Semantically-Enhanced Pre-Filtering for Context-Aware Recommender Systems

Victor Codina
Universitat Politècnica de Catalunya - BarcelonaTech (UPC)
Jordi Girona 1-3, Ed. 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
Several research works have demonstrated that if users' ratings are truly context-dependent, then Context-Aware Recommender Systems can outperform traditional recommenders. In this paper we present a novel contextual pre-filtering approach that exploits the implicit semantic similarity of contextual situations. For determining such a similarity we rely only on the available users' ratings and we deem as similar two syntactically different contextual situations that are actually influencing in a similar way the user's rating behavior. We validate the proposed approach using two contextually tagged ratings data sets showing that it outperforms a traditional pre-filtering approach and a state-of-the-art context-aware Matrix Factorization model.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – information filtering.

General Terms
Algorithms, Experimentation, Performance

Keywords
Recommender Systems, Contextual Pre-filtering, Semantic Similarity, Collaborative Filtering, Matrix Factorization

1. INTRODUCTION
Context-Aware Recommender Systems (CARSs) differ from traditional recommenders because when they estimate the rating of a target user u for an item i they do not only use a data set of ratings (of users for items), but they also exploit both the knowledge of the contextual conditions under which the ratings were acquired and the contextual situation of the target user asking for a recommendation. In Adomavicius and Tuzhilin [2], three different paradigms for incorporating contextual information into the recommendation process are presented: (1) contextual pre-filtering, where context is used for selecting the relevant set of rating data before computing predictions; (2) contextual post-filtering, where context is used to adjust predictions generated by a traditional model; and (3) contextual modeling, in which contextual information is directly incorporated in the prediction model, usually by extending a traditional rating prediction model.

Terminology-wise, 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 for a contextual factor (e.g. sunny). The term *contextual situation* refers to a specific set of these contextual conditions that describe the context in which the user consumed the item.

Among the three paradigms, pre-filtering is especially appealing because it has a straightforward justification: when context matters, use in the recommendation process only the data acquired in the same contextual situation of the target user, because only this data is relevant for predicting the user preferences. However, pre-filtering is not always the best option: in an experimental comparison of traditional pre-filtering versus post-filtering approaches, presented by Panniello et al. [12], their relative merits and issues are indicated. In fact, the main limitation of contextual pre-filtering comes from the difficulty to obtain ratings in all the possible contextual situations in order to build a robust and contextualized rating prediction model.

As proposed by Adomavicius et al. [1], one solution to that sparsity problem is to use contextual segments, which are usually supersets of the target context (e.g. the *weekend* segment may be used when the target contextual situation is *Saturday* or *Sunday*). The limitation of this solution is that useful contextual segments need to be identified in the space of all possible segments and the computational cost of such an operation depends on the size of the search space. Moreover, this solution may result in a too coarse aggregation of contextual situations if the chosen contextual segments are too broad.

Another solution has been proposed by Baltrunas and Ricci [6], and is based on the idea of "item splitting" which consists of identifying in a selective way the relevant contextual conditions for the rating prediction of each single item. Hence, here ratings' filtering is selectively carried out item by item.

However, both previously mentioned solutions do not exploit the potential similarities of contextual situations. For instance, a mild temperature in a cloudy day may have, on the user evaluation of a place of interest, the same effect of a cooler but sunny day. In this paper we propose a pre-filtering approach that is based on the intuition that ratings acquired in contextual conditions *similar* to the target one may still be used to improve the rating prediction accuracy. Similarly to a traditional Collaborative Filtering (CF) approach, which predicts the target user's ratings with a local regression on similar users' ratings, we propose here to predict a context-dependent rating by regressing predictions computed in contextual situations similar to the target one.

In our approach the similarity between contextual conditions is estimated by identifying the "meaning" of a condition defined by its implicit semantics: that is, the meaning of a concept (here, a contextual condition) is captured by the usage of the concept. Hence, contextual conditions are similar if they co-occur and produce a similar effect on the user's rating behavior. For example, in the recommendation of places of interest, bad weather conditions such as *cold* and *rainy* will have a positive effect on users' ratings for indoor places like museums, and a negative effect for outdoor places like castles. In this case, given that *cold* and *rainy* are similarly influencing user's ratings on the same type of items, we consider them as semantically similar.

In a recent work [7], we demonstrated that this type of implicit semantics can be useful to improve the prediction accuracy of content-based models. In that work we acquired the implicit semantics of item attributes considering two attributes as similar if they co-occur often in the content-based user profiles (previously learnt from user ratings) and with a similar degree of interest.

These similarities are then exploited to enhance a traditional content-based prediction method relying on item-user matching. Similarly, in this paper we show that the implicit semantics of contextual conditions can be exploited to implement an effective contextual pre-filtering approach.

The remainder of this paper is organized as follows. Section 2 presents the method for acquiring the implicit semantics of contextual conditions based on rating data. Section 3 describes our novel pre-filtering approach that exploits the acquired implicit semantics. Section 4 presents the experimental evaluation of the proposed approach. Finally, section 5 draws the main conclusions and foresees some future work.

2. IMPLICIT SEMANTICS ACQUISITION

In order to calculate the semantic similarities between contextual conditions we use the Vector Space Model (VSM). Our approach is similar to that used in *information retrieval* to calculate semantic similarities between two terms from a given corpus of documents [11]. It relies on the distributional hypothesis, i.e., terms co-occurring many times in the same linguistic context (e.g. document, paragraph, sentence) tend to have similar meaning. Calculating semantic similarities between terms is divided in three steps: (1) term-document matrix generation, where each entry stores the frequency, or a frequency-based weight (e.g. entropy), of a term in a document; (2) term-document matrix transformation, which usually consists of reducing the matrix dimensionality, and (3) vector matching calculation, where term (semantic) vectors are compared using an appropriate vector-based similarity measure. A well-known example of this approach is Latent Semantic Analysis (LSA) [8], where Singular Value Decomposition (SVD) is applied to the term-document matrix for decomposing it, the term weights are computed with the entropy weighting function, and the cosine similarity is used to calculate vector correspondence.

We adapt these techniques to learn the semantic similarities of contextual conditions from contextually tagged rating data by estimating two *contextual conditions as similar if they co-occur and influence the users' ratings in a similar way*. To accomplish this task, we generate a "term-document" matrix where terms correspond to contextual conditions and documents are the domain items. Each element of the matrix stores a weight representing the general influence, which can be positive, negative or neutral, of the contextual condition on the item rating. The influence weight for a specific contextual condition *c* and item *i* (w_{ci}) is calculated as the average of normalized set of ratings associated to the item in the contextual condition, denoted by R_{ic}:

$$w_{ci} = \sum_{r_{uic} \in R_{ic}} (r_{uic} - b_{ui}) \frac{1}{|R_{ic}|} \qquad (1)$$

where r_{uic} is the rating of user u for item i in the contextual condition *c*; and $b_{ui} = \mu + b_u + b_i$ is the baseline predictor that is used to normalize all the item ratings by subtracting the rating prediction that does not take into account any contextual condition [10]. (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*.) Once we have generated the condition-item matrix, we reduce its dimensionality using SVD. That is, if M is *n*-by-*m*, it is decomposed into VU, where V is *n*-by-*f* and U is *f*-by-*m*. The *n* rows of V model the contextual conditions in a *f*-dimensional abstract factor space (*f* is the selected number of factors.) and the U columns model the items in the same *f*-dimensional factor space. In our experimentation, we have obtained the best results by setting $f = n$, when $n <= m$, and not applying SVD in the other case. We believe that this is related to the fact that in general the number of contextual conditions *(n)* is already small, and by using less than *n* factors causes a too large information loss when computing the similarities between conditions (M rows). Finally, we use the cosine measure to calculate the similarity between the semantic vectors of two contextual conditions.

3. SEMANTICS EXPLOITATION

The previously described semantic similarities between contextual conditions can be exploited to implement a new pre-filtering approach. Given a data set of ratings provided by users for items in contextual situations, and given a target contextual situation, we first extract from the data set the rating data acquired in that situation. Then, for each *user* and *item* pair for which we miss the corresponding rating in the target contextual situation we determine if in the data set at least a rating in a similar situation is available. If this is the case, we also extract this rating, and finally we provide all the extracted ratings to a standard Matrix Factorization (MF) model to generate predictions for the still missing ratings in that target situation. A convenient threshold defines the minimum similarity value for a contextual situation to be considered as usable in a target context. We have developed two variants of the method: (1) *Semantic-Prefiltering-gt*, which uses a global threshold for all targets' contextual situations; and (2) *Semantic-Prefiltering-dt,* which employs a different threshold for each target's contextual situation. We note that using the proposed method, when we identify ratings acquired in similar contextual situations, it is possible to identify, for a given user and item, more than one rating given by that user to that item in contextual conditions similar to the target one. In this case, we simply average these ratings, in order to generate a unique rating for a given *user, item* and *target contextual situation* triple.

4. EVALUATION AND RESULTS

We have compared the performance of the proposed semantically enhanced, contextual pre-filtering approach to two baseline techniques: (1) a traditional pre-filtering approach that generates rating predictions for a target contextual situation by considering only the ratings acquired in that situation; (2) *CAMF-CC,* a context-aware MF approach proposed in [5]. We have considered two real-world data sets about music and tourism. Both data sets were collected using a web-based interface for simulating contextual situations. Table 1 shows the main characteristics of the data sets.

Table 1. Data sets' characteristics

Data set	ratings	users	items	factors	conditions
Music	4013	43	139	8	26
Tourism	1679	25	20	14	53

The music data set contains eight contextual factors such as *driving style*, *mood* and *landscape*; the tourism one contains 14 factors such as *weather*, *companion* and *travel goal*. Each factor has multiple contextual conditions. (E.g., *active*, *passive*, *happy* and *sad* are possible conditions of the *mood* factor.) The ratings were acquired by requesting the user to imagine a target contextual condition, e.g., "Imagine that today is sunny; how would you rate this song while traveling by car?"; hence these data sets contain multiple ratings for the same item and user, but in different contextual conditions. Moreover, the data sets also provide the ratings acquired without specifying the target's context conditions, e.g., "How would you rate this song while traveling by car?"; in which case the rating measures to what extent a user likes an item without considering any special contextual condition that may hold during the evaluation of the item. More details about the rating acquisition of these data sets can be found in [3] and [4]. The baseline 2-dimensional rating prediction model that we have used in our experiments is the bias-based MF. Here the rating estimation for user *u* and item *i* is given by:

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T p_u \quad (2)$$

where q_i is the factor vector of item i, and p_u is the factor vector of user u. To learn the model parameters we minimized the *regularized error* using stochastic gradient descent, which has been proved to be an effective approach [10]. In our experiment, we used the typical values for the hyper-parameters[1]. With this hyper-parameter configuration the model converges within 200 iterations in both data sets. To measure the rating prediction accuracy of the considered models we have employed a per-user evaluation schema where, for each user, five ratings were randomly chosen as test set to compute the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE). We minimized the *Absolute loss* function when using the MAE and the *Squared loss* function when using the RMSE. In order to confidently evaluate the considered CARSs, we only selected as testing ratings those provided in contextual conditions in which the user rated additional items (at least two, as we found experimentally). In addition, initially, we required that all test ratings of a user were about items not present in the training set. As a result of this data partition, only some users could be finally considered for testing (18 in the music data set and eight in the tourism data set). The remaining users and ratings were used to train the prediction models. We also tried other data partitions, but we obtained quite similar results for all of them.

Table 2 shows the rating prediction performance of the considered models in terms of RMSE and MAE in both data sets:

- *User-Item-AVG* refers to the non-personalized baseline prediction model based only on the user and item biases. The estimation function is defined by: $\hat{r}_{ui} = \mu + b_u + b_i$.
- *MF* refers to the 2-dimensional bias-based MF model without incorporating context (see definition in Equation 2). Multiple ratings for the same item and user, in alternative contextual conditions, are averaged.
- *CAMF-CC* is the context-aware MF model proposed by Baltrunas et al. [5]. This MF model learns one additional baseline parameter for each contextual condition and items' category pair. The tourism and music data sets contain 5 and 10 item categories respectively. In the music domain, examples of categories are: *Pop, Classical, Rock* and *Disco*.
- *Traditional-Prefiltering* is a contextual approach, which, similarly to the other approaches compared here, is used together with *MF* and consists of selecting only the ratings whose context exactly matches the target context. Then, for each target context a *MF* prediction model is trained using the selected ratings. Given that in our data sets there are ratings without any particular contextual condition, we use them only when no rating is available in the target context.
- *Semantic-Prefiltering-gt* is the proposed approach using a global, optimal similarity threshold, which is learned from the training set by cross-validation. In the music data set it is 0.36 and in the tourism data set 0.28. Table 3 shows, as an example, the top-5 conditions more similar to *happy* and *cold,* two conditions used in the music and tourism data set respectively. Their similarity values are calculated using the method presented in Section 2. After having selected the rating data, acquired in the target and similar contexts, we have built a *MF* model for each target context. Ratings without context were used when no rating in the target or similar contexts was available.
- *Semantic-Prefiltering-dt* is the proposed semantically enhanced pre-filtering approach using a specific similarity threshold for each target context (see Section 3 for more details). The optimal threshold was selected for each contextual situation with cross-validation on the training set. The threshold values range between (0, 0.5) in our experimentation. The same predictive model used in *Semantic-Prefiltering-gt* was then applied.

Table 2.Results of considered models in the music and tourism data sets. (Values below the non-contextual models are in bold and the best of each metric and data set is underlined.)

Prediction Model	Music		Tourism	
	RMSE	MAE	RMSE	MAE
User-Item-AVG	1.186	.978	1.193	1.030
MF	.825	.573	1.124	.959
CAMF-CC	.858	.599	**.966**	**.775**
Traditional-Prefiltering	**.769**	**.503**	**.986**	**.729**
Semantic-Prefiltering-gt	**<u>.755</u>**	**<u>.480</u>**	**<u>.944</u>**	**<u>.673</u>**
Semantic-Prefiltering-dt	**.791**	**.483**	**.988**	**.715**

Table 3. Top-5 similar contextual-conditions to the condition *happy* (music data set) and *cold* (tourism data set). Their semantic similarities are enclosed in parentheses.

Happy	Cold
urban (0.44)	cloudy (0.84)
awake (0.42)	working-day (0.81)
relaxed driving (0.30)	rainy (0.79)
coast line (0.29)	winter (0.75)
serpentine (0.17)	night-time (0.71)

[1] Hyper-parameters: *learning-rate* = 0.1; *num-factors* = 100; *factor-reg.* = 0.005; *item-bias-reg.* = 0.01; *user-bias-reg.* = 0.03.

In the music data set, *MF* is clearly better than *User-Item-AVG*, with an improvement of 41% (p<0.001) for MAE and of 30% (p<0.001) for RMSE. The performance of *MF* is further improved by all the contextual pre-filtering approaches. *Traditional-Prefiltering* achieves an improvement of 12% for MAE (p=0.28), and of 6.7% for RMSE (p=0.86), but this improvement is not statistically significant.

Conversely, *Semantic-Prefiltering-gt* is the best model achieving a significant improvement of 16% for MAE (p=0.03), but not for RMSE (8.2% with p=0.6). *Semantic-Prefiltering-dt* performs worse than *Semantic-Prefiltering-gt* and, in terms of RMSE, even than *Traditional-Prefiltering.*

The performances of the tested context-aware approaches in the tourism data set are quite similar to those illustrated above for the music data set. But the improvements achieved in this case are considerably more significant. *MF* slightly improves the performance of *User-Item-AVG* (6.9% in MAE with p=0.05 and 5.8% in RMSE with p=0.46).

All the contextual pre-filtering approaches improve the performance of *MF* and even in this case the semantically-enhanced variants are clearly the best approaches in terms of MAE. *Traditional-Prefiltering* achieves an improvement of 24% for MAE (p=0.02) and of 12% for RMSE (p=0.05) with respect to *MF*. *Semantic-Prefiltering-gt* is again the best model with an improvement of 16% for RMSE (p=0.03) and 30% for MAE (p=0.01). Again, results of *Semantic-Prefiltering-dt* are worse than *Semantic-Prefiltering-gt*'s ones.

Comparing the semantically-enhanced pre-filtering variants with *CAMF-CC* we can see that both of them are better (above all in terms of MAE): *CAMF-CC* achieves an improvement of 19% for MAE and of 14% for RMSE with respect to *MF* in the tourism data set, while in the music data set *CAMF-CC* is worse than *MF*. These results are not the same as reported in [5] for the same data sets; we believe that these differences are due to the different composition of the training and testing data.

5. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a novel recommendation approach that overcomes the major limitation of contextual pre-filtering, i.e.: using only user data acquired in the target context when the system is requested to make a recommendation in that target situation.

Our solution relies on the identification and usage of ratings acquired in contextual situations that are similar to the target one. We have introduced a novel notion of similarity among contextual conditions that is based on how similarly two different contextual conditions are influencing the users' rating behavior.

The experimental results on two contextually tagged ratings data sets show that the proposed pre-filtering approach, concretely the variant *Semantic-Prefiltering-gt* that uses a global similarity threshold, outperforms a traditional pre-filtering one and a state-of-the-art model-based approach.

In a future work we plan to further test the proposed approach on larger data sets. Furthermore, we will investigate alternative methods for acquiring the semantic similarities among contextual conditions; for instance, by generating a condition-user matrix instead of the condition-item matrix we presented here.

In addition, we will carry out a more extensive comparison of the proposed approach and other state-of-the-art model-based approaches, like the one proposed by Karatzoglou et al.[9], based on tensor factorization. Similarly as in our approach, this work also employs a latent factor representation of contextual information, but differs in how this representation is obtained and exploited during model learning.

6. 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 (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.

7. REFERENCES

[1] G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin. Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems*, 23(1):103-145, 2005.

[2] G. Adomavicius and A. Tuzhilin. Context-aware recommender systems. In F. Ricci, L. Rokach, B. Shapira and P. B. Kantor (eds.) *Recommender Systems Handbook*, pages 217-250. Springer, 2011.

[3] 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 *EC-Web'11*, 89-100, 2011.

[4] L. Baltrunas, B. Ludwig, S. Peer and F. Ricci. Context relevance assessment and exploitation in mobile recommender systems. *Personal and Ubiquitous Computing*, 16(5):507-526, 2012.

[5] L. Baltrunas, B. Ludwig and F. Ricci. Matrix factorization techniques for context aware recommendation. In *RecSys'11*, pages 301-304, 2011.

[6] L. Baltrunas and F. Ricci. Context-dependent items generation in collaborative filtering. In *RecSys'09*, pages 245-249, 2009.

[7] V. Codina and L. Ceccaroni. Semantically-Enhanced Recommenders. In *CCIA'12,* 248: 69-78, IOS Press, 2012.

[8] S. Dumais. LSA and information retrieval: Getting back to basics. In *LSA: A Road to Meaning*, 293-321, 2006.

[9] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Olivier. Multiverse Recommendation: N-dimensional Tensor Factorization for Context-aware Collaborative Filtering. In *RecSys'10*, pages 79-86, 2010.

[10] Y. Koren and R. Bell. Advances in Collaborative Filtering. In F. Ricci, L. Rokach, B. Shapira and P. B. Kantor (eds.) *Recommender Systems Handbook*, pages 145-186 Springer, 2011.

[11] C. D. Manning, P. Raghavan and H. Schütze. Introduction to Information Retrieval. Cambridge University Press, 2008.

[12] U. Panniello, A. Tuzhilin, M. Gorgoglione, C. Palmisano, A. Pedone. Experimental comparison of pre- vs. post-filtering approaches in context-aware recommender systems. In *RecSys'09*, pages 265-268, 2009.