix datasets

Fig. 3. Novelty and serendipity for a) Movielens dataset; b) Netflix dataset

6 Conclusions

This paper compares the performance of a standard collaborative algorithm against a content-collaborative hybrid approach that explores an enhanced profile that reflects the value of metadata categories. We confirmed experimentally that constructing a metadata profile improves the results when compared to using single ratings. Besides improving the performance, other advantages can be identified. The first concerns the ability to consider proximities between users even if they did not rate any common item. The second is related to the scalability of the algorithm. While for the CF approach the information to be processed (matrix of ratings for old and new items) grows continuously with time, the profiles' matrixes grow more slowly, influencing then positively the computational costs. This can be illustrated by the genre metadata profile, as usually this element has, from the beginning, a fixed and small number of categories (comedy, terror, drama, etc.). This enables the profile length to be kept with a fixed size over time, enhancing the scalability of the approach.

Results also show that the metadata element used to construct the users' profiles has some influences on the recommendation's output. Used individually, the actors demonstrate to be the better choice. However, the combination of metadata demonstrated, in general, a more stable performance for all the tested scenarios than any metadata considered individually. The different possible combinations between two metadata elements also showed, for most scenarios, better performance than the metadata used individually. For that reason, the results seem to demonstrate that combining different information allows representing more accurately users' preferences.

The results provided by other metrics (diversity, novelty and serendipity) show that an improvement of our approach can also be noticed when comparing to the standard user-based CF algorithm. The combination of genre, actors and directors, enables, for both datasets, the best and most consistent results. However, the gain in performance cannot be said to provide a substantial benefit, as it implies greater computational costs. If used individually, actors and directors can be said to have the disadvantage over genre by growing along time.

Future work includes the integration, in the recommendation framework, of other collaborative filtering methods that also utilize metadata and then evaluating its performance against results presents in this paper.

References

1. Su, X., Khoshgoftaar, T.M.: A Survey of Collaborative Filtering Techniques. Adv. in Artif. Intell. 2009, 4:2–4:21 (2009)
2. Lops, P., Gemmis, M., Giovanni, S.: Content-based Recommender Systems: State of the Art and Trends. In: Recommender Systems Handbook. pp. 73–105. Springer US (2010)
3. Soares, M., Viana, P.: Tuning metadata for better movie content-based recommendation systems. Multimed Tools Appl. 74, 7015–7036 (2014)
4. Burke, R.: Hybrid Recommender Systems: Survey and Experiments. User Model User-Adap Inter. 12, 331–370 (2002)
5. Kabassi, K.: Personalizing recommendations for tourists. Telematics and Informatics. 27, 51–66 (2010)
6. Viana, P., Soares, M.: A Hybrid Recommendation System for News in a Mobile Environment. In: 6th International Conference on Web Intelligence, Mining and Semantics. p. 3:1–3:9. ACM, New York (2016)
7. Lakiotaki, K., Matsatsinis, N.F., Tsoukias, A.: Multicriteria User Modeling in Recommender Systems. IEEE Intelligent Systems. 26, 64–76 (2011)
8. Al-Hassan, M., Lu, H., Lu, J.: A semantic enhanced hybrid recommendation approach: A case study of e-Government tourism service recommendation system. Decision Support Systems. 72, 97–109 (2015)

9. Wang, R.-Q., Kong, F.-S.: Semantic-Enhanced Personalized Recommender System. In: Int. Conf. on Machine Learning and Cybernetics. pp. 4069–4074. IEEE Press, New York (2007)

10. Cantador, I., Bellogín, A., Castells, P.: A multilayer ontology-based hybrid recommendation model. AI Communications - Recommender Systems. 21, 203–210 (2008)

11. Kim, B.M., Li, Q., Park, C.S., Kim, S.G., Kim, J.Y.: A new approach for combining content-based and collaborative filters. J Intell Inf Syst. 27, 79-91 (2006)

12. Lee, W.-P., Kaoli, C., Huang, J.-Y.: A smart TV system with body-gesture control, tag-based rating and context-aware recommendation. Knowledge-Based Systems. 56, 167-178 (2014)

13. Manzato, M.G.: Discovering latent factors from movies genres for enhanced recommendation. In: 6th ACM Conf. on Rec. Syst. pp. 249–252. ACM, New York (2012)

14. Symeonidis, P., Nanopoulos, A., Manolopoulos, Y.: Feature-weighted user model for recommender systems. In: Conati, C., McCoy, K.F., and Paliouras, G. (eds.) User Modeling. pp. 97–106. Springer (2007)

15. Breese, J.S., Heckerman, D., Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering. In: 14th Conf. on Uncertainty in Artificial Intelligence. pp. 43–52. Morgan Kaufmann Publishers Inc., USA (1998)

16. Herlocker, J., Konstan, J.: Evaluating collaborative filtering recommender system. ACM Transactions on Information Systems. 22, 5–53 (2004)

17. Adomavicius, G., Kwon, Y.: New recommendation techniques for multicriteria rating systems. IEEE Intelligent Systems. 22, 48–55 (2007)

18. Cremonesi, P., Turrin, R., Lentini, E., Matteucci, M.: An evaluation methodology for collaborative recommender systems. In: Int. Conf. on Automated Solutions for Cross Media Content and Multi-channel Distribution. pp. 224-231. IEEE, New York (2008)

19. Murakami, T., Mori, K., Orihara, R.: Metrics for evaluating the serendipity of recommendation lists. In: Satoh, K., Inokuchi, A., Nagao, K., and Kawamura, T. (eds.) New Frontiers in Artificial Intelligence. pp. 40–46. Springer, Berlin (2008)

20. Vargas, S., Castells, P.: Rank and relevance in novelty and diversity metrics for recommender systems. In: 5th ACM Conf. Rec. Syst. pp. 109–116. ACM, New York (2011)

21. Hurley, N., Zhang, M.: Novelty and diversity in top-N recommendation–analysis and evaluation. ACM Transactions on Internet Technology (TOIT). 10, 14:1-14:30 (2011)

22. Kompatsiaris, Y., Merialdo, B., Lian, S.: TV Content Analysis - Techniques and Applications Multimedia Computing, Communication and Intelligence. CRC Press, New York (2012)

23. Smyth, B., McClave, P.: Similarity vs. Diversity. In: 4th Int. Conf. on Case-Based Reasoning: Case-Based Reasoning Research and Development. pp. 347–361. Springer-Verlag, UK (2001)

24. Bar-Ilan, J., Keenoy, K., Yaari, E., Levene, M.: User rankings of search engine results. J Assn Inf Sci Tec. 58, 1254–1266 (2007)

25. Niemann, K., Wolpers, M.: A new collaborative filtering approach for increasing the aggregate diversity of recommender systems. In: 19th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining. pp. 955–963. ACM, New York (2013)

26. McNee, S.M., Riedl, J., Konstan, J.A.: Being accurate is not enough: how accuracy metrics have hurt recommender systems. In: CHI '06 Extended Abstracts on Human Factors in Computing Systems. pp. 1097–1101. ACM, New York (2006)

27. Ge, M., Delgado-Battenfeld, C., Jannach, D.: Beyond accuracy: evaluating recommender systems by coverage and serendipity. In: 4th ACM Conf. on Recommender Systems. pp. 257–260. ACM, New York (2010)

28. Qiuxia, L., Tianqi, C., Weinan, Z., Diyi, Y., Yong, Y.: Serendipitous personalized ranking for top-N recommendation. In: IEEE/WIC/ACM Int. Joint Conf. on Web Intelligence and Intelligent Agent Technology. pp. 258–265. IEEE Computer Society, New York (2012)

29. Apache Mahout, http://mahout.apache.org/