PERSONALIZED RECOMMENDATION FOR COLD START USERS

Bhagya Ramesh[1], Reeba R[2]

[1]*P G scholar, Dept. of Computer Science And Engineering, Sree Buddha College of engineering, Pattoor*
[2]*Asst. professor, Dept. of Computer Science And Engineering, Sree Buddha College of engineering, Pattoor*

---***---

Abstract - *Recommendation is an information filtering process. The huge collection of data on the internet day by day leads to the development of methods to extract only the useful information. The methods and algorithms to extract this information are dealt in data mining. The process of recommendation is being utilized in e-commerce to improve the sales and to popularize the product. Personalizing the recommendation includes involving the factors such as user interest, interpersonal interest, and interpersonal interest similarity. Methods for recommendation are collaborative filtering, content-based method, and hybrid method. Matrix factorization techniques have been used in the different recommendation process. Trust is an important factor when considering the friend circle to do recommendation in a more personalized way. The different attacks aimed at recommendation can be classified on the basis of the attacker's intention to either promote the particular product or demote it.*

***Key Words*: Attacks, Collaborative filtering, content-based method, a hybrid method, Matrix factorization, personal interest, recommender system**.

1.INTRODUCTION

The recommendation is a process of information filtering. The process of recommendation is being utilized in e-commerce to improve the sales of the product. A recommender system is actually providing only useful information from the overwhelming information. The most commonly used method to do recommendation is collaborative recommendation [1] which involves the usage of the rating records of people who has similar characteristics when compared to the user. Content aware recommendation [1][2], it uses historical records of the user. Using the previously observed patterns predict the future interest of the user can be predicted. Next is the hybrid recommendation [3] which is a combination of the above two. Matrix factorization is also used in many ways to perform recommendation. Trust based systems are important for the personalized recommendation since the person trusted will always provide a useful recommendation to the user. The factor of trust is the measure of how the recommended product will help the user. If any attack occurs then the chance of disappointment to the user will be more. There can be different types of attack on the recommender system like random attacks, average attacks, probe attack, bandwagon attack and segment attack. The attacks on recommender system can be either to promote or demote the product

2. TYPES OF RECOMMENDATION

2.1 Content-Based Method

The content is used as the key source in which the purchase history of a particular user helps to identify the area of interest and thus proceed to provide further recommendations. It can be either the purchase history or the history of browsing which can include the documents, items etc. Based on these aspects interest of a particular user is being identified. There can be both structured as well as unstructured data in it. The use of Vector Space Model (VSM) [1] helps to deal with unstructured data. The cosine similarity is employed in VSM to identify the matching item for the user based upon the history. It is user independent since only the user's interest is being considered. Recommendations can be provided by listing content searched by the user that caused an item to be recommended. The problem of new item recommendation can be done which doesn't have a rating yet. It also has several shortcomings like limited content analysis because only the contents related to the individual is being considered. Only the recommendation based on the individual can be done.

2.2 Collaborative Filtering Technique

This technique is not user independent. It uses the related users to make the recommendation. The users who have a common interest are considered and based on this the related interests will be useful to make the recommendation to the user. The similarities between the users are found by the item rated by the user, purchase history etc. Here the network attributes are chosen in which the people belonging to one network is considered to share similar interest. They will be having many interests in common. The interests that are not common among them but it should be an interest of the person sharing a similar interest can be placed as a recommendation to the user. In this case, the recommendation of a new product can become difficult. Also, the users who don't share the common interest will not receive any recommendation. There are different collaborative filtering methods like user-user collaborative filtering, item-item collaborative filtering. The collaborative filtering[1] algorithm initially finds the nearest neighbors set for each user and then infers the like degree for an unseen item based on the nearest neighbors behavior. Sparsity problem is a limitation of this system. It explains the situation when the number of ratings obtained is usually very small compared to the number of ratings to be predicted. Grey sheep problem explains the problem related to the unusual user. The unusual users are those people whose opinions don't consistently agree or disagree with any group of people. Those who have high similarity with many other users and therefore it will be easy to find recommendations for that person, it falls under the category white sheep. CF systems require data from a large number of users before providing effective recommendation so scalability remains a major problem. Lack of transparency problem is one of the major issues since a user is given no indicators to consult in order to decide when to trust a recommendation and when to doubt one. There are mainly two types of data collection needed for building a user profile namely explicit data and implicit data [18]. Explicit data involves asking the user to rate an item on a scale and present two items to the user to choose the best one and ask a user to create a list of items that they like. Implicit data is found by observing the user views and analyzing item/user viewing items and also keeps a record of the items that the user purchased online. The matrix factorization for collaborative filtering is employed.

2.3 Hybrid Recommendation

This is a combination of both collaborative method and content-based method. They combine these methods to overcome the drawbacks in each recommendation system and to strengthen the model. The hybridization methods are weighted, switching, mixed, feature combination, cascade, feature augmentation and meta-level. Weighted depends on the score given to recommended item. Switching uses some methods to switch between different algorithms. In a mixed recommendation from different methods are presented at the same time. Feature combination includes a combination of recommendation features from different recommendation sources. In cascade method refining of the recommendation is done to get the perfect result. Feature augmentation uses the output of one algorithm as the input for the other. Meta-level focus on the model, the model learned by the recommender is used as input for the other. Also, there are other types of recommendation methods involved such as demographic recommender systems which consider the demographic attributes for the recommendation purpose. Another type is a knowledge-based recommendation system. The major challenge included in recommendation system is cold start problem which defines the lack of purchase history. Another problem is data sparsity [3] which involves the user who has less number of rating records, it includes people who are less active on the social sites. Next problem is scalability since the data on the internet is increasing day by day to the massive amount, the way in which the recommender system handle the data to extract useful information is a challenge. Since the major challenges in recommendation system are a cold start, data sparsity and scalability the use of probabilistic matrix factorization [6] is considered. The trust between the users is considered and in the case of sparse data a reasonable prediction is made. Based on the study of different recommender system [4] it has proved that the social relationship among the users is of great importance. The social factors like user interest, interpersonal interest similarity, and interpersonal influence in order to improve the accuracy of our recommendation system are considered. The factors of trust with another user, interest similarity with other user and interest on a particular set of the item is being considered here. The measures designed to evaluate the quality of the recommendations were considered while some focused on the interface issues like quality of recommendation, overall satisfaction with the recommendation and with online recommendation system, time measures, and interface issues.

3. TRUST BASED SYSTEM

Recommendation system based on trust and social circle have great differences [5]. The trust-based system is unilateral, when a person adds another person to his trust circle then he will be able to view his actions. In the case of the social circle, a relation confirmation process is involved. Also, users related in a social circle can have different interests. The low-rank matrix factorization is used to implement the recommendation more effectively which involves the user-item matrix. The methods used to implement the social regularization are average based regularization, individual based regularization, similarity functions. To overcome the problems of recommender system a model of social trust ensemble [8] was made which assumes that each user is independent, the user can be influenced easily by their trusted friends and user's final decision is based on users interest and trusted friends. In order to do social trust ensemble initially user feature through Bayesian inference have to be learned. Based on conditional distribution the social trust ensemble model is executed. Social ensemble involves the formulation of the social trust restriction on the recommendation system. A circle based recommender system [10] explains the group based on trust on different categories. For this, the expertise level of a user in a particular category is being measured based on their rating records. Trust value assignment can be based on equal trust or expertise based trust or trust split. Through this, the best subset of users for a particular category is obtained. Trust value assignment is a factor with which the models are being classified, it involves mainly 3 models circle con1, circle con 2, circle con 3. Mainly the memory based recommenders are using the concept of trust between the users. SocialMF [12] have been proposed which uses the trust concept in matrix factorization technique. In this model, the feature vector of the user is directly dependent on the feature vector of his neighbor. This improves the accuracy of the system. The concept of trust propagation is also considered here which improve the accuracy in the recommendation to the cold start user. The recommendation framework is based on probabilistic matrix factorization technique. The user adopts an item based on the aspects like what is the item, what item does the user like and who are the senders. The network data can be composed to produce information about the trust between two individuals without any direct connection. The properties of trust used in the development of the algorithm are transitivity and personalization [16]. Also, the asymmetry of trust is very important it can be unidirectional in many cases. The person can maintain two types of trust with a person that is trust in the person and trust in the person's recommendation of other people. To implement the trust inference the rounding algorithm and non-rounding algorithm are implemented. Trust mail is another work describing the trust factor. Trust score is calculated through the inference factor and propagation. The similarity measure between the users is weighting factor. To incorporate the trust into the recommendation process Resnick's prediction strategies is performed [17]. It involves trust based weighting and trust-based filtering. The computational model of trust involves the item level and profile level trust. The combination of these will help to filter the profile according to the trust value and trust value of highly trusted profiles are combined with profile similarity during prediction process.

4. MATRIX FACTORIZATION MODELS

The factors like individual preference and interpersonal influence are considered as social contextual factors [11]. A probabilistic matrix factorization has been used to include these factors into the recommendation. For an item, the user can rank it based on his particular interest or the user's strength of the relationship with the item sender. The user prefers the item based on what is the item, the item preferred by the user and who is the sender. Matrix factorization methods for recommender systems [22] considers the set of users, set of items and the set of partial ratings given for some users as well as some items, and outputs the items with top ratings for a selected user. This can be decomposed into sub-problems like finding the unknown ratings associated with users and items and sorting the ratings to select the top items. Bayesian Matrix Factorization includes factor analysis and simple probabilistic models. Bayesian Model would integrate or sum over all values of the parameters. To deal with the scalability problems in the recommendation system a probabilistic matrix factorization [13] is proposed. PMF scales linearly with the datasets and also deals with the problems like cold start and data sparsity problem. Two derivatives of the model are constrained PMF and PMF with adaptive priors. In constrained PMF an additional way of constraining user-specific feature vectors is that it has a strong effect on infrequent users. PMF with a learnable prior and constrained PMF are the two derivatives explained. These models can be efficiently trained and applied large datasets. Efficiency in PMF models comes from finding only point estimates of model parameters, instead of inferring the full posterior distribution over them. Content-based filtering and collaborative filtering are important strategies. Matrix factorization model [14] deals with the latent factors in RS. It characterizes both user and the item based on their rating behavior. This model is memory efficient and provides more accuracy. In order to avoid over-fitting probabilistic matrix factorization is used. The learning algorithm involved in matrix factorization is stochastic gradient descent approach and alternating least square approach. This technique is more flexible in handling different forms of data. Matrix factorization model maps both user and item to a joint latent factor space. Efficient retrieval of recommendation in a matrix factorization framework [19] proposes two efficient ways to do the searching process. A binary spatial partitioning metric tree and bound algorithm with a novel scheme are used to efficiently obtain the exact solution. Spherical clustering is used to index the users on the basis of their preference. Recommendations are pre-computed for the representative users of each cluster to obtain extremely efficient approximation solution. MF can be easily used to incorporate additional information. RoR in a trained MF model involves finding the set of k items for a user u with maximum predicted rating. The nearest neighbor problem in this can be rectified by using proper sensitive hashing. Metric trees are binary space –partitioning trees that are widely used for the task of indexing datasets.

5.RECOMMENDATION INVOLVING SOCIAL FACTORS

The interpersonal influence includes mainly three factors receiver interest, item quantities, and interpersonal influence. The first two items describe the behavior of the attributes of the receiver and the item. The last factor describes the relation between the sender and the receiver. In the social utility model, each recommendation is labeled accepted or refused. These factors give rise to the approach social utility [7]. The relation between the users is determined by the last factor while the former will describe the receiver and the item. This is different from the factor of interpersonal similarity. Social utility refers to the measure of the usefulness of the item recommended by the sender to the receiver. This system focus on how the receiver is willing to accept the social recommendation. In order to overcome the data prediction accuracy in rating, an integrated complementary model [9] was proposed that focus the patterns at different levels of local scale, high regional scale. At the local scale, the use of neighborhood-based technique will be more appropriate. This method poses some disadvantage like it doesn't account for the interaction among the neighborhood. It also points out the disadvantage of the neighborhood-based method as it is heuristic in nature and it doesn't account for interaction among the neighborhood. The components crucial for this model are incorporating interaction among user and neighborhood. Removal of global components and focusing on the residual helps to improve the system. To avoid over-fitting of parameters it uses shrinkage. It includes the local, neighborhood-based analysis into a regional factorization model. Solving recommendation problem by using social factors and user location [21] uses the bi-clustering and fusion methods to effectively reduce the cold start problem. This system will save the user location at the time of sign up i.e. user location is taken from the user when the user first-time register to the system. Once the user location is saved, the system will use this location to give location based recommendations to the user. The system will collect user's location data and user's interests from various locations are compared and recommendations are given to the users having similar interests in that particular location area. Also, it focuses on the three social factors like user personal interest which define the independent interest of a user, interpersonal influence, and interpersonal interest similarity.

6. ATTACKS ON RECOMMENDER SYSTEM

An Analysis of Attack Models and Algorithm Robustness [23] explain that the most discuss collaborative recommender system faces the problem of security. The profile injection attack is one of the major threats to security. An attacker interacts with the recommender system to build within it a number of profiles with the aim of altering the system's output. This attack can be categorized based on the knowledge required by the attacker to perform the attack, the intent of a particular attack, and the size of the attack. There are two types of knowledge-based attack like high-knowledge attack and low-knowledge attack. An attack against a collaborative filtering recommender system consists of a set of attack profiles; each contains altered rating data associated with some user and including target items. The item that the attacker wishes the system to recommend more highly is known as push attack and those wishes to prevent the system from recommending is known as nuke attack. Mounting a nuke attack is more complex than simply inverting the rating for the target item. In the case of push attacks, two models proposed are a random attack and average attack models. Chiron: A Robust Recommendation System with Graph Regularized [24] familiarize the shilling attacks in which attackers try to manipulate a recommendation system by introducing fake users, and subsequently fake ratings. The model developed is Chiron which is resistant to this attack. The push attacks are usually more successful than nuke attack.The method used for evaluating recommendation systems is Mean Absolute Error (MAE) which is used to measure accuracy in predicting ratings. Two important metrics that are used for evaluation of different shilling attacks are the attack size and the filler size. Another metrics used for evaluation of different shilling attacks is the filler size. The filler size is the set of items which are voted in the attacker's profile. The most effective attack models are derived by reverse engineering. The most common recommendation systems attack methods are random, average, and bandwagon. Bandwagon attack is similar to the random attack. Creating random ratings within an interval will allow the attacker to have a high impact in the decision-making process. Recommender Systems: Attack Types and Strategies [25] mainly focus on the attacks caused to the recommender system based on the domain knowledge. The set of items considered plays a major role in attack detection. The attack costs can be minimized by choosing popular items from the subdomain which include the items which have received many ratings from users. The advantages in choosing popular items when building attack profiles are the likelihood of a high number of rated items between the genuine and attack profiles is increased. It is desirable that each attack profile has a high probability of being located in the neighborhood of many genuine users, thereby minimizing the cost of attack in terms of a number of attack profiles that need to be created, and popular items tend to receive consistently and predictable. Probe Attack Strategy is utilized here. The rating strategy adopted is tailored to the particular ACF algorithm that is used. It makes an assumption that full knowledge concerning both item popularity and item likeability is available to attackers.

7. CONCLUSIONS

Different types of recommendation system like collaborative systems, content-based system, and hybrid system have been explained along with the drawbacks in a different system. The recommender systems are vulnerable to different types of attacks which can affect the recommendation process in a severe way. The different models with which we can implement the recommender system are matrix factorization models, trust-based models, recommendation involving social factors are considered.

ACKNOWLEDGEMENT

I am extremely grateful to my PG Coordinator Prof. Minu Lalitha Madhavu , Assistant Professor of Computer Science and Engineering department and my guide Prof. Reeba R, Assistant Professor of Computer Science and Engineering department, Sree Buddha College of Engineering, Alappuzha for her valuable guidance, timely suggestions and for providing all the vital facilities like providing the Internet facilities and important books, which were essential in the completion of this work . Finally, I would like to thank every individual who gave me even the slightest support to make my project a success.

REFERENCES

[1] Chuong C. Vo, Torab Torabi, Seng W. Loke "A Survey of Context-Aware Recommendation Systems "

[2] "Context-Aware Recommender Systems" Gediminas Adomavicius, Alexander Tuzhilin

[3] Gediminas Adomavicius Alexander Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions ", Member, IEEE, and Member, IEEE

[4] Rashmi Sinha and Kirsten Swearingen SIMS "Comparing Recommendations Made by Online Systems and Friends", University of California Berkeley, CA 94720.

[5] Chao Liu "Hao Ma "Recommender Systems with Social Regularization The Chinese University of Hong Kong Shatin, N.T .hk.Dengyong, Microsoft Research One Microsoft Way Redmond, WA 98052.

[6] Hao Ma, Haixuan Yang, Michael R. Lyu, Irwin King "SoRec: Social Recommendation Using Probabilistic Matrix Factorization " The Chinese University of Hong Kong Shatin, N.T., Hong Kong

[7] JunmingHuang1 and Xue-QiCheng1 and JiafengGuo1 and Hua-WeiShen1 and KunYang2 "Social Recommendation with Interpersonal Influence"

[8] Hao Ma, Irwin King and Michael R. "Learning to Recommend with Social Trust Ensemble" Liu Department of Computer Science and Engineering The Chinese University of Hong Kong Shatin, N.T.

[9] Robert M. Bell, Yehuda Koren and Chris Volinsky "Modeling Relationships at Multiple Scales to Improve Accuracy of Large Recommender Systems" AT&T Labs – Research 180 Park Ave, Florham Park, NJ 07932

[10] Xiwang Yang Yong Liu "Circle-based Recommendation in Online Social Networks". ECE Department Polytechnic Institute of NYU Brooklyn, New York x Bell Labs Alcatel-Lucent Murray Hill, New Jersey. ECE Department Polytechnic Institute of NYU Brooklyn, New York

[11] Meng Jiang†, Peng Cui Rui Liu, Qiang Yang, Fei " Social Contextual Recommendation". Wang‡ Wenwu Zhu Shiqiang Yang†Beijing Key Laboratory of Networked Multimedia, Tsinghua University, Beijing, ChinaHealthcare Transformation Group, IBM T, USA

[12] Mohsen Jamali Martin Ester "A Matrix Factorization Technique with Trust Propagation for Recommendation in Social Networks". School of Computing Science Simon Fraser University Burnaby, BC, Canada mohsen_. School of Computing Science Simon Fraser University Burnaby

[13] Ruslan Salakhut div and Andriy Mnih "Probabilistic Matrix Factorization". Department of Computer Science, UniversityofToronto 6 King's College

[14] Yehuda Koren Chris Volinsky "Matrix Factorization Techniques For Recommender Systems", Yahoo Research Robert Bell and, AT&T Labs

[15] .Xin Liu and Karl Aberer École "SoCo: A Social Network Aided Context-Aware Recommender System" Polytechnique Fédérale de Lausanne Batiment BC

[16] Jennifer Golbeck and James Hendler "Inferring Binary Trust Relationships in Web-Based Social Networks." University of Maryland, College Park

[17] John O'Donovan & Barry Smyth "Trust in Recommender Systems". Adaptive Information Cluster Department of Computer Science University College Dublin Belfield, Dublin 4 Ireland.

[18] Prodan "Implementation Of A Recommender System Using Collaborative Filtering". Andrei-Cristian.

[19] Noam Koenigstein Parikshit Ram "Efficient Retrieval of Recommendations in a Matrix Factorization Framework". School of Electrical Engineering Tel Aviv University. Computational Science & Engineering Georgia Institute of Technology. Yuval Shavitt School of Electrical Engineering Tel Aviv University

[20] YizeLi, JiazhongNie, YiZhang Baoshi Yan "Contextual Recommendation based on Text Mining "School of Engineering University of California Santa Cruz Bingqing Wang School of Computer Science Technology Fudan University., Fuliang Weng Research and Technology Center Robert Bosch LLC

[21] Sagar Kothimbire Vina M Lomte "Recommendation Problem Solving Using Social Factors and User Locations", Prof. PG Student, Dept. of Computer Engineering, Savitribai Phule Pune University, Pune, India Dept. of Computer Engineering, Warje, Savitribai Phule Pune University, Pune,

[22] Prof. Dr. Patrik Eklund Dr. Jerry Eriksson "Matrix Factorization Methods for Recommender Systems Shameem Ahamed Puthiya Parambath Master's Thesis in Computing Science,

[23] Bamshad Mobasher " Towards Trustworthy Recommender Systems: An Analysis of Attack Models and Algorithm Robustness", robin burke, runabhaumik,and chadwilliams Center for Web Intelligence School of Computer Science, Telecommunication and Information Systems DePaul University, Chicago,

[24] Regularizer Saber Shokat Fadaee1, Mohammad Sajjad Ghaemi2, Ravi Sundaram3, and Hossein Azari Soufiani "Chiron: A Robust Recommendation System with Graph "College of Computer and Information Science Northeastern University

[25] Michael P. O'Mahony and Neil J. Hurley and Gu´enol´ " Recommender Systems: Attack Types and Strategies "e C.M. Silvestre University College Dublin Belfield, Dublin 4 Ireland Michael