© 2018 IJSRSET | Volume 4 | Issue 9 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

Exploration of Users Rating on Reputed Items on Recommender Systems

Madhavi Darsinala[1], D. Varalakshmi[2]

[1]M.Tech Scholar, Department of CSE, NRI Institute of Technology Visadala (V&M), Guntur (Dt), Andhra Pradesh, India

[2]Assistant Professor, Department of CSE, NRI Institute of Technology Visadala (V&M), Guntur (Dt), Andhra Pradesh, India

ABSTRACT

Data mining is a subscription-based service where the networked storage space and computer resources can be obtained. Data mining economically enables the paradigm of data service outsourcing. However, to protect data privacy, sensitive DATA MINING data have to be encrypted before outsourced to the commercial public DATA MINING, which makes effective data utilization. In the proposed system, the problem of effective secure ranked keyword search over encrypted DATA MINING data is done. Ranked keyword search greatly enhances the system usability by returning the matching files in a ranked order. The existing technique resolves the optimization complexities in ranked keyword search and its effective utilization of remotely stored encrypted DATA MINING data. But it limits the further optimizations of the search results by preventing DATA MINING server to interact with DATA MINING users to maintain the integrity of actual owner's keyword and the data associated with it. The aim is to define a framework which enhances the accuracy of the ranked keyword search by secured machine learning, which does not affect the data integrity.

Keywords : Efficient Ranked Keyword Search, Search engine in DATA MINING, Security in Search engine, confidential data, searchable encryption.

I. INTRODUCTION

Data mining is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. However, data mining technology challenges many traditional approaches to datacenter and enterprise application design and management. The effectiveness and efficiency of traditional protection mechanisms are being reconsidered as the characteristics of this innovative deployment model can differ widely from those of traditional architectures. An alternative perspective on the topic of DATA MINING security is that this is but another, although quite broad, case of "applied security" and that similar security principles that apply in shared multi-user mainframe security models apply with DATA MINING security. Secure search over encrypted data has recently attracted the interest of many researchers. Song et al.[3] first define and solve the problem of secure search over encrypted data They propose the conception of searchable encryption, which is a cryptographic primitive that enables users to perform a keyword-based search on an encrypted dataset, just as on a plaintext dataset. Searchable encryption is further developed by [4], [5], [6], [7], [8]. However, these schemes are concerned mostly with single or Boolean keyword search. Extending these techniques for ranked multi keyword search will incur heavy computation and storage costs. Secure search over encrypted DATA MINING data is first defined by Wang et al. [9] and further developed by [10], [11], [12]. These researches not only reduce the computation and storage cost for secure keyword search over encrypted DATA MINING data. To protect data privacy, sensitive data has to be encrypted before outsourcing so as to provide end-to-end data confidentiality assurance in the DATA MINING and beyond. Thus, exploring privacy-preserving and effective search service over encrypted DATA MINING data is of paramount importance. Data owners may share their data with large number of on-demand data users and huge amount of outsourced data documents in DATA MINING, this problem is particularly challenging as it is extremely difficult to meet also the requirements of performance, system usability and scalability. On the one hand, to meet the effective data retrieval need, large amount of documents demand DATA MINING server to perform result relevance ranking, instead of returning undifferentiated result. Such ranked search system enables data users to find the most relevant information quickly. Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data. The other hand, to improve search result accuracy as well as enhance user searching experience, it is also crucial for such ranking system to support multiple keywords search, as single keyword search often yields far too coarse result. As a common practice indicated by today's web search engines (e.g., Google search), data users may tend to provide a set of keywords instead of only one as the indicator of their search interest to retrieve the most relevant data.

II. PROBLEM STATEMENT

In this Paper we identify three problems to deal with. The first two problems are related to reputation systems, while the third is related to the use of product reputation in recommender systems. The three problems are detailed here.

Problem 1: Sparse Dataset A sparse dataset is a dataset with the majority of its items having a small rating count. Datasets in many domains may have this characteristic as the nature of the domain does not attract users to provide feedback or it has a very large number of items. Reputation systems depend on historical feedback to generate reputation scores for items. When the available feedback is sparse it becomes more difficult to produce accurate reputation scores for items. How do we then use statistical data in the rating aggregation process to• enhance the accuracy of reputation scores over a sparse dataset?

Problem 2: Item Popularity The second problem we address here is the different popularity of items in the dataset. Most of the current reputation models pay insufficient attention to item popularity when generating reputation scores. An item is considered popular if it has a large rating count in relation to other items' count of ratings in the same dataset. The following example explains the problem in detail. Suppose we have two items in a dataset that contains ratings on the scale of [1 – 5], with the first item having the frequencies of each rating level of {2, 1, 3, 3, 1}, and the second item having the frequencies {10, 5, 15, 15, 5}. The first item has 10 ratings and the second item has 50 ratings, which indicates that the second item is more popular than the first. However, we notice that both items have the same proportional distribution of ratings over the rating levels. In such cases, the two items will have similar reputation scores using any reputation model that does not explicitly consider the number of ratings per item in the rating aggregation. A reputation score should reflect the popularity of an item; specifically, unpopular items are less likely to have high reputation scores. How do we then reflect the item popularity, presented by the count of• ratings of an item in the reputation calculation process? Problem

3: Reputation and Recommendation The third problem can be rewritten as follows: recommender systems, in general, are item reputation agnostic (O'Donovan & Smyth, 2005), which means that item reputation is not considered as part of the recommendation process and that a recommender system may produce a recommendation for an item with a low reputation score that is not likely to be consumed satisfactorily by the user.

III. LITERATURE REVIEW & RELATED WORK

In this section, we survey recent work related to our approach. Firstly, we review some approaches based on collaborative filtering (CF). Then, we review the often utilized rating prediction/recommendation methods based on matrix factorization. A. Collaborative Filtering Collaborative filtering (CF) is an important and popular technology for recommender systems. The task of CF is to predict user preferences for the unrated items, after which a list of most preferred items can be recommended to users. The methods are classified into user-based CF and item-based CF. The basic idea of user-based CF approach is to find out a set of users who have similar favour patterns to a given user (i.e., „neighbors"of the user) and recommend to the user those items that other users in the same set like, while the item-based CF approach aims to provide a user with the recommendation on an item based on the other items with high correlations (i.e., „neighbors" of the item). In all collaborative filtering methods, it is a significant step to find users" (or items") neighbors, that is, a set of similar users (or items). Currently, almost all CF methods measure users" similarity (or items" similarity) based on co-rated items of users (or common users of items). Collaborative filtering and content based filtering have been widely used to help users find out the most valuable information. B. Matrix Factorization based Approaches 1) Basic Matrix Factorization Matrix factorization is one of the most popular approaches for low-dimensional matrix decomposition. Matrix factorization based techniques have proven to be efficient in recommender systems when predicting user preferences from known user-item ratings. Matrix can be inferred by decomposing item reviews that users gave to the items. Matrix factorization methods have been proposed for social recommendation due to their efficiency to dealing with large datasets. several matrix factorization methods have been proposed for collaborative filtering. The matrix approximations all focus on representing the user-item rating matrix with low-dimensional latent vectors. 2) Social Recommendation In real life, people"s decision is often affected by friends" action or recommendation. How to utilize social information has been extensively studied. Yang et al. [6] propose the concept of "Trust Circles" in social network based on probabilistic matrix factorization. Jiang et al. [7] propose another important factor, the individual preference. some websites do not always offer structured information, and all of these methods do not leverage users" unstructured information, i.e. reviews, explicit social networks information is not always available and it is difficult to provide a good prediction for each user. For this problem the sentiment factor term is used to improve social recommendation. C. Reviews based Applications There are also many reviews based work for the task of recommendation. Wang et al. [1] propose a review rating prediction method by incorporating the social relations of a reviewer. In addition, they classify the social relations of reviewers into strong social relation and ordinary social relation. In addition, they classify the social relations of reviewers into strong social relation and ordinary social relation. Luo et al. [10] define and solve a new problem: aspect identification and rating, together with overall rating prediction in unrated reviews. D. Sentiment based Applications Sentiment analysis can be conducted on three different levels: review-level, sentence-level, and phrase-level. Review-level analysis, [2] and sentence-level analysis [11] attempt to classify the sentiment of a whole review to one of the predefined sentiment polarities, including positive, negative and sometimes neutral. While phrase-level analysis [3], attempt to extract the sentiment polarity of each feature that a user expresses his/her attitude to the specific feature of a specific product. There are many approaches leveraging sentiment analysis for personalized recommendation [4], [3], [5]. Zhang et al. [4] propose a self-supervised and lexicon-based sentiment classification approach to determine sentiment polarity of a review that contains both textual words and emotions. And they use sentiment for recommendation. By analyzing the user ratings, they can recommend special experts to a target user based on the user population. The information contained in user-service interactions can help predict friendship propagations and vice versa. They use data from both user item interactions and user-user relations.

IV. REPUTATION MODELS

Reputation has been defined as "what is generally said or believed about a person's or thing's character or standing" (Jøsang et al., 2007). Reputation is closely related to trust; however, it reflects global opinions and is not personalized, while trust is subjective and represents a measure between two agents (Bhuiyan, 2011; Yao Wang & Vassileva, 2007). Some work suggests that reputation represents a collection of trust values (Jøsang, Bhuiyan, Xu, & Cox, 2008; Yao Wang & Vassileva, 2003). Reputation systems are used for many objects, such as WebPages, items, services, and users, as well as in peer-to-peer networks. An item's reputation (product reputation) is calculated based on ratings given by many users through a specific aggregation method. Garcin, Falting, and Jurca (2009) describe ratings aggregators including mean, weighted mean, median, and mode. In general, we discuss five methods used in reputation models – the weighted mean, Bayesian models, fuzzy model, flow models, and probabilistic models.

V. RECOMMENDER SYSTEMS

Recommender systems nowadays represent an essential component of many websites. To understand why, we need to look at the definition of a recommender system and what its purpose is. Resnick and Varian (1997) suggest that recommender systems work similarly to the recommendations received from other people as word of mouth, where they are used to assist and augment this natural social process. Recommender systems are usually built based on two methods: collaborative filtering and content-based filtering, while a third hybrid method has emerged that combines both methods (Adomavicius & Tuzhilin, 2005). Before proceeding to recommender systems' approaches, we introduce users' profiling literature, which is considered the first stage to web personalisation and the input for recommender systems. 2.2.1 Users' Profiling Users' profiling has been widely used in recommender systems, and other personalization systems on the web, in order to understand users by collecting information about them that can be used to deliver personalized offerings. As defined by Zhou et al. (2012), "User profiling is the process of acquiring, extracting and representing the features of users." The profile can be used to present more relative content to each user and they usually contain users' basic information, such as age, gender, country, and keywords or concepts that represent users' interests. More sophisticated profiles may contain users' behavior information, such as the sequence of clicks and time spent on pages, which can also be useful in personalization. Recently, some researchers have suggested using users' social information in building users' profiles, such as social connections with other users, groups and pages, and social behaviors like shares, clicks, and likes or thumbs ups between users (Abel, Gao, Hoban, & Tao, 2011; Chen, Nairn, Nelson, Bernstein, & Chi, 2010; Tao, Abel, Gao, & Hoban, 2012). Social information is believed to be useful in enhancing many predictive results of different applications (Jawdat, Obeidat, & Aljanaby, 2011a;

Mezghani, Zayani, Amous, & Gargouri, 2012; L. Yu, Pan, & Li, 2011).

VI. REPUTATION SYSTEM COMPONENTS

People are increasingly dependent on information online in order to decide whether to trust a specific object or not. Reputation systems are therefore an essential part of any e-commerce or product review website, where they provide methods for collecting and aggregating users' ratings in order to calculate the overall reputation scores for products, users, or services (Resnick et al., 2000). The existence of reputation scores in these websites helps people making decisions about whether to buy a product or to use a service, or not. Reputation systems play a significant role in users' decision-making process. Reputation systems consist of three major components, as we illustrate in Figure 3.1 (Jøsang et al., 2007). The first component is the feedback collection from users. In this stage, reputation systems describe the methods used for collecting users' feedback; that is, centralized or distributed. They also describe what sort of feedback to be collected, such as user ratings, textual reviews, or critics' and experts' reviews. This stage may involve opinion-mining techniques to detect opinion polarity and strength in textual reviews and then represent them as numerical scores (Abdel Hafez & Xu, 2013a). The output of the feedback collection stage is a set of ratings towards items to be used in the reputation engine for generating reputation scores. The second component is the reputation engine, where the reputation system explains how the users' feedback on an item will be converted into a reputation score for that item. This stage is explained in more detail in Figure 3.2. The final component for reputation systems is the reputation presentation. The most popular presentation method is the five-star rating. Other methods include a percentage of 100, or a floating point score. Apart from numerical presentation of reputation scores, some methods proposed use icons that are easier for users to read, such as the Rotten Tomatoes method.

VII. DETAILED EXAMPLE AND DISCUSSION

This chapter provides a thorough evaluation and analysis of the proposed normal distribution-based reputation model. In this section we provide a detailed example to analyze and understand the different behaviors of several ratings aggregators. The chosen aggregation methods include the arithmetic mean, median, and harmonic mean calculated using Equation (3.8), geometric mean calculated using Equation (3.9), quadratic mean calculated using Equation (3.10), the trimmed-mean, which is the arithmetic mean of values after a certain proportion of the highest and lowest values have been discarded, with the trim value of 0.4, a fuzzy model (Bharadwaj & Al-Shamri, 2009), and the Dirichlet reputation system (Jøsang & Haller, 2007). Other reputation models, such as IMDb 10 , Trusti (Yingjie Wang et al., 2015), and PerContRep (Yan et al., 2014), use other data related to the number of ratings for other items in the dataset. Therefore, we include two aggregated values for each one of these methods, with two different assumptions about the dataset-related factors. In the first we assume that this item has a relatively large number of ratings compared with other items in the dataset. The second assumption is that the item has a low number of ratings. $Hm = n \sum 1\ ri\ n\ i=1$ (3.8) $Gm = (\prod ri\ n\ i=1\ )\ 1\ n$ (3.9) $Qm = \sqrt{}\ 1\ n \times \sum ri\ 2\ n\ i=1$ (3.10).

In Table 3.2 we provided the results for several aggregation methods, which are used to aggregate 10 ratings. In this table we provided 8 different ratings occurrences in order to see how the reputation scores change with different rating values. All the aggregation methods implemented in this table do not include a user's related data, ratings tendency, or time factor. In the table we highlight different aggregators with similar colors in order to emphasize a similar changing trend in reputation score when compared with the arithmetic mean method. The methods with unique colors have unique trends. In Table 3.2 the harmonic mean and geometric mean provide lower scores than the arithmetic mean in all of the illustrated cases. The PerContRep method produces similar behavior to the harmonic mean and geometric mean methods, where it generates scores lower than the arithmetic mean in all cases. This behavior is justified, as the PerContRep is a discount method that punishes reputation scores based on the number of ratings of an item compared with the total number of users. The PerContRep is calculated using Equations (2.13), (2.14), and (2.15). We use 10 http://www.imdb.com/chart/top 59 σ = 5 as a scale parameter and K = {10,100} as the total number of users. When the number of users is larger, such as K = 100, the PerContRep method generates even lower reputation scores. This method does not consider the uncertainty factor and does not employ rating distribution in its aggregation process.

VIII. CONCLUSION

In this paper, a recommendation model is proposed by mining sentiment information from social users" reviews. We propose social users sentiment measurement approaches based on the mined sentiment words and sentiment degree words from users" reviews. We fuse user sentiment similarity, interpersonal sentiment influence, and item reputation similarity into a unified matrix factorization framework to achieve the rating prediction task. This rating product recommendation system which can be also used as the social relation collaboration model which can be used to identify the social relation between the users. The genuine reviews will give the rating prediction easy and user will easily get the result in the desired time. This prediction based on rating also decides the products or items purchasing quality whether it is good or bad. This recommendation will also help us to identify the products reputation on the basis of good and bad reviews. This type of accurate recommendation system can be used to identify the items or products on mobile also. Also the products or items which do not have any type of rating or do not have any reviews will also be recommended to the person or user for the good decision. This use will help the users to do their work in less time and with good quality products for healthy life.

IX. REFERENCES

[1]. B. Wang, Y. Min, Y. Huang, X. Li, F. Wu, "Review rating prediction based on the content and weighting strong social relation of reviewers," in Proceedings of the 2013 international workshop of Mining unstructured big data using natural language processing, ACM. 2013, pp. 23-30.

[2]. D. Tang, Q. Bing, T. Liu, "Learning semantic representations of users and products for document level sentiment classification," in Proc. 53th Annual Meeting of the Association for Computational Linguistics and the 7thInternational Joint Conference on Natural Language Processing, Beijing, China, July 26-31, 2015, pp. 1014–1023.

[3]. Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, S. Ma, "Explicit factor models for explainable recommendation based on phrase-level sentiment analysis," in proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, 2014.

[4]. W. Zhang, G. Ding, L. Chen, C. Li , and C. Zhang, "Generating virtual ratings from Chinese reviews to augment online recommendations," ACM TIST, vol.4, no.1. 2013, pp. 1-17.

[5]. X. Lei, and X. Qian, "Rating prediction via exploring service reputation," 2015 IEEE 17th International Workshop on Multimedia

[6]. X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks, " in Proc. 18th ACM SIGKDD Int. Conf. KDD, New York, NY, USA, Aug. 2012, pp. 1267–1275.

[7]. M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W. Zhu, and S. Yang, "Social contextual recommendation," in proc. 21st ACM Int. CIKM, 2012, pp. 45-54.

[8]. Z. Fu, X. Sun, Q. Liu, et al., "Achieving Efficient DATA MINING Search Services: Multi-Keyword Ranked Search over Encrypted DATA MINING Data Supporting Parallel Computing," IEICE Transactions on Communications, 2015, 98(1):190-200.

[9]. Y. Ren, J. Shen, J. Wang, J. Han, and S. Lee, "Mutual Verifiable Provable Data Auditing in Public DATA MINING Storage," Journal of Internet Technology, vol. 16, no. 2, 2015, pp. 317-323.

[10]. W. Luo, F. Zhuang, X. Cheng, Q. H, Z. Shi, "Ratable aspects over sentiments: predicting ratings for unrated reviews," IEEE International Conference on Data Mining (ICDM), 2014, pp. 380-389.

[11]. T. Nakagawa, K. Inui, and S. Kurohashi, "Dependency tree-based sentiment classification using CRFs with Hidden Variables," NAACL, 2010, pp.786-794.

[12]. Xiaojiang Lei, Xueming Qian, Member, IEEE, and Guoshuai Zhao, "Rating Prediction based on Social Sentiment from Textual Reviews," IEEE Transactions On Multimedia, MANUSCRIPT ID: MM-006446