Recommending Items with Conditions Enhancing User Experiences Based on Sentiment Analysis of Reviews

Konstantin Bauman
Stern School of Business
New York University
kbauman@stern.nyu.edu

Bing Liu
University of Illinois
at Chicago (UIC)
lub@cs.uic.edu

Alexander Tuzhilin
Stern School of Business
New York University
atuzhili@stern.nyu.edu

ABSTRACT

In this paper we propose a new method of recommending not only items of interest to the user but also the conditions enhancing user experiences with those items, such as recommending to go to a restaurant for seafood. This method is based on the sentiment analysis of user reviews, predicts sentiments that the user might express about the aspects determined in an application, and identifies the most valuable aspects of user's potential experience with the item. Furthermore, our method recommends the items together with those most important aspects over which the user has control and can potentially select them, such as the time to go to a restaurant, e.g. lunch vs. dinner, or what to have there, such as seafood. We tested our method on three applications (restaurants, hotels and beauty&spas) and experimentally showed that those users who followed our recommendations of items with their corresponding conditions had better experiences, as defined by the overall rating, than others.

CCS Concepts

•Information systems → Recommender systems; Sentiment analysis; •Computing methodologies → *Non-negative matrix factorization;*

Keywords

Recommender systems, user reviews, sentiment analysis, user experience, conditions of user experience

1. INTRODUCTION

Over the last five years, there has been much interest in leveraging user reviews for providing personalized recommendations based on these reviews [2]. Most of this work focuses on trying to improve estimations of user ratings based on the review and other relevant information [2] and also trying to explain why particular recommendations are provided to the user based on the review [15].

CBRecSys 2016, September 16, 2016, Boston, MA, USA.

These approaches aimed at predicting and explaining ratings in terms of the user and item characteristics without taking into the consideration additional factors, such as circumstances of consuming the item and personal choices of the user. For example, consider the user choosing between doing a manicure or a haircut in a salon. Depending on what the user would choose to do during her visit, she can give different ratings to the salon. Therefore, user experience of a particular item can be improved by recommending some appropriate conditions of consuming that item, such as doing a pedicure in that specific salon.

In this paper, we address this issue by recommending not only particular items, but also the *conditions* (i.e. the most important aspects) of potential experiences that the user can control, such as having haircut or pedicure in a salon. Furthermore, we can recommend certain actions to the management of an establishment (item) that would personalize experiences of the user when consuming the item (e.g., visiting the establishment). For example, we may recommend to the management of the Gotham Bar & Grill restaurant in New York to suggest the duck dish to a user because our method estimated that the user would particularly like that dish in that restaurant.

In this paper, we make the following contributions. First, we propose a novel approach of enhancing functionality of recommender systems by recommending not only the item itself but the item with the corresponding conditions enhancing user experience of the item. Further, we propose that the management of an establishment reviews the crucial aspects of the user's experience of the establishment that were identified by our method and uses them to personalize user experiences, thus improving the overall user ratings. Second, we developed a method for identifying the most valuable aspects of future experiences of the users that is based on the sentiment analysis of user reviews. Third, we tested our method on actual reviews and showed that users who followed our recommendations rate their experiences significantly higher than others.

2. LITERATURE REVIEW

Over the last few years, several papers tried to improve estimation of unknown ratings by using user reviews [5, 12, 16, 17]. For example, authors of [5] found six aspects in restaurant reviews and trained classifiers to identify them in text. Further they use this information to improve rating prediction quality. As another example, [11] uses the LDA-based approach combined with Matrix Factorization to better predict the unknown ratings. Furthermore, [3, 8] use more extensive graphical models than [11] to predict unknown ratings that are based on collaborative filtering and topic modeling of user reviews. As a result they capture interpretable aspects and the sentiments of each aspect of a review. Moreover, [19] proposes another method to improve rating prediction based on learning users' interests and items' characteristics. In [18] authors proposed a tensor factorization approach for the rating inference. Operating on the tensor composed of the overall and the aspect ratings, this approach is able to capture the intrinsic relationships between users, items, and aspects of user reviews, and provide reasonably good predictions for unknown ratings. Finally, [10] determines aspect weight preferences of the user and estimates user satisfaction with the item based on these weights.

In contrast to these papers, in our work, we focus not only on rating predictions but also on *determining the most important aspects* having the highest impact on ratings measuring user's potential future experience of an item. Moreover, we provide recommendations not only of items of interest to the user but also the *conditions enhancing user experiences* with those items.

Furthermore, [1] constructed an aspect ontology for the Digital Camera application, developed a set of rules for identifying those aspects in text along with their sentiments. Based on the collected data, [1] aggregated profiles of these cameras and presented simple recommendations using knowledge based recommendation techniques. Also, [14] proposed an approach of extracting aspect-specific ratings from the reviews and then recommending new reviews to the users based on these ratings. In contrast to [1, 14], we focus on recommending the most important aspects over which the user has control and that she can potentially select.

Finally, there are a few publications on explanations of recommendations based on user reviews. In particular, [20] proposes the Explicit Factor Model (EFM) to generate explainable recommendations while keeping high prediction accuracy. In addition, [7] provides reasons why an item has been recommended vs. its various alternatives based on the aspects extracted from the reviews. In contrast to this work, our goal is not only to provide explanations but also to recommend the most important aspects over which the user has control and that she can potentially select.

3. OVERVIEW OF THE METHOD

In this section we present a method of identifying the most valuable aspects of future user experiences of the items and recommend the items together with the most important aspects over which the user has control and that she can potentially select. This method consists of sentiment analysis of user reviews, identification of relations between different aspects of the reviews, predicting sentiments that the user might express about these aspects, and calculating personal impact factors that each aspect contributes to the overall rating for the user. Moreover, we identify and recommend to the management of an establishment to consider potentially important aspects of the user experience with the establishment and use them to personalize user experiences there, thus improving the overall user rating of the establishment. The proposed method consists of 5 steps described below.

(1) Extracting aspects from the reviews.

In this step we utilized the state-of-the-art "industrial-strength" sentiment analysis system *Opinion Parser* [9] that uses Double Propagation method [13] for extracting aspects from texts. The sentiment classification algorithm is the lexicon-based method [9], which has its roots in [4, 6] but with significant improvements.

We applied *Opinion Parser* to the set of reviews R for a given application (e.g. restaurants) in order to build a set of aspects $\mathbb{A}$ occurring in R. Furthermore, for each review $r \in R$, *Opinion Parser* identifies a set of aspects A_r occurring in r and corresponding sentiments s^t_{ui} opinions of user u about aspects $t \in A_r$ of experience with item i.

(2) Training the sentiment predicting model.

In a typical review, set A_r contains only a small part of all the aspects $\mathbb{A}$, and thus for each particular aspect the matrix of known sentiments for (user, item) pairs is more sparse than matrix of ratings. Therefore, the problem of predicting sentiments of the aspects becomes even more challenging than prediction of the overall rating of review r.

Since various aspects of a review can be correlated, such as "desert" and "fruits", we propose an approach of using this correlation information to train the sentiment prediction model. In particular, in review r for each aspect $t \in \{\mathbb{A} - A_r\}$ we estimate its (unknown) sentiment s^t_{ui} as a weighted average of the explicitly specified sentiments of its k nearest neighbors. More formally,

$$\hat{s}^t_{ui} = \frac{\sum_1^k w_{tj} \cdot s^j_{ui}}{\sum_1^k w_{tj}},$$

where the weights w_{tj} are computed as Spearman correlations between aspects t and j. In case if the user did not express sentiments about a certain number of the correlated aspects, we leave the sentiment value of the aspect t blank.

Next, for each aspect $t \in \mathbb{A}$ we train the Matrix Factorization model in the following form:

$$\hat{s}^t_{ui} = \mu^t + b^t_u + b^t_i + p^t_u \cdot q^t_i.$$

This approach takes into account user's personal preferences, items individual characteristics and their interaction for each particular aspect $t \in \mathbb{A}$.

(3) Building regression model to predict ratings.

To build the regression model for predicting ratings, we first fill in the remaining missing values of sentiments for those aspects $t \in \{\mathbb{A} - A_r\}$ that were left blank in Step 2. We fill it with the average values of the aspects' sentiments $avg(s^t_{ui})$ taken across explicitly specified values in R. Note, that if the user did not mention the aspect in the review, we assume that she had the average satisfaction of the aspect and it did not affect her overall rating significantly.

We next build the regression model predicting rating r_{ui} of the review based on the estimated sentiments $\hat{s}^t_{ui}$ for aspects from $\mathbb{A}$. More specifically we estimate the overall rating with the regression model in the following form:

$$r_{ui} = (A + B_u + C_i) \cdot S_{ui} \quad (1)$$

where, $A = (a_0, \ldots, a_n)$ is a vector of general coefficients, $B_u = (b^u_0, \ldots, b^u_n)$ is a vector of coefficients pertaining to user u, $C_i = (c^i_0, \ldots, c^i_n)$ is a vector of coefficients pertaining to item i, and $S_{ui} = (\hat{s}^0_{ui}, \ldots, \hat{s}^n_{ui})$ is a vector of estimated values of sentiments corresponding to aspects from $\mathbb{A}$ in the particular review. Further, we avoid over-fitting by using the regularized model:

$$\min_{a_*, b_*, c_*} \sum_{(u,i) \in R} (r_{ui} - \overline{r}_{ui})^2 + \lambda \cdot \left(\sum_{t=0}^{n} \left(a_t^2 + \sum_u (b^u_t)^2 + \sum_i (c^i_t)^2 \right) \right)$$

As it follows from Eq.(1), the proposed model estimates individual preferences of user B_u and the individual characteristics of item C_i.

(4) Calculating impacts of aspects on rating.

In this step we apply the models built in Steps 2 and 3 for determining most important aspects of user's potential experiences with the establishment, where the importance of an aspect is determined by its weight in the regression model described in Step 3. More specifically, for a new potential review we predict sentiment $\bar{s}^t_{ui}$ of each of the aspects $t \in \mathbb{A}$. Then we compute the impact of each sentiment being potentially explicitly expressed in the review as follows. By construction, our regression model takes an average sentiment value in case when the user did not express her opinion. Therefore, we calculate the aspects' impact as product of the difference between the predicted value of sentiment and the average sentiment value for the particular aspect, with the corresponding coefficient in the regression model (Eq.(1)):

$$impact^t_{ui} = (a_t + b^u_t + c^i_t) \cdot (\bar{s}^t_{ui} - avg(s^t_{ui})). \qquad (2)$$

These calculated impacts reflect the importance of each aspect on the overall rating and they might be positive or negative.

(5) Recommending items and conditions.

Next, we identify two groups of aspects in $\mathbb{A}$ over which (a) the user has control and (b) the management of the establishment has control, in order to recommend most important of them in both cases. Furthermore, we identify the conditions that we want to recommend to the user or the management together with the item, where the condition is a suggestion to experience (positive) or not to experience (negative) a particular aspect. Finally, we recommend an item and its corresponding conditions to the user, or the most important conditions to the management.

For example, if our system identified aspect "fish" as having high *positive* impact on the rating, we will recommend this restaurant *and* the condition of ordering fish in that restaurant to the user. Similarly, if aspect "desert" has strong *negative* impact on the rating, we may still recommend to the user to visit that restaurant under the condition *not* to order desert there. Similarly, we can recommend such conditions to the management. For example, we can recommend to the management of a health&beauty salon to provide a complementary drink to the user (since it will improve her overall experience) and don't chat to her too much while in session.

In summary, we proposed a method for identifying and recommending the most valuable aspects of potential user experiences of the items. This method consists of sentiment analysis of user reviews, identification of relations between different aspects of the reviews, predicting sentiments that the user might express about these aspects, and calculating personal impact factors that each aspect contributes to the overall rating for the user. In Section 4, we show the results of applying the proposed method to real data from three applications.

4. EXPERIMENTS AND RESULTS

To demonstrate how well our method works in practice, we tested it on the Yelp dataset[1] for restaurants, hotels and beauty&spas applications for the reviews collected in several US cities over a period of 6 years. In this study we used the reviews of 24,917 restaurants produced by 384,821 users (1,344,405 reviews in total), 1,424 hotels produced by 65,387 users (96,384 reviews) and for the 6,536 beauty&spas produced by 71,422 users (104,199 reviews in total).

[1]https://www.yelp.com/dataset_challenge/dataset

Meat	Fish	Dessert	Money	Service	Decor
beef	cod	tiramisu	price	bartender	design
meat	salmon	cheesecake	dollars	waiter	ceiling
bbq	catfish	chocolate	cost	service	decor
ribs	tuna	dessert	budget	hostess	lounge
veal	shark	ice cream	charge	manager	window
pork	fish	macaroons	check	staff	space

Table 1: Examples of words pertaining to aspects in the restaurant application.

	Restaurants	*Hotels*	*Beauty & Spas*
Regression	1.256	1.275	1.343
Matrix Factorization	1.244	1.273	1.328

Table 2: RMSE of predicted ratings for our method vs. standard MF across three applications.

We applied the 5-step method presented in Section 3 to this Yelp data. As a result, we managed to extract 68 aspects for restaurants, 44 aspects for hotels, and 45 aspects for beauty&spas applications in Step 1 of our method using *Opinion Parser*. Table 1 presents several aspects pertaining to the restaurant application with examples of corresponding words.

Further, the set of reviews R in each application is partitioned into train and test sets in the ratio of 80% to 20%. After determining the sets of aspects in the reviews and aggregating their sentiments as described in Step 1, we filled in the missing sentiment values and trained the sentiment prediction models for each aspect, as described in Step 2.

We compared the performance of our approach with the standard MF models built only on the sets of explicitly expressed sentiments. Our results show that the proposed approach works better than standard one in terms of RMSE for most of the aspects across all the three applications. In particular, our method significantly outperformed standard MF for 43 aspects (out of 68) for restaurants, for 19 aspects (out of 44) for hotels, and for 33 aspects (out of 45) for beauty&spas. For *all* the remaining aspects, the differences between our method and the standard MF approach are not statistically significant. As it was expected, our approach works better for those aspects that have several close neighbors frequently mentioned in the reviews. For example, in a restaurant application aspect "music" is close to "atmosphere" and "interior", and therefore, our approach outperformed the standard one in predicting sentiments of the "music" aspect by 7.8% in terms of RMSE.

In Step 3 of our method, we obtained the rating prediction model and compared its performance with the standard MF approach to rating predictions. The results of this comparison (in terms of RMSE) are presented in Table 2, from which we conclude that the two methods are very similar in terms of their performance. This means that our regression model (Eq.(1)) predicts ratings reasonable well (comparably to the MF approach).

		Restaurants		*Hotels*		*Beauty & Spas*	
		users	*managers*	*users*	*managers*	*users*	*managers*
Positive Recommendations	Followed	3.818	3.816	3.410	3.537	4.176	4.167
	Other cases	3.734	3.737	3.320	3.324	4.051	4.053
Negative Recommendations	Not followed	3.482	3.473	3.105	2.869	3.740	3.744
	Other cases	3.784	3.787	3.342	3.429	4.126	4.127

Table 3: Average ratings for the users who followed (or not) our positive/negative recommendations of items with conditions.

We next applied our models to the test data and determined the most important aspects for each user and item pair, as described in Step 4. Finally, we produced recommendations both for the users and the management, as described in Step 5.

We next compared the ratings of the users and the management who followed our positive recommendations (i.e., mentioned the recommended aspects in their reviews) against other cases (i.e., that did not mention them) across the restaurants, hotels and beauty&spas applications. Further, we also did similar comparisons for negative recommendations (such as, do not order the desert in restaurant X), i.e., we compared users or managers who did not follow our negative recommendations (i.e., mentioned the negatively recommended aspects in their reviews) with others.

The results of these comparisons are presented in Table 3, where numbers in cells represent average ratings for users and managers across the three applications. The rows in Table 3 represent different conditions of whether or not the users followed our recommendations. As Table 3 shows, *our recommendations of items and conditions lead to higher evaluation ratings in those cases when users followed them* vs. other cases, and all the differences are statistically significant.

5. CONCLUSION

In this paper, we presented a new method of recommending not only items of interest to the user but also the conditions enhancing user experiences with those items. This method is based on the sentiment analysis of user reviews, predicts sentiments that the user might express about the aspects determined in an application, such as restaurants or hotels, and identifies the most valuable aspects of user's potential experience with the item. We tested it on three Yelp applications (restaurants, hotels and beauty&spas) and showed that our recommendations lead to higher evaluation ratings when users followed them vs. others.

This new approach to providing recommendations helps users to customize and enhance their experiences when consuming items, such as deciding what they should order in a particular restaurant.

6. REFERENCES

[1] S. Aciar, D. Zhang, S. Simoff, and J. Debenham. Informed recommender: Basing recommendations on consumer product reviews. *Intelligent Systems*, 2007.

[2] L. Chen, G. Chen, and F. Wang. Recommender systems based on user reviews: the state of the art. *User Modeling and User-Adapted Interaction*, 2015.

[3] Q. Diao, M. Qiu, C.-Y. Wu, A. J. Smola, J. Jiang, and C. Wang. Jointly modeling aspects, ratings and sentiments for movie recommendation. KDD, 2014.

[4] X. Ding, B. Liu, and P. S. Yu. A holistic lexicon-based approach to opinion mining. WSDM '08. ACM, 2008.

[5] G. Ganu, Y. Kakodkar, and A. Marian. Improving the quality of predictions using textual information in online user reviews. *Inf. Syst.*, 38(1):1–15, Mar. 2013.

[6] M. Hu and B. Liu. Mining and summarizing customer reviews. KDD '04. ACM, 2004.

[7] A. Lawlor, K. Muhammad, R. Rafter, and B. Smyth. Opinionated explanations for recommendation systems. *Research and Dev. in Int. System*, 2015.

[8] G. Ling, M. R. Lyu, and I. King. Ratings meet reviews, a combined approach to recommend. RecSys '14, pages 105–112, New York, NY, USA, 2014. ACM.

[9] B. Liu. *Sentiment Analysis and Opinion Mining.* Morgan and Claypool Publishers, 2012.

[10] H. Liu, J. He, T. Wang, W. Song, and X. Du. Combining user preferences and user opinions for accurate recommendation. *Electron. Commer. Rec. Appl.*, 2013.

[11] J. McAuley and J. Leskovec. Hidden factors and hidden topics: Understanding rating dimensions with review text. RecSys '13. ACM, 2013.

[12] J. McAuley, J. Leskovec, and D. Jurafsky. Learning attitudes and attributes from multi-aspect reviews. ICDM '12, pages 1020–1025, 2012.

[13] G. Qiu, B. Liu, J. Bu, and C. Chen. Opinion word expansion and target extraction through double propagation. *Comput. Linguist.*, 2011.

[14] V. Suresh, S. Roohi, and M. Eirinaki. Aspect-based opinion mining and recommendationsystem for restaurant reviews. RecSys '14. ACM, 2014.

[15] N. Tintarev and J. Masthoff. Explaining recommendations: Design and evaluation. *Recommender Systems Handbook 2nd edt.*, 2015.

[16] I. Titov and R. McDonald. A joint model of text and aspect ratings for sentiment summarization. In *Proceedings of ACL-08: HLT*, 2008.

[17] H. Wang, Y. Lu, and C. Zhai. Latent aspect rating analysis without aspect keyword supervision. KDD '11. ACM, 2011.

[18] Y. Wang, Y. Liu, and X. Yu. Collaborative filtering with aspect-based opinion mining: A tensor factorization approach. In *IEEE ICDM*, 2012.

[19] Y. Xu, Z. Chen, J. Yin, Z. Wu, and T. Yao. Learning to recommend with user generated content. *WAIM*, 2015.

[20] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *ACM SIGIR*, 2014.