

Local Context Modeling with Semantic Pre-filtering

Victor Codina

Universitat Politècnica de Catalunya-
BarcelonaTech

Girona 1-3, K2M 201
08034 Barcelona, Spain
vcodina@lsi.upc.edu

Francesco Ricci

Faculty of Computer Science
Free University of Bozen-Bolzano

Piazza Domenicani 3
39100 Italy
fricci@unibz.it

Luigi Ceccaroni

Barcelona Digital Technology Centre
Roc Boronat 117

08018 Barcelona, Spain
lceccaroni@bdigital.org

ABSTRACT

Context-Aware Recommender Systems locally adapt to a specific contextual situation the rating prediction computed by a traditional context-free recommender. In this paper we present a novel semantic pre-filtering approach that can be tuned to the optimal level of contextualization by aggregating contextual situations that are similar to the target one. The similarities of contextual situations are derived from the available contextually tagged users' ratings according to how similarly the contextual conditions influence the user's rating behavior. We present an extensive evaluation of the performance of our pre-filtering approach on several data sets, showing that it outperforms state-of-the-art context-aware Matrix Factorization approaches.

Categories and Subject Descriptors

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

Keywords

Context-Aware Recommender Systems; Implicit Semantics; Collaborative Filtering; Matrix Factorization

1. INTRODUCTION

Context-Aware Recommender Systems (CARSs) differ from traditional recommenders because they estimate the rating of a target user u for an item i , not only by using a data set of ratings (of users for items), but also exploiting the knowledge of the contextual situations under which the ratings were acquired and the contextual situation of the target user asking for a recommendation. In this paper we use the term *contextual factor* when referring to a specific *type* of contextual information (e.g. weather), and *contextual condition* when referring to a specific value of a contextual factor (e.g. sunny). The term *contextual situation* refers to a combination of contextual conditions that describe the context in which the user experienced the item.

The main intuition supporting CARS techniques is that, when context matters, the ratings acquired in the target contextual situation should be more relevant for predicting what to recommend in that situation. As we will describe in the next section, in the last years several approaches for exploiting

contextual knowledge have been proposed. We view CARS as locally adapted prediction models, i.e., adapted to a local contextual situation. Hence, as in other machine learning tasks, there is the need to judge to what extent the predictive model must fit the local situation. In other words, one must detect the optimal middle point between a global model based on all user preferences (i.e. personalization without context) and a local model just fitting the user preferences in a specific context. This means that one must find the optimal level of contextualization for a given rating data and contextual situation. For instance, a high degree of contextualization may not be appropriate when the contextual situation is too specific and not relevant, or when there are not enough ratings in that situation for generating reliable predictions. Actually, here we present an effective approach to accomplish this task: to find the right level of contextualization.

In a recent work [5] we proposed a flexible pre-filtering approach that exploits the semantic similarities between conditions during rating aggregation in order to find the optimal level of contextualization. Our approach follows a two-step process: (1) we first estimate the similarity of apparently unrelated contextual conditions, by focusing on their effect on user's rating behavior; (2) we exploit these condition-to-condition similarities to estimate the situation-to-situation similarities which are then used to expand the set of ratings for a target contextual situation aggregating the ratings obtained in the most similar situations.

In this work we improve the accuracy and computation efficiency of the proposed semantic pre-filtering approach by redefining its key component: the situation-to-situation similarity computation method. We show that our pre-filtering approach combined with standard Matrix Factorization (MF) is able to outperform, on several contextually tagged data sets, context-free MF (i.e. the global model), *exact* pre-filtering MF (i.e. the local MF models based only on ratings acquired exactly in the recommendation's target situation), and state-of-the-art MF prediction models based on the contextual modeling paradigm. Furthermore, we provide a detailed performance analysis on the considered data sets using different contextualization levels; ranging from context-free MF (global model) to *exact* pre-filtering (perfect local models).

The remainder of this paper is organized as follows. Section 2 positions our work with respect to the state of the art. Section 3 presents our novel semantic pre-filtering approach to CARS. Sections 4 and 5 present the evaluation and its results, and Section 6 draws the main conclusions and describes the future work.

2. RELATED WORK

CARS have been classified into three paradigms [1]: (1) pre-filtering, where context is used for selecting the relevant ratings before computing predictions with a traditional context-free model; (2) contextual post-filtering, where context is used to adjust predictions generated by a context-free model; and (3)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
RecSys '13, October 12 - 16 2013, Hong Kong, Hong Kong
Copyright 2013 ACM 978-1-4503-2409-0/13/09...\$15.00.
<http://dx.doi.org/10.1145/2507157.2507218>

contextual modeling, in which contextual information is directly incorporated in a context-free model as additional parameters.

Exact pre-filtering rigidly implements the basic intuition behind CARS: when context matters, generate prediction using only the data acquired in the target contextual situation. Its main limitation is the lack of flexibility: it always uses the maximum level of contextualization, a perfect local model. In the literature, several solutions to this limitation have been proposed. The most popular one is *generalized* pre-filtering [1] that exploits the available hierarchical relations between contextual conditions when searching for the optimal contextual segment (i.e. aggregations of semantically-related situations) for the target contextual situation. However, this solution is computationally expensive and its performance depends on the quality of the context ontology that is often limited to general hierarchical relationships between conditions of the same contextual factor.

Recent solutions have focused on integrating contextualization into a global MF model instead of building several local models. An example is *item splitting* [2], a pre-filtering variant that identifies in a selective way the relevant contextual conditions for the rating prediction of each item. Here ratings' filtering is selectively carried out item by item, for the most relevant contextual condition, and then a prediction model is built using all the data. In the category of contextual modeling, *tensor factorization* [7] extends standard MF to an n dimensional model according to the number of contextual factors. Its main limitation is the computational complexity, being applicable only when the number of contextual factors (i.e. dimensions) is small. The authors presented a more scalable approach in [4], which consists of extending standard MF model with contextual baseline predictors modeling the relationships between conditions and items. However, it is effective only when the target contextual situations are described by one or few conditions. Furthermore, all these approaches do not exploit the potential similarities between conditions, and, as we will show in Section 5, this knowledge may be useful to improve the performance of CARS.

3. SEMANTIC PRE-FILTERING

Similarly to *generalized* pre-filtering, we exploit semantic relations between contextual situations to find the optimal level of contextualization for building the local models. But instead of using ontology-based semantic relations, our approach looks for the implicit semantics of contextual conditions that can be estimated from rating data. We deem two contextual conditions *similar* if they influence the users' ratings in a similar way. To compute these similarities from rating data, we adapted Latent Semantic Analysis (LSA) [6] to the task of reducing a context-by-item or context-by-user matrix where entries are real values that measure the influence of conditions on user's rating behavior.

We first measure the influence of a condition on the users' preferences as the aggregated deviation between the actual rating value when the condition holds, and a predicted context-free rating. This aggregation can be based on different perspectives. Here we propose two variants: *user-based*, in which we use a per-user aggregation; and *item-based*, where a per-item aggregation is used. If additional data about the items or the users are available, then other aggregations may be employed, such as those based on item types (e.g. movie genres) or users' groups (e.g. age).

Let us denote with R_{ic} the set of ratings, acquired in contextual condition c , for item i , and with R_{uc} the set of ratings given by a user u in condition c . The importance weight of the condition c and item i (w_{ci}) or user u (w_{cu}) is calculated as the average of the

deviations between the ratings r_{uic} in R_{ic} or R_{uc} , respectively, and the context-free baseline predictions b_{ui} :

$$\text{Item-based:} \quad w_{ci} = \frac{1}{|R_{ic}| + \beta} \sum_{r_{uic} \in R_{ic}} (r_{uic} - b_{ui}) \quad (1)$$

$$\text{User-based:} \quad w_{cu} = \frac{1}{|R_{uc}| + \beta} \sum_{r_{uic} \in R_{uc}} (r_{uic} - b_{ui}) \quad (2)$$

where β is used to reduce the estimated weights when R_{ic} or R_{uc} are small. We set the value of β experimentally; in our experiments $\beta \in [0,10]$. In this work we used as baseline predictor the context-free model presented in [8]: $b_{ui} = \mu + b_u + b_i$, where μ is the overall rating average, b_u is the bias associated to the user u , and b_i the bias associated to the item i . We also tested more sophisticated predictors but no significant differences were observed, hence we decided to use a simple one.

Then, we reduce the dimensionality of the condition-item or condition-user matrix M by using SVD. As a result, we obtain a l -dimensional vector of latent features that represents the implicit semantics of a condition c :

$$y_c = [f_{c1}, \dots, f_{cl}] \quad (3)$$

The optimal dimensionality l depends on data sparsity and was also identified experimentally (see Section 4 for more details about how we have set the model hyper-parameters and Table 3 where we show the optimal l values for the considered data sets).

Assume now that there is a training set of rating data provided by users for items in various contextual situations S , and, a situation s is defined by a vector $s = [c_1, \dots, c_k]$, where, k is the number of contextual factors, $c_j = 0, 1, \dots, z_j$ are the contextual conditions for the j -th factor, and $c_j=0$ means that the j -th condition is unknown.

The semantic pre-filtering algorithm is a two-step process. Firstly, given a target contextual situation s^* , for each possible contextual situation s in S , we measure how similar s is to the target situation s^* , $Sim(s, s^*)$, by computing the cosine similarity between the vector representations of the two situations y_s and y_{s^*} . If s is defined by one condition c , then $y_s = y_c$, i.e., it is equal to the l -dimensional representation of the contextual condition into the reduced latent feature model representing it, as described in Eq. 3. While if s is described by multiple conditions, we define y_s as the average of the vectors y_c of its known conditions C_s :

$$y_s = \frac{1}{|C_s|} \cdot \sum_{c \in C_s} y_c \quad (4)$$

Secondly, all the ratings acquired in a contextual situation s where $Sim(s, s^*) \geq t$ (being t a similarity threshold) are then aggregated to the training set X_{s^*} , which is used for building the local rating prediction model for the target situation s^* . Denoting with R_s the set of ratings acquired in situation s , and with t a similarity threshold we have that:

$$X_{s^*} = \bigcup_{s: Sim(s, s^*) \geq t} R_s \quad (5)$$

The similarity threshold t is a global parameter of the data set: the larger the threshold, the stronger the contextualization level of the local contextual models is, i.e., the local model of each target contextual situation is built aggregating less contextual situations more similar to the target one. When $t=1$ the local models are those built by *exact* pre-filtering, and only the ratings acquired exactly in the target contextual situation are used. Conversely, when the threshold is set to -1 the local models are all equal and are built using all the rating data, i.e., they are not local anymore and coincide with the context-free model. The goal of our approach is to find the level of contextualization (i.e. a threshold between -1 and 1) that yields the best prediction accuracy. In our

experiments all the predictive models are based on MF; they only differ in the set of ratings used for training.

Note that if there is more than one rating for a given user and item acquired in some contextual situations similar to the target one, these ratings are aggregated (averaged) in order to generate a unique rating for user, item and target contextual situation.

4. EXPERIMENTAL SETUP

In order to evaluate our semantic pre-filtering technique we have considered several contextually tagged data sets with different characteristics (see Table 1). Music and Tourism data sets [3] contain ratings for items and users in contextual situations described by one condition only. They also contain ratings not tagged with a specific condition (context-free). Comoda [10] and Adom [1] data sets contain at most one rating per user and movie in a contextual situation that is described by the conjunction of several conditions (e.g. *summer*, *home* and *alone*). LibraryThing and MovieLens are data sets about books and movies, respectively, which contain ratings augmented with user tags (e.g. *funny*), extracted from their respective websites¹. Here the contextual situation is described by the tags assigned by the users to the rated items: they provide a contextual clue of why the item is important for the user. But given the inherent noise of user-generated tags, we only used those tags that have a statistically significant impact on the user’s rating behavior and have been used by a minimum number of users (5 users in MovieLens and 200 in LibraryThing). In particular, as significance test we used Pearson’s chi-squared that proved to be an effective method to identify relevant contextual information [10]. We selected the tags with significant dependency on the ratings (99% confidence).

As suggested by [11] we measured the prediction accuracy of the considered models by conducting a per-user evaluation by randomly selecting, for each user, n ratings as test set ($n=5$ in LibraryThing and $n=3$ in the others). The remaining ratings were used to acquire the implicit semantics of contextual conditions and train the MF models. We then measured the Mean Absolute Error (MAE) of the predictions on the test items. All the reported results are averages of per-user evaluations, and the statistical significance of the differences between evaluated models has been calculated using the paired Wilcoxon sign rank test.

We evaluated the performance of the two variants of the semantic pre-filtering described in Section 3: using the *item-based* implicit semantics perspective according to Eq. 1 (*Pref-IB*), and using the *user-based* one according to Eq. 2 (*Pref-UB*). To build the local prediction models, with the training ratings identified by each pre-filtering approach, we used a standard MF approach that generates the rating estimation for the user u and item i as the sum of the user and item biases and the dot product between their corresponding vectors of latent features (p_u) and (q_i):

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T p_u \quad (6)$$

For comparisons, we also evaluated two variants of the contextual modeling approach presented in [4] and also used in [10], which extends MF by adding baseline parameters that model the influence of the conditions on the items (CAMF-CI) or on the users (CAMF-CU). The parameters of the MF-based models are optimized using stochastic gradient descent. The numeric hyperparameters of the models were optimized on a single validation set by using the Nelder-Mead search algorithm [9].

¹See [www.librarything.com] and [www.movielens.org] (accessed August 5th, 2013)

Table 1. Data sets’ statistics

Data set	ratings	users	items	density	factors	conditions	scale
Music	4013	43	139	16%	8	26	1-5
Tourism	1679	25	20	48%	14	53	1-5
Adom	1464	84	192	9.1%	4	14	1-13
Comoda	2296	121	1197	1.4%	12	49	1-5
Movie	2190	428	1115	0.4%	29	29	1-10
Library	609K	7192	37K	0.2%	149	149	1-10

5. RESULTS AND DISCUSSION

Here we evaluate the performance of the two variants of the semantic pre-filtering approach (*Pref-IB* and *Pref-UB*) against the context-free model (*MF*), *exact* pre-filtering (*Exact*) and two contextual modeling variants (*CAMF-CI* and *CAMF-CU*). Table 2 shows that the proposed prediction models achieve the best accuracy in most of the data sets. We can observe that in the data sets where contextual situations are described by one or few conditions (viz. Tourism, Music and LibraryThing) the semantic pre-filtering performs similarly to the contextual modeling approach, especially in terms of MAE. However, in data sets where situations are defined by several conditions (viz. Comoda and Adom) our approach clearly outperforms it. Similar results were obtained for Normalized Discounted Cumulative Gain (NDCG) (not shown here for space reasons).

As it can be observed in Table 3 (last two columns) the proposed semantic pre-filtering technique always outperforms *MF* and *Exact*. We can also observe that when the optimal level of contextualization is used (using the optimal threshold shown in the second data column), the average number of ratings increases by several factors with respect to the local models used by *Exact* (sub-column “vs. *Exact*” in “Expansion”), and considerably differs among data sets. In general, the proposed method produces the best results, compared with both *Exact* and *MF*, when the optimal level of contextualization causes a significant increase of the number of ratings that are used to build the local models. This occurs for Music, Tourism, Comoda and LibraryThing data sets where the local models built by our approach are on average 7.9, 7.2, 20 and 677 times larger than those built by *Exact*.

We can also observe that the value of the similarity threshold t is not important, what matters most is the amount of ratings that must be aggregated. For instance, a threshold of 0.9 is much more restrictive in Adom compared to Comoda. This is related to the average similarity of contextual situations in a data set. In fact, in Adom the average similarity between pairs of situations is small, 0.38, while in Comoda is much larger, 0.86; hence in Adom there are much less pairs of situations that are more similar than 0.9, which is the optimal threshold for both data sets.

Table 2. MAE of evaluated models: results in bold are significantly better (95% confidence) than context-free *MF* and results underlined are also significantly better than *Exact*

Models	Tourism	Music	Adom	Comoda	Movie	Library
<i>MF</i>	1.229	1.001	2.249	.760	1.274	1.261
<i>Exact</i>	1.116	1.210	2.213	.828	1.131	1.193
<i>CAMF-CI</i>	1.002	.942	2.291	.749	1.249	1.257
<i>CAMF-CU</i>	1.031	.976	2.230	.767	1.186	1.140
<i>Pref-IB</i>	1.047	.934	2.193	<u>.675</u>	1.100	<u>1.142</u>
<i>Pref-UB</i>	1.009	.966	2.185	<u>.651</u>	1.116	<u>1.136</u>

Table 3. MAE performance in optimal contextualization level
($l = y_c$ dimensionality (Eq. 3); t = similarity threshold)

Data set	l	t	Model size		Expansion		Improvement	
			#ratings	vs. <i>global</i>	#situations	vs. <i>Exact</i>	vs. <i>Exact</i>	vs. <i>MF</i>
Music	10	0	876	90%	21	7.9	23%	7%
Tourism	10	.24	180	64%	17	7.2	10%	18%
Adom	10	.90	26	1.8%	1.3	1.4	1.5%	3%
Comoda	49	.90	896	39%	730	20	21%	14%
Movie	29	.06	192	8.7%	13	3.3	3%	14%
Library	130	.40	38534	6.3%	2275	677	5%	10%

We note that the effect of the situation-to-situation similarity threshold on the prediction accuracy is similar to the effect of the user-to-user similarity in user-based Collaborative Filtering (CF): using a lower similarity threshold yields to a larger neighbor set and it is well known that it is important to find the right level of locality (neighbor size) in user-based CF.

Figure 1 illustrates the MAE and the level of contextualization in terms of ratings increase (bottom) as functions of the similarity threshold in the Comoda data set. The minimum MAE is obtained when the threshold is 0.9, which causes an aggregation of (on average) 730 situations and produces an increase of 20 times the average model size used by *Exact*.

Compared to the *all-pairs* situation-to-situation similarity strategy presented in [5], the one proposed here always performs better and has a reduced time complexity: its cost is at most $O(n)$ rather than $O(n^2)$, n being the number of contextual factors.

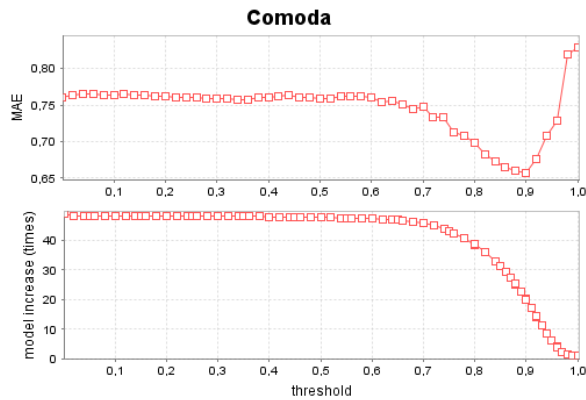


Figure 1. MAE and level of contextualization as functions of the similarity threshold

6. CONCLUSIONS AND FUTURE WORK

We have presented a semantic pre-filtering approach to CARS that can be tuned to find the right level of contextualization. This approach employs a global situation-to-situation similarity threshold that can be used to identify the ratings acquired in situations similar to the target one and that can be effectively used in the given target contextual situation to train the prediction model. The key component of our approach is a novel method for accurately measuring the similarity between contextual situations that is based on how similarly two different situations influence the users' ratings. Therefore, our method does not depend on external knowledge sources as in *generalized* pre-filtering.

The prediction accuracy of the proposed approach has been experimentally compared to state-of-the-art techniques on several contextually tagged data sets. We have shown that the proposed pre-filtering method outperforms both context-free MF and other CARS techniques based on MF as well.

To further improve the proposed semantic pre-filtering technique we will investigate a variant able to fine tune the level of contextualization per contextual situation. We believe that in some data sets where the specificity and relevance of contextual situations considerably differ, this fined-grained tuning may produce even better prediction accuracy.

7. ACKNOWLEDGEMENTS

The research described in this paper is partly supported by the SuperHub and the Citclops European projects (FP7-ICT-2011-7, FP7-ENV-308469), and the Universitat Politècnica de Catalunya-BarcelonaTech (UPC) under an FPI-UPC grant. The opinions expressed in this paper are those of the authors and are not necessarily those of SuperHub or Citclops projects' partners.

8. REFERENCES

- [1] Adomavicius, G. and Tuzhilin, A. 2011. Context-aware recommender systems. In *Recommender Systems Handbook*. Springer, 217–256.
- [2] Baltrunas, L. and Ricci, F. 2009. Context-based splitting of item ratings in collaborative filtering. In *Proceedings of RecSys '09*, 245–248.
- [3] Baltrunas, L., Ludwig, B., Peer, S. and Ricci, F. 2012. Context relevance assessment and exploitation in mobile recommender systems. *Pers. Ubiquit. Comput.* 16(5):507–526.
- [4] Baltrunas, L., Ludwig, B. and Ricci, F. 2011. Matrix factorization techniques for context aware recommendation. In *Proceedings of RecSys '11*, 301–304.
- [5] Codina, V., Ricci, F. and Ceccaroni, L. 2013. Exploiting the Semantic Similarity of Contextual Situations for Pre-Filtering Recommendation. In *Proceedings of the 21st Conference on User Modeling, Adaptation and Personalization*, 165–177.
- [6] Dumais, S. 2006. LSA and information retrieval: Getting back to basics. In Landauer, T.-K., McNamara, D.-S., Dennis, S., Kintsch, W. (Eds.) *LSA: A Road to Meaning*, Lawrence Erlbaum, 293–321.
- [7] Karatzoglou, A., Amatriain, X., Baltrunas, L. and Olivier, N. 2010. Multiverse Recommendation: N-dimensional Tensor Factorization for Context-aware Collaborative Filtering. In *Proceedings of RecSys '10*, 79–86.
- [8] Koren, Y. and Bell, R. 2011. Advances in Collaborative Filtering. In *Recommender Systems Handbook*, 145–186.
- [9] Nelder, A. and Mead, R. 1965. A simplex method for function minimization. In *The Computer Journal*, 7(4).
- [10] Odić, A., Tkalčič, M., Tasič, J. and Košir, A. 2013. Predicting and Detecting the Relevant Contextual Information in a Movie Recommender System. *Interact. Comput.* 25(1):74–90.
- [11] Shani, G. and Gunawardana, A. 2011. Evaluating Recommendation Systems. In *Recommender Systems Handbook*, Springer, 257–297.