International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 4 | Issue 2 | ISSN : 2456-3307

E-Commerce for Cold Start Product Suggestion Using Micro Blogging Data Through Connecting Social Media

G. Prashanti[1], Karlakunta Narendra[2]

ABSTRACT

In recent years, the boundaries between e-commerce and social networking became progressively blurred. Several e-commerce websites support the mechanism of social login wherever users will sign in the websites victimization their social network identities like their Facebook or Twitter accounts. Users also can post their fresh purchased product on microblogs with links to the e-commerce product websites. During this paper we have a tendency to propose a unique answer for cross-site cold-start product recommendation, that aims to advocate product from e-commerce websites to users at social networking sites in “coldstart” things, a haul that has seldom been explored before. A serious challenge is the way to leverage data extracted from social networking sites for cross-site cold-start product recommendation.We propose to use the coupled users across social networking sites and e-commerce websites (users United Nations agency have social networking accounts and have created purchases on e-commerce websites) as a bridge to map users’ social networking options to a different feature illustration for product recommendation. In specific, we have a tendency to propose learning each users’ and merchandises’ feature representations (called user embeddings and product embeddings, respectively) from information collected from e-commerce websites victimization continual neural networks so apply a changed gradient boosting trees methodology to remodel users’ social networking options into user embeddings. We have a tendency to then develop a feature-based matrix factorisation approach which might leverage the learnt user embeddings for cold-start product recommendation. Experimental results on an oversized dataset made from the biggest Chinese microblogging service SINA WEIBO and also the largest Chinese B2C e-commerce web site JINGDONG have shown the effectiveness of our planned framework.

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

I. INTRODUCTION

In these days, product recommendation is a very important area to concentrates in increased sales for any ecommerce website. For example, Netflix has re-leased an interesting fact that about 75% of its subscriber’s watches are from recommendations system. There are many algorithms which focus on connecting the social media to ecommerce but none are focused on product recommendation by leveraging the social media information like demographic, micro-blogs, location, 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 need of an active user by collecting rating information from other similar users or items.

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 start-up e-commerce website.

II. LITERATURE SURVEY

2.1Opportunity Models for E-commerce
Recommendation: Right Product, Right Time Author: Jian Wang, Yi Zhang This paper studies the new problem: how to recommend the right product at the right time? We adapt the proportional hazards modeling approach in survival analysis to the recommendation research field and propose a new opportunity model to explicitly incorporate time in an e-commerce recommender system. The new model estimates the joint probability of a user making a followup purchase of a particular product at a particular time.

This joint purchase probability can be leveraged by recommender systems in various scenarios, including the zero-query pull-based recommendation scenario (e.g. recommendation on an e-commerce web site) and a proactive push-based promotion scenario (e.g. email or text message based mar keting). We evaluate the opportunity modeling approach with multiple metrics.

Experimental results on a data collected by a real-world ecommerce website(shop.com) show that it can predict a user's follow-up purchase behavior at a time with descent accuracy. In addition, the opportunity model significantly improves the conversion rate in pull-based systems and the user satisfaction/utility in push-based systems

2.2 We Know What You Want to Buy: A Demographic-based System for Product Recommendation On Microblogs Author: Wayne Xin Zhao1, YanweiGuo Product recommender systems are often deployed by ecommerce websites to improve user experience and increase sales.

However, recommendation is limited by the product information hosted in those e-commerce sites and is only triggered when users are performing e-commerce activities. In this paper, we develop a novel product recommender system called METIS, a Merchant Intelligence Recommender System, which detects users' purchase intents from their microblogs in near real-time and makes product recommendation based on matching the users' demographic information extracted from their public profiles with product demographics learned from microblogs and online reviews. METIS distinguishes itself from traditional product recommender systems in the following aspects: 1) METIS was developed based on a microblogging service platform. As such, it is not limited by the information available in any specific e-commerce website. In addition, METIS is able to track users' purchase intents in near realtime and make recommendations accordingly. 2) In METIS, product recommendation is framed as a learning to rank problem. Users' characteristics extracted from their public profiles in microblogs and products' demographics learned from both online product reviews and microblogs are fed into learning to rank algorithms for product recommendation. We have evaluated our system in a large dataset crawled from Sina Weibo. The experimental results have verified the feasibility and effectiveness of our system. We have also made a demo version of our system publicly available and have implemented a live system which allows registered users to receive recommendations in real time

2.3 Retail Sales Prediction and Item Recommendations Using Customer Demographics at Store Level Author: Michael Giering

This paper outlines a retail sales prediction and product recommendation system that was implemented for a chain
of retail stores. The relative importance of consumer demographic characteristics for accurately modeling the sales of each customer type are derived and implemented in the model. Data consisted of daily sales information for 600 products at the store level, broken out over a set of non-overlapping customer types. A recommender system was built based on a fast-online thin Singular Value Decomposition. It is shown that modeling data at a finer level of detail by clustering across customer types and demographics yields improved performance compared to a single aggregate model built for the entire dataset. Details of the system implementation are described and practical issues that arise in such real-world applications are discussed. Preliminary results from test stores over a oneyear period indicate that the system resulted in significantly increased sales and improved efficiencies. A
brief overview of how the primary methods discussed here were extended to a much larger data set is given to confirm and illustrate the scalability of this approach.

III. PROPOSED SYSTEM

We propose to use the coupled users across social networking sites and e-commerce websites (users United Nations agency have social networking accounts and have created purchases on e-commerce websites) as a bridge to map users' social networking options to latent options 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 knowledge collected from ecommerce websites exploitation continual neural networks then apply a changed gradient boosting trees methodology to rework users' social networking options into user embeddings. We have a tendency to then develop a featurebased matrix factoring approach which might leverage the learnt user embeddings for cold-start product recommendation. It target text attribute, network attribute and temporal attribute

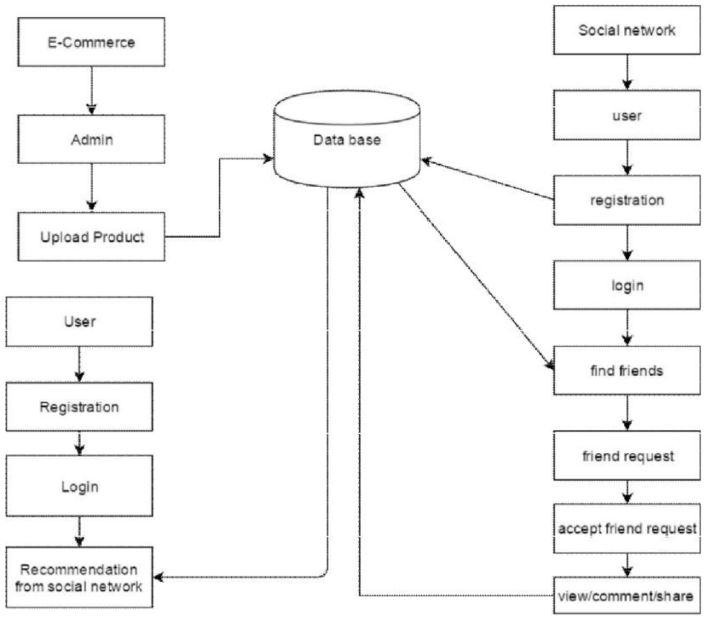

Figure 1. System Architecture

Advantages of Proposed System:

1. We have a tendency to propose a changed gradient boosting trees methodology to rework users' microblogging attributes to latent feature illustration which may be simply incorporated for product recommendation.

2. We have a tendency to propose and instantiate a feature-based matrix resolving approach by incorporating user and products options for cold-start product recommendation.

   1. The results show that our projected framework is so effective in addressing the cross-site cold-start product recommendation drawback.

IV. EXTRACTING AND REPRESENTING MICROBLOGGING ATTRIBUTES

Our solution to microblogging feature learning consists of three steps:

Prepare a list of potentially useful microblogging attributes and construct the microblogging feature vector au for each linked user u 2UL;

Generate distributed feature representations fvugu2U using the information from all the users U on the ecommerce website through deep learning; Learn the mapping function, f(au)->vu, which transforms the microblogging attribute infor mation auto the distributed feature representations v u in the second step. It utilizes the feature representation pairs {a u ,vu} of all the linked users u € U L as training data.

4.1 MICROBLOGGING FEATURE SELECTION

In this section, we study how to extract rich user information from microblogs to construct au for a microblogging user. We consider three groups of attributes.

4.1.1DEMOGRAPHIC ATTRIBUTES

A demographic profile (often shortened as "a demographic") of a user such as sex, age and education can be used by ecommerce companies to provide better personalized services. We extract users' demographic attributes from their public profiles on SINA WEIBO.

Demographic attributes have been shown to be very important in marketing, especially in product adoption for consumers. Following our previous study, we identify six major demographic attributes: gender, age, marital status, education, career and interests.

4.1.2 TEXT ATTRIBUTES

Recent studies have revealed that microblogs contain rich commercial intents of users. Also, users' microblogs often reflect their opinions and interests towards certain topics.

As such, we expect a potential correlation between text attributes and users' purchase preferences. We perform Chinese word segmentation and stop word removal before extracting two types of text attributes below. Topic distributions. Seroussi et al. proposed to extract topics from usergenerated text using the Latent Dirichlet Allocation (LDA) model for recommendation tasks.

Follow the same idea, we first aggregate all the microblogs by a user into a document, and then run the standard LDA to obtain the topic distributions for each user. The benefits of topics distributions over keywords are twofold. First, the number of topics is usually set to 50 200 in practice, which largely reduces the number of dimensions to work with. Second, topic models generate condense and meaningful semantic units, which are easier to interpret and understand than keywords. Word embeddings. Standard topic models assume individual words are exchangeable, which is essentially the same as the bag-of-words model assumption. Word representations or embeddings learned to use 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. We employ the Skip-gram model implemented by the tool word2vec4 to learn distributed representations of words. Finally, we average the word vectors of all the tokens in a user's published document as the user's embedding vector.

4.1.3 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 parse out latent user groups by the users' following patterns assuming that users in the same group share similar purchase preferences. Latent group preference. Since it is infeasible to consider all users on WEIBO and only keeping the top users with the most followers would potentially miss interesting information, we propose to use topic models to learn latent groups of following sasin [10].Wetreatafollowinguserasatokenandaggregateallt hefoll owingsofauserasanindividualdocument.In this way, we can extract latent user groups sharing similar interests (called "following topics"), and we represent each user as a preference distribution over these latent groups

4.1.4TEMPERAL ATTRIBUTES

Temporal activity patterns are also considered since they reflect the living habits and lifestyles of the microblogging 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. 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. We summarize all types of features in above **table**.

TABLE 1
Categorization of the Microblogging Features

Categories	Features
Demographic Attributes	Gender (2), Age (6), Marital status (10), Education (7), Career (9), Interests (6)
Text Attributes	Topic distributions (50), Word embeddings (50)
Network Attributes	Latent group preference (50)
Temporal Attributes	Daily activity distribution (24), Weekly activity distribution (7)

The number of feature dimensions are shown in parentheses.

V. 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 Skipgram 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 valuations, 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 ecommerce. 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.

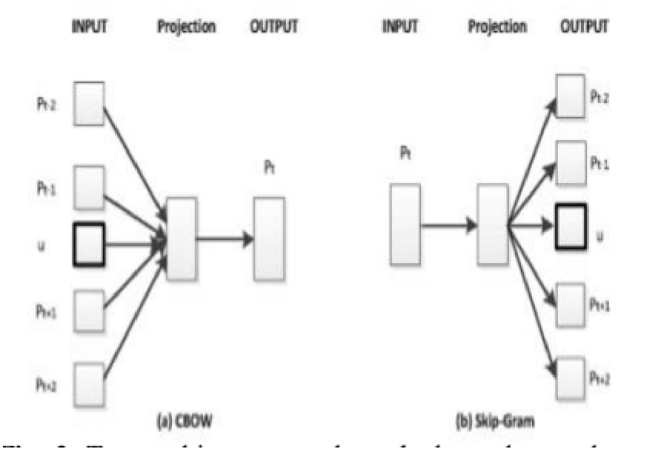

Figure 2. Two architectures to learn both product and user embeddings. Here u denote a user ID. The major difference between para2vec and word2vec lies in the incorporation of userID as additional context

VI. FUNCTIONAL MATRIX FACTORISATION

Now we consider constructing the interview process for cold-start collaborative filtering. Assume that a new user registers at the recommendation system and nothing is known about her. To capture the preferences of the user, the system initiates several interview questions to query the responses from the user. Based on the responses, the system constructs a profile for the user and provides recommendations accordingly. In the plain matrix factorization model described in Section 3.1, the user profile ui is estimated by optimizing the ℓ2 loss on the history ratings rij. This model does not directly apply to cold-start settings because no rating is observed for the new user prior to the interview process. To build user profiles adaptively according to the user's responses in the course of the interview process, we propose to parameterize the user profile ui in such a way that the profile ui is tied to user i's responses in the form of a function, thus the name functional matrix factorization (FMF). More precisely, assume there are P possible interview questions. We assume that an answer to a question takes value in the finite set {0,1, Unknown}, representing "Dislike", "Like" and "Unknown", respectively. Furthermore, let ai denote the P dimensional vector representing the answers of user i to the P questions. And we tie the profile to the answers by assuming ui = T(ai), where T is a function that maps the responses ai to the user profile ui Rk. To make recommendations for user i, we simply use rij = vT j T(ai).

Our goal is to learn both T and vj from the observed ratings K. To this end, substituting ui = T(ai) into the lowrank matrix factorization model, we have the following optimization problem:

$$T, V = \underset{T \in \mathcal{H}, V}{\operatorname{argmin}} \sum_{(i,j) \in O} (r_{ij} - v_j^T T(a_i))^2 + \lambda \|V\|^2, \quad (1)$$

where V = (v1, ...,vM) is the matrix of all item profiles, H is the space from which the function T(a) is selected an d the second term is the regularization term. Several issues need to be addressed in order to construct the interview process by the above functional matrix factorization. First, the number of all possible interview questions can be quite large (e.g. up to millions of items in movie recommendation); yet a user is only patient enough to answer a few interview questions. Second, the interview process should be adaptive to user's responses, in other words, a follow-up question should be selected based on the selection process should be efficient to generate interview questions in real time after the function T(a) is constructed. In addition, since we allow a user to choose "Unknown "to the interview questions, we need to deal with such missing values as well. Following prior works of [8,20], we use a ternary decision tree to represent T(a). Specifically, each node of the decision tree corresponds to an interview question and has three child nodes. When the user answers the interview question, the user is directed to one of its three child nodes according to her answer. As a result, each user follows a path from the root node to a leaf node during the interview process. A user profile is estimated at each leaf node based on the users' responses, i.e., T(a). The number of interview questions presented to any user is bounded by the depth of the decision tree, generally a small number determined by the system. Also, non-responses to a question can be handled easily in the decision tree with the introduction of a "Unknown" branch.

VII. CONCLUSIONS

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.

VIII. REFERENCES

[1]. J. Wang and Y. Zhang, "Opportunity model for E-commerce recommendation: Right product; right time," in Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2013, pp. 303– 312.

[2]. Opportunity Models for E-commerce Recommendation: Right Product, Right Time Jian Wang, Yi Zhang School of Engineering University of California Retail Sales Prediction and Item Recommendations Using Customer Demographics at Store Level Michael Giering

[3]. Connecting Social Media to E-Commerce: Cold-Start Product Recommendation Using Microblogging Information.Wayne Xin Zhao, Member, IEEE, Sui Li, Yulan He, Edward Y. Chang, Ji-Rong Wen, Senior Member, IEEE, and Xiaoming Li, Senior Member, IEEE

[4]. Connecting Social Media to ECommerce System. Prof. Milind Hegade, Shital Arjun Salke, Snehal Mohan Shinde, Priyanka Gautam More, Samruddhi Vinod Shinde

[5]. K. Zhou, S. Yang, and H. Zha, "Functional matrix factorizations for Cold-start recommendation," in Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2011, pp. 315–324

AUTHOR DETAILAS

G.PRASHANTI is working an assistant professor in VIGNAN'S LARA INSTITUTE OF TECHNOLOGY & SCIENCE.

Vadlamudi-522213 Guntur Dist.She has Experience in the teaching field For 7 years and her interested in research areas are network security Steganography and data mining.

KARLAKUNTA NARENDRA he is Currently pursuing MCA in MCA Department,Vignan's Lara Institute Of Technology&Science,Vadlamudi, Guntur, Andhra Pradesh, India.

He received his Bachelor of science from ANU