Repeat Consumption Recommendation Based on Users Preference Dynamics and Side Information

Dimitrios Rafailidis
Aristotle University of Thessaloniki
Thessaloniki, Greece
draf@csd.auth.gr

Alexandros Nanopoulos
Catholic University of Eichstät-Ingolstadt
Ingolstadt, Germany
alexandros.nanopoulos@ku.de

ABSTRACT

We present a Coupled Tensor Factorization model to recommend items with repeat consumption over time. We introduce a measure that captures the rate with which the preferences of each user shift over time. Repeat consumption recommendations are generated based on factorizing the coupled tensor, by weighting the importance of past user preferences according to the captured rate. We also propose a variant, where the diversity of the side information is taken into account, by higher weighting users that have more rare side information. Our experiments with real-world datasets from last.fm and MovieLens demonstrate that the proposed models outperform several baselines.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Miscellaneous

Keywords

Repeat consumption; preference dynamics; coupled tensor factorization

1. INTRODUCTION

User preferences can be fairly dynamic, since users tend to exploit a wide range of items and modify their taste accordingly over time. Temporal collaborative filtering with matrix [4, 7] and tensor factorization techniques [3] can generate accurate recommendation, since they can capture the drifts in the rating behavior and the changes of user preference over time in a collaborative-filtering fashion. However, users may repeatedly interact with items over time in several applications [2]; for instance, visiting the same web sites, buying retail items from Amazon or implicitly interacting, such as artist listenings on last.fm or movie viewing from a specific genre on MovieLens. To account also for the fact that users' side information, such as demographics, can improve the recommendation accuracy [5], we present a basic CTF model and its variant W-CTF, where the diversity of the side information is taken into account, by higher weighting the users that have more rare side information.

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2742732.

2. THE PROPOSED CTF MODEL

Our basic CTF model [6] captures the shared latent factors across the time varying user preference data and the side information about the users. CTF consists of the following steps:

Modeling User Preferences and Side Information: Given the sets U, I and T of users, items and tags, as well as a tensor $\mathcal{X} \in \mathbb{R}^{|U|\times|I|\times|T|}$, each non-empty tensor cell $\mathcal{X}(u,i,t)$ contains the number of interactions of user u to item i at the time period t. The time period can be days, months, or whole years, corresponding to the $|T|$ different time slices $|U| \times |I|_t$ of tensor $\mathcal{X}$, with $t \in 1 \dots |T|$. The test time period, e.g., the test month where the personalized recommendations have to be generated, is included in the last time slice $|U| \times |I|_{|T|}$, e.g., the current 6-month period, of the tensor $\mathcal{X}$. With respect to the users' side information the u-th, $u \in U$, row of the auxiliary matrix Y corresponds to the set D of attributes of user u. In case of numerical attributes, e.g., age, we perform an equal-width binning, where for each numerical entry in the matrix Y we store the respective number of the bin. In the case of categorical attributes, such county and gender, we calculate the c distinct categorical values and then we create a binary vector $\mathbf{c}_i \in \mathbb{R}^{c\times 1}$, where 1 denotes the categorical attribute of user u_i. Finally, we append the transformed numerical and categorical attributes to generate the final $|D|$ attributes in matrix Y.

Down-weighting past preferences: Given (i) a test period t within the current/last time slice $|U| \times |I|_{|T|}$ of the tensor $\mathcal{X}$; (ii) $I^u_{cur} \subseteq I$,the set of items that user u has interacted at the current/last time slice $|U| \times |I|_{|T|}$ of $\mathcal{X}$; and (iii) $I^u_{prev} \subseteq I$ the set of items that user u has interacted at all the previous time slices $|U| \times |I|_t$, with $t = 1 \dots |T| - 1$; for each user u we calculate its User Preference Dynamic (UPD_u) value:

$$UPD_u = 1 - \frac{|I^u_{cur} \cap I^u_{prev}|}{|I^u_{cur} \cup I^u_{prev}|}$$

where a high UPD_u value indicates that u has significantly shifted his current preferences. Then, we calculate a personalized smoothing factor $sf_u = 1 - UPD_u$ and for each user $u \in U$ we downweigh his past preferences $\mathcal{X}(u,i,t) := sf_u * \mathcal{X}(u,i,t)$, with $t = 1, \dots, |T| - 1$ and $i \in I^u_{prev}$.

Coupled Tensor Factorization: To generate the final top-k recommendations we calculate the low rank R approximation of tensor $\mathcal{X}$ by minimizing the objective function

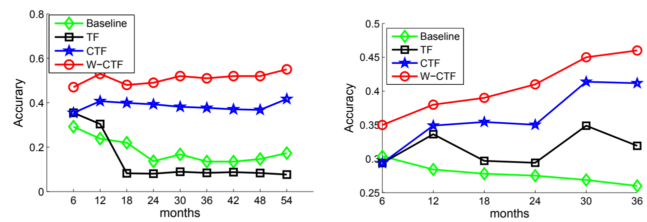
$f(A,B,C,V) = ||\mathcal{X} - [A,B,C]||^2 + ||Y - AV^T||^2_F$, where $||\cdot||^2_F$ indicates the Frobenius norm; matrices $A \in \mathbb{R}^{|U|\times R}$, $B \in \mathbb{R}^{|I|\times R}$ and $C \in \mathbb{R}^{|T|\times R}$ are the factor matrices of $\mathcal{X}$, extracted by CANDECOMP/PARAFAC decomposition; matrices A and $V \in \mathbb{R}^{|D|\times R}$ are the factor matrices extracted from matrix Y (with users' side information) by a matrix factorization technique. Hence, we calculate the matrices A, B, C and V by minimizing the approximation error with the help of the CMTF-OPT algorithm [1], which also handles missing values.

W-CTF: In the first step of CTF, the modeling of users' side information with the $|D|$ attributes in the coupled matrix Y may negatively affect the performance of the model, i.e. $Y(u,d) = 1$, if user u has an attribute $d \in D$ and 0 otherwise; for instance, users that have a unique attribute that rarely appears in set D, are not considered in the calculation of coupled matrix Y. To handle this problem we propose a variant of CTF, namely W-CTF. Given a subset $C \subseteq D$ of attributes, that indicate the same demographic data, e.g. the set of the $|C|$ distinct counties, $\forall$ $d \in C$ the initial $|U|$ weights $Y(u,d)$ are recalculated as $Y(u,d) := Y(u,d) * |U|/f(d)$ and then are normalized to [0 1], where $|U|$ is the number of distinct users and $f(d)$ is the number of users that have 1s for the d-th attribute in the initial Y matrix. In doing so, we higher weigh the users that have more rare side information in the d-th attribute. Provided that W-CTF follows a collaborative filtering strategy, in the extreme case of $f(d)$=1, we filter out the d-th attribute, since in this case there is only one user with the d-th attribute, and consequently the model cannot identify any other user to base the recommendation according to the side information of the d-th attribute.

3. EXPERIMENTS

Datasets: In our experiments, we used the last.fm-1K [9] and MovieLens-1M [10] datasets. Last.fm contains 992 users, 176,948 artists, and 19,150,868 listening events over 54 months. The side information includes users' age, gender and county. We split the dataset into 9 different 6-month periods, each represented as a tensor slice. Given a set of training months the goal is to perform artist recommendation for a user at a test month. We used a time window equal to 6 months, where as training set we considered all the past months of the previous periods and the first five months of the current period. The goal is to predict the artists that each user is going to listen at the last (6-th) test month of the current period. MovieLens-1M has 1,000,209 ratings of 3,952 movies of 6,040 users over 36 months, where we also split the data set into 6-month periods. The side information includes users' age, gender and occupation type. Each movie can belong to more than one movie-genre (18 in total). In our case of repeat consumptions, we consider how many times a user has watched a movie of a specific movie-genre within the time slot. The goal is to predict the genres of movies that each user is going to watch at the last (6-th) test month of the current ongoing period. More details for both datasets and the evaluation protocol can be found in Section 1 of our online technical report [8].

Results: In Figures 1(a)-(b), we compare the proposed CTF and W-CTF models against TF [3] and a baseline method, where we recommend the most-popular artists/movie-genres for each user within the training months. The proposed models outperform TF and the baseline, because both capture users' preference dynamics in a personalized way and also incorporate users' side information. An important observation is that after the critical point of 18 months (see Figure 3 at [8]), where users start to significantly shift their preferences in last.fm, the accuracy of TF starts to decrease fast, even lower than the baseline method. This happens because TF neither handles the users' personalized preference dynamics nor use auxiliary information. Finally, W-CTF achieves higher recommendation accuracy than CTF, by considering the diversity of the side information and higher weighting the users that have more rare side information. In Table 1 at [8], we evaluate the performance of the examined methods on 3 users UPD groups and observe that W-CTF outperforms CTF. For instance, for group $UPD \geq 0.75$ at the 54-th and 36-th months in last.fm and MovieLens, W-CTF achieves accuracy $[0.55 \pm 0.07, 0.50 \pm 0.03]$, while CTF $[0.43 \pm 0.09, 0.48 \pm 0.04]$.



(a) Accuracy (last.fm) (b) Accuracy (MovieLens)

Figure 1: Experimental results.

4. CONCLUSIONS

We proposed the CTF and W-CTF models for recommendation of items with repeat consumption, by capturing the rate with which the current preferences of each user shift over time and by exploiting side information in a coupled tensor factorization technique. Future work includes the incorporation of items' inherent *quality of attractiveness* into our models, which may increase the recommendation accuracy according to the analysis of the dynamics of repeat consumption by Anderson et al. in [2].

5. REFERENCES

[1] E. Acar, T. G. Kolda, and D. M. Dunlavy. All-at-once optimization for coupled matrix and tensor factorizations. *CoRR*, abs/1105.3422, 2011.

[2] A. Anderson, R. Kumar, A. Tomkins, and S. Vassilvitskii. The dynamics of repeat consumption. In *WWW*, pages 419–430, 2014.

[3] D. M. Dunlavy, T. G. Kolda, and E. Acar. Temporal link prediction using matrix and tensor factorizations. *TKDD*, 5(2):10, 2011.

[4] Y. Koren. Collaborative filtering with temporal dynamics. In *KDD*, pages 447–456, 2009.

[5] A. Narita, K. Hayashi, R. Tomioka, and H. Kashima. Tensor factorization using auxiliary information. In *ECML/PKDD*, pages 501–516, 2011.

[6] D. Rafailidis and A. Nanopoulos. Modeling the dynamics of user preferences in coupled tensor factorization. In *RecSys*, pages 321–324, 2014.

[7] C. Zhang, K. Wang, H. Yu, J. Sun, and E. Lim. Latent factor transition for dynamic collaborative filtering. In *SDM*, pages 452–460, 2014.

[8] https://github.com/drafail/docs/raw/master/TechReport.pdf.

[9] www.dtic.upf.edu/~ocelma/MusicRecommendationDataset.

[10] http://grouplens.org/datasets/movielens.