ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 6, June 2017

Implementing Cold-Start Product Recommendation System by using Micro blogging Information

[1] Marjana Khathima, [2] AV, Mr. Ravindra Prasad S
[1] M.Tech Student, Department of Computer Science and Engineering, RR Institute of Technology, Bangalore.
[2]Associate Professor, Department of Computer Science and Engineering, RR Institute of Technology, Bangalore, India

***Abstract*— Nowadays, it is possible to access many E-commerce websites by using users social network accounts like facebook, twitter etc. Users of social networks can able to post their newly purchased products in the microblogs, and can give links to the E-commerce web pages from where they are purchased. This paper presents a solution for cross-site cold start product recommendation. A major challenge for this problem is how to use the information taken from social networking sites for cold start product recommendation. This paper proposes, by using neural networks extract user features or user embeddings and product feature or product embeddings from the data collected from E-commerce websites. Then by using gradient boosting tree method on the social networking sites, collect user features and combine this with the user embeddings. Then by using matrix factorization method use these user embeddings for the cold start product recommendation.**

***Index Terms*— social network, micro blogs, Cold-start, gradient boosting tree method, matrix factorization method.**

I. INTRODUCTION

Recommendation systems are mainly used by e-commerce companies like Amazon.com, for promoting sales to potential customers by discovering items to customers that they might not have found by themselves. A good recommendation system is able to provide customers the most relevant products. Cold start problem is that problem, where system is not able to recommend items to the users. For every recommender system, it is required to build user profile by considering user preferences and likes. User profile is developed by considering user activities and behaviours that he perform with the system. On the basis of user previous history and activities system make decisions and recommend items consequently. The problem arises when a new user or new item enters the system, for such user/items system don't have enough information to make a decision. For example, a new user has not rated some items and not yet visited/viewed some items then it would be difficult for the system to build a model on that basis.

Nowadays, there is a cross connection between e-commerce and social networking websites. E-commerce websites such as eBay have features of social networks, such as real-time status updates and interactions between its buyers and sellers. Some e-commerce websites also allow the mechanism of social login, thereby new users can 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 that allow users to buy products directly from their websites by clicking a "buy" button to purchase items in adverts or other posts.

In this paper, we have an interesting problem called "cold-start problem". We are resolving it by recommending products from e-commerce websites to users at social networking sites who do not have historical purchase records, i.e., in "cold-start" environment". We called this recommendation as "cross site cold-start product recommendation".

II. LITERATURE SURVEY

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

This paper has taken the proportional hazards modelling method for literature survey and proposed the "opportunity model. This model calculates the joint probability of a user making a follow-up purchase of a particular product at a particular time. This joint purchase probability can be utilized by recommender systems in various scenarios, for e.g. recommendation on an e-commerce web site an email or text message based marketing etc. The opportunity modelling approach has evaluated with multiple metrics. This model can predict a user's follow-up purchase behaviour at a particular time with descent accuracy.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 6, June 2017

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

This paper was implemented for a chain of retail stores Data collected from daily sales information for 600 products broken out over a set of customer types. Recommender system was built based on a fast online thin singular value decomposition. It provides improved performance than single aggregate model built for the entire database. This model was implemented both as a product recommender and as an customer analysis tool. The predictability accuracy of this recommender was 1.5-5 times greater for the items of interest as measured by r-squared error statistics.

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

Recommendation algorithms are mainly used in e-commerce Websites, where they take customer's interests to generate a list of recommended item. Most of the applications use only the items that customers purchase and rate to represent their interests. But they can also use other attributes, including items viewed, demographic data, subject interests, and favourite artists. At Amazon.com, they use recommendation algorithms to personalize the online store for each customer. The store continuously changes based on customer interests, for e.g.: showing programming titles to a software engineer and baby toys to a new mother.

D. "We know what you want to buy: a demographic-based system for product recommendation on micro blogs," W. X. Zhao, Y. Guo, Y. He, H. Jiang, Y. Wu, and X. Li, in SIGKDD, 2014.

E-commerce websites develops product recommender systems mainly for improving user experience and increase sales. However, recommendation is limited by the product information hosted in those e-commerce sites and is only activated when users are performing e-commerce activities. This paper, develops a product recommender system called METIS, a MErchanT Intelligence recommender System, which detects users' purchase intents from their micro blogs in near real-time and makes product recommendation based on matching the users' demographic information taken from their public profiles with product demographics learned from micro blogs and online reviews. METIS diffentiates itself from traditional product recommender systems in the following aspects: 1) METIS was developed based on a micro blogging service platform, because of that, it is not limited by the information available in any specific e-commerce website. Also, METIS is able to track users' purchase intents in near real-time and make recommendations accordingly. 2) In METIS, product recommendation is made as a learning to rank problem. Users' characteristics extracted from their public profiles in micro blogs and products' demographics learned from both online product reviews and micro blogs are fed into learning to rank algorithms for product recommendation.

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

The availability of the very large amount of online product feedbacks or reviews provides demographic information of product adopters from the review documents. This paper proposes a novel approach to the extraction of product adopter mentions from online reviews. The extracted product adopters are then categorise into a number of different demographic user groups. The aggregated demographic information of many product adopters can be used to characterise both products and users, which can be incorporated into a recommendation method.

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

Accurate rating prediction is essential for recommendations. This paper addresses the new user problem by introducing several extensions to the basic matrix factorization algorithm which takes user attributes in to account when generating user predictions. Here both

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 6, June 2017

demographic attributes and attributes inferred from user generated texts are considered.

III. EXISTING SYSTEM

Below are the challenges occurred when there is an interaction between users on social network and ecommerce sites:

- Social networks are private and so direct access may cause negation by the users. This leads to damaging the social network platforms, as users might stop accessing the site to avoid access of their privacy.
- Main aim of brands is that social media interaction should limit in customers and their retention.
- The other challenge is to remain customer friendly during the changing trends and competition. Customers will not give much effort when they want to buy something online and this is more impacting for a new shopper who comes up on a ecommerce site because of a social network recommendation. The intent is very volatile and can go in case of complex application.

Whatever be the modes in the application, the UI needs to be completely effortless. Our work mostly addresses the new trend of social commerce connecting social and e-commerce domains. A very deep study of the growth and success of a social commerce site was performed. The investigation is finalized to the use of micro blogs to target the customers. The following three concepts work concurrently to create a global community that has started to take the place of traditional commerce and socialization: Web technology, E-commerce, and social media. Research findings indicate that social commerce is very profitable because of the various offers given to users as they connect with others in spite of their identity and location. The focus of this paper is to augment understanding on swiftly developing Web based social media and their later effects on the evolving social commerce. Majority of the existing models use various methods for product recommendation to the users present on both social and commerce domains.

IV. PROPOSED SYSTEM

The boundary between e-commerce and social networking sites has become blurred. E-commerce websites such as e-Bay has many of the traits of social networks, including real-time updates and interaction between buyers and sellers by using their micro blogs. Some e-commerce websites also support the mechanism of social login, which allows users to login with their existing login information from social networking.
There is no such system that has adopted the use of micro-blogging and other demographic information for cold start situation where a customer to e-commerce site is offered suggestion of the products. Here we are focused on the details of the micro-blogging information, demographic information, location information, etc for the product recommendation. In this paper, we face the problem of recommending products to users who do not have any historical purchase records, i.e., in “cold-start” situations. We called the solution to this problem as “cross site cold-start product recommendation”.
We propose to use the coupled users across social networking sites and e-commerce websites (users of the social networking accounts and have done purchases on e-commerce websites) as a bridge to map users’ social networking features to latent features for product recommendation. In specific, we have a tendency to propose learning each users’ and products’ feature representations (called user embeddings and product embeddings, respectively) from the information collected from the ecommerce websites by using neural networks then apply a modified gradient boosting trees method to transform users’ social networking features into user embeddings. Then by applying a feature based matrix factoring approach which might utilize the learnt user embeddings for cold-start product recommendation.

1. Extracting and Evaluating Micro-blogging Attributes and Features

Our proposed solution to micro-blogging attribute learning has three steps:
- Create a list of useful micro-blog attributes and create the micro-blogging feature map.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 6, June 2017

- Generate feature maps using the information from all the users on the e-commerce website through intensive learning;
- Learn the mapping function, which transforms the micro-blogging information to the features in the second step.

2. *Microblogging Feature Selection*

In this section, we study how to extract user information from micro blogs. We have three groups of attributes.

a. *Demographic Attributes*

A demographic profile (often shortened as "a demographic") of a user such as gender, age and education can be used by e-commerce companies to provide better customized services. Demographic attributes have been shown to be very important in marketing, especially in product adoption for consumers. As per our previous study, we identify six major demographic attributes: gender, age, marital status, education, career and interests.

b. *Text Attributes*

Recent studies tell that microblogs contain rich commercial information of users. Also, users' microblogs often display their opinions and interests towards certain areas. As such, we expect a potential relation between text attributes and users' purchase preferences. We first collate all the microblogs by a user into a document, and then run the analysis function. The benefits of topics distributions over keywords are double. Word embeddings, Standard topic models assume individual words can be exchanged, which is essentially the same as the bag-of-words model assumption. Word representations or embeddings learned using neural language models help addressing the problem of traditional bag-of-word approaches which fail to capture words' contextual semantics. In word embeddings, each dimension represents a latent feature of the word and semantically similar words are close in the latent space. Finally, we average the word maps of all the tokens in a user's published document as the user's embedding vector.

c. *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. As such, we can find out useful user groups by the users' following shopping patterns assuming that users in the same group share similar purchase preferences. Latent group preference, we treat a following user as a token and aggregate all the followings of a user as an individual document. Thus, we can extract latent user groups having same interests (called "following topics").

d. *Temporal Attributes*

Temporal activity patterns are also utilized as they show the habits and lifestyles of the microblogging users to some extent. There are some relations between temporal activities patterns and users' purchase preferences. Temporal activity distributions, we analyze two types of temporal activity distributions, daily and weekly activity distributions. The daily activity distribution of a user is characterized by a distribution of 24 ratios, and the ith -ratio indicates the average proportion of tweets published within the ith hour of a day by the user; similarly weekly activity distribution of a user is characterized by a distribution of seven ratios, and the ith -ratio indicates the average proportion of tweets published within the ith day of a week by the user.

All the attributes can be summarized into below table:

Categorization of the micro-blogging categories and features:

Categories	**Features**
Demographic Attributes	Gender, Age, Marital status, Education, Career, Interests

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 6, June 2017

Text Attributes	Topic distributions , Word embeddings
Network Attributes	Latent group preference
Temporal Attributes	Daily activity distribution ,Weekly activity distribution

3. Distributed Representation Learning With Recurrent Neutral Networks

We use recently proposed methods in learning word embeddings using recurrent neutral networks to learn user embeddings or distributed representation of user. We first discuss how to learn product embeddings and in the later part the word embeddings.

There are two simple recurrent neutral architectures to train product embeddings, the Continuous Bag-Of-Words model (CBOW) and the Skip-gram model . The major difference between these two architectures is in the direction of prediction: CBOW predicts the current product using the surrounding context, while Skip-gram predicts the context with the current product. In our evaluations, the context is defined as a window of size 4 surrounding a target product which contains two products purchased before and two after. With product embeddings, if we can learn user embeddings in a similar way, then we can explore the related representations of a user and products for product recommendation. The purchase history of a user is like a "sentence" having of a sequence of product IDs as word tokens. A user ID is placed at the beginning of each sentence, and both user IDs and product IDs are treated as word tokens in the learning process. During training, for each sentence, the sliding context window will always include the first word (i.e., user ID) in the sentence. In this way, a user ID is essentially always associated with a set of her purchase records (of 4 products at a time).

Advantages:

- Gain customer information like what they are, what they like, etc. which can transform our business.
- Increase brand awareness i.e. targets more people to our e-commerce.
- Run customer targeted ads with real time results.
- Generate valuable leads i.e. transform ad viewer to a customer.
- Increase website traffic and search ranking.
- Find out information about how competitor is performing and change ourselves according to that.
- Share content faster and easier.

V. CONCLUSION

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.

ACKNOWLEDGMENT

With immense pleasure, we are publishing this paper as a part of the curriculum of M.Tech, Computer Science and Engineering. It gives us proud privilege to complete this paper work under the valuable guidance of Principal for giving all facilities for the smooth progress of paper work. We would also like to thank all the faculty Members of Computer Engineering Department, Management, friends and family members, Who have directly or indirectly guided and helped us for the preparation of this paper.

REFERENCES

1. Wayne Xin Zhao, Sui Li, Yulan He, Edward Y. Chang, Ji-Rong Wen and Xiaoming Li, "Connecting Social Media to E-Commerce: Cold-Start Product Recommendation using Microblogging Information", 2015
.
2. M. Zhang, J. Tang, X. Zhang, and X. Xue, "Addressing cold start in recommender systems: a semi-supervised co-training algorithm," in SIGIR, 2014.
3. Ma, T. C. Zhou, M. R. Lyu, and I. King, "Improving recommender systems by incorporating social contextual in-formation," ACM Trans. Inf. Syst., vol. 29, no. 2, 2011.

4. J. Wang, W. X. Zhao, Y. He, and X. Li, "Leveraging product adopter information from online reviews for product recommendation," in ICWSM, 2015.
5. J. Lin, K. Sugiyama, M. Kan, and T. Chua, "Addressing cold-start in app recommendation: latent user models constructed from twitter followers," in SIGIR, 2013.

6. Steffen Rendle, "Social Network and Click-through Prediction with Factorization Machines" Social Network Analysis University of Konstanz 78457 Konstanz, Germany,2012.

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

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

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

10. V. A. Zeithaml, "The new demographics and market fragmentation," Journal of Marketing, vol. 49, pp. 64–75, 1985.

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

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

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

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

15. B. Hollerit, M. Kr¨ oll, and M. Strohmaier, "Towards linking buyers and sellers: Detecting commercial intent on twitter," in WWW Companion, 2013.