Switching Hybrid for Cold-Starting Context-Aware Recommender Systems

Matthias Braunhofer
Free University of Bolzano
Piazza Domenicani 3
Bolzano, Italy
mbraunhofer@unibz.it

Victor Codina
Technical University of Catalonia
Jordi Girona 1-3
Barcelona, Spain
vcodina@lsi.upc.edu

Francesco Ricci
Free University of Bolzano
Piazza Domenicani 3
Bolzano, Italy
fricci@unibz.it

ABSTRACT

Finding effective solutions for cold-starting Context-Aware Recommender Systems (CARSs) is important because usually low quality recommendations are produced for users, items or contextual situations that are new to the system. In this paper, we tackle this problem with a switching hybrid solution that exploits a custom selection of two CARS algorithms, each one suited for a particular cold-start situation, and switches between these algorithms depending on the detected recommendation situation (new user, new item or new context). We evaluate the proposed algorithms in an off-line experiment by using various contextually-tagged rating datasets. We illustrate some significant performance differences between the considered algorithms and show that they can be effectively combined into the proposed switching hybrid to cope with different types of cold-start problems.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*information filtering*

Keywords

Context-Aware Recommender Systems; Cold-Start Problem; Switching Hybrid System

1. INTRODUCTION

Context-Aware Recommender Systems (CARSs) are a special type of Recommender Systems (RSs) that aim at generating more accurate recommendations by exploiting not only the traditional user and item dimensions but also relevant contextual information (e.g., time, weather, location) in the recommendation process [1]. Context management aggravates the cold-start, i.e., dealing with new users and new items, which is affecting traditional RSs; in CARSs it does not suffice to have enough ratings for users and items, it is also necessary a sufficient number of ratings in the various contextual situations the system may be requested to operate: the new context problem.

In this paper, we analyse the cold-start situations that can be observed in CARSs. In particular, we deal with: (i) the new user problem, which occurs when a new user is added to a CARS and there is not enough information about the user's preferences to make good suggestions; (ii) the new item problem, which refers to new items introduced into a CARS and that can not be recommended until some users have rated them; and (iii) the new context problem, which was not specifically considered so far, and relates to the difficulty of accurately recommending items to users under new contextual situations.

We have tackled the above mentioned problems with two novel CARS algorithms namely, demographics-based CAMF-CC and content-based CAMF-CC, which are hypothesised to perform better in one or more of the above considered cold-start cases, and a switching hybrid combination [5], named SHCA. SHCA switches between these two depending on the target recommendation situation. These three algorithms are compared with two state of the art CARS algorithms: CAMF-CC, which is a variant of Context-Aware Matrix Factorisation (CAMF) [3] and SPF (Semantic Pre-Filtering) [6]. We have evaluated all the considered algorithms in an off-line experiment on some contextually-tagged rating datasets, i.e., datasets of ratings for items tagged with the contextual situations of the user while experiencing the rated item. Results from this study show that, as conjectured, there exist some significant performance differences between the algorithms across the individual cold-start cases, and that the proposed switching hybrid provides a more robust solution to the identified cold-start situations.

2. RATING PREDICTION MODELS

This section describes the four CARS algorithms (two state of the art and two novel) and the switching hybrid solution that we have tested in cold-start situations. The $|Users| \times |Items| \times |Context|$ rating tensor is denoted as R, where $|Users|$ is the number of users, $|Items|$ is the number of items, and $|Context|$ is the number of contextual situations. The tensor entry $r_{uic_1,\dots,c_k}$ is the rating of user u on item i in the contextual situation described by the contextual conditions $c_1, \dots, c_k$, and c_j is the index of the contextual condition for factor j (e.g., the index of condition "sunny" for the "weather" contextual factor).

RecSys'14, October 6–10, 2014, Foster City, Silicon Valley, CA, USA.

http://dx.doi.org/10.1145/2645710.2645757.

2.1 CAMF-CC

CAMF-CC (Context-Aware Matrix Factorisation for item categories) is a variant of CAMF [3] that extends standard Matrix Factorisation (MF) by incorporating baseline parameters for each contextual condition and item category. It models how the ratings of items belonging to a specific category are affected by the contextual conditions. We have chosen that particular CAMF variant since it was shown to outperform other CAMF variants in terms of Mean Absolute Error (MAE) [3]. Given a user u, an item i, a contextual situation $c_1, ..., c_k$ and the set of categories $T(i)$ associated to item i, it predicts ratings using the following rule:

$$\hat{r}_{uic_1,...,c_k} = q_i^\top p_u + \mu + b_i + b_u + \sum_{t \in T(i)} \sum_{j=1}^{k} b_{tc_j} \qquad (1)$$

where q_i and p_u are the latent factor vectors associated to the item i and user u, respectively. μ is the overall average rating, and b_i, b_u and b_{tc_j} are the baseline parameters for item i, user u, and item category-contextual condition pair tc_j, respectively. We learned the model parameters by minimising the associated regularised squared error function.

2.2 Semantic Pre-Filtering

Semantic Pre-Filtering (SPF) [6] is a pre-filtering method [1] that, given a target contextual situation, uses a standard MF model learnt on all the ratings tagged with contextual situations equal or similar to the target one. SPF is conjectured to better perform in cold-start situations because it is able to reuse ratings tagged with contexts that, although syntactically different, they influence the users' ratings similarly. Thus, SPF can build more robust local MF models. In our work, we used the particular SPF variant described in [6], which models contextual conditions as their per-item aggregated impact in combination with Singular Value Decomposition (SVD).

2.3 Content-based CAMF-CC

We conjecture that CAMF-CC may fail to provide accurate recommendations for items that have received few or no ratings (i.e., new item problem). The reason is that CAMF-CC, in order to create profiles for items, relies exclusively on explicit ratings. Content-based CAMF-CC, which we propose here, addresses this problem by taking into account item attributes, for instance, the type of a place of interest (POI), the actors or the genre of a movie, in addition to user ratings. A distinct factor vector x_a for each item's attribute $a \in A(i)$ models how an item's attribute influences the item's rating $\sum_{a \in A(i)} x_a$:

$$\hat{r}_{uic_1,...,c_k} = (q_i + \sum_{a \in A(i)} x_a)^\top p_u + \mu + b_i + b_u + \sum_{t \in T(i)} \sum_{j=1}^{k} b_{tc_j} \qquad (2)$$

2.4 Demographics-based CAMF-CC

Content-based CAMF-CC doesn't explicitly tackle the new user problem. Hence, we propose also a demographics-based CAMF-CC variant that addresses this problem by profiling users through known users attributes, such as gender, age group, personality traits, and so on. A distinct factor vector y_a is used to describe a user through the set of user-associated attributes $\sum_{a \in A(u)} y_a$. By enhancing the user representation demographics-based CAMF-CC predicts ratings using the following rule:

$$\hat{r}_{uic_1,...,c_k} = q_i^\top (p_u + \sum_{a \in A(u)} y_a) + \mu + b_i + b_u + \sum_{t \in T(i)} \sum_{j=1}^{k} b_{tc_j} \qquad (3)$$

2.5 Switching Hybrid - SHCA

We finally propose here SHCA, a switching hybrid algorithm that combines the conjectured strengths of the two previously described novel CARS algorithms: content-based CAMF-CC and demographics-based CAMF-CC. SHCA uses content-based CAMF-CC when it is requested to predict a rating for a new item, while it uses demographics-based CAMF-CC when predicting a rating for a new user or a new contextual situation. When a mixture of the elementary cold-start situations is detected, SHCA computes the predicted rating by averaging the predictions of the two constituent algorithms. For instance, the predictions for a rating that is both for a new user and a new item is the average of the ratings estimated by demographics-based CAMF-CC and content-based CAMF-CC. SHCA is expected to better tackle all kinds of cold-start situations found in CARSs. More adaptive combinations of elementary CARS algorithms will be considered in future work.

3. EXPERIMENTAL EVALUATION

We present here the evaluation of the aforementioned CARS algorithms, compared with a baseline Matrix Factorisation (MF) algorithm [10]. We first describe the datasets used in the evaluation, then the experimental design, and finally we present and discuss the obtained results.

3.1 Datasets

We have used three contextually-tagged rating datasets with different characteristics: *STS* [8]; *CoMoDa* [11]; *Music* [2]. Table 1 provides some descriptive statistics of these datasets. We note that there are only a few contextually-tagged rating datasets available, and these are amongst the more popular ones.

Table 1: Datasets' characteristics

Dataset	STS	CoMoDa	Music
Domain	POIs	Movies	Music
Rating scale	1-5	1-5	1-5
Ratings	2,534	2,296	4,012
Users	325	121	43
Items	249	1,232	139
Contextual factors	14	12	8
Contextual conditions	57	49	26
Contextual situations	931	1,969	26
User attributes	7	4	10
Item features	1	7	2

3.2 Evaluation Procedure

Our experiment involved three separate test cases, one for each of the three cold-start situations being addressed. In the first test case, we measured the algorithms' performance for new users by carrying out a ten-fold cross-validation scheme as follows. First, we shuffled the set of users in the whole rating dataset and split it into ten (roughly) equally sized subsets. Then, in each cross-validation iteration, we used the ratings coming from nine merged user subsets as training set to build the models and the remaining ones as testing set. In this way, the testing set contains users with no ratings in the training set (new users). Finally, for each of the ten iterations, we computed the Mean Absolute Error (MAE) of the models, which were then averaged over all iterations to obtain more reliable results.

In addition, we measured the normalised Discounted Cumulative Gain_1 (nDCG_1) for each evaluation run, using the test procedure described in [12]. We also calculated nDCG_k, $k = 2, \ldots 5$, but for lack of space and since they gave similar results we do not report them.

An identical procedure was used when assessing models' performance on new items and new contextual situations, respectively. The only difference was that here cross-validation was performed on the set of items and contextual situations rather than the set of users. In principle, our evaluation procedure could also be used for assessing the models' performance on mixtures of cold-start cases, by cross-validating on user-item, user-context, item-context and user-item-context tuples. This would be interesting since the proposed hybrid algorithm SHCA is expected to handle this kind of mixed cases. However, we did not assess these performances because the available rating datasets are not large and dense enough to perform this fine-grained analysis. Another aspect not considered here is that there exist various degrees of coldness; but we focused on really cold users, items and contextual situations, i.e., without any observed rating.

3.3 Evaluation Results

The obtained results are summarised in Table 2. Not considering SHCA, which is discussed at the end of this section, it can be noted that demographics-based CAMF-CC performed very well in the new user situation. This supports our hypothesis that in this case CAMF-CC can be improved by using user attributes. The weaker performance of demographics-based CAMF-CC on the *Music* dataset may be explained by noting that, differently from the other datasets, where attributes like age, gender, user personality and zip code are available, here only the users' evaluations of the music genres (e.g., user u likes pop music, giving it five out of five stars) are available.

Regarding the new item scenario, we can observe, as conjectured, the excellent performance of content-based CAMF-CC, if compared to the other non-hybrid algorithms. It is effective in this situation because it profiles items also using the available metadata. In the new context scenario, again the demographics-based CAMF-CC is the best. We note that SPF performs very badly in the *Music* dataset because here contextual situations are made up of only one contextual condition. This makes it impossible for SPF to correctly compute similarities between contextual situations.

Finally, if we now focus on the results obtained by the proposed switching hybrid approach SHCA, it can be seen that it is able to correctly exploit the strengths (and avoid the weaknesses) of the elementary algorithms in all the individual cold-start cases. In all cases, except for the new user scenario in the *Music* dataset, it is the best or second best performing algorithm. This illustrates that SHCA, and likely a more adaptive version of it, can become a viable and flexible solution for cold-starting CARSs if an adequate set of elementary (non-hybrid) models are available.

4. RELATED WORK

So far, only little research has been conducted specifically on the cold-start problem for CARSs. One direct approach to this problem is to acquire additional information about the users, items or contextual situations. However, acquiring such information usually requires some additional effort of the users and/or service providers. This effort can be mitigated by employing an Active Learning (AL) strategy [8] or a cross-domain recommendation technique [9]. Another alternative, which we considered in our work, is to exploit a set of metadata describing the users and items (e.g., demographics and item descriptions), and to utilise them in a hybrid CARS in order to overcome the new user and new item problem, respectively [14].

Switching hybrid, as described, e.g., in [7], was used in traditional RSs, i.e., non context-aware. From these approaches we borrowed the idea of selecting and applying the best performing recommender among a set of candidate models, based on the current recommendation situation, and hence avoiding problems specific to a particular recommendation algorithm.

Instead of acquiring new preference data or metadata, one can try to better process the existing preference data. In fact, an example of such approach is SPF [6], which we have evaluated in this paper. SPF tries to deal with the cold-start problem in CARSs by using in the recommendation process, not only contextual ratings that exactly match the target context, but also those that were provided in semantically equivalent contexts. But, as we have shown in this paper, SPF may fail to reliably assess the situation-to-situation similarity in cold-start scenarios.

5. DISCUSSION

We must observe that the off-line evaluation of cold-start performance of CARSs is a complex task. This is mainly due to the lack of large (enough) contextually-tagged rating datasets, which hinders a one-to-one application of the cold-start evaluation protocols proposed before for traditional RSs. For instance, the evaluation schema presented in [12] that splits user ratings into four partitions (i.e., recommendation on existing items for existing users, recommendation on existing items for new users, recommendation on new items for existing users, and recommendation on new items for new users) by randomly selecting half of the users (items) as new users (items) and the rest as existing users (items) can only be applied to very large and dense contextually tagged rating datasets. Otherwise the partitions (i.e., eight in the case of a CARS) will be filled with no or few ratings, or even worse, with "wrong" ratings as the corresponding training ratings might not be available.

Similar limitations apply to the testing methodology illustrated in [13] when trying to extend it and evaluate the system performance not only on new items (the case considered in their work) but also on new users and new contexts. Unfortunately, the standard large-scale datasets used for non context-aware RS research (e.g., MovieLens[1] and LibraryThing[2]) are of limited use here since they do not provide either contextual tags, or user/item attributes, or both.

[1]MovieLens: http://grouplens.org/datasets/movielens/
[2]LibraryThing: https://www.librarything.com

Table 2: Results in terms of MAE and $nDCG_1$ (boldface indicates significant improvements over the baseline MF model with $p < 0.05$; underline means same or better performance of SHCA over all CARS algorithms)

Cold-start case	Algorithm	STS		CoMoDa		Music	
		MAE	$nDCG_1$	MAE	$nDCG_1$	MAE	$nDCG_1$
Recommendation for new users	MF	1.035	0.650	1.044	0.633	1.349	0.351
	CAMF-CC	**0.999**	0.646	0.895	0.548	**1.263**	**0.442**
	SPF	**1.000**	**0.682**	**0.815**	0.630	**1.259**	**0.480**
	Content-based CAMF-CC	1.020	0.641	1.054	**0.692**	2.327	0.397
	Demogr.-based CAMF-CC	**0.997**	0.679	**0.803**	**0.711**	1.307	**0.421**
	SHCA	**0.997**	**0.681**	**0.837**	0.694	1.307	**0.421**
Recommendation for new items	MF	1.056	0.322	0.985	0.242	1.227	0.188
	CAMF-CC	**0.970**	**0.541**	**0.884**	**0.605**	**1.195**	**0.493**
	SPF	**0.984**	**0.597**	**0.837**	**0.288**	1.258	**0.545**
	Content-based CAMF-CC	**0.935**	**0.655**	**0.783**	**0.747**	**1.187**	**0.515**
	Demogr.-based CAMF-CC	**0.941**	**0.552**	**0.868**	**0.743**	1.434	**0.494**
	SHCA	**0.930**	**0.649**	**0.779**	**0.749**	**1.186**	**0.515**
Recommendation under new contexts	MF	0.907	0.711	0.893	0.653	0.817	0.735
	CAMF-CC	**0.868**	0.714	**0.820**	0.680	0.796	0.715
	SPF	**0.854**	**0.745**	**0.807**	0.653	1.009	0.679
	Content-based CAMF-CC	**0.863**	**0.746**	**0.745**	**0.773**	1.108	0.630
	Demogr.-based CAMF-CC	**0.848**	**0.768**	**0.810**	**0.759**	**0.759**	0.747
	SHCA	**0.848**	**0.766**	**0.771**	**0.771**	**0.759**	0.747

6. CONCLUSIONS AND FUTURE WORK

In this study we focused on cold-starting CARSs, i.e., dealing with a lack of ratings for users, items or contextual situations. We have compared (off-line) four CARS algorithms, two state of the art and two novel, which we conjectured to perform well in these scenarios, and a novel switching hybrid approach. We have shown that the two novel CARS algorithms and their hybrid combination (SHCA) successfully deal with all the considered cold-start problems. We have also shown that the new context scenario is the hardest and more research is required to effectively solve it.

Cold-start in CARSs is still a new and poorly explored topic, and there are several research questions that deserve future work. For instance, can the elementary (non-hybrid) CARS algorithms be adaptively combined based on ratings gathered from users [4]? Additionally, how will the proposed algorithms perform on larger contextually-tagged rating datasets? Our results offer an initial support to the conjecture that a switching hybrid approach can be a viable solution to the cold-start problem in CARSs.

7. REFERENCES

[1] G. Adomavicius, B. Mobasher, F. Ricci, and A. Tuzhilin. Context-aware recommender systems. *AI Magazine*, 32(3):67–80, 2011.

[2] L. Baltrunas, M. Kaminskas, B. Ludwig, O. Moling, F. Ricci, A. Aydin, K.-H. Lüke, and R. Schwaiger. Incarmusic: Context-aware music recommendations in a car. In *E-Commerce and Web Technologies*, pages 89–100. Springer, 2011.

[3] L. Baltrunas, B. Ludwig, and F. Ricci. Matrix factorization techniques for context aware recommendation. In *Proceedings of RecSys'11*, pages 301–304. ACM, 2011.

[4] M. Braunhofer. Hybridisation techniques for cold-starting context-aware recommender systems. In *Proceedings of RecSys'14, Doctoral Symposium*. ACM, 2014.

[5] R. Burke. Hybrid web recommender systems. In *The Adaptive Web*, pages 377–408. Springer Berlin / Heidelberg, 2007.

[6] V. Codina, F. Ricci, and L. Ceccaroni. Local context modeling with semantic pre-filtering. In *Proceedings of RecSys'13*, pages 363–366. ACM, 2013.

[7] M. Ekstrand and J. Riedl. When recommenders fail: predicting recommender failure for algorithm selection and combination. In *Proceedings of RecSys'12*, pages 233–236. ACM, 2012.

[8] M. Elahi, M. Braunhofer, F. Ricci, and M. Tkalcic. Personality-based active learning for collaborative filtering recommender systems. In *AI* IA 2013: Advances in Artificial Intelligence*, pages 360–371. Springer, 2013.

[9] M. Enrich, M. Braunhofer, and F. Ricci. Cold-start management with cross-domain collaborative filtering and tags. In *E-Commerce and Web Technologies*, pages 101–112. Springer, 2013.

[10] Y. Koren and R. Bell. Advances in collaborative filtering. In *Recommender Systems Handbook*, pages 145–186. Springer, 2011.

[11] A. Odić, M. Tkalčič, J. F. Tasič, and A. Košir. Predicting and detecting the relevant contextual information in a movie-recommender system. *Interacting with Computers*, 25(1):74–90, 2013.

[12] S.-T. Park and W. Chu. Pairwise preference regression for cold-start recommendation. In *Proceedings of RecSys'09*, pages 21–28. ACM, 2009.

[13] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 253–260. ACM, 2002.

[14] W. Woerndl, M. Brocco, and R. Eigner. Context-aware recommender systems in mobile scenarios. *International Journal of Information Technology and Web Engineering (IJITWE)*, 4(1):67–85, 2009.