Asian Journal of Engineering and Applied Technology (AJEAT)
Vol.3.No.1 2015 pp 6-12
available at: www.goniv.com
Paper Received : 08-04-2015
Paper Accepted: 20-04-2015
Paper Reviewed by: 1. R. Venkatakrishnan 2. R. Marimuthu
Editor : Prof. P.Muthukumar

MINING EMERGING SKILLS FROM SOCIAL NETWORK SOURCED BIG DATA SET

[1]I.JENITTO and [2]R.V. BELFIN
[2]*Asst. Professor*
Department of Computer Science and Engineering
Karunya University
Coimbatore, Tamil Nadu
Jonus1843@gmail.com

ABSTRACT

Collaborative filtering is one of the most important and popular technology for recommender systems. This methods predict the preferences of active users on items based on the preferences of other similar users or items. But collaborative filtering methods have certain demerits like data sparsity and big error in predictions .So in this paper an SVM classifier based recommendation method is introduced. A distinct feature of SVM based recommender system is that it recommends precise items based on the request made by the user. Here the SVM classifier is used to overcome problems found in collaborative based methods. SVM produce better performance compared to other methods. The objective of recommender system is to provide users with a series of personalized suggestions for certain types of items.SVM based recommendation system improves recommendation accuracy by7.25 percent (in terms of MAE) in employee data set. The result shows that the SVM based recommendation system has comparatively lower time cost and thereby outruns many CF recommendation methods.

Index terms **–CF,recommendation,SVM classifier**

I.INTRODUCTION

Collaborative filtering (CF) is a technique used by some recommender systems. In general, collaborative filtering is the process of filtering for information or patterns using techniques involving collaboration among data source, viewpoints, multiple agent etc. Typically involving very large data sets is one of the main application of collaborative filtering. Collaborative filtering methods have been applied to many different kinds of technologies including: sensing and monitoring data, mineral exploration, environmental sensing, and financial service institution; or in electronic commerce and web applications etc. In order to select an item that is more favourable to a user, sufficiently large amount of information is needed. Technologies enable us to easily get more information, especially on the Internet. For instance, if a user needs to rent a movie online, there are numerous choices available. However, too much information can lead to decision-making inefficient, leading to information overload.Personalizationtechnologies and recommender systems help to overcome this problem by providing personalized suggestions regarding which information is most relevant to users. Now a day's recommender systemis used in most of the online shopping sites and many other applications. The most popular examples include Netflix, which suggest movies, and Amazon.com, which suggests books, CDs, and various other products. If users offer their feedback on purchased or consumed items, the task of recommender systems is to predict user preferences for the yet unseen items based on user's prior feedback and activities and, subsequently, to recommend the item(s) with the highest estimated relevance to the user. Recommender system is responsible for providing users with a series of personalized suggestions for certain types of items. The recommendation methods are of three types. It is mainly classified into collaborative filtering (CF), content based (CB), and hybrid methods. Although in E-Commerce the recommendation methods are widely

used, a number of inadequacies have been found including:

Data sparsity. Data sparsity means it is having too many ratings so it is difficult to find correlation between user and item. Its main issue is in recommendation methods.

*Other problems.*Recommendation inaccuracy and big error in predictions is another major issue in recommendation methods. It mainly affects the recommendation accuracy. So in this paper a SVM classifier based recommendation method is introduced. Here the SVM classifier is used to overcome such problems. The mechanism of SVM based recommendations are as follows:

1. Cluster all items into several item groups.
2. Form a user group corresponding to each item group (i.e., a set of users who like items of a particular item group).
3. Build a user-typicality matrix and measure users' similarities based on users' typicality degrees in all user groups so as to select a set of "neighbors" of each user.
4. Predict the unknown rating of a user on an item based on the ratings of the "neighbors" of the user on the item using SVM CLASSIFIER.

SVM advantage:

1. SVM does not have the over-fitting problem for the high dimensional data.
2. SVM produce results in the form of binary data.
3. Better performance compared to other methods
4. Reduces noise.

Rest of the paper is organized as follows: Section 2 introduces the background and related work. The SVM-based CF method is introduced in detail in Section 3. In Section 4, we evaluate the proposed SVM-based CF method using the employee data set and compared it with previous methods. The differences among SVM and existing recommendation methods are discussed in Section 5. And the final Section 6 describes the conclusion.

2.BACKGROUND AND RELATED WORK

2.1 Prototype View and Typicality

Jing Wang et al. [1] used a transfer learning method for collaborative filtering. In this method the user feature subspace learned from the auxiliary data is transferred to the target domain and also this approach can overcome the sparsity problem. An iterative algorithm is also proposed for solving the optimization problem. This model mainly focus on transferring the user feature subspace to help improve the prediction performance in the target domain.

Jingyu Zhou et al. [2]used atwo-stage mining algorithm (GAUP). This approach used to mine the most influential nodes in a network. GAUP having two stages .first computes user preferences with a latent feature model based on SVD or a model based on vector space and Then to find top-K nodes in the second stage. Our result about the expert finding shows that GAUP performs better than the state-of-the-art greedy algorithm, SVD-based collaborative filtering, and HITS.

ChhaviRana et al. [3] used a Dynamic Recommender system (DRS).this approach is based on evolutionary clustering algorithm This clustering algorithm makes clusters of similar users and evolves them depicting accurate and relevant user preferences over time and also this approach to improve the quality of recommendations and computation time.

Thomas Verbraken et al. [4] proposed a Network-based classifiers, this approach tend to perform especially well for highly durable goods .The main contribution of this paper is that it shows that knowledge of a person's social network can be valuable to help predict the person's acceptance of the online channel for buying different products.

Wei Deng et al. [5] used a well-known SVD (Singular Value Decomposition) algorithm. This algorithm contain two contributions. First, they derive groups by overlapping the communities with the clusters, and feed them as implicit feedback to SVD++.after that to introduce a matrix, Called difference matrix. This approach is one of the important approach for time complexity. Meng-Lun Wu et al. [6] proposed a co-clustering with augmented matrices (CCAM). CCAM is based on information theoretic co-clustering but further considers augmented data matrices like user profile and item description. This approach can overcome both alleviate the data sparsity and simultaneously consider both user feature and item feature information.

Fernando Ortega et al. [7] proposed Pareto dominance to perform a pre-filtering process eliminating less representative users from the k-neighbour selection process while retaining the most promising ones. The method is applicable to all memory- based collaborative filtering recommender systems, is easy to implement and is effective in producing improved results.

Kenneth Wai-Ting Leung [8] proposed a Collaborative Location Recommendation (CLR). CLR is capable of generating more precise and refined recommendations to the users compared to the existing methods. This framework which makes location recommendations based on users' GPS trajectory data.

Lars Backstrom et al. [22] used a Supervised Random Walks, this approach combines the information from the network structure with node and edge level attributes. SRW is a new learning algorithm for link prediction and link recommendation. Supervised Random Walks can be applied to many problems such that anomaly detection, missing link etc. All these works mentioned above focuses on the developing ofthose methods that calculate the object typicality within the concepts. There has been no work on integrating typicality in collaborative filtering recommendation.

2.2 Recommendation system

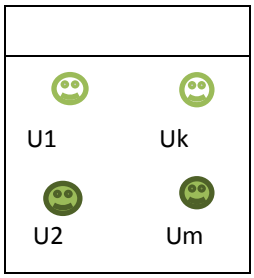
Recommender system is a subclass of information filtering system Recommendation methods are widely used in variety of applications.Recommendation system are of different types such as content based (CB), collaborative filtering (CF), and hybrid methods.

2.2.1 Content-Based Recommender Systems

The simplest approach to content-based recommendation is that recommendations are made by comparing the user profile with candidate items expressed in the same set of features.CBRSystems that recommend an item to a user based on the description of the item and the profile of user's interests. Content-based recommendation systems may be used in a variety of domains ranging from recommending web pages, news articles, restaurants, television programs, and items for sale. The description of items are analysed to identify interesting items for users in CB recommender systems. Another simplest approach to content-based recommendation is recommendations are made by comparing the user profile with candidate items expressed in the same set of features. The top-*k* best matched or most similar items are recommended to the user. Content-based recommendation is used to compute the similarity of the user profile with each item.

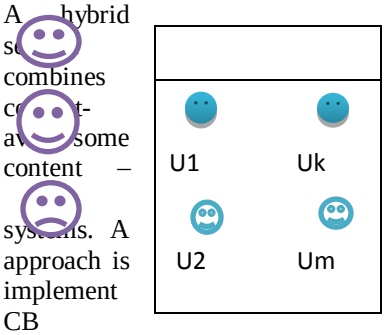
2.2.2 Collaborative Filtering

CF recommendation methods predict the preferences of active users on items, based on the preferences of other similar users or items. The reason that well-structured item not required descriptions is they are more for CF methods, implemented often methods, and than CB collaborative many systems are developed in industry and academia. There are two types of CF methods, such as user-based CF approach and item-based CF approach.



2.1.3 Hybrid Recommender Systems

A hybrid approach is used by recommender. It is combines collaborative and content-based methodsto avoid some limitations of content – based and collaborative systems. A naive hybrid approach is used separately to implement collaborative and CB methods.Then combine their predictions by a combining function, such as a linear combination of ratings, a voting scheme or other metrics.



2.3SVM classifier

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms This algorithm analyse data and recognize patterns, used for regression analysis and classification analysis. In 1992 The Support Vector Machine (SVM) which is a state-of-the-art classification method was introduced by Boser, Guyon, and Vapnik. SVMs belong to the general category of kernel methods. The SVM classifier is widely used (and other disciplines) due to its high accuracy, ability to work with high-dimensional data, ex. gene expression, and flexibility of data, high accuracy in Bioinformatics. SVMs belong to the general category of kernel methods.

A kernel method depends on the data only through dot-products. Kernel method is also an algorithm. When this is the case, the dot product can be replaced by a kernel function which computes a dot product in some possibly high dimensional feature space. This is having two advantages: First, the strength or knowledge to generate non-linear decision boundaries designed for linear classifiers using methods. Second, the use of kernel functions allows the user to apply a classifier to data that have no obvious fixed-dimensional vector space representation [20]. The prime example of such data in bioinformatics are sequence, and protein structure, either DNA or protein,SVM has been used successfully in many real-world problems

- Text (and hypertext) categorization
- Image classification
- Bioinformatics (Protein classification, Cancer classification)

3. SVM BASED RECOMMENDATION METHOD

1. Cluster all items into several item groups.

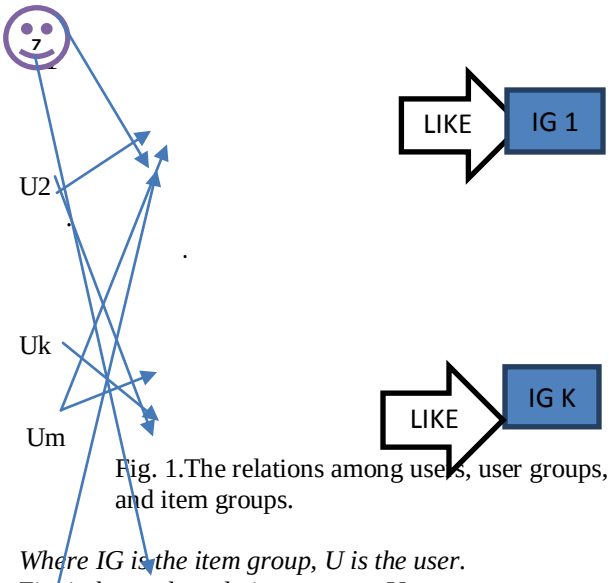


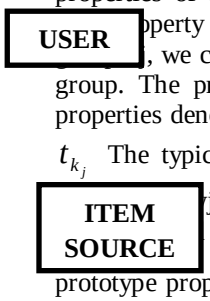
Fig. 1.The relations among users, user groups, and item groups.

Where IG is the item group, U is the user.

Fig 1 shows the relations among Users, user groups, and item groups.Users have different typical degrees in different user groups: the darker a user is in Fig.1, the more typical it is in that user groups. For examples, U1 and Uk are typical in user group IGk but not typical in ig1, while U2 and Um are typical in IG1 but not typical in IGk.

2. Form a user group corresponding to each item group (i.e., a set of users who like items of a particular item group).

For example, keywords, tittle, salary, and producers are properties of a skills and these properties can form an property vector to represent a skill. For each item , we can extract a prototype to represent the item group. The prototype of kj is represented by a set of properties denoted by the prototype property vector of kj, t_{k_j} The typicality of an item Oy in an item group kj, j;y, depends on the similarity between the the prototype of kj. where $\longrightarrow_{t_{k_t}}$ is the prototype property vector of item group kj, $\longrightarrow_{p_{o_y}}$ is the item property vector of Oy, and Sim is a similarity function.



$$w_{j,y} = Sim(\longrightarrow_{t_{k_j}}, \longrightarrow_{p_{o_y}})$$

3. Build a user-typicality matrix and measure users' similarities based on users' typicality degrees in all user groups so as to select a set of "neighbors" of each user [21].

The following function shows the similarity between prototype of a concept c and object:

$$\text{Sim: } P \times T \longrightarrow [0,1],$$

Where T is the set of all prototype vectors and P is the set of all object property vectors. For the dissimilarity between the unique prototype of a concept c and an object a in this method, it is defined as the complement of similarity, as follows:

Dissimilar

$$(\longrightarrow_{p_a}, \longrightarrow_{t_c}) = 1 - Sim(\longrightarrow_{p_a}, \longrightarrow_{t_c})$$

4.Predict the unknown rating of a user on an item based on the ratings of the "neighbors" of at user on the item using SVM CLASSIFIER.

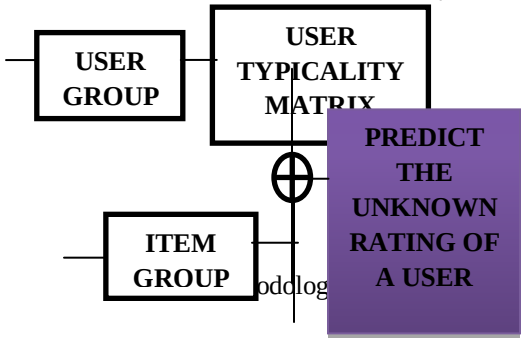
We select a fuzzy set of "neighbors" of user Uj, denoted by $\longrightarrow_{N_J}$, by choosing users who are sufficiently similar to Uj

$$\longrightarrow_{N_J} = \{U_i \mid Sim(U_i, U_j) \geq \gamma\}$$

Before prediction the neighbour selection is a very important step because the prediction ratings of an active user on items will be inaccurate if the selected neighbors are not sufficiently [10] similar to the active user. The distance between two users depends on matching of their corresponding properties. The similarity between Ui and Uj is measured as follows:

$$\text{Sim}(U_i, U_j) = \exp\left(-\sqrt{\sum_{y=1}^{n} |u_{i,y} - u_{i,y}|^2} \right)$$

Where n is the number of user groups, and $\sqrt{\sum_{y=1}^{n} |u_{i,y} - u_{j,y}|^2}$ t is the Euclidean distance between $\longrightarrow_{U_i}, \longrightarrow_{U_J}$



4.EXPERIMENTAL SETTING

4.1 Data Set Description

Employee skill data set has been widely used in previous papers such as [16].To evaluate recommendation method, employee skill data set is used in this experiment. From the employee skills data set 17,428 ratings are obtained which are assigned by 17, 428 users on1000 different skills. Each user has rated at least 3 skills. And data set contain employee details also.

4.2 Metrics

To measure statistical accuracy, the mean absolute error (MAE) metric is used which is the average absolute difference between predicted ratings and actual ratings [2]. MAE is a measure of deviation of recommendations from real user-rated ratings, and MAE is most commonly used and it is very easy to interpret. Lowervalue of MAE indicates that recommendation method can more accurately predict user rating.

4.3 Experiment Process

First, cluster all items into several item groups. Second, form a user group corresponding to each item group (i.e., a set of users who like items of a particular item group). Third, build a user-typicality matrix and measure users' similarities based on users' typicality degrees in all user groups so as to select a set of "neighbors" of each user. Then, finally predict the unknown rating of a user on an item based on the ratings of the "neighbors" of at user on the item using SVM classifier.

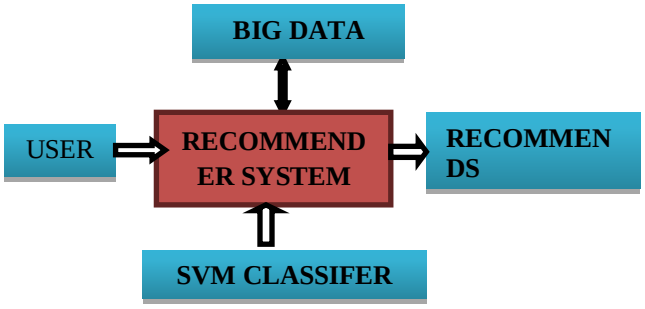


Fig. 3. Architecture

The proposed architecture provides a distinct methodology that can provide the exact needed information. For example Fresh graduates who are unaware of job opportunities can be guided with deeper knowledge on specializing courses that can lead them with better job opportunities.

1. User send the request to the enhanced recommender system regarding the required job skills
2. The recommender system search for the requested job skills in big data after that SVM classifier filters the related exact information [14].

5. EXPERIMENTAL RESULT

Fig 4. Result of SVM Based Recommendation

Using collaborative filtering based recommendation method large data can be retrieved from big data. However accuracy of the retrieved data is too low. So by using SVM based recommendation more accurate data can be retrieved.

Table 1. Recommended Skills Comparison

Recommendation Method	**Recommended Skills**	**Accuracy**
Without Using SVM	119	Low
With SVM	49	High

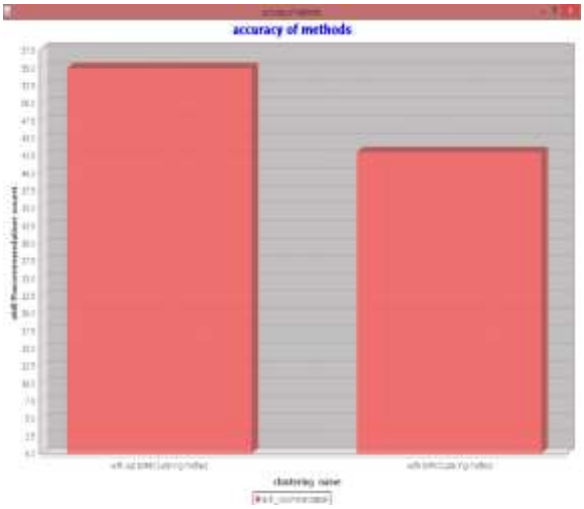


Fig 5. Performance Analysis

One skill (ex: CSE, IT) has been selected from the data set .After that perform SVM based recommendation. SVM retrieved related job skills from the data set. Accuracy is based on that result retrieved by the SRS (SVM- based recommendation SKILLS).if more recommendation are being made then the accuracy is low, otherwise accuracy is high.

CONCLUSION

Recommender system is responsible for providing users with a series of personalized suggestions for certain types of items. Typicality based recommender system can be act as a recommender system. In TYCO method sometimes there may be chances for filtering unwanted, insufficient information. So we use SVM classification to produce exact and actual information from big data.

REFERENCE

[1] Jing Wang n, LiangwenKe(2014), "Feature subspace transfer for collaborative filtering," ELSEVIER-Neurocomputing.

[2] Jingyu Zhou, Yunlong Zhang, JiaCheng(2014), "Preference-based mining of top-K influential nodes in social networks" ELSEVIER- Future Generation Computer Systems.

[3]ChhaviRana a, SanjayKumarJain(2014), " An evolutionary classification algorithm based on temporal features for dynamic recommender system" ELSEVIER-ChhaviRanaa,n, SanjayKumarJain.

[4]ThomasVerbrakenFrankGoethals ,WouterVerbeke , (2014) ," Predicting online channel acceptance with social network" ELSEVIER-Decision Support Systems.

[5] Wei Deng, RajvardhanPatil, LotfollahNajjar, Yong Shi, Zhengxin Chen (2014), "Incorporating Community Detection and Classification Techniques into Collaborative Filtering Model" ELSEVIER- Information Technology and Quantitative Management .*

[6] Meng-Lun Wu, Chia-Hui Chang , Rui-Zhe Liu(2014) "Integrating content-based filtering with collaborative filtering Using co-classification with augmented matrices" ELSEVIER- Expert Systems with Applications.

[7] O.Fernando, José-LuisSánchez, JesúsBobadilla, Abraham Gutiérrez,(2013) "Improving collaborative filtering-based recommender systems results using Pareto dominance," ELSEVIER- Information Sciences.

[8] Kenneth Wai-Ting Leung, DikLunLee(2011), "CLR: A Collaborative Location Recommendation Framework based on Co-Classification" ACM SIGIR Conf. Research and Development inInformation Retrieval (SIGIR '11).

[9] Y. Koren, R. Bell, and C. Volinsky(2008), " Functional matrix factorization for cold-start recommendation" Computer.

[10]H. Ma, C. Liu, I. King, and M.R. Lyu(2011), " Probabilistic Factor Models for Web Site Recommendation"Research and Development in Information Retrieval.

[11] Xiwang Yang a,YangGuo b, Yong Liu a, HaraldSteck(2014), "A survey of collaborative filtering based social recommender systems," ELSEVIER- Expert Systems with Applications.

[12] Tianqing Zhu, YongliRen,Wanlei Zhou(2014) ,"An effective privacy preserving algorithm for neighborhood-based collaborative filtering" ELSEVIER- Future Generation Computer Systems.

[13] Qian-Ming Zhanga, Ming-Sheng Shanga, Wei Zenga, Yong Chena, LinyuanL¨ub(2010),"Empirical comparison of local structural similarity indices forcollaborative-filtering-based recommender systems" ELSEVIER- Physics Procedia.

[14] Heung-Nam Kim a, AbdulmajeedAlkhaldia, Abdulmotaleb El Saddika,c, Geun-Sik Jo b (2011),"Collaborative user modeling with user-generated tags for social recommender systems,"ELSEVIER-Expert Systems with Applications.

[15] J. Bobadilla , F. Ortega, A. Hernando, A. Gutierrez(2013), "Recommender systems survey" ELSEVIER- Knowledge-Based Systems.

[16] M. RostamNiakanKalhori , M.H. FazelZarandi , I.B. Turksen(2014), "A new credibilistic classification algorithm" ELSEVIER- Information Sciences.

[17] JinohOh ,Sungchul Kim , Jinha Kim , Hwanjo Yu, "When to recommend: A new issue on TV show Recommendation" ELSEVIER- Information Sciences.

[18]RuzhiXua,b,ShuaiqiangWanga,b,XuweiZhengb(2014) ,"Distributed collaborative filtering with singular ratings for large scalerecommendation" The Journal of Systems and Software.

[19] N. Mehrbakhsh,I. Othman bin Ibrahim and I. Norafida(2014) , "Multi-criteria collaborative filtering with high accuracy using higher order singular value decomposition and Neuro-Fuzzy system," ELSEVIER-Knowledge-Based Systems.

[20] Szu-Yin Lina, Chin-HuiLaia, Chih-HengWub, Chi-Chun Lob(2014), "A trustworthy QoS-based collaborative filtering approach for webservice discovery," ELSEVIER-The Journal of Systems and Software.

[21] Le Hoang Son(2014), "HU-FCF: A hybrid user-based fuzzy collaborative filtering method in Recommender Systems," ELSEVIER- Expert Systems with Applications.

[22] Xiwang Yang a,YangGuo b, Yong Liu a, HaraldSteck(2014), "A survey of collaborative filtering based social recommender systems," ELSEVIER- Expert Systems with Applications.

[23] CihanKaleli(2014), "An entropy-based neighbor selection approach for collaborative filtering," ELSEVIER- Expert Systems with Applications.