Connecting Social Media to Ecommerce: Cold-Start Product Recommendation using Micro blogging Information

Yande M.S[1], Wakchaure M.A[2]

[1] *Student, Department of computer engineering, Avcoe, Maharashtra, India*
[2] *Professor, Department of computer engineering, Avcoe, Maharashtra, India*

ABSTRACT

In recent years, the edge between e-commerce and social networking have become increasingly blurred. Many e-commerce websites support the mechanism of social login where users can sign on the websites using their social network identities such as their Facebook or Twitter accounts. Users can also post their newly purchased products on microblogs with links to the e-commerce product web pages. In this paper we represent a novel solution for cross-site cold-start product recommendation, which aims to recommend products from e-commerce websites to users at social networking sites in "coldstart" situations, a problem which has rarely been explored before. A major threat is how to leverage knowledge extracted from social networking sites for cross-site cold-start product recommendation. We propose to use the linked users across social networking sites and e-commerce websites (users who have social networking accounts and have made purchases on e-commerce websites) as a bridge to map users' social networking features to another feature representation for product recommendation. In specific, we propose learning both users' and products' feature representations (called user embeddings and product embeddings, respectively) from data collected from e-commerce websites using recurrent neural networks and then apply a modified gradient boosting trees method to transform users' social networking features into user embeddings. We then develop a feature-based matrix factorization approach which can leverage the learnt user embeddings for cold-start product recommendation. Experimental calculation on a large dataset build from the largest Chinese micro blogging service SINA WEIBO and the largest Chinese B2C e-commerce website JINGDONG have given the effectiveness of our proposed framework.

Keyword : *e-commerce, product recommender, product demographic, microblogs, recurrent neural networks etc….*

---

1. INTRODUCTION

1.1 Motivation:

In recent years, the boundaries between e-commerce and social networking have become increasingly blurred. E-commerce websites such as eBay features many of the characteristics of social networks, including real-time status updates and interactions between its buyers and sellers. Some e-commerce websites also support the mechanism of *social login*, which allows new users to sign in with their existing login information from social networking services such as Facebook, Twitter or Google+. Both Facebook and Twitter have introduced a new feature last year that allow users to buy products directly from their websites by clicking a "buy" button to purchase items in adverts or

other posts. In China, the e-commerce company ALIBABA has made a strategic investmentin SINA WEIBO1 where ALIBABA product adverts can be directly delivered to SINA WEIBO users. With the new trend of conducting e-commerce activities on social networking sites, it is important to leverage knowledge extracted from social networking sites for the development of product recommender systems. In this paper, we study an interesting problem of propose products from e-commerce websites to users at social networking sites who do not have historical purchase records, i.e., in "cold-start" situations. We called this problem *cross-site cold-start product production.* Although online product recommendation has been extensively studied before [1], [2], [3], most studies only focus on constructing solutions within certain e-commerce websites and mainly utilize users' historical transaction records. To the best of our knowledge, *cross-site cold-start product recommendation* has been rarely studied before. In our problem setting here, only the users' social networking information is available and it is a challenging task to transform the social networking information into latent user features which can be effectively used for product recommendation. To address this threat, we represent to use the linked users across social networking sites and e-commerce websites (users who have social networking accounts and have made purchases on e-commerce websites) as a bridge to map users' social networking features to latent features for product recommendation. In specific, we represent learning both users' and products' feature representations (called user embeddings and product embeddings, respectively) from data collected from ecommerce websites using recurrent neural networks and then apply a modified gradient boosting trees method to transform users' social networking features into user embeddings. We then develop a featurebased matrix factorization approach which can leverage the learnt user embeddings for cold-start product recommendation. We built our dataset from the largest Chinese microblogging service SINA WEIBO2 and the largest Chinese B2C e-commerce website JINGDONG3, containing a total of 20,638 linked users. The experimental results on the dataset have shown the feasibility and the effectiveness of our proposed framework.

Our major contributions are summarised below:

• We formulate a novel problem of recommending products from an e-commerce website to social networking users in "cold-start" situations. To the best of our knowledge, it has been rarely studied before.

• We propose to apply the recurrent neural networks for learning correlated feature representations for both users and products from data collected from an e-commerce website.

• We propose a modified gradient boosting trees method to transform users' microblogging attributes to latent feature representation which can be easily incorporated for product recommendation.

• We propose and instantiate a feature-based matrix factorization approach by incorporating user and product features for cold-start product recommendation.

.

2. METHODS AND MATERIAL

Extracting and Representing Micro blogging Activities Three steps: Prepare a list of potentially useful micro blogging attributes and construct the micro blogging feature vector for each linked user. Learn the mapping function, which transforms the micro blogging attribute information to the distributed feature representations in the second step. It utilizes the feature representation pairs. B. Micro blogging-Feature Selection We study about how to extract information from micro blogging from rich user. By this micro blogging feature representation can be constructed. C. Demographic Attributes A demographic profile is often called as demographic. It is very important in marketing and mainly in product adoption. Users information such as gender, age and education can be used by e-commerce to provide personalized service. We extract users demographic attributes from their public profiles on SINA WEIBO. By studying it earlier, we have identifies six major demographic attributes: Gender, age, marital status, education, career and interest. 1. Text Attributes In this user often reflect their opinions and interest about certain topics. Unabsorbed products will be asked to take a look. 2. Network Attributes In the online social media space, it is often observed that users connected with each other (e.g., through following links) are likely to share similar interests. 3. Temporal Attributes Temporal activity patterns are also considered since they reflect the living habits and lifestyles of the micro blogging users to some extent. As such, there might exist correlations between

temporal activities patterns and users' purchase preferences. Temporal activity distributions. We consider two types of temporal activity distributions, namely daily activity distributions and weekly activity distributions for product recommendation. D. Existing System The existing is the novel problem of recommending the products from an e-commerce website to social networking users in "cold- start" situations. The recurrent neural network is used which is used for learning correlated feature representations for both users and products. It is the connection between units form a directed cycle, which allows it to exhibit dynamic temporal network. And modified gradient boosting tress method to transform user's micro blogging attributes to latent feature representation which can be easily incorporated for product recommendation. It is a machine learning technique for regression and classification problems. Regression is the measure of the relation between the mean value of one variable (eg:output) and corresponding value of other variable(eg: time and cost) or a return to a former or less developed state. A feature – based matrix factorization approach is instantiated by incorporating user and product feature for cold-start product recommendation.

3. LITURATURE SURVEY

we highlight related important approaches: 1) traditional collaborative recommender, 2) trust-enhanced recommender, and 3) reviews-based recommender. First, the traditional collaborative filtering approaches can be either memory-based or model-based. These methods are based on the rating history from users. In the memory-based methods, similarity computation is a primary element. They use a heuristic utility of similarity between users' vectors such as Pearson Correlation Coefficient (PCC) or cosine similarity measure (VCC) [1], [5], [13], [14]. On the other hand, the model-based methods employ machine learning models to predict product ratings [15], [16]. For example, Sarwar et al. [4], [17] implemented clustering algorithms to identify groups of customers who rated similar products and these clusters can be seen as likeminded neighbors. Since k clusters are created, recommendation prediction can be computed by averaging the ratings in that cluster. Miyahara and Pazzani [5] proposed a RS based on Naive Bayes classifier and they only considered items which co-rated between users. They manipulated two classes: like and don't like and features are selected in a preprocessing step. Recent proposals focused in the accuracy of predictions such as matrix factorization for collaborative filtering. The approach proposed in [8] involved social connections data in providing recommendation by assigning social regularization terms in order to constraint matrix factorization objective function. They assumed that friends rate products and hence they used PCC and VCC to measure similarity as intermediate step. Second, more studies have focused on trust-enhanced recommenders. Some studies applied trust by building trust net- work based on the assumption that users can obtain more accurate recommendation from people they trust [18], [19]. These kinds of methods used direct evaluations of trust from users. Golbeck et al. [18] propagated trust from trust network so- called Web Of Trust WOT. Only friends whose trust evaluation exceeds a threshold will be involved in recommendation encounter. Recommendations are obtained by weighted average of ratings along with the trust value using Film Trust dataset. In another context, Massa et al. [19] used trust to filter the set of neighbors and only their rated items would be considered in predicting ratings to an active user. After filtering neighbors, they applied the traditional recommendation algorithm. The experiments were based on Opinion dataset which contains both users' ratings and the direct trust values from users towards each other. Third, recent researches have been done to exploit the sentiment in the textual reviews to augment ratings in collaborative recommenders [11], [12], [20]. Authors in [11] tried to improve the RSs by leveraging topic and sentiment information at sentences level. They inferred ratings from text reviews written by users about restaurants in multi-point rating scale rather than only positive or negative polarities. They applied text regression model to estimate scaled sentiment ratings. They are the first who integrated the useful information in reviews into RSs. Lenug et al. [12] proposed a probabilistic sentiment inference framework. They applied natural language techniques to compute sentiment orientation in reviews. They built their rating inference model based on the Naive Bayes classifier. Then, they integrated between the inference ratings from reviews and a CF algorithm to increase users' preferences and achieved encouraging results. Esparza et al. in [21] investigated how to obtain recommendation from online micro blogging services. They proposed a solution to exploit short posts written by users as product reviews. These posts are used to build user-item profile. Then a query search algorithm is applied to retrieve relevant item profiles based on a twitter-like review service called blipper.com. This study is similar to our work in using micro blogging as a source of recommendation. Some inherent drawbacks still have not been solved in the above mentioned methods. Most of these approaches require users to produce some structured data first such as trust evaluations and ratings to allow the corresponding systems to work properly. In fact, this is not practical and

usually not available. Nevertheless, the weaknesses of sparsity and cold start problems appear in the case of trust network as it is in the user-item rating matrix. On the other hand, review based recommenders require a user to write reviews and rate products to generate the suggestions. Unlike existing studies, our novel approach ISTS overcomes the need of ratings or written reviews by users and reflects the real hidden social trust relations. In our work we personalize recommendations from micro loggings using sentiment analysis and trust between friends.

4. MICROBLOGGING SERVICES

In this section we introduce our target social network Twitter. Users in the micro blogger Twitter can publish short posts in 140 character limit so-called tweets. Today, Twitter users can generate more than 300 Million tweets each day [22] about different topic and interest. For example, people can generate brief posts about their personal experience in reading books, watching movies, breaking news or even the release of new electronic gadgets. In addition, users have the choice to establish relationships among each others for social links, seeking information or identifying following/followers friends.

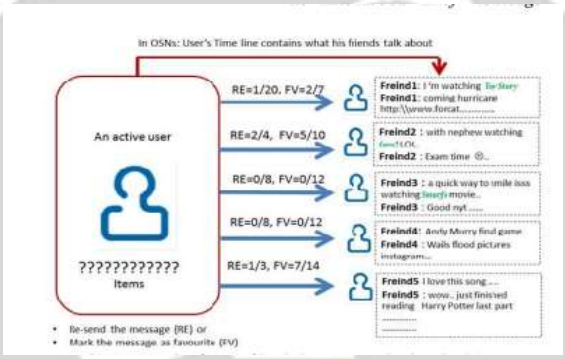

Fig.1: Example of user's friends intercommunications in OSNs. Measuring the different level of hidden and subjective trust relationships between friends in Twitter is crucial in our research. Therefore, we developed a tool to automatically collect social network data by using Twitter API. This tool extracts the required interactions between friends, and it is called Twitter Interaction Extractor (TIE).

5. CONCLUSIONS

In this paper, we have studied a novel problem, *cross-site cold-start product recommendation*, i.e., recommending products from e-commerce websites to micro blogging users without historical purchase records. Our main idea is that on the e-commerce websites, users and products can be represented in the same latent feature space through feature learning with the recurrent neural networks. Using a set of linked users across both e-commerce websites and social networking sites as a bridge, we can learn feature mapping functions using a modified gradient boosting trees method, which maps users' attributes extracted from social networking sites onto feature representations learned from e-commerce websites. The mapped user features can be effectively incorporated into a feature-based matrix factorization approach for cold start product recommendation. We have constructed a large dataset from WEIBO and JINGDONG. The results show that our proposed framework is indeed effective in addressing the *cross-site cold-start product recommendation* problem. We believe that our study will have profound impact on both research and industry communities. Currently, only a simple neutral network architecture has been employed for user and product embeddings learning. In the future, more advanced deep learning models such as Convolution Neural Networks13 can be explored for feature learning. We will also consider improving the current feature mapping method through ideas in transferring learning [30].

6. ACKNOWLEDGEMENT

The authors thank the anonymous reviewers for their valuable and constructive comments. The work was partially supported by National Natural Science Foundation of China under the grant number 61502502, the National Key

Basic Research Program (973 Program) of China under the grant number 2014CB340403 and the Innovate UK under the grant number 101779.

7. REFERENCES

[1]J. Wang and Y. Zhang, "Opportunity model for e-commerce recommendation: Right product; right time," in *SIGIR*, 2013.

[2] M. Gyring, "Retail sales prediction and item recommendations using customer demographics at store level," *SIGKDD Explor. Newsl.*, vol. 10, no. 2, Dec. 2008.

[3] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: Item-to-item collaborative filtering," *IEEE Internet Computing*, vol. 7, no. 1, Jan. 2003.

[4] V. A. Zeithaml, Sarwar et al "The new demographics and market fragmentation," *Journal of Marketing*, vol. 49, pp. 64–75, 1985.

[5] W. X. Zhao, Y. Guo, Y. He, H. Jiang, Y. Wu, and X. Li, Miyahara and Pazzani "We know what you want to buy: a demographic-based system for product recommendation on microblogs," in *SIGKDD*, 2014.

[6] J. Wang, W. X. Zhao, Y. He, and X. Li, "Leveraging product adopter information from online reviews for product recommendation," in *ICWSM*, 2015.

[7] Y. Seroussi, F. Bohnert, and I. Zukerman, "Personalised rating prediction for new users using latent factor models," in *ACM HH*, 2011.

[8] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *NIPS*, 2013.

[9] Q. V. Le and T. Mikolov, "Distributed representations of sentences and documents," *CoRR*, vol. abs/1405.4053, 2014.

[10] J. Lin, K. Sugiyama, M. Kan, and T. Chua, "Addressing coldstart in app recommendation: latent user models constructed from twitter followers," in *SIGIR*, 2013.

[11] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *CoRR*, vol. abs/1301.3781, 2013.

[12] RSs. Lenug et al. ,Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, Aug. 2009.

[13] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, pp. 1189–1232, 2000.

[14] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Monterey, CA: Wadsworth and Brooks, 1984.

[15] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, Oct. 2001.

[16] K. Zhou, S. Yang, and H. Zha, "Functional matrix factorizations for cold-start recommendation," in *SIGIR*, 2011.