Job Recommendation Algorithm Based on Knowledge Graph

Hongxing Wang
University of Chinese Academy of Sciences
Shenyang Institute of Computing Technology
Chinese Academy of Sciences
Shengyang, China
wang17854284776@163.com

Zhenhao Gu*
Shenyang Institute of Computing Technology
Chinese Academy of Sciences
Shengyang, China
guzhenhao@sict.ac.cn

Mengyu Gu
University of Chinese Academy of Sciences
Shenyang Institute of Computing Technology
Chinese Academy of Sciences
Shengyang, China
gumengyu20@mails.ucas.ac.cn

2022 IEEE 8th International Conference on Computer and Communications (ICCC) | 978-1-6654-5051-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCC56324.2022.10065776

***Abstract*—Aiming at the problem that the traditional collaborative recommendation algorithm ignores the semantic information of post matching personnel in the post recommendation system, this paper proposes a post recommendation algorithm based on a knowledge map and collaborative filtering. Firstly, the knowledge map of employees is constructed by using the data of employees of a company, and the improved knowledge representation algorithm is used to embed the semantic information of employees into a low-dimensional space. Then, the similarity between employees and the score similarity between positions and employees are calculated respectively. Finally, the two similarities are fused and applied to collaborative filtering recommendation. The algorithm in this paper uses the semantic relationship between the recommended objects to alleviate the cold start and data sparsity problems, so that the recommendation results are more accurate. The experimental results on the employee data set of employees of a company show that the proposed algorithm has higher accuracy and recall rate than other traditional recommendation algorithms.**

Keywords*—Convolutional neural networks; anomaly detection; network security.

I. Introduction

With the advent of the era of big data, a variety of data volume began to explosive growth, which brings convenience to people to analyze and solve problems from multiple perspectives. At the same time, complex data also led to information overload. Recommendation system [1] can help people quickly screen useful information from complex information, such as film recommendation [2], music recommendation [3], recipe recommendation [4], etc. The direction of the post recommendation system in the recommendation system has attracted wide attention in all walks of life because it can achieve rapid and efficient matching between personnel and jobs.

In the recommendation system, the recommendation system based on collaborative filtering (CF) [5] is the most widely used, which mainly uses the similarity of users or items for recommendation. In recent years, researchers have continuously improved the CF-based recommendation system from various aspects and have achieved varying degrees of results. Some researchers use item click factor to improve the similarity of items, optimize the similarity matrix [6], so as to improve the recommended coverage rate and increase the diversity of recommended items. Some researchers use matrix decomposition to reduce the sensitivity of the recommendation system to data sparsity and improve the recommendation effect [7]. However, in the collaborative filtering-based recommendation system, it is inevitable to be affected by data sparsity and cold start. Knowledge map is a new form of data organization. It describes the concept, entity and their relationship in the objective world in a structured way, which can enrich the description of items and enhance the correlation between items.

Therefore, this paper takes the knowledge map as the auxiliary information, considering the requirements of the positions of the company and the similarity between the staff of the company, a fusion algorithm of knowledge map and collaborative filtering is proposed. The main idea of the algorithm is to learn the entity characteristics of the staff knowledge map, map the staff entity to a low-dimensional dense vector, and calculate the semantic similarity between the staff of the company. Then, the scoring matrix of positions and employees is constructed. Finally, the semantic similarity between employees and the post-staff score are fused to recommend.

II. Related Work

A. *Collaborative Filtering*

Collaborative filtering algorithm is the most widely used algorithm in recommendation system, which does not need to rely on the labels of users and items, but digs out users ' potential favorite items through user interaction information. The core idea is to recommend similar items to users. Collaborative filtering algorithm is generally divided into user-based collaborative filtering [8] and item-based collaborative filtering [9]. Item-based collaborative filtering is to predict users ' evaluation of items by calculating the similarity between items. The main process is generally divided into three steps [10]:

1) Build a user-item scoring matrix

Set the collection of users to $U=\{u_1,u_2,\cdots,u_m\}$,The collection of items to $V=\{v_1,v_2,\cdots,v_n\}$, Then the user-item rating matrix is:

Supported by LiaoNing Revitalization Talents Program (XLYC2019019).

$$R_{m\times n}=\begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ R_{21} & R_{22} & \cdots & R_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ R_{m1} & R_{m2} & \cdots & R_{mn} \end{bmatrix} \tag{1}$$

Among them, rows represent users and columns represent items, R_{ij} is user u_i rating if item v_j.

2) Calculate the similarity between items

Think of the user's rating of an item v_j as an m-dimensional vector, namely $S_i=\{s_{1i},s_{2i},\cdots s_{mi}\}$. Take cosine similarity as an example, Set the user's rating vector for items v_i, v_j separately is S_i, S_j. The cosine similarity of the item is:

$$sim(v_i,v_j)=\cos(S_i,S_j)=\frac{\sum s_{ui}\cdot\sum s_{ui}}{\sqrt{\sum s_{ui}^2}\cdot\sqrt{\sum s_{uj}^2}} \tag{2}$$

When the value of $sim(v_i,v_j)$ is larger, the more similar vector S_i and vector S_j are, so the more similar items v_i and v_j are.

3) Selection of the item's immediate neighbors

Find the nearest neighbor of items, predict user preferences. The *K* nearest neighbor selection method is used to find the most similar *K* items to each historical item as the candidate set by traversing all the historical items of the user, predict the user' s score of the candidate items, sort and generate the recommendation list.

B. Knowledge Graph Representation Learning

In May 2012, Google put forward the concept of knowledge map in order to improve the efficiency of search engines [11]. Knowledge map is a method of structured representation of facts, which uses the ``entity-relation-entity'' triple to describe the relationship between entities and entities in the real world. Nowadays, knowledge map has been widely used in intelligent question answering [12], recommendation systems [13], search engines [14] and other fields.

The knowledge graph is composed of the knowledge ternary head entity h, the relation r, and the tail entity t, which can be expressed as (h,r,t) for each triplet. Knowledge representation learning is the projection of entities and relationships in the knowledge graph into a low-dimensional continuous vector space, which can quickly calculate semantic similarities between entities. The most classic of the knowledge graph vector representation methods is the TransE [15]. The TransE model employs the idea of translational invariance to take the relation r vector as a translation of the head entity h to the tail entity t, and employs a distance-based fractional function to assess the confidence of the triplet, as below:

$$f_r(h,t)=\|h+r-t\|_2^2 \tag{3}$$

When there is only one relationship between entities, the TransE model shows good results, but when dealing with multiple relationships, namely one-to-many, many-to-one, and many-to-many, since the entities and relationships of TransE are embedded in a plane, there are deficiencies in reflecting the multi-relationship model between entities. In the knowledge map of the company staff, the relationship between the staff is multi-to-multi, so this paper uses the improved translation model TransHR [16], as shown in Fig. 1.

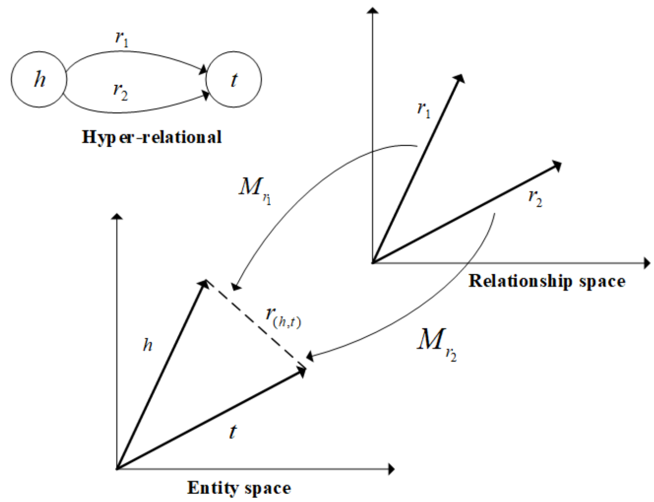

Fig. 1. TransHR Model

Different from the TransE model, the TransHR model embeds the head and tail entities into the Entity Space and embeds the relationship into the Relation Space. When the two entities have n relationships, the ith relationship is set as $r_i(i=1,2,\cdots n)$. For each relationship, a transfer matrix M_{r_i} is used to map the relationship vector from the relationship space to the entity space. Therefore, the relationship vector of TransHR can be expressed as:

$$r_{(h,t)}=r_iM_{r_i}\;(i=1,2,\ldots n) \tag{4}$$

Through the transfer matrix, the vector r_i of the relation space is mapped to the vector $r_{(h,t)}$ of the entity space, and the new vector $r_{(h,t)}$ is used as the link of the head and tail entities. Then the head and tail entity relationship in TransHR model is:

$$h+r_{(h,t)}\approx t \tag{5}$$

Therefore, TransHR's score function is:

$$f_r(h,t)=\|h+r_{(h,t)}-t\|_2^2 \tag{6}$$

The TransHR model can not only retain the integrity of entities and relationships, but also deal with multiple relationships between entities. Compared with the TransE model, the TransHR model can obtain higher similarity accuracy in dealing with the entity relationship of company staff.

III. Recommendation Algorithm Integrating Knowledge Map

In this paper, the knowledge map is embedded into the recommendation algorithm based on collaborative filtering as the auxiliary information, and a person-post recommendation algorithm integrating knowledge map and CF is proposed.

Firstly, the knowledge map of the staff of the company is constructed, and the entities in the knowledge map are represented by low-dimensional vectors using the TransHR model. The similarity between staff is calculated, and the similarity matrix based on the knowledge map is formed. Then, the rating matrix is constructed for the post of the company and the scores of all employees in the post, and the similarity of employees based on the rating is calculated. Finally, the fusion ratio is set, and the two matrices are fused to form a recommendation list. The process of TransHR-CF algorithm is shown in Fig. 2.

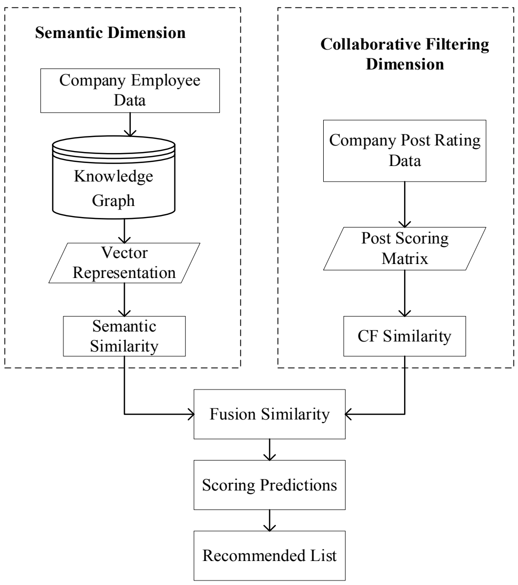

Fig. 2. The Algorithm Flow of TransHR-CF

A. Construction of A Company Personnel Knowledge Map

Based on the analysis of the field of post recommendation and the actual information of company personnel, this paper constructs a top-down knowledge map. The data are mainly from the existing database data within the company and the electronic data generated by the company personnel in their work and study, which reflect their experience and ability. Entity extraction is mainly based on the employee's resume and job requirements. Entity is divided into 12 types of entities, including staff, origin, position, political position, education background and professional title, and then 12 types of relationship extraction between various types of entities is carried out while confirming the entity,For example, Zhang San-education-undergraduate. The obtained entities and relationships are sorted out to obtain a triplet of constructing a knowledge graph, and finally complete the construction of the knowledge graph of the company personnel.

B. Semantic Similarity Calculation

Embedding entities and relationships in staff knowledge maps into a d-dimensional semantic space using TransHR, the semantic vector of staff entities is expressed as:

$$I_i = (E_{1i}, E_{2i}, \cdots, E_{di})^T \tag{7}$$

In the formula, I_i denotes the semantic vector of staff i, E_{mi} is the value of the semantic vector on the m-dimensional. The loss function of TransHR algorithm in model training is Euclidean distance, so in order to maintain consistency, the semantic similarity of staff is also measured by Euclidean distance. The calculation formula is :

$$d(I_i, I_j) = \sqrt{\sum_{m=1}^{d} (E_{mi} - E_{mj})^2} \tag{8}$$

In order to avoid the influence of the magnitude of similarity on the final results, the distance between the two employees is normalized to (0, 1] in order to standardize the calculation. The formula is as follows :

$$sim_{sg}(I_i, I_j) = \frac{1}{1 + d(I_i, I_j)} \tag{9}$$

In the constructed knowledge map of company employees, the relationship between the two employees is divided into direct relationship and indirect relationship. For example, there are three same relationships between the two employees, namely native place, graduate school and language ability so their direct relationship is 3, namely $C_1(I_i, I_j) = 3$. The corresponding number of relationships between two entities with an interval of 1 is represented by $C_2(I_i, I_j) = 1$. When calculating entity similarity, direct relation should have higher weight than indirect relation, so similarity between staff can be translated into :

$$sim_{sg}(I_i, I_j) = \frac{1}{1 + d(I_i, I_j)} [\lambda C_1(I_i, I_j) + \lambda^2 C_2(I_i, I_j)] \tag{10}$$

where λ is the weight factor and the value range is (0,1).

After the similarity between employees is obtained according to Equation (10), the similarity matrix of employees is generated, as shown in Table 1. Since employees who have already served in the same position are not recommended, the diagonal is 0.

TABLE I. EMPLOYEE ENTITY SIMILARITY MATRIX

Employee	E_1	E_2	···	E_j	···	E_n
E_1	0	···	$\cdots$	a_{1j}	$\cdots$	a_{1n}
E_2	$\cdots$	0	$\cdots$	$\cdots$	$\cdots$	a_{2n}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
E_i	a_{i1}	···	···	a_{ij}	$\cdots$	a_{in}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
E_n	a_{n1}	···	···	a_{nj}	$\cdots$	0

C. Similarity Fusion

The traditional collaborative filtering recommendation algorithm uses ratings for recommendation, which is greatly affected by subjective factors. At the same time, the data of ratings are generally sparse, and the accuracy rate is relatively low when calculating the nearest neighbor of users or items, thus reducing the recommendation quality of the recommendation system. The semantic similarity of items based on knowledge map can calculate the similarity by using the information of items themselves from the perspective of objective items characteristics. In order to improve the accuracy of staff recommendation in companies, this paper considers from the subjective and objective perspectives, calculates the score similarity $sim_{sg}(I_i, I_j)$ based on the traditional collaborative filtering algorithm, and then calculates the semantic similarity $sim_{cf}(I_i, I_j)$ between staff through TransHR vector quantization based on knowledge map. Finally, the two similarities are fused by linear weighting to make full use of the advantages of the two similarities.

$$sim(I_i, I_j) = \alpha \cdot sim_{sg}(I_i, I_j) + (1-\alpha) sim_{cf}(I_i, I_j) \qquad (11)$$

Where α is the weight coefficient, the range is [0,1].

D. Form A List of Recommendations

In this paper, P_{ui} is used to represent the prediction score of staff i on position u. Calculate the predicted score of a position for all staff not serving on that position by P_{ui}, and then the top k were sorted to recommend the position. The calculation formula of the prediction score P_{ui} is:

$$P_{ui} = \frac{\sum_{j \in N(u) \cap S(i,k)} sim(I_i, I_j) \times R_{u,j}}{\sum_{j \in N(u) \cap S(i,k)} sim(I_i, I_j)} \qquad (12)$$

Among them, $sim(I_i, I_j)$ is the similarity between staff I_i and staff I_j, $R_{u,j}$ is the score of staff j on position u, $N(u)$ is the set of staff serving on position u, $S(i,k)$ represents the k staff most similar to staff I_i.

IV. EXPERIMENT

A. Data Set

The data set used in this paper to verify the TransHR-CF algorithm mainly includes the rating data of employees in a companies in previous positions and the knowledge map data of employees in a companies.

The scoring data of employees in companies in previous positions are linearly weighted by the questionnaire score and the higher-level leadership score. Finally, 15,6543 valid scoring data are collected, of which 112,365 are scored by the questionnaire, and 44,178 are scored by the higher-level leadership. The scoring range is [0,5]. The higher the score is, the more matching the staff and the position is. The construction data of knowledge map are from the existing structured data and semi-structured data of companies. The main access paths are the data after digital processing of paper archives, the electronic data directly generated by employees' learning experience and work, and the existing electronic archives information within the company. Finally, 12 types of entities are abstracted, such as origin, position and educational background, with a total of 88985 entities. And 12 types of relations such as schools, rewards and job relations, with a total of 141752 relations.

B. Evaluating Indicator

Based on the recommendation results of the TransHR-CF recommendation algorithm, this paper uses the mainstream evaluation indicators, Precision and recall in the recommendation algorithm to measure the performance of the algorithm. The calculation of the two is as follows:

$$Precision = \frac{TP}{TP+FP} \qquad (13)$$

$$Recall = \frac{TP}{TP+FN} \qquad (14)$$

Where TP is the number of positive examples of correct prediction, FN is the number of negative cases of correct prediction, and FP represents the number of negative cases of false prediction.

C. Experimental Results and Analysis

In this paper, the fusion ratio of collaborative filtering similarity and knowledge map similarity is changed by changing the fusion weight factor α. When the nearest neighbor k is 50 and the embedding dimension is 100, the value of α is increased from 0 to 1, and the step length is 0.1. Each group of experiments do 10 times to observe the Recall and Precision. The experimental results are shown in Fig. 3 and Fig. 4.

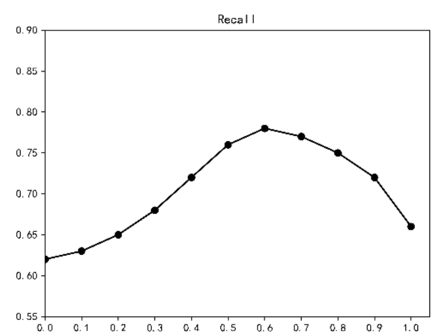

Fig. 3. Results of Recall

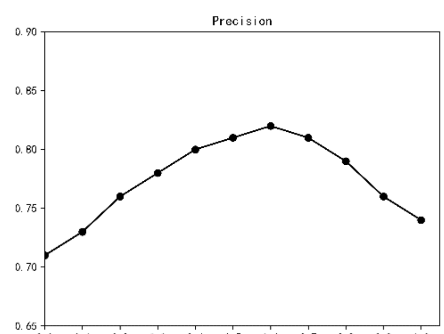

Fig. 4. Results of Precision

It can be seen from Fig. 3 and Fig. 4 that when the fusion factor α is 0.6, the Recall rate and Precision of the algorithm are the best.

In order to verify the effectiveness of the proposed algorithm, a comparative experiment is conducted with the benchmark algorithms RippleNet, TransE-CF and TransH-CF. The nearest neighbor number k was set to 10, 30, 50, 70 and 100, respectively, with the fusion factor $\alpha = 0.6$ that made the Recall rate and Precision the highest. Each group of experiments was averaged ten times. The experimental results of different algorithms are shown in Fig. 5 and Fig. 6.

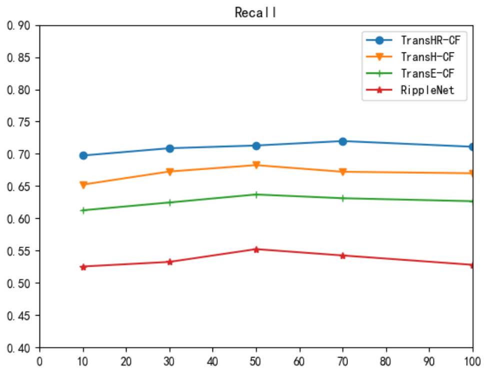

Fig. 5. Recall Comparison of Different Algorithms

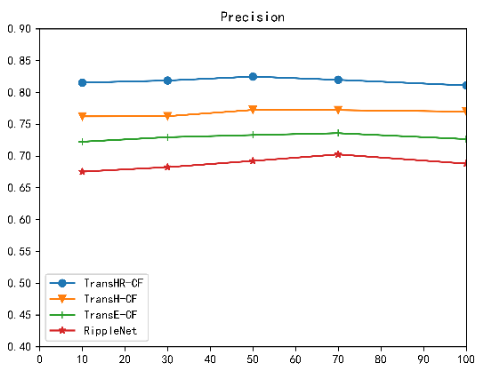

Fig. 6. Precision Comparison of Different Algorithms

Experimental results show that the proposed algorithm has better results than the benchmark algorithm in terms of accuracy and recall rate when selecting different neighbor numbers, and can also show the superiority of using this algorithm in the job recommendation of companies.

V. CONCLUSIONS AND FUTURE WORK

Aiming at the defects that the traditional collaborative filtering recommendation algorithm can not meet the post recommendation of company staff, this paper first constructs the knowledge map of company staff information, and embeds the knowledge map as auxiliary information into the algorithm based on collaborative filtering. The two are fused, and an improved collaborative filtering algorithm is proposed, which can accurately and efficiently realize the matching of personnel and posts in personnel changes. Experiments show that the proposed algorithm can improve the cold start and data sparsity of traditional methods, and improve the accuracy and quality of recommendation. However, the transHR used in this paper in the process of knowledge representation does not consider the multi-step relationship between entities. In the future, we will continue to try to introduce the reasoning algorithm of multi-step relationship to mine the deep relationship between entities for similarity calculation.

REFERENCES

[1] Fuzheng Zhang,Nicholas Jing Yuan,Defu Lian,Xing Xie,Wei-Ying Ma. Collaborative Knowledge Base Embedding for Recommender Systems[P]. Knowledge Discovery and Data Mining, 2016.

[2] Weizhuang Han. Research on film recommendation algorithm based on knowledge graph[D].Beijing University of Technology, 2020.DOI:10.26935/d.cnki.gbjgu.2020.000197.

[3] Zhuoyuan Li, Dan Zeng, Zhijiang Zhang. Research on Music Recommender Systems Based on Collaborative Filtering and Music Emotion[J]. Industrial Control Computer, 2018.

[4] Tian Yijun,Zhang Chuxu,Metoyer Ronald,Chawla Nitesh V.. Recipe Recommendation With Hierarchical Graph Attention Network [J]. Frontiers in Big Data, 2022.

[5] Yifan Liu, Yongjun Fan, Mingyue Sun , Junke Chen. Algorithm Design of Intelligent Music Track Recommendation System based on Collaborative Filtering[J]. International Core Journal of Engineering,2021,7(4).

[6] Chen Yi, Ren Zhengyun, Wen Pan. An improved collaborative filtering recommendation method based on item content characteristics: CN108664564A[p],2018.

[7] Li Xiaohong,Li Maolin. The application of the matrix decomposition recommendation algorithm based on attribute coupling is studied in the internet of things[J]. Journal of Intelligent \& Fuzzy Systems,2019,37(5).

[8] Wang Hulong,Shen Zesheng,Jiang Shuzhen,Sun Guang,Zhang Ren Jie. User-based Collaborative Filtering Algorithm Design and Implementation[J]. Journal of Physics: Conference Series,1757(1), 2021.

[9] Zheng G , Yu H , Xu W . Collaborative Filtering Recommendation Algorithm with Item Label Features[J]. International Core Journal of Engineering, 2020, 6(1):160-170.

[10] Wang Y , Deng J Z . User Similarity Collaborative Filtering Algorithm Based on KL Divergence[J]. Journal of Beijing University of Posts and Telecommunications. university,2021.DOI:10.27232/d.cnki.gnchu.2021.002563, 2017.

[11] Singhal A . Official Google Blog: Introducing the Knowledge Graph: things, not strings, 2012.

[12] Trung Phan,Phuc Do. Building a Vietnamese question answering system based on knowledge graph and distributed CNN[J]. Neural Computing and Applications,2021,33(21).

[13] Yinghui Wang. A Novel Chinese Traditional Medicine Prescription Recommendation System based on Knowledge Graph[J]. Journal of Physics Conference Series,2020,1487(1).

[14] Bo Peng,Guohua Chen,Yong Tang,Saimei Sun,Yuxia Sun. Semantic Navigation of Keyword Search Based on Knowledge Graph[P]. Computer Supported Cooperative Work and Social Computing,2017.

[15] Bordes A , Usunier N , Garcia-Duran A , et al. Translating Embeddings for Modeling Multi-relational Data[C]// Neural Information Processing Systems. Curran Associates Inc.2013.

[16] Zhang C, Miao Z, Xiao H, et al. Knowledge graph embedding for hyper-relational data[J]. Tsinghua Science \& Technology, 2017, 22(2):185-197.