Research on System of Data Mining Technology Based on Computer

Yuan Zhu¹⁺, Jingjie Chen²

¹School of Trade and Economics Taizhou Vocational and Technical College Zhejiang, China
²No.2 Vocational and Technical School of Jiaojiang District, Zhejiang, China

*Corresponding author e-mail: tzzy246@163.com

Abstract. The existing implicit feedback collaborative algorithm directly uses sparse binary social trust information to assist recommendation, which has serious data sparse problems and does not have the influence of deep integration of social trust information. In response to the above problems, the paper proposes an algorithm that uses a denoising autoencoder to deeply integrate user implicit feedback data and social information. At the same time, a cloud computing-based social trust information data mining system architecture is constructed, and its modules are described. This framework is conducive to solving the problems of high data complexity and large amount of data encountered in current data mining, and can significantly improve the performance of data mining.

Keywords: Computer, data mining technology, system analysis, analysis performance, noise reduction autoencoder, social data.

1. Introduction

Social network (CN) is an emerging field that was inspired by social radio (CR) technology and was only proposed after 2000. Its core idea is to enable the network to perceive changes in the internal and external environment, adjust the configuration of the network in real time, and adapt dynamically and intelligently. The external environment changes and guides the autonomous decision-making of the network. Due to its higher intelligence and autonomy, it has shown great advantages in the fields of end-to-end performance and QoS guarantee, and has become an important trend in future network development [1]. However, since social networking is an emerging field, the current research on its QoS guarantee is still in the initial stage. Due to the randomness of network node access and the dynamic nature of the load, the QoS guarantee for real-time interactive streaming media services on social networks is still huge challenges.

Most of the existing recommendation algorithms use explicit rating information for recommendation. In fact, most users on the platform only generate implicit interactive information, such as user browsing and clicking, which makes traditional recommendation algorithms based on rating predictions unsatisfactory. The need for this type of platform. In recent years, recommendation algorithms based on users' implicit historical feedback have received extensive attention from the academic community. Although existing research has proposed a variety of algorithms that utilize
implicit user interaction information for recommendation, these algorithms still have two key problems. First of all, most of these works use binary social trust data, making the algorithm vulnerable to data sparseness. Secondly, existing research has not integrated user implicit feedback information and social information at a deep level. Based on this, this paper proposes an algorithm for comprehensively mining social trust relationships, and uses deep learning to fully measure the impact of social data [2]. At the same time, the paper proposes a social trust information data mining system architecture adapted to cloud computing. On the one hand, it conforms to the current development trend of cloud computing, on the other hand, it can also solve the problems encountered in the actual operation of data mining to a certain extent.

2. Implicit feedback recommendation algorithm based on denoising autoencoder

This paper comprehensively considers the trust data of user trust and user trust, respectively calculates the corresponding user trust to obtain a more accurate user trust value, and finally uses the noise reduction autoencoder to mine the implicit correlation information of the trust data to further improve top-N recommended quality.

2.1. Matrix factorization

The paper assumes that the trust matrix extracted from a social network with m nodes is $T = [T_{ik}]_{m \times m}$, where $T_{ik}$ represents the trust degree between user i and user k, which is usually binary information obtained from the actual data set. In terms of using implicit trust information of users, related scholars have proposed some efficient methods. Some scholars use matrix factorization technology to decompose the user trust matrix, calculate the user's social trust characteristics as a whole, and effectively mine the implicit trust relationship between users, but they have not further used the user's social trust characteristics to calculate the trust similarity of users. Some scholars have further explored the user's trust characteristics from the perspective of user trust and trust, but also did not use these characteristics to calculate more accurate user trust similarity [3]. In order to mine the user's implicit trust information from the overall trust relationship, this paper, based on previous work, makes full use of user trust and trust characteristics to calculate the implicit user trust similarity, and uses this to characterize the trust strength between users.

We use matrix decomposition to decompose each user i in the trust matrix into two different potential feature vectors, and use $B_i$ and $W_i$ to represent the trust feature vector of user i and the trusted d-dimensional feature vector respectively. $B_i$ and $W_i$ respectively represent the behavior of trusting others and being trusted by others, and the inner product of $B_i$ and $W_k$ can be used to represent the trust value $T_{ik}$. The characteristic matrices $B \in \mathbb{R}^{d \times m}$ and $W \in \mathbb{R}^{d \times m}$ can be obtained by minimizing the following loss function:

$$
I = \sum_{i=1}^{m} \sum_{k=1}^{m} (B_i^T W_k - T_{ik})^2 + \lambda (\|B\|_F^2 + \|W\|_F^2) \quad (1)
$$

Among them, $\|\cdot\|_F$ represents the Frobenius paradigm, and $\lambda$ is a parameter to avoid model overfitting.

2.2. Implied similarity of trust relationship

Users will influence each other in the process of making choices. Whether to choose will depend on the opinions of the users’ trustees. At the same time, the user’s decision will inevitably affect the choices of the trustees. In summary, the final observed user selection results are actually produced under the mutual influence between users [4]. This paper uses matrix decomposition technology to divide the trust matrix into two perspectives: user trust and trust, respectively calculate the trust
similarity between users, integrate trust and trust information, and comprehensively improve the accuracy of the recommendation algorithm. When acting as a trusted at the same time, the implicit similarity between user u and user v is:

\[ S_{u,v}^B = B_u^T B_v \]  (2)

When acting as a trusted person at the same time, the implicit similarity between user u and user v is:

\[ S_{u,v}^W = W_u^T W_v \]  (3)

The user's initiative to choose a trusted and being trusted by others are different user trust behaviors, both of which can be used to reveal the trust relationship between users. This article starts from these two aspects and fully explores the implicit trust interaction. In addition, matrix decomposition can also consider the relationship between users as a whole, and finally obtain a more accurate trust strength.

2.3. Estimation of the strength of trust relationship

This article compromises and considers the influence of \( S_{u,v}^B \) and \( S_{u,v}^W \), and the final trust strength is:

\[ T_{u,v} = (S_{u,v}^B + S_{u,v}^W) / 2 \]  (4)

It can be seen from equation (4) that the trust relationship similarity proposed in this paper comprehensively considers the horizontal and vertical information of the trust matrix, and fully excavates the trust relationship data. Even if users do not have a common trustee, they can be Relevant information is obtained from trustees, which effectively improves the impact of data sparseness. In addition, the similarity is obtained through the whole, which can comprehensively consider the global relationship.

2.4. Algorithm flow

Based on the hypothesis of social convergence, the interaction preference information of related people in social networks influences each other's decision-making. To this end, this paper proposes a recommendation algorithm that fully mines the implicit similarity of trust data. It considers the trust relationship between users from the two aspects of trust and trust, and obtains a more accurate trust value measurement and recommendation algorithm [5]. Inspired by Ma's research on using matrix factorization technology to share user social features and item preference features to improve recommendation quality, this section proposes a new method of using denoising autoencoders to share user social features and item preference features at a deeper level, and in order to improve the quality of implicit feedback recommendations. The algorithm flow introduced in this section is shown in Figure 1. Utilize the implicit similarity \( S_{u,v}^B \) when the users are both trustees and the implicit similarity \( S_{u,v}^W \) when they are the trustees at the same time to obtain more accurate trust relationship data of the users, and finally use the noise-reducing autoencoder to make in-depth decision-making.
In the algorithm of this paper, the implicit feedback information of user items and the trust information between users are deeply fused in the middle layer through the denoising autoencoder, and finally the final predictive voting is carried out by reconstructing the input. First, the encoder layer is used to map the input to the low-dimensional space. The coding layer is represented by equation (5):

$$h = f(W_1(\tilde{X}_u, \tilde{T}_u) + b)$$

Among them, $h$ is the feature of the intermediate hidden layer; parameter $W_1$ is the weight parameter of the network coding layer and the bias $f(\cdot)$ B is $sigmoid$. Functions $\tilde{X}_u$ and $\tilde{T}_u$ are the rating data of the user $u$ after noise processing and the trust user rating data of $u$, $(\tilde{X}_u, \tilde{T}_u)$ is the combined vector of $\tilde{X}_u$ and $\tilde{T}_u$. Let the network reconstruct the input data from the noisy data to prevent over-fitting and learn more robust data features. The most commonly selected noises are Gaussian noise and random erasure noise. Random erasure noise is used in this algorithm. Finally, the original input data is reconstructed through the decoder layer. The decoder is represented by equation (6):

$$\tilde{X}_u, \tilde{T}_u = f(W_2h + c)$$

Among them, $\tilde{X}_u$ and $\tilde{T}_u$ are the implicit feedback data and the predicted value of the trust data of user $u$ respectively, and the parameters $W_2$ and $c$ are the weight parameters and biases of the decoding layer. The network loss function is defined as:

$$L = l(X, \tilde{X}) + a(l(T, \tilde{T}) + \lambda_\Omega \Omega(W_1, W_2, b, c))$$

Figure 1. Algorithm flow
Among them, $l(\bullet)$ is the loss function to calculate the reconstruction error, $\lambda T$ is the parameter to prevent over-fitting, $\alpha$ is the attenuation parameter, and $\Omega(\bullet)$ is the regularization term, defined as Equation (8):

$$
\Omega(\cdot) = \|W_1\|^2 + \|W_2\|^2 + \|\Theta\|^2 + \|b\|^2
$$

In order to fully consider the influence of auxiliary data, we introduce the attenuation parameter $\alpha$ in equation (7) to control the influence of trust data.

3. System Design

3.1. Cloud computing platform

At present, typical cloud computing platforms mainly include Google's cloud computing platform, Amazon's Amazon Web Services cloud computing platform, Microsoft's Windows Azure cloud computing service platform, and IBM's blue cloud computing platform. In general, the cloud computing platform is explained as the following architecture, as shown in Figure 2.

![Cloud computing platform](image)

**Figure 2.** Cloud computing platform

The bottom layer is IaaS, which provides cloud services for basic hardware such as CPU, network, and storage. At this layer, cloud storage services are mainly provided to users. The next level is PaaS, which provides services and management similar to the operating system level. For example, Google GAE. You can throw your own Java applications (or Python) into Google’s GAE and run it. GAE is like a "cloud". Operating system, as far as you are concerned, you don't need to care about which machine your program runs on. The last layer is SaaS, which is the familiar software as a service. SaaS emphasizes paying on demand. Cloud computing abstracts computing and storage resources and dynamically allocates them to users who need to use it. It has high scalability, high reliability, underlying transparency and friendly monitoring and maintenance interfaces [6]. As shown in Figure 3, when developing applications on the "cloud", you only need to call the required resources in accordance with the specifications of the application program interface, so that users only need to care about the logic implementation of the business. For data mining, we can deploy various algorithms to the cloud computing platform and then set the target response time through the cloud computing platform control panel or interface to obtain satisfactory results.
3.2. Architecture of social trust information data mining system based on cloud computing

3.2.1. Traditional social trust information data mining architecture. Traditional social trust information data mining is based on a data warehouse, which can perform statistical query, multi-dimensional analysis, various complex charts and visualizations on data, so as to provide supporting basis for enterprise decision-making. The abstract structure of the system is shown in Figure 4.

![Figure 4. Data mining architecture](image)

3.2.2. Social trust information data mining system based on cloud computing. The social trust information data mining system based on cloud computing is built on the "cloud", which transparently provides interface services for various terminal users; provides an open interface for programs developed based on the system, and users can call the system through other applications. Open
interface to indirectly use various services provided by the system, as shown in Figure 5. Users do not need to understand how the system is implemented, nor do they need to worry about the computing and storage capabilities of the system. They only need to select the appropriate algorithm to process the data, and finally deploy it to the system area in a task manner to obtain the results of social trust information data mining. The social trust information data mining system based on cloud computing can be carried out by paying on demand. Enterprises or individuals can directly obtain certain services through this platform, so that they do not need to purchase expensive software; after the arrival of the cloud era, enterprises Most of the data of individuals and individuals are stored and stored in the cloud, making it possible to use a cloud computing platform-based data mining tool for social trust information.

**Figure 5.** Social trust information data mining system model

4. Experimental analysis

In order to focus on the core issues of social trust information data mining, we use simulation methods to simulate social trust information data mining and recovery mechanisms between different nodes in the network to test the social trust information data mining and service waiting after node failure. Pros and cons. In order to investigate the efficiency of data mining of social trust information for multiple types of failures in a complex environment, the experiment achieved a total of three types of network node failures, namely CPU failure, memory failure, and service communication failure [7]. According to the classic failures in trusted computing- Error-failure theory, failure occurs due to a failure and then an error, and the error is transmitted to the service level. Therefore, the experiment has realized three types of failure injection. In the experiment, the DAG service graph is randomly generated. For two service failure recovery solutions for comparison, one is the service waiting recovery scheme, referred to as the WR scheme, and the other is the social trust information data mining scheme proposed in this article, referred to as the TM scheme. The WR scheme is a traditional service recovery scheme, which refers to when a failure occurs After the service is terminated, when the node is repaired, the service is executed again.

Unless otherwise specified, the number of nodes included in this experiment is changed. For example, DAG-20 refers to 50 nodes. The purpose of this is to minimize the impact of changes in the number of nodes on DAG reconstruction. The average service failure rate of each node is 1.5103~2.5103 equally spaced per hour, and the downtime of each failure is randomly selected, ranging from 2 to 4 hours. The time cost of error recovery is fixed at 2 hours. The load of each node is fixed. In order to facilitate comparison, the parameter settings of the WR scheme and the TR scheme are the same, that is, the average service failure rate of each node is 1.5103~2.5103 equally spaced per
hour, each failure downtime and error recovery time are the same. The DAG data mining experiment generates a total of 3 types of DAG services, which are divided into three service maps DAG-20, DAG-40 and DAG-100 according to the number of services, and tested separately the time overhead of DAG data mining with the number of layers of 2, 5, and 10 respectively, as shown in Figure 6.

![Figure 6. DAG data mining time overhead](image)

It can be seen from Figure 6 that the time overhead of DAG social trust information mining increases with the increase in the number of services. Among them, the hierarchical reconstruction time of DAG-20 does not exceed 5s; while DAG-40 increases to more than 10s; DAG-100's hierarchical reconstruction time is the longest, but it does not exceed 110s. Considering that DAG service networks in actual applications generally have concurrent execution capabilities, the reconstruction time will be greatly shortened, and the service hierarchical reconstruction time overhead is much less than that of process migration or Checkpoint backup recovery time.

The experiment also examined the relationship between the data mining effect of social trust information and the accuracy of failure detection, as shown in Figure 7. There are two types of failure detection in failure detection, namely, failure detection error response in the absence of failure and after failure the failure detection does not respond. The lower the detection accuracy rate after the failure occurs, the longer the recovery time after the network failure, resulting in a corresponding extension of the service execution time. In extreme cases, that is, the failure detection false detection rate reaches 90%, the execution time of the DAG-100 service set has increased by 22%. Considering that this is only a single point of network failure, the uncertainty of the occurrence of failures in the actual network will inevitably extend the service execution time.

![Figure 7. The relationship between service execution and failure false detection rate](image)

5. Conclusion
This paper proposes an implicit feedback recommendation data mining algorithm based on the noise-reducing autoencoder for the implicit feedback recommendation problem. It considers the data mining
trust value between users from many aspects, improves the algorithm’s ability to deal with data sparse problems, and uses the denoising autoencoder deeply integrates social trust information to further improve the recommendation performance.

References
[1] Peng, Y.Yang, X.& Xu, W. Optimization research of decision support system based on data mining algorithm. Wireless Personal Communications, vol. 102, pp.2913-2925,April 2018.
[2] Massaro, A.Vitti, V.Lisco, P.Galiano, A.& Savino, N. A business intelligence platform implemented in a big data system embedding data mining: a case of study. International Journal of Data Mining & Knowledge Management Process (IJDKP), vol. 9, pp. 1-20,January 2019.
[3] Sugiyarti, E.Jasmi, K. A.Basiron, B.Huda, M.Shankar, K.& Maseleno, A. Decision support system of scholarship grantee selection using data mining. International Journal of Pure and Applied Mathematics, vol. 119, pp.2239-2249,Fifteen 2018.
[4] Subramaniyaswamy,V.& Logesh,R. Adaptive KNN based recommender system through mining of user preferences. Wireless Personal Communications, vol. 97, pp.2229-2247,February 2017.
[5] Sun, H.Study on application of data mining technology in university computer network educational administration management system. Journal of Intelligent & Fuzzy Systems, vol.37, pp.3311-3318,March 2019.
[6] Li, G.& Wang, F. Research on art innovation teaching platform based on data mining algorithm.Cluster Computing, vol. 22, pp.13867-13872,June 2019.
[7] Li, T.Zeng, C.Zhou, W.Xue, W.Huang, Y.Liu, Z.& Zhu, X. FIU-Miner (a fast, integrated, and user-friendly system for data mining) and its applications. Knowledge and Information Systems, vol. 52, pp. 411-443,February 2017.