|| Volume 3 || Issue 3 || March 2018 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

Product Recommendation for Cold-Start User using Microblogging Information by Connecting Social Media to E-commerce

Sangale Sulochana[1], Dhokane Gokila[2], Gade Vidya[3], Tupe K. A.[4]
Student, Shri Chhatrapati Shivaji College of Engineering, Rahuri Factory Maharashtra ,India[1 2 3]
Assistant Professor Shri Chhatrapati Shivaji College of Engineering, Rahuri Factory Maharashtra ,India[4]
sangalesulochana133@gmail.com[1] ,gokiladhokane30@gmail.com[2] ,vidyagade156@gmail.com [3],Kapiltupe11@gmail.com[4]

Abstract— **Nowadays, 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. Social commerce is a subset of electronic commerce that involves social media, online media that supports social interaction and user contributions to assist online buying and selling of products and services. Anyone can login to e-commerce websites using their social accounts like Facebook or Google+. Also they share their recent purchase details on the social media using the sites to the product pages of e-commerce sites. Focus on the product recommendation to the users on e-commerce sites by leveraging the information or knowledge gained from the user's 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 ecommerce website for the first time and do not 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. Suppose, we will be applying data mining algorithms to access the microblogs the user has creating and extracting the useful information and hence this data from the microblogs becomes the basis for the product recommendation in cold-start situations.**

Keywords: *E-commerce, Social media, Product recommender, Product demographic, Microblogs, Cold-Start User, Information Search.*

I INTRODUCTION

In early days, social media, e-commerce site becomes popular because of increasing trend in people. Users typically access social media services via web-based technologies on desktop, computers, and laptops, or download services that offer social media functionality to their mobile devices. When starting with these services, users can create highly interactive platforms for individuals, communities and the organizations can share, create, discuss, generate and modify user-generated content posted online on social site. They discover large and universal changes to the communication between businesses, organizations, communities and individuals. Interesting problem about recommending products from e-commerce websites to users at social networking sites who do not have historical purchase records, such that in "cold-start" situations, called it as cross-site cold-start product and other recommendation. As a huge number of mobile applications (apps) are easily available, users have difficulty in Recognize apps that are applicable to their likes. Recommender systems that depend on existing user ratings (i.e., collaborative filtering, or CF) can situation this problem for apps that have enough ratings from existing users. But for applications that are newly free, collaborative filtering doesn't have any user ratings to base recommendations on, which leads to the cold-start situation problem. Although online product recommendation has been extensively studied before the most studies of only focus on the constructing solutions within certain e-commerce websites and mainly utilize the user's historical transaction records. To best of theour knowledge, cross-site cold-start product recommendation has been rarely studied before this. Individuals can likewise express their sentiments on different subjects of the hobbies. A wide mixture of subjects, extending from current occasions and political civil argument, to games and diversion, are in effect effectively talked about on these social discussions, for the instance, Facebook users could remark on or "like" campaign posted by an organization.

Twitter users could send tweets with a most extreme length of 140 characters to immediately impart and convey their insights on the games, motion pictures, and so forth. With the increasing popularity of the online e-commerce services, more and more people buy products online sites. As such, a large volume of online reviews have been constantly generated by the users. Since review data contain rich information about the user's feedback and opinions towards products they purchased, mining online reviews has attracted much interest. In existing system the problem is only the user's 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 the product recommendation. To address this informing, Linked user across social networking sites and the e-commerce sites as a bridge to map users' social networking features to the latent features for the product recommendation. In specific, learning both users and product's feature representations from the data collected from e-commerce websites. Next, how to extracting the microblogging information features and transform them into a distributive feature representation before the presenting a feature-based matrix factorization approach, which manufacture the learned distributive feature representations for the product recommendation. Prepare a list of the potentially useful microblogging attributes and construct the microblogging feature.

II RELATED WORK

J. Wang and Y. Zhang[1]:The majority of existing e-commerce recommender systems try to recommend the right product with a user, determined by whether or not the user is likely to purchase or like a product. Conversely, the potency of recommendations also depends upon the time of the recommendation. Allow us to have a user who just got a new laptop as one example. She may buy a replacement battery by 2 years (if the laptop's original battery often does not deal with that time) and buy a fresh laptop in another 24 months. In cases like this, it is not a good idea to recommend a fresh laptop or a replacement battery soon after that the user purchased the new laptop. It could possibly hurt the user's satisfaction in the recommender system if she receives a potentially appropriate product recommendation at the wrong time. We believe that something shouldn't only recommend probably the most relevant item, but additionally recommend at the perfect time. Anther than a single aggregate model built for the whole dataset. Information the system implementation are described and practical conditions arise in such real-world applications are discussed.

M. Giering[2]: A retail sales prediction and product recommendation system that has been implemented for a chain of stores. The relative need for consumer demographic characteristics for accurately modelling the sales of each one customer type are derived and implemented in the model. Data contained daily sales information for the 600 products at the store level and broken out over the some of the non-overlapping customer types. A recommender system was built based on a fast online thin Singular Value Decomposition. The modelling of the data at a finer level of details by the clustering across of customer types and demographics to improved that performance.

G. Linden, B. Smith, and J. York [3]:The recommender system used various algorithms are the better use on the e-commerce sites, this algorithm use as a input about a customer's interests and to generate a list of the recommendation product. Many applications use only the product that customers purchase and explicitly rate to represent their interests, but they can also use other attributes, including product viewed, demographic data, subject interests. At Amazon.com, we are use recommendation algorithms to personalize the online store for the each customer. The store radically changes based on the customer interests, showing programming titles to a software engineer and baby toys to a new mother. we are compare these methods with our algorithm, which we call product-to-product collaborative filtering.

V. A. Zeitham [4]: The underlying premise of this article is that exchange the demographics attribute will lead to a splintering of the markets for grocery products recommendation. A field study investigated the relationships between five demographic factors gender, female working status, education, interest, age, income, and marital status and a wide range of variables associated with preparation for and execution of supermarket shopping. Results indicate that the demographic groups of attribute various in significant ways from the traditional supermarket shopper. Discussion centers on the ways that changing demographics attribute and family roles may affect retailers and manufacturers of the grocery products recommendation.

W. X. Zhao, Y. Guo, Y. He, H. Jiang, Y. Wu.[5] : Product recommender systems are often deployed by e-commerce 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. they has develop a novel product recommender system called METIS, a MErchanT Intelligence recommender System, which detects users' can be purchase the product intents from their microblogging information real-time and makes product recommendation based on matching the users' demographic attribute information extracted from their public profiles with product demographics learned from microblogs and online user reviews. They following are: 1) METIS was developed based on a microblogging service platform, it is not limited by the information available in any of the specific e-commerce site. METIS is able to track users' purchase intents in near about real-time and make recommendations accordingly. 2) In METIS, product recommendation is framed as a learning to ranking the problem. Users' characteristics extracted from their public profiles in microblogging and products' demographics attribute information to be learned from both online product reviews and microblogging information are fed into learning to rank algorithms for product recommendation.

Jovian Lin, Kazunari Sugiyama, Min-Yen Kan, Tat-Seng Chua,[6] :Milions of mobile applications (apps) are available, but users have difficulty in identifying apps that are relevant to their interests. Earlier recommender method that depend on previous user ratings (i.e., collaborative filtering, or CF) can address this problem for apps that have sufficient ratings from past users. But for newly released applications, CF(i.e., collaborative filtering) does not have any user ratings to base recommendations on, which leads to the cold-start user problem. A new method which uses twitter followers as a base for application recommendation, is used which can address the cold- start situations.

Mi Zhang, Jie Tang, Xuchen Zhang, XiangyangXue[7]: The cold-start user problem is addressed by proposed a semi-supervised learning co-training method. The method has several unique advantages over the standard product recommendation techniques for addressing the cold-start user problem. First, it defines a fine-grained context that is more accurate and perfectly for modeling the user-item preference .Second, the method can naturally support supervised and semi-supervised learning algorithm, which provides a flexible way to incorporate the unlabeled data.

III PROPOSED SYSTEM

In proposed system having the more exactness for analysing the both technology. In this system user can use both websites at same location. If any user can purchase any product from ecommerce website, he/she can send review of the product on his/her social site. Once Cold Start Product Recommendation using Social Data and Micro-blogging Information user send that review then that post is updated on social site for product recommendation to his/her friends. In this project, we are going to create two websites namely social site and e-commerce site. No. of users are connected to both sites. Social site have functions like Create profile, Update profile, Sending friend request, giving feedback, and sharing the product information. E-commerce site also has features like Check product, Buy product, Feedback, Ranking the product. Mining the results from both sites user can get to know appropriate product recommendation and sell of e-commerce also get increased by receiving feedback from users. Fig. shows the architecture of proposed system for Social Commerce System. In order to recommend the products in the cold start situations, first we have to extract the attributes from the micro-blogging websites and transform them in to feature map representation to recommend the products. The process is explained step by step.

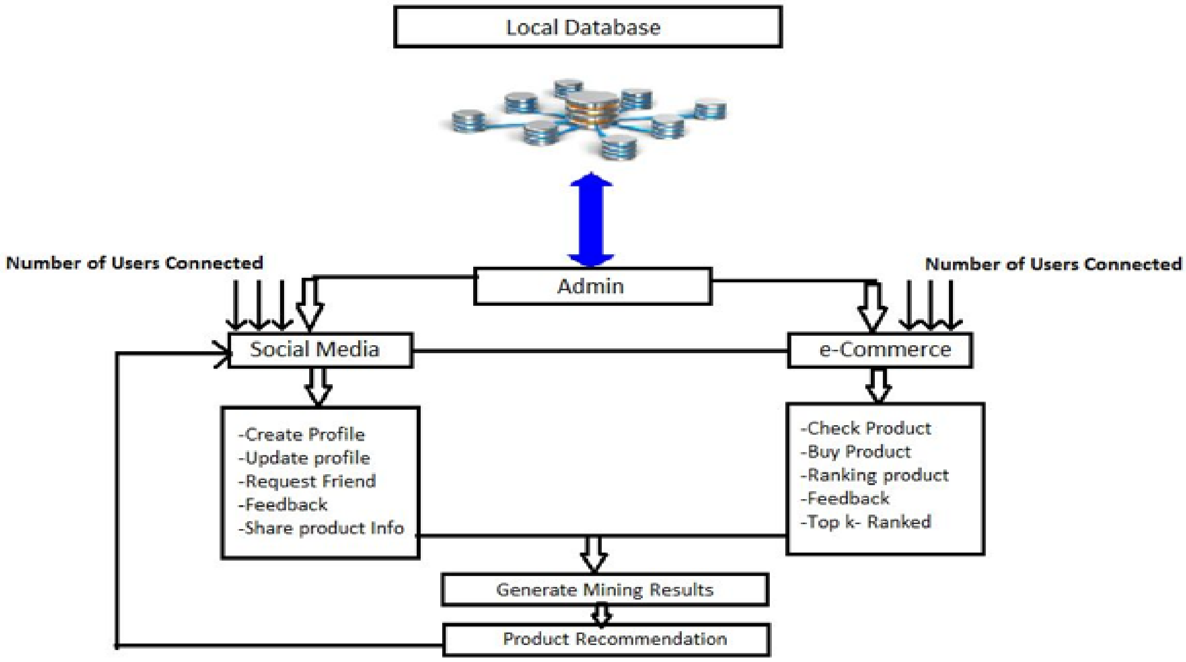


Figure 1 Proposed System Architecture

A. Micro-blogging Feature Selection

For a particular micro-blogging user now we will see how to extract information from the micro-blogging website. According to our knowledge the micro-blogging attributes are divided in to four categories. Demographic characteristic, text characteristic, and network characteristic, temporal characteristic a user can give the interests of the other user. A demographic profile of the user such as gender, marital status, career interests etc. can be utilized by the internet business organizations to give better customized administrations. To extracted the text attributes topic distributions, word embedding techniques can be useful. It is clear that users associated together with some links, hence extracting network attributes also used for product recommendations. The temporal attributes such daily activity and weekly activity distributions of a user can give the interests of the user, which can be helpful in the product recommendation.

1. Demographic Attributes:- A demographic account of a user, contain gender, age and Education can be used by ecommerce companies to give better individual services. 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: gender, age, marital status, education, career and interests.

2. Text Attributes:- Current studies have release that microblogs contain rich business intents of users. Also, user's microblogs often show their view and interests towards certain topics. As such, we expect a potential association between text attributes and users purchase priority. We perform Chinese word partitioning and stop word removal before remove two types of text attributes below.

Topic Distributions:-

To remove topics from user created text using the Latent Dirichlet Allocation (LDA) model for recommendation tasks. Follow the same idea, we first composite all the microblogs by a user into a document, and then run the standard LDA to obtain the topic distributions for each user. The advantages of topics transportation over keywords are double. First, the number of topics is usually set to 50 200 in practice, which largely decrease the number of dimensions to work with. Second, topic models generate compress and relevant semantic units, which are easier to interpret and understand than keywords.

- **Word embedding:-**

Standard topic models assume individual words are changeable, which is essentially the same as the bag-of-words model conclusion. Word representations or embedding learned to use neural language models help tackle the problem of traditional bag-of-word approximate which fail to catch words adjective. In word embedding, each measurement represents to an idle element of the word and semantically comparative words are close in the latent space.

3) Network Attributes:- In the online web-based social networking space, it is regularly watched that users associated with each other (e.g. through after connections) are probably going to have comparable interests. In that capacity, we can parse out latent user groups by the user's following examples accepting that users in a similar gathering share similar purchase interest. Since it is in feasible to consider all users just keeping the best users with the most followers would possibly miss interesting data, we propose to utilize subject models to learn latent groups. We regard a following user as a token and total every one of the followings of a user as an individual document. Along these lines, we can remove latent user groups having same interests (called "following topics"), and we represent to every user as an preference distribution over these latent groups.

4) Temporal Attributes:- Temporal activity designs are likewise considered since they mirror the living habits and ways of life of the microblogging user to some degree. In that capacity, there might exists relationships between temporal activities and user's purchase preferences.

Temporal activity distributions

We think about two kinds of temporal activity distributions, to be specific day by day activity distributions and week after week activity distributions. The day by day activity distributions of a user is described by an appropriation of 24 ratios, and the ith proportion shows the normal extent of tweets distributed inside the ith hour of a day by the user; correspondingly week after week activity distributions of a user is portrayed by a circulation of seven ratios, and the ith proportion demonstrates the normal extent of tweets distributed inside the ith day of seven days by the user.

IV PROPOSED SYSTEM FLOW DIAGRAM

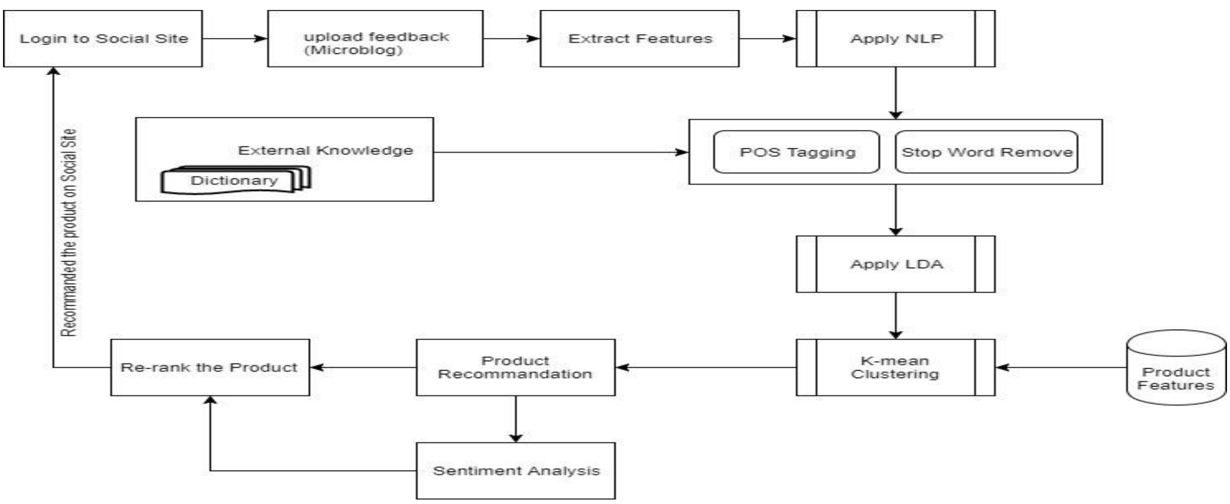


Figure 2: Proposed System Flow Diagram

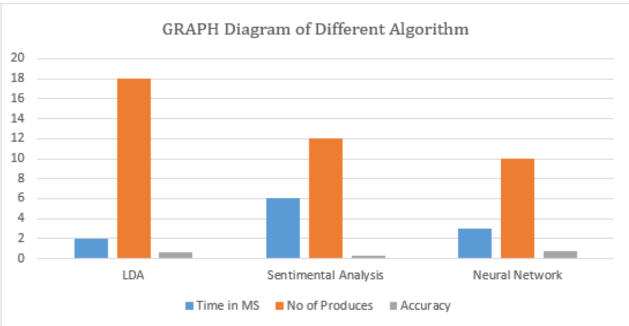
|| Volume 3 || Issue 3 || March 2018 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

V SYSTEM ANALYSIS

Result:

Table 1: Performance of Different Algorithm

Algorithm	Time in Ms	No. of produces	Accuracy
LDA	2	18	60%
Sentimental Analysis	6	12	30%
Neural Network	3	10	70%



Graph 1: Graph Diagram of Different Algorithm

System Requirements

1) Software Requirement:-

- Operating System : Microsoft Windows 7 /8
- IDE : Net Beans 8.2
- Programming Language : Java, HTML, CSS, JavaScript, Jsp
- Database : MySQL 5.5
- Application Server: Tomcat 5.0
- Web Server: Java web server
- Jdk:1.8

2) Hardware Requirement:-

- Processor : Intel core i3, i5
- Speed: 1.1 Ghz
- RAM : 4 GB (min)
- Hard Disk : 500 MB hard disk drive free space minimum (1 GB or more recommended)
- Keyboard : Standard Keyboard and Mouse

Mathematical Model:-

- **System S** = { I, O, C }
  Where, I = set of input
  O = set of outputs
  C = set of constraints
- **Input : I** = { Login, Request }
  Login = {Username, Password}
  Request = {Search product, search product by value, View comments, Apply rating and review, Mining the data, List apps, View History}
- **Users** = { User, Service provider }
- **Username** = { Username1, Username2, Username n }
- **Password** = { Password1, Password2, password n }
- **Output : O** = Show the right product at right time
- **Constraint : C** = User should login to the system
- **Success Conditions** : Success system when Correct mining result get from our system.
- **Failure Conditions** : Our system fails when no any result found to the given input.

Advantages:-

- Gain user data like what they are, what they like, and so forth which can change our business.
- Increase brand awareness i.e. targets more individuals to our internet business.
- Run user focused ads with real time outcomes.
- Generate significant leads i.e. change ad watcher to a user.
- Increase site activity and search ranking.
- Find out data about how contender is performing and change ourselves as per that.
- Share content quicker and simpler.

VI ALGORITHM

Algorithm 1 : K-means Clustering Algorithm

Let W = {w1,w2,w3,……..,wn } be the set of watches (data points) and

V = {Sony, FastTrack, sonata, titanic, quartz, citizen…….,n } be the set of company products (set of data centers).

Steps:

1. Randomly select the product of watches in cluster centers.
2. Calculate the distance between each watches from the cluster centers in product.
3. Assign the watches to the cluster center whose distance from the cluster center is minimum of all the cluster centers.
4. Recalculate the new cluster center using:
Mean of the existing cluster, that mean considered as a new cluster center of each cluster.
5. Recalculate the distance between each watches and new obtained cluster centers.
6. If no watches (data point) was reassigned then stop, otherwise repeat from step 3 because it was iterative process

B. Wu-Palmer algorithm:

The Wu & Palmer calculates relatedness by considering the depths of the two synsets in the WordNet taxonomies, along with the depth of the LCS (Least Common Subsumer).

The formula is

Score = 2 * depth (lcs) / (depth (s1) + depth (s2)).

This means that 0 < score <= 1. The score can never be zero because the depth of the LCS is never zero (the depth of the root of taxonomy is one). The score is one if the two input concepts are the same.

Steps:-

1. First , each sentence is partitioned into a list of tokens

|| Volume 3 || Issue 3 || March 2018 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

2. Part of speech for tagging
3. Stemming word
4. Find the most appropriate sense for every word in a sentence
5. Finally, compute the similarity of the sentence based on the similarity pairs of words

C. LDA:- Latent Dirichlet Allocation

1) For every document :

a) Randomly select distribution over topic (a multinomial of length K)

b) for each word in the document:

i) Probabilistically find one of the K topic from obtained distribution over topics (i) say topic βj

ii) Probabilistically find one of the V words from βj.

P(w/d)=p(w/z)p(z/d)

Where,

z=Set of life style i.e ”n” No. of habits.

w=Activity i.e daily work that we perform

d=Document.

VII CONCLUSION

The outcome of project has recommending products to e-commerce websites from microblogging users without historical purchase records. Our main proposal is cold-start user product recommendation, connecting cross site social media website and e-commerce website showed that it works on a data collected by a user from social media website and e-commerce website show that it can predicate a user's follow-up purchased behaviour at a particular time descent accuracy ,using a set of linked user across both e-commerce website and social media website as a bridge ,we can learn attributes of multiple user and recommend the product using microblogging.

REFERENCE

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

[2] M. Giering, “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, “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, “We know what you want to buy: a demographic-based system for product recommendation on microblogs,” in SIGKDD, 2014.

[6] 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.

[7] M. Zhang, J. Tang, X. Zhang, and X. Xue, “Addressing cold start in recommender systems: a semi-supervised co-training algorithm,” in SIGIR, 2014.

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

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

[10] H. Ma, T. C. Zhou, M. R. Lyu, and I. King, “Improving recommender systems by incorporating social contextual information,” ACM Trans. Inf. Syst., vol. 29, no. 2, 2011

[11] S. J. Pan and Q. Yang, “A survey on transfer learning,” IEEE TKDE, vol.22, no. 10, pp. 13451359, 2010.

[12] S. Rendle, “Factorization machines with libfm,” ACM Trans. Intell. Syst. Technol., vol. 3, no. 3, May 2012.

[13] B. Xiao and I. Benbasat, “E-commerce product recommendation agents: Use, characteristics, and impact.” MIS Quarterly, vol. 31, pp. 137209, 2007.

[14] Y. Shi, X. Zhao, J. Wang, M. Larson, and A. Hanjalic, “Adaptive diversification of recommendation results via latent factor portfolio,” in SIGIR, 2012.

[15] L. Hong, A. S. Doumith, and B. D. Davison, “Co-factorization machines: Modeling user interests and predicting individual Decisions in twitter,” in WSDM, 2013