Yelp Review Rating Predictions Based on Collaborative Filtering with Phrase Level Similarity in Reviews

Yifei Xu, Xu Feng, Kai Liu, Wensi Fu

Business review websites are helping customers to learn about products by providing information in the form of ratings and user-generated reviews. As the number of products increases, recommender systems are built to help the customers to make buying decisions. Many recommender systems are based on collaborative filtering (CF) approach [1], which recommends items by considering other users with similar preference. Traditional collaborative filtering methods use the users' numeric rating information to build the recommender system. Recently, studies show that recommender systems with information from reviews text could provide more accurate recommendation [2, 3, 4, 5] as the review text could provide valuable information about users' sentiments and preferences of features.

Several methods have been developed to combine star ratings and reviews text in the recommender system. To the best of our knowledge, the first work is the study by Leung et al. in 2006 by using a model-based CF [6]. They extracted the sentimental orientation and the intensity of the opinion based on a probabilistic inference model, and then used the results of the model as new rating scores to replace the original scores in CF. Later, Ganu et al. [2] made the improvement based on Leung et al.'s work by summarizing all the features and their roles in the model, and by aiming to predict a 5-point rating scale. In later study, Jacob et al. [3] came up with a model-based hybrid CF system, in which they created the user profiles by using multi-relational model that reflects the interactions between users' opinion polarity and movies. Studies found that memory-based CF generally provides better performance in terms of prediction accuracy [4]. Two later studies used memory-based CF to incorporated text review, focusing on the cold start problems of movie [5] and hotel recommender systems [7]. They directly integrated the feature-oriented opinions in the user-item matrix to better identify the genuine similar neighbors. Recently, Zhang et al. [8] proposed a user-review enhanced collaborative filtering (urCF) by integrating text reviews into traditional score-based KNN. In their work, instead of incorporating the reviews directly into CF, they used them alongside with the user-item rating matrix to enhance CF performance. They proposed a FF-IRF (Feature Frequency-Inverse Review Frequency) method to weigh feature priority and used two methods to incorporate the information gained from the text review into user similarity measurement. In the paper they claimed the urCF performed better than model-based CF in [2, 4].

Based on [8], in our project, we incorporate the text review information into the calculation of similarity in KNN method, and try to improve the accuracy of this technique further in the following ways:

1) We will extract features and sentiments from reviews in sub-sentence level. Even though using sentence as the unit is the most popular method to identify features and sentiments, it is often the case that each sentence contains more than one features of the business in several clauses. In addition, one clause can express different sentiment than other clause in the same sentence. For example, in the sentence "The food here is rather good, but only if you like to wait for it" [2], the first half of the sentence indicates positive sentiment in feature "food", while the second half shows opposite sentiment in feature "waiting time". In traditional methods, such sentences is categorized into conflict category, which is not used

in the prediction [2]. Therefore, extracting features and sentiments in clause-level allows us associate potential sentiments with features more accurately.

2) We will use each reviewer's rating calculated based on his/her review, instead of the star rating, in the KNN. This way will solve the problem that different users could have various standards in rating the same business [6]. For instance, two users both agree a restaurant is good, but one may think "good" deserves a five-star while the other may use four star to represent "good". In our proposed method, similar sentimental words from different reviews will be given same sentimental orientations.

3) We will make changes in the similarity calculation. In the method mentioned in [8], the similarity between two users are calculated based only on movies rated by both users. Since the reviews for restaurants are relatively much sparser than that for movies, we will changes the way of similarity calculation, and use all the reviews of the users. Following the assumptions that the more frequently a user mentions a specific feature[1, 2], the more important this feature is for the overall rating, we will calculate a user-feature attention matrix similar to [9] to describe the priority of different features in users' overall opinion and proposed a formula for similarity calculation based on this matrix. If two users have common rated restaurants, the method in [8] will also be used and combined with our method.

The proposed approach is applied to a Yelp dataset for rating prediction. The dataset contains 1.6M reviews and 500K tips provided by 366K users for 61K businesses. We only focus on food-related business categories (58 categories, i.e. food, taco, Chinese et al.), which have two thirds of total reviews. The traditional memory-based CF will be used as a baseline for the purpose of accuracy comparison.

References

[1] Y. Koren et al., “Matrix Factorization Techniques for Recommender Systems,” *Computer (Long. Beach. Calif).*, vol. 42, no. 8, pp. 42–49, 2009.

[2] G. Ganu et al., “Beyond the Stars : Improving Rating Predictions using Review Text Content,” *Text*, no. WebDB, pp. 1–6, 2009.

[3] N. Jakob et al., “Beyond the stars: exploiting free-text user reviews to improve the accuracy of movie recommendations,” in *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, 2009, pp. 57–64.

[4] G. Adomavicius et al., “Personalization and recommender systems,” *State-of-the-art Decis. Mak. tools information-intensive age*, pp. 55–100, 2008.

[5] A. Levi et al., “Finding a needle in a haystack of reviews: cold start context-based hotel recommender system,” in *Proceedings of the sixth ACM conference on Recommender systems*, 2012, pp. 115–122.

[6] C. W. K. Leung et al., “Integrating collaborative filtering and sentiment analysis: A rating inference approach,” in *Proceedings of the ECAI 2006 workshop on recommender systems*, 2006, pp. 62–66.

[7] Y. Moshfeghi et al., “Handling data sparsity in collaborative filtering using emotion and semantic based features,” in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, 2011, pp. 625–634.

[8] Z. Zhang et al., “urCF : User Review Enhanced Collaborative Filtering,” in *Twentieth Americas Conference on Information Systems, Savannah*, 2014, pp. 1–11.

[9] Y. Y. Zhang et al., “Explicit Factor Models for Explainable Recommendation based on Phrase-level Sentiment Analysis Categories and Subject Descriptors,” in *Sigir*, 2014, pp. 83–92.

[10] Y. Lu et al., “Automatic construction of a context-aware sentiment lexicon: an optimization approach,” in *Proceedings of the 20th international conference on World wide web*, 2011, pp. 347–356.

[11] Y. Zhang et al., “Do users rate or review?: boost phrase-level sentiment labeling with review-level sentiment classification,” in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, 2014, pp. 1027–1030.