ISSN 2395-1621

Connecting Social Media to E-Commerce System

#1Prof. Milind Hegade, #2Shital Arjun Salke, #3Snehal Mohan Shinde, #4Priyanka Gautam More, #5Samruddhi Vinod Shinde

1milindrhegade@gmial.com
2shital.a.salke@gmail.com
3snehal.m.shinde@gmail.com
4pmore662@gmail.com
5shindesam30@gmail.com

#12345Department of Computer Engineering,
JSPM's Bhivarabai Sawant Institute Of Technology and Research (BSIOTR)
Savitiribai Phule Pune University India

ABSTRACT

In recent years, the border between Online Media and E-Commerce is diminishing. Almost every single person in a metropolitan daily uses both social media like Facebook, Twitter, etc. for networking and uses internet to make huge purchases using e-commerce sites like Flipkart, Amazon, etc. We often login to e-commerce websites using our social accounts like FB or G+. We can also share our recent purchase details on the social media using the links to the product pages of e-commerce sites. We are focusing on the product recommendation to the users on e-commerce sites by leveraging the information or knowledge gained from the users' social accounts. This will enable to assess the needs of the user in cold start situations. Cold Start is a state when user logs in to the e-commerce website for the first time and we don't have any information about the history of purchases, shopping trends, etc. as it is not yet created or available. When we have users social account information (no confidential information will be accessed) like posts, friends, shares, etc. then we can harness this to our benefit. For example, we will be applying data mining algorithms to access the micro-blogs the user has created and extract the useful keywords and hence this data from the micro-blogs becomes the basis for product recommendation in cold start situations.

Keywords: Cold start, Product Recommendation, E-commerce, Micro-blogs, Product Demography, Data mining, Information Search.

ARTICLE INFO

Article History

Received: 25th September 2016

Received in revised form :

25th September 2016

Accepted: 30th September 2016

Published online :

18th October 2016

I. INTRODUCTION

Now-a-days, product recommendation is a key area to focus for increased sales for any e-commerce website.For example, Netflix has re-leased an interesting fact that about 75% of its subscribers watchare from recommendations [2].There are many algorithms which focus on connecting the social media to e-commerce but none are focused on product recommendation by leveraging the social media information like demographic, micro-blogs, location [1], etc.

Recommender systems currently used, focus on solving the information overload problem, by providing users with personalized and accurate information services. Typically, recommendation systems which use collaborative filtering, can automatically predict the needof an active user by collecting rating information from other similar users oritems[3].

Another way of recommending products is based on online reviews a purchaser leaves after a purchase and has his/her feedback. The information from the product reviews can be used by analyzing the knowledge hidden in it. But, this technique cannot address the Cold Start situations when there are no purchases or very less purchases for a startup e-commerce website [4].

App recommender uses the information from the twitter followers of any app. There are millions of apps on Google Play store or Apple store, but if someone wants to use a certain app, then there are very rare chances that he/she will be able to find it if they don't know the name. Hundreds of similar products will line up and hence it is difficult to find the right app for the users' need. All of the apps has twitter accounts and hence the number of

followers, can be used to recommend the best app out of a confusing list. Here as well cold start problem cannot be addressed and hence we cannot use this technique for product recommendation [5].

II. LITERATURE SURVEY

[1] Steffen Rendle, "Social Network and Click-through Prediction with Factorization Machines", The two tasks of KDDCup 2012 are to predict the followersof a microblogger (track 1) and to predict the click-throughrate of ads (track 2).KDDCup 2012 [4] consists of two predictiontasks from the micro-blogging website. The first task is to predict which micro-blogger a user is following. For each user a set of recommendedmicro-blogs is given and the prediction task is to rank themicro-blogs of this set by the chances that the user followsa micro-blog. The two main variables in this problem are user and micro-blog which are variablesof large domain. Other than this, other information is available: like user attributes such as gender and age, the social network about followers/ followees and time information abouteach recommendation of a micro-blog. The second task isto predict the clickrate of ads given a user and aquery. The main variables in this prediction problem arethe ad, the user and the query which again are variables oflarge categorical domains. This task also includes additionalinformation, e.g. age, gender, query tokens, or the positionof an ad.

[2] Mi Zhang, Jie Tang, Xuchen Zhang, XiangyangXue, "Addressing Cold Start in Recommender Systems", In this paper the cold-start problem is addressed by proposing a context-aware semi-supervised co-training method.The method has several unique advantages over the standard recommendation techniques foraddressing the cold-start problem. First, it defines a fine-grainedcontext that is more accurate for modeling the user-item preference.Second, the method can naturally support supervised learning andsemi-supervised learning, which provides a flexible way to incorporate the unlabeled data.

[3] Hao Ma, Tom Chao Zhou, Michael R. Lyu, Irwin King, "Improving Recommender Systems by Incorporating SocialContextual Information", Here we consider recommender systems which are basedon collaborative filtering, a technique that automatically derives the interest of any user by collecting and analyzing rating information from other similar users oritems.

[4] Jinpeng Wang, Wayne Xin Zhao, Yulan He, Xiaoming Li, "Leveraging Product Adopter Information from Online Reviewsfor Product Recommendation", The availability of the hugeamount of online product feedbacks or reviews provides demographicinformation of product adopters from reviewdocuments. In this paper we extract product adopter information from onlinereviews. The extracted product adopters are then categorized into many demographic groups which can later be used for product recommendation.

[5] Jovian Lin, Kazunari Sugiyama, Min-Yen Kan, Tat-Seng Chua, "Addressing Cold-Start in App Recommendation:Latent User Models Constructed from Twitter Followers", Milions of mobile applications (apps) are available, but users have difficulty in identifying apps that are relevantto their interests. Earlier recommender method that depend on previoususer ratings (i.e., collaborative filtering, or CF) can address thisproblem for apps that have sufficient ratings from past users. Butfor newly released apps, CF does not have any user ratingsto base recommendations on, which leads to the cold-start problem. In this paper, a new method which uses twitter followers as a base for app recommendation, is used which can address the cold start situations.

III. RELATED WORK

Our work is mostly address the new trend of social commerce as electronic commerce and online social media. The infusions of new technologies on the connect users in their homes and workplaces, thus transforming social formations and business transactions. An in-depth study of the growth and success of a social commerce site was conducted. The investigation is finalized with a triad relational model which reflects socioeconomic life in the Internet today. The following three concepts work jointly to form a global community that has already started to take the place of traditional commerce and socialization: Web technology, E-commerce, and online social media. A discussion of the research findings indicates that social commerce networks are sustainable because of the various incentives given to users as they collaborate with others regardless of their identity and location. The focus of this paper is to increase understanding on quickly developing Web based social media and their subsequent effects on the emerging social commerce.

IV. EXISTING SYSTEM

Below are the challenges involved when there is an interaction between users on social media and e-commerce sites:

- Social networks are private and hence direct access may lead to negation by the users. This can be damage the social network platforms as well, as users might stop accessing the site to avoid access of their privacy. At the same time, brands cannot ignore a platform which provides access to zillions of inter-connected users.
- Main aim of brands' social media interaction should limit in customers and their retention.
- The other challenge is remain customer friendly during changing trends and competition. Consumers will not give much effort when they want to buy something online and this is more impacting for an new shopper who comes up on

a e-commerce 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.

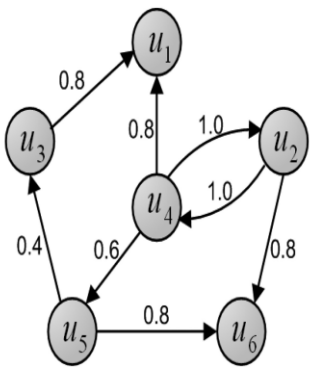
Our work mostly addresses the new trend of social commerce connecting social and e-commerce domains. A very deep study of the growth and success of a social commerce site was performed. The investigation is finalized to the use of micro-blogs to target the customers. The following three concepts work 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.

Now, we discuss some of the existing systems present for product recommendations before moving on to our proposed system.

Recommendation with Social Trust Network [3]

We first demonstrate our recommendation framework using a simple but illustrative toy example. Then we introduce the recommendation framework by factor analysis using probabilistic matrix factorization.

A Toy Example.



(a) social network graph

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	5	2		3		4		
u_2	4	3			5			
u_3	4		2				2	4
u_4								
u_5	5	1	2		4	3		
u_6	4	3		2	4		3	5

(b) user-item matrix

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	5	2	**2.5**	3	**4.8**	4	**2.2**	**4.8**
u_2	4	3	**2.4**	**2.9**	5	**4.1**	**2.6**	**4.7**
u_3	4	**1.7**	2	**3.2**	**3.9**	**3.0**	2	4
u_4	**4.8**	**2.1**	**2.7**	**2.6**	**4.7**	**3.8**	**2.4**	**4.9**
u_5	5	1	2	**3.4**	4	3	**1.5**	**4.6**
u_6	4	3	**2.9**	2	4	**3.4**	3	5

(c) predicted user-item matrix

Let us first consider the typical social trust network graph in Figure 1(a). There are 6 users in total (nodes, from u_1 to u_6) with 8 relations (edges) between users in this graph, and each relation is associated with a weight w_{ij} in the range [0, 1] to specify how much user ui knows or trusts user u_j. In an online social network Web site, the weight w_{ij} is often explicitly stated by user u_i. As illustrated in Figure 1(b), each user also rates some items (from i_1 to i_8) on a 5-point integer scale to express the extent of favor of each item. The problem we study in this article is how to predict the missing values of the user-item matrix effectively and efficiently by employing two different data sources. As mentioned in Section 1, motivated by the intuition that a user's social trust connections will affect this user's behaviors on the Web, we therefore factorize the social trust graph and user-item matrix simultaneously and seamlessly using UTZ and UTV, where the shared low-dimensional matrix U denotes the user latent feature space, Z is the factor matrix in the social network graph, and V represents the low-dimensional item latent feature space. If we use 5 dimensions to perform the matrix factorization for social recommendation, we obtain where U and V are the column vectors and denote the latent feature vectors of user u_i and item v_j, respectively.

Note that the solutions of **U and V** are not unique.

$$U = \begin{bmatrix} 1.55 & 1.22 & 0.37 & 0.81 & 0.62 & -0.01 \\ 0.36 & 0.91 & 1.21 & 0.39 & 1.10 & 0.25 \\ 0.59 & 0.20 & 0.14 & 0.83 & 0.27 & 1.51 \\ 0.39 & 1.33 & -0.43 & 0.70 & -0.90 & 0.68 \\ 1.05 & 0.11 & 0.17 & 1.18 & 1.81 & 0.40 \end{bmatrix},$$

$$V = \begin{bmatrix} 1.00 & -0.05 & -0.24 & 0.26 & 1.28 & 0.54 & -0.31 & 0.52 \\ 0.19 & -0.86 & -0.72 & 0.05 & 0.68 & 0.02 & -0.61 & 0.70 \\ 0.49 & 0.09 & -0.05 & -0.62 & 0.12 & 0.08 & 0.02 & 1.60 \\ -0.40 & 0.70 & 0.27 & -0.27 & 0.99 & 0.44 & 0.39 & 0.74 \\ 1.49 & -1.00 & 0.06 & 0.05 & 0.23 & 0.01 & -0.36 & 0.80 \end{bmatrix},$$

Recommendation using twitter followers [5]

For two months since its release, the Evernote app did not have any ratings. However, its Twitter account had active tweets and followers. This shows that despite the no-ratings on the app or we can say a cold-start, there is still

information present about the app, mainly on social networking services like Twitter.

V. PROPOSED SYSTEM

The boundary between e-commerce and social networking has become blurred. E-commerce websites such as Bay has many of the traits of social networks, including real-time updates and interaction between 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.

None of the e-commerce systems have 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. We are focused on the details of the microblogs, demographic information, location information, etc. to address the product recommendation. In this paper, we address the problem of recommending products to users who do not have any purchase records, i.e., in "cold-start" situations. We called it cold-start product recommender.

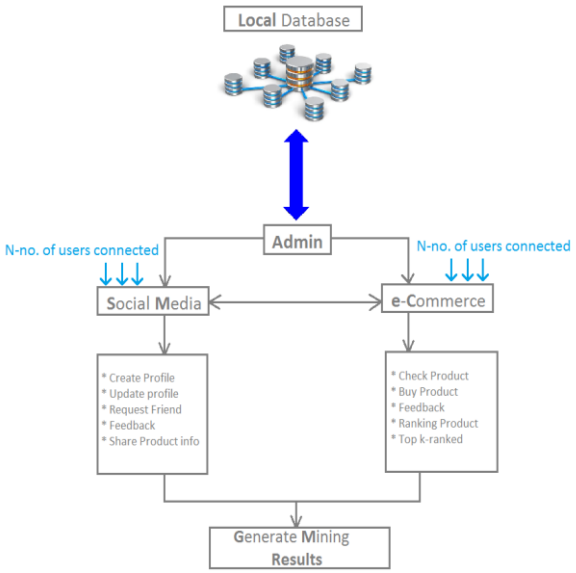
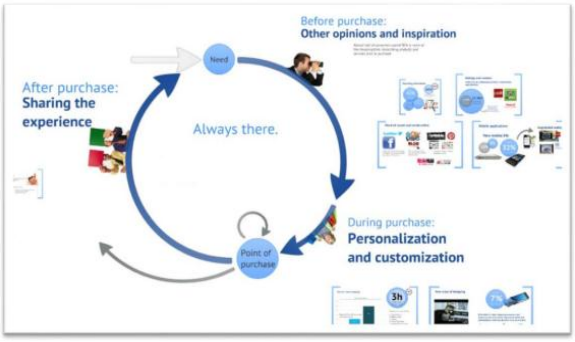


Fig 1. System architecture

The above fig 1 shows that combining the socio and ecommerce. This system gives the more accuracy for analysing the both technology. In this system user can user both website same location. If any user can purchases the any product from e-commerce website. But user use that product and he allow to give the review of the product, like how it is, how work functionality etc. so he can send review of the product. Once user send that review then that post is updated on social to recommendation friends.

Due to the heterogeneous type of the data in the social network posts, information extracted from micro-blogs cannot be used directly for product recommendation on e-commerce websites [7] [8] [9]. Therefore, one huge challenge is to transform users' micro-blogging information into another meaningful representation, which can be used more effectively for product recommendation.

This cycle on the right explains the typical cycle we are going to follow for recommending products to the customer. When customer logs in for the first time to the e-commerce site then his/her social media information is used like posts, age, gender, location, profession, etc. to suggest product in cold start. Later after the purchases this information can be posted on to their social media accounts which can attract more customers from his/her friend circle. This history of purchase can later be used in conjunction with the microblogs to suggest more effectively.



Below are the main steps involved in the data processing and analysis of microblogs to extract useful information and knowledge from the social media account of the user.

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.
- 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 microblogs. 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
Text Attributes [12] [13] [14]	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 [1]. 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.

VI. CONCLUSION

We study the new problem: how to recommend the right product at the right time? Experimental results on a data collected by a user e-commerce website show that it can predict a user's follow-up purchase behavior at a particular time with descent accuracy. Using a set of linked users across both e-commerce websites and social

networking sites as a bridge, we can learn feature prediction of multiple users.

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] H. 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] G. 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.

