A Cross-Modal Warm-Up Solution for the Cold-Start Problem in Collaborative Filtering Recommender Systems

Behnoush Abdollahi
Knowledge Discovery & Web Mining Lab,
University of Louisville
b.abdollahi@louisville.edu

Olfa Nasraoui
Knowledge Discovery & Web Mining Lab,
University of Louisville
olfa.nasraoui@louisville.edu

ABSTRACT

We present a cross-modal recommendation engine that leverages multiple domains of data while performing matrix factorization. We show how our approach has the potential to alleviate the cold-start problem for new items, one of the notorious limitations of Collaborative Filtering (CF) techniques.

Categories and Subject Descriptors

G.1.3 [**Sparse, structured, and very large systems (direct and iterative methods)**]: Numerical Linear Algebra; H.3.3 [**Information filtering**]: Information Search and Retrieval

Keywords

Collaborative Filtering; Non-negative Matrix Factorization

1. INTRODUCTION

Non-negative Matrix Factorization (NMF) [5] has recently gained popularity as a powerful technique for generating Collaborative Filtering (CF) recommendations [2]. Using matrix factorization, items, users and their attributes can be represented in a latent space and the recommendation task can be performed in this latent space. NMF has been used as a recommender system technique [3] to map both users and items to a joint latent space of k factors, where each user p_u, and item q_i, is represented as a vector of k elements. The user-item interaction can then be modeled as the dot product of their representation in the latent space, which estimates the interest of user p_u in item q_i, as follows:

$$\hat{r} = \mathbf{p}_u \mathbf{q}_i^T \tag{1}$$

However, the rating matrix is usually very sparse, and when new users enter the system there is not enough information regarding their interest. Also, new items may have insufficient or even no ratings at all. The latter case is known as the new item cold-start problem, which is a prevalent problem in CF recommender systems. Using multiple domains of data is one way to overcome this problem, typically within the scope of hybrid recommender systems. Combining information from multiple domains (ratings, item attributes, etc) has been used in cross-domain recommendation engines, such as [6, 8] which can transfer a learned model from the rating data to another rating domain. Koren et al. [4] used movie attributes and demographic data to build their NMF-based recommender systems. The latter model is a mixed model, that considers all data domains simultaneously. In contrast to the mixed model, our approach is based on an asymmetric multi-domain NMF-based model [1] that was recently proposed for multimodal image retrieval and annotation. Our NMF-based cross-modal CF recommender system leverages multiple domains of data to recommend even new items, thus overcoming the new item cold-start problem.

WebSci'14, June 23–26, 2014, Bloomington, IN, USA.
ACM 978-1-4503-2622-3/14/06.
http://dx.doi.org/10.1145/2615569.2615665 .

2. NON-NEGATIVE MATRIX FACTORIZATION (NMF)

Non-negative matrix factorization (NMF) is a family of algorithms where the non-negative data R is factorized into two non-negative matrices P and Q [5]

$$\mathbf{R}_{n \times m} \simeq \mathbf{P}_{n \times k} \mathbf{Q}_{m \times k}^T \tag{2}$$

where k is the rank of matrices P and Q. In order to estimate P and Q, the Frobenius norm of the errors between the approximation PQ^T and the data matrix R is used as the cost function to be minimized

$$J_{NMF} = ||\mathbf{E}||_F^2 = ||\,\mathbf{R} - \mathbf{P}\mathbf{Q}^\mathbf{T}\,||_F^2 \tag{3}$$

where $||\mathbf{E}||_F^2 = \sum_i \sum_j E_{ij}^2 = tr(\mathbf{E}^T\mathbf{E})$. This optimization problem can be solved using gradient descent algorithms.

3. PROPOSED APPROACH

3.1 Asymmetric NMF for Collaborative Filtering

In CF, the item attributes can be a valuable source of information to integrate into building a multi-domain model. In this work, we use multi-domain Asymmetric NMF to exploit multi-modal interactions between user ratings and other domains such as movie genre (for movie recommendation), and hence improve recommendations while solving the new item cold start problem. In Asymmetric NMF, the latent semantic space is first derived using the first domain, then follows an adaptation phase in which the second domain is utilized to fit the former latent space. This results in a common space where all domains co-exist.

For movie recommendation, the two domains are: Movie Genre (R_1 $(g \times m)$) and Movie Ratings (R_2 $(n \times m)$), where g is the number of genres, m is the number of movies, and n is the number of users. The two main steps of the algorithm are as follows:

1. Build a latent space model based on the Movie Genre domain: This step applies NMF to the Movie Genre matrix: $R_1 = P_1 Q^T$. In this formula, P_1 $(g \times k)$ is the basis matrix for transforming the Genre data to the latent space and Q $(k \times m)$ is the representation of the items (movies) in latent space.

2. Adapt the basis data in the latent space: $R_2 = P_2 Q^T$. In this step, Q is transferred as fixed latent factor coefficients from the first step and the ratings matrix is used to adapt the basis P_2 $(n \times k)$ for the second domain consisting of the user-item ratings, and thus construct a combined latent space based on both the Movie Ratings and the Movie Genres. As a consequence, Q spans the semantic space of both ratings and item attributes.

4. EXPERIMENTS

We tested our approach on the MovieLens dataset [1]. The movies fall into 19 different genres, such as comedy, drama, etc. The ratings data has 100,000 ratings, on a scale of 1 to 5, for 1700 movies by 1000 users. We performed two experiments. In the first experiment, where we test the recommendation accuracy, a different percentage of the ratings data (from 5% to 50%) is selected randomly out of all ratings (1700×1000) and changed to unrated to serve as test data. The remaining percentage of data, in addition to the genre domain, are used in training. In the second experiment, we focus on the new item cold start problem. The test set is selected by sampling varying percentages from the 1700 movies (from 5% to 50%). Then the ratings of the selected test movies are changed to unrated to test the impact of new items. In addition to varying the percentage of test data, the number of factors, k, was varied. All the experiments were repeated 10 times and the average metrics are reported. We compared our method with the classical NMF as the baseline on the data used in Experiment 1. Figure 1 shows the Mean Square Error (MSE) for both experiments and the classical NMF with different percentages of test data. We also illustrate our results in Table 1 which shows the top 5 rated movies for two sample users and their top 5 recommended movies. The underlined recommended movies are new movies with no ratings available in the training data. The results show that Asymmetric NMF increases the accuracy of recommendations compared to the classical NMF. In the new movie cold-start case, our approach performed even better, while classical NMF cannot be applied for lack of any ratings.

[1] http://grouplens.org/datasets

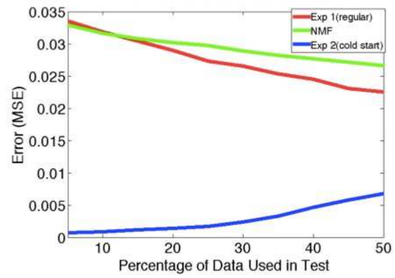

Figure 1: MSE vs. percentage of data as test

Table 1: Top 5 rated and recommended movies to 2 sample users. Underlined movies are unrated movies.

Model	Test user	Top 5 rated movies by the user	Top 5 recommended movies
Expt 1	User#1	**Braveheart (1995), Star Wars (1977), Pulp Fiction (1994),** Stargate (1994), Crow The (1994)	**Star Wars (1977), Braveheart (1995),** Raiders of the Lost Ark (1981), Terminator The (1984), **Pulp Fiction (1994)**
Expt 2	User#2	Star Wars (1977), Godfather The (1972), **Bottle Rocket (1996),** Twelve Monkeys (1995), **When We Were Kings (1996)**	**Bottle Rocket (1996),** Shallow Grave (1994), 2 Days in the Valley (1996), **When We Were Kings (1996),** Blue in the Face (1995)

5. CONCLUSION

Cross-modal Asymmetric NMF provides an easy and accurate solution to the cold-start problem, a notorious problem plaguing traditional collaborative filtering recommender systems. We have also used this approach in other applications, such as cross-modal annotation for hashtag completion in tweets [7].

6. REFERENCES

[1] Ben-Abdallah, J., Caicedo, J. C., González, F. A., and Nasraoui, O. Multimodal image annotation using non-negative matrix factorization. In *Web Intelligence* (2010), pp. 128–135.

[2] Herlocker, J. L., Konstan, J. A., Borchers, A., and Riedl, J. An algorithmic framework for performing collaborative filtering. In *SIGIR* (1999), pp. 230–237.

[3] Koren, Y. Factorization meets the neighborhood: A multifaceted collaborative filtering model. *Proceedings of the 14th ACM SIGKDD* (2008), 426–434.

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

[5] Lee, D. D., and Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature 401* (1999), 788–791.

[6] Li, B., Yang, Q., and Xue, X. Transfer learning for collaborative filtering via a rating matrix generative model. *Proceedings of the 26th Annual International Conference on Machine Learning* (2009), 617–624.

[7] Nutakki, G. C., Nasraoui, O., Abdollahi, B., Badami, M., and Sun, W. Distributed lda-based topic modeling and topic agglomeration in a latent space. In *SNOW 2014 Data Challenge, 23rd International World Wide Web Conference (WWW 2014)* (2014), pp. 17–24.

[8] Pan, W., Xiang, E., Liu, N., and Yang, Q. Transfer learning in collaborative filtering for sparsity reduction. *Proceedings of the 24rd AAAI Conference on Artificial Intelligence* (2010), 425–434.