Knowledge Map Reasoning Technology Based on Complex Network

Xiang DAI [a,1] and Yongchao WEI [b]

[a] *Southwest China Institute of Electronic Technology, Chengdu 610036, China*
[b] *Civil Aviation Flight University of China, GuangHan, 618307, China*

Abstract: In recent years, with the rapid development of artificial intelligence technology, knowledge mapping research based on data mining, machine learning and knowledge engineering has attracted the attention of a large number of researchers. Knowledge mapping has important theoretical value and practical significance for the development of artificial intelligence and machine cognitive intelligence. In recent years, researchers have achieved a lot of excellent research results in the field of knowledge mapping technology, but there are still some problems to be solved. This paper studies the construction technology of knowledge map based on social network, mainly discusses the problems of entity acquisition, entity disambiguation, relationship completion, relationship reasoning and knowledge conflict resolution. The research results of this paper can provide reference for intelligent applications such as intelligent information retrieval, dialogue system and intelligent recommendation.

Keywords: Knowledge Mapping, Complex Network, Entity Disambiguation, Relational Reasoning

1. Introduction

Knowledge mapping is composed of concepts, entities, events of the objective world and their relationships, which are usually represented by a triple. It can provide a large amount of factual knowledge, and structured storage method for unstructured data, which helps to discover the potential association between entities. By establishing the mapping from string description to structured semantic description, it can provide technical support for natural language understanding, machine reading, semantic based information retrieval and other technologies. It has broad application prospects in intelligent search, recommendation system, question answering system, intelligent investment advisory, intelligent marketing, complex network and other scenarios [1].

Knowledge mapping construction is a challenging work, which attracts great attention in academic and industrial fields. Although there are many excellent researches, but still many problems need to be studied. The traditional knowledge comes from domain experts, which has high accuracy but limited quantity, and cannot adapt to the explosive growth of data. Therefore, the research on complex media data has become a hot topic. Aiming at the ubiquitous knowledge defects, PtranSparse is proposed to complete the knowledge graph based on metric space and relational paths, which combines TranSparse model to deal with the heterogeneity and imbalance of

---

[1] Corresponding Author, Xiang Dai, Southwest China Institute of Electronic Technology, Chengdu, 610036, China; E-mail: 54831015@qq.com.

entities and relationships, and PTransE to make full use of the semantic information of the relationship path, the weight is used to distinguish the relationship type when the entity is projected [2]. The system based on social network knowledge graph includes five modules: the modules of data source, knowledge graph discovery, distributed graph database, and core sub-algorithm are connected to multi-graph network engine module, the distributed graph database module is connected to core sub-algorithm module [3]. The concept graph data developed by Microsoft from web pages and search logs contains 5.4 million core concepts, can provide common sense knowledge and concepts for machine by mapping concepts to vector space, which is used in many regions.

This paper mainly uses machine learning, data mining, relational statistical learning and deep learning to solve the problem of knowledge mapping. Data-driven method and machine learning are used to obtain the entity data; Then, the Markov logic network is proposed to disambiguate double named entities in the knowledge map; thirdly, the entity relationship completion and reasoning of the knowledge map are carried out; finally, the method based on deep learning is proposed to complete the knowledge map, which resolves the knowledge conflict in the process of the change between the two countries.

2. Related Concepts

2.1. Task Association Method Based on Complex Network

The topological structure of nodes is an important feature of complex networks, degree centrality, clustering coefficient and eigenvector centrality are three features of which used to associate people. Degree centrality represents its direct influence on neighbor nodes. The greater the degree centrality is, the greater its influence on neighbor nodes is. We use the normalized node degree to represent the degree centrality DC (i)[4].

$$DC(i) = \frac{k_i}{k_{max}} \tag{1}$$

Clustering coefficient measures the degree of close relationship between neighbors. The value is obtained by dividing the number of edges between its neighbors [5]:

$$C_i = \frac{2n}{k_i(k_i-1)} \tag{2}$$

$$EC(i) = X = \lambda^{-1} \sum_{j=1}^{n} a_{ij} x_j \tag{3}$$

2.2. Graph Segmentation Based on Spectrum

It is time-consuming to calculate node clustering coefficient and eigenvector centrality; graph segmentation can divide large graph into small subgraphs. Spectral graph partition is a common method. The spectrum is the eigenvalue of a matrix and its eigenvector. G is divided into two disjoint sets U and W, so that the number of edges $min(\frac{|E(G) \cap (U \times W)|}{|U| \cdot |W|})$ is the least [6].

2.3. Bipartite Graph Matching to Associate Cross Network Tasks

In addition to the user's portrait and complex structure, user's network behavior such as online time, writing style and user interest can also be used to describe the similarity of users. Based on the location trajectory, online time and user generated content, Kong et al. proposed a multi network anchor link method to find related accounts in different platforms, and used a stable matching method to map the matching problem of user pairs. Goga and others use location tags, time information and writing style to solve the task of user identification. Although writing style works well in the long text, it is not suitable for the microblog network with short text [7]. Zafarani used knowledge limitations and typing patterns to model user behavior patterns. Mishari pointed out that the author's online comment writing style can be used as feature to connect the same users. Iofciu used tag to match accounts. Goga used activity information to match accounts.

Existing methods are limited to use certain features, and cannot be extended to large-scale complex networks. In this section, the problem of people linking between complex networks is transformed into a weighted bipartite graph matching problem. Kuhn Munkres (KM) algorithm is used to solve the problem, which use user portraits, online time distribution and interest features to measure the similarity between users [8].

2.4. User Portraits and Complex Structural Features

User profile information, such as nickname, gender, hobbies, educational background, can also be used to identify the same users. Jain designed a system to match Facebook and twitter accounts based on user portraits, user generated content and complex relationships, the disadvantage that it needs to be redesigned for different complex networks. Tan pointed out that 50% of users use the same user name, but some are owned by different users [9]. Malhotra proposed a digital footprint based on name, nickname, location, photo and number of friends to associate a user's information. Jaro Winkler method can measure user ID and account name effectively, but can't associate some people with missing information. Zafarani and Liu proposed Mobius method to find out the relationship between users, based on the user's behavior pattern in complex media. Mobilus uses different content information to correlate user. Vosecky uses user portraits represented by vectors to associate people. Carmagnola used user profile attributes to associate users. Goga constructed a similarity matrix to identify 80% of the character pairs in twitter, Facebook and Google + datasets.

The above research mainly focuses on specific networks, and cannot be applied to academic complex networks. This section uses the complex structure and user portrait characteristics to associate the same users in complex networks [10].

$$sim_{profile}(A,B) = \frac{A_{profile} * B_{profile}}{\|A_{profile}\| * \|B_{profile}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}} \qquad (4)$$

In order to integrate portrait similarity and network structure similarity, Mapme introduces a balance factor to adjust the weight, as shown in equation 6, where the smoothing factor $\alpha \in [0,1]$, when α increases, the weight of portrait similarity is larger, and when α decreases, the weight of structure similarity is larger.

$$sim(A,B) = \alpha * sim_{profile}(A,B) + (1-\alpha) * sim_{structure}(A,B) \qquad (5)$$

2.5. Network Entity Disambiguation Based on Markov Logic

In Markov logic network, the weight of rules can be calculated by the prior probability. When the new rule challenges, the weight is greater than that of the old rule, a new relationship is established. It can effectively deal with the problem of relationship conflict in knowledge base, which combines the advantages of probability graph model and first-order logic. Probability graph model can deal with the learning problem, and first-order logic can deal with the reasoning problem. Markov logic network can not only provide a compact description, but also have the ability to integrate modular knowledge flexibly; on the other hand, Markov logic network can provide uncertain reasoning ability for first-order predicate logic, and tolerate the lack and conflict of knowledge.

Eq. (6) gives the probability distribution of a possible benchmark rule X.

$$p = (X = x) = \frac{1}{z} exp\{\sum_i w_i n_i(x)\} = \frac{1}{z} \Pi_i \varphi_i(x_{\{i\}})^{n_i(x)} \tag{6}$$

$n_i(x)$ Indicates the number of corresponding benchmark rules, $x_{\{i\}}$ represents the state of rule X.

Entity disambiguation based on tensor decomposition:

$$H_{a\mu} \leftarrow H_{a\mu} \frac{(W^T V)}{(W^T W H)_{a\mu}}, W_{ia} \leftarrow W_{ia} \frac{(V^T H)_{ia}}{(W H H^T)_{ia}} \tag{7}$$

If W and H are on the fixed distance, the Euclidean distance is invariant in the renewal case. If the initial matrix is nonnegative, all intermediate results are nonnegative.

2.6. Relationship Cmpletion Based on Link Prediction

Lao and Cohen developed a path ranking algorithm (PRA) for knowledge base reasoning. PRA randomly walks on the graph according to the graph structure to find the order (or path) of the relationship, thereby predicts the new relationship between entities. It has stronger expressive ability than other reasoning methods, and can be easily extended to very large graphs for real-time reasoning, but has weak representation ability. So, we use path feature and vector space representation to represent the relationship between entities.

If the relationship path $P=R_1R_2...R_T$ not empty，Let $P'=R_1R_2...R_{T-1}$, define the path characteristics of node e:

$$h_{s,P}(e) = \sum_{e' \in range(P')} h_{s,P'}(e') * P(e|e'; R_l) \tag{8}$$

Give path P, the path feature $h_{s,Pl}(e)$, and the nodes can be sorted by linear model $\theta_1 h_{s,P_1}(e) + \theta_2 h_{s,P_2}(e) + ... \theta_n h_{s,P_n}(e)$.

The ranking of candidate nodes e related to query node s can be given:

$$score(e; s) = \sum_{P \in P_l} h_{s,P(e)} \theta_P \tag{9}$$

2.7. Relationship Reasoning Based on Probabilistic Soft Logic

The occurrence times of each relationship can be counted, which represent the support or strength of the relationship. Based on the method proposed by Bhatia, a relational

ranking method is proposed to find out the corresponding target entities and their relationships. We use the following formula to rank the relationships

$$P\ (r, e_t|e_s) = P(e_t|e_s)P(r|e_t, e_s) = P(r|e_s, e_t)\frac{P(e_t)P(e_s|e_t)}{P(e_s)} \tag{10}$$

Entity class identity thinks that the more and greater the entity relationship between a target and an input, the closer and important the relationship between the two entities is, the higher their entity class identity is. The similarity degree between entities is shown:

$$P(e_s|e_t) = \frac{\sum_{r_i \in R(e_s|e_t)} w(r_i)}{\sum_{r_i \in R(|e_t)} w(r_i)} \tag{11}$$

2.8. Uncertain Time Series Knowledge Map

Uncertain time series knowledge graph is the combination of time series and probability graph, which can represent both uncertain knowledge and time characteristics. The semantic temporal relationship based on the joint probability distribution of entity relationship is proposed by Bhatia, finds out the corresponding target entities and their relationships. Given the entity e_s and the relation set $\mathrm{R}e_s$ ($\mathrm{R}e_s \subseteq \mathrm{R}$), Using the following formula to sort these two relations by relevance:

$$P(r, e_t|e_s) = P(e_t|e_s)P(r|e_t, e_s) = P(r|e_t, e_t)\frac{P(e_t)P(e_s|e_t)}{P(e_t)} \tag{12}$$

The relationship probability between an entity with a popular entity is higher than with a rare entity.

$$P(e_t) = relCount(e_t) \tag{13}$$

We mainly consider three constraints types: time disjoint, antecedent and time mutual exclusion. The time disjoint is to limit any two-time intervals with the same relationship. The following template is used to describe it:

$$R_1\ (e_i, e_j, t_1) \wedge R(e_i, e_k, t_2) \wedge (e_j \neq e_k) \rightarrow disjio\ int(t_1, t_2) \tag{14}$$

Semantic, path and time embedding representation are used to resolve the knowledge conflict, and predict the relationship between entities, so that the knowledge map can change with entities.

3. Experimental Results and Analysis

The platform is Intel i7 2.80GHZ with centos6.4 and C++ program language. The accuracy(P), recall(R) and F1 value were used to evaluate the proposed method. P and R are the proportion of correct node pairs in the found pairs and all correct pairs separately.

Data collected from aminer, sina weibo and science blog from January to June 2015 are the experiment. aminer, sina weibo and science blog contains 17000,17000 and 16000 user portraits, and 112000, 769000 and 89000 complex relationships separately. The common users are sina weibo 6512, science blog 4096, and aminer 3953.

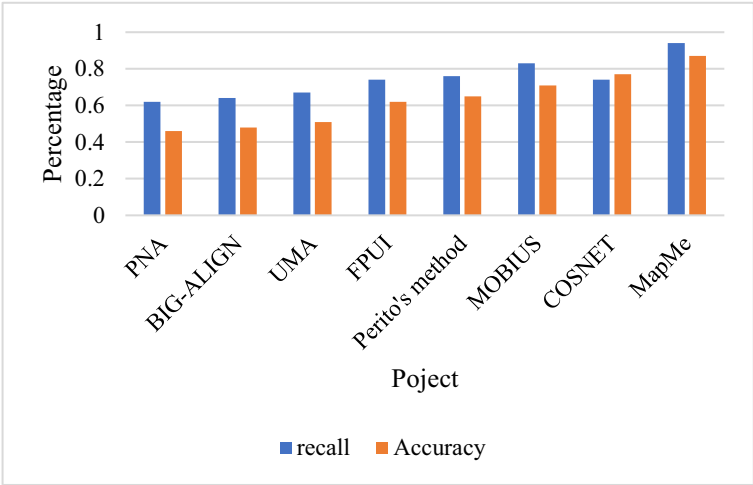(1) Result Analysis Using User Portraits and Complex Structural Feature

Figure 1. Comparison of recall and accuracy rates

Figure 1 shows the recall and accuracy of mapme and baseline method on Weibo aminer dataset, from which we can see that mapme has achieved the best results. The recall rate of mapme is 10% higher than that of cosnet and 21% higher than that of baseline method, and the accuracy rate of mapme is 10% higher than cosnet and 27% higher than that of baseline method.
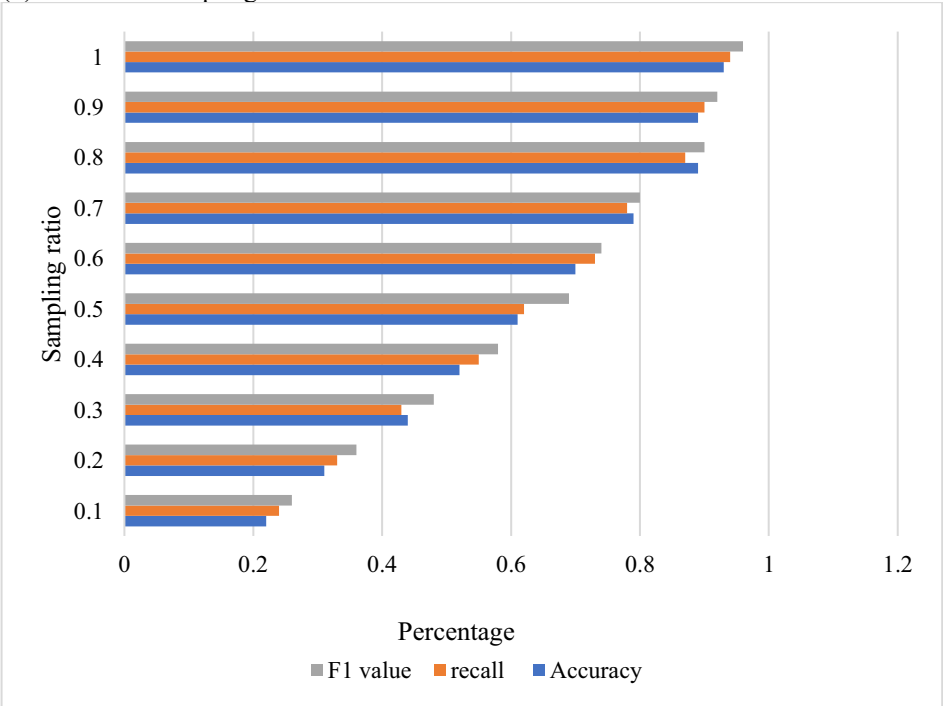
(2) Different Sampling Ratio Test

Figure 2. Comparison results of accuracy, recall and F1 values

Figure 2 shows P, R and F1 value of mapme method under different sampling ratios on science dataset. The sampling ratio η is the ratio of the training set to the total data. As can be seen from Figure 2, P of mapme increases with the sampling ratio.

Table 1. Comparison of the accuracy, recall and F1

	PNA	ALIGN	UMA	FRUI	Perito	MOBIUS	COSNET	MapMe
P	0.394	0.443	0.536	0.632	0.703	0.806	0.812	0.937
R	0.401	0.438	0.546	0.641	0.724	0.801	0.823	0.946
F1	0.398	0.439	0.539	0.639	0.716	0.801	0.816	0.941

Table 1 shows the comparison results between mapme and baseline method when the sampling ratio is 0.5 on aminer dataset, and mapme has achieved the best effect.

(3) Bipartite Graph Matching Method

Table 2. Inter-network Matching

OSN	OSN	Matching account pairs
Facebook	Tumblr	392783
Tumblr	Twitter	488667
Twitter	Instagram	413675
Instagram	Facebook	376528
Twitter	Facebook	432846
Tunmblr	Instagram	375654

Table 2 shows the statistical information. Instagram is a photo sharing network, includes 1105492 users and 34387876 relationships. Facebook is a complex network with the largest user's number, includes 1101324 users and 27693893 relationships. Twitter is a micro blog network, includes 1260752 users and 46083383 relationships. Tumblr is a light blog based complex network, includes 1007136 users and 29571930 relationships.

On Tumblr & twitter and instagram & Facebook, our method is the best. P, R and F1 value were 90.12%, 73.82% and 81.08% respectively. Compared with NS, Mobius, cosnet and hydra, P, R and F1 are 11%, 17% and 29% average higher respectively.

(4) Entity Disambiguation Experiment Based on Markov Logitech Network

Table 3. Entity disambiguation experiment

Data set	knowledge domain	Method	F1	P	R
Y A	YAGO3	Our method	0.711	0.815	0.631
		DoSeR	0.659	0.706	0.617
	DBpedia	Our method	0.693	0.803	0.609
		DoSeR	0.656	0.729	0.597
	Wikidata	Our method	0.704	0.812	0.621
		DoSeR	0.634	0.737	0.556
D B	YAGO3	Our method	0.710	0.821	0.625
		DoSeR	0.646	0.742	0.572
	DBpedia	Our method	0.712	0.817	0.631
		DoSeR	0.647	0.742	0.573
	Wikidata	Our method	0.706	0.831	0.614
		DoSeR	0.630	0.753	0.542
W D	YAGO3	Our method	0.704	0.827	0.613
		DoSeR	0.651	0.749	0.576
	DBpedia	Our method	0.715	0.824	0.631
		DoSeR	0.661	0.759	0.585
	Wikidata	Our method	0.726	0.831	0.645
		DoSeR	0.647	0.742	0.573

The results given in Table 3, show that our method is better than DoSeR. In WD and wikidata, the F1 value is 8% higher; in YA, DB, WD data sets and wikidata knowledge, the F1 value of is 7% higher on average; in wikidata and WD, P is 10% higher, and 5% higher on average. Markov logic network combines the advantages of Markov network

and first-order logic, which makes the statistical relation model based on Markov logic network very suitable for entity disambiguation.

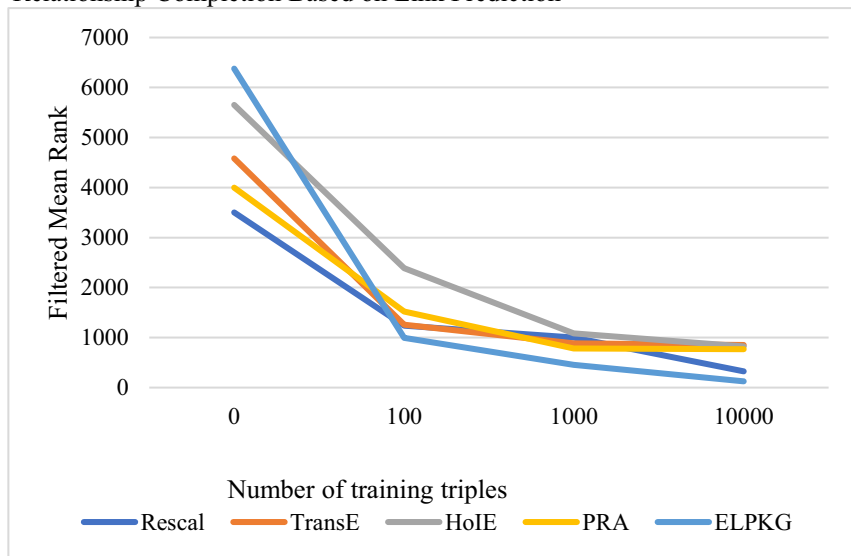(5) Relationship Completion Based on Link Prediction

Figure 3. FMR comparison result

Figure 3 shows the comparison results. Elpkg can learn new relationships with fewer samples. When the number of new relational tuples increases from 10 to 100, the learning speed of elpkg is the fastest, and the FMR decreases rapidly from 6378 to 996. The link prediction method elpkg proposed solved missing relationship problem, which can mine the potential relationship between entities more accurately and can be applied to question answering system and precision recommendation system.

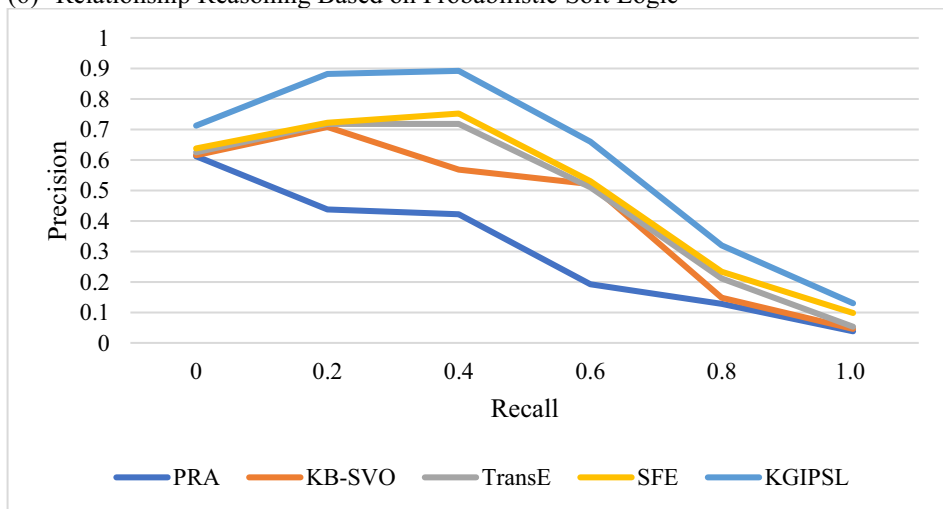(6) Relationship Reasoning Based on Probabilistic Soft Logic

Figure 4. Comparison between the proposed method and the baseline method

Figure 4 shows the P and R curves on the Yago dataset. which shows that the kgipsl method proposed is obviously better than others. When R is 0.4, the maximum P is 0.89, which is higher than SFE and kb-svo respectively. When R is from 0.1 to 1, the average P of kgipsl is 9% higher than others. kgipsl can mine the potential relationship between entities, and improve the knowledge coverage and completeness of the knowledge map.

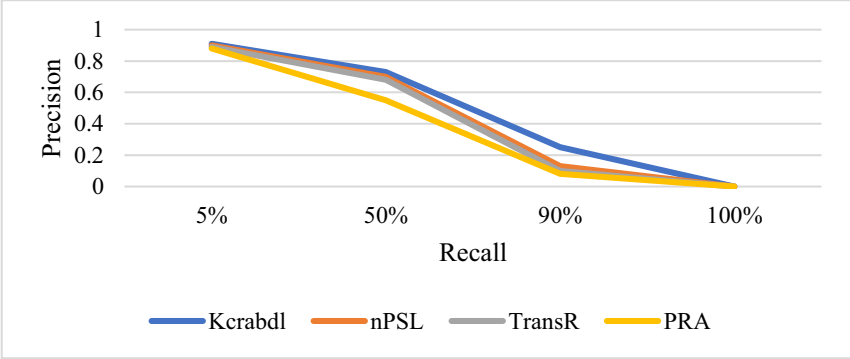(7) Knowledge Conflict Resolution Based on Time Constraint

Figure 5. Recall rate of proposed method Kcrabdl and baseline method

Figure 5 shows the comparison results between the proposed kcrabdl method and several other methods on Yago dataset. It can be seen that kcrabdl is superior to other methods. When R is 5%, P of kcrabdl, npsl, TRANSR and PRA is 0.91, 0.9, 0.89 and 0.88 respectively; when R is 50%, P is 0.73, 0.7, 0.68 and 0.55 respectively; when R is 90%, P is 0.25, 0.13, 0.1 and 0.08 respectively. Experiments show that this method is superior to the current mainstream methods.

4. Conclusion

The social network-oriented knowledge mapping technology studied in this paper belongs to the cross domain of knowledge engineering, social network, data mining and machine learning, and has become an important content in the field of artificial intelligence. However, due to the large scale of knowledge mapping, wide range of data sources, complex data types, fast knowledge updating and other characteristics, the research has great challenges. This paper obtains knowledge from social media, knowledge base and other data sources, and studies the key technologies of knowledge mapping construction technology. Different from the traditional method of extracting entities from Wikipedia and knowledge base to build knowledge map, this paper first obtains the person entities from social networks; then proposes an entity disambiguation method based on Markov logic network to solve the entity ambiguity problem in the process of knowledge map fusion; and then proposes a link prediction method combining entity relationship semantic embedding and path embedding to solve the problem of knowledge map fusion, and a method based on probabilistic soft logic is proposed to solve the problem of relational reasoning in knowledge map; finally, a deep learning method based on recurrent neural network is proposed to solve the problem of knowledge conflict in the process of knowledge map changing with time.

Reference

[1] Niu D, Liu L, Lu S. Augmenting Negation Normal Form with Irrelevant Variables[J]. IEEE Access, 2019, PP(99):1-1.

[2] Gao J, Chen H, Zhang Y, et al. Knowledge-Based Approaches to Accurate Mapping of Mangroves from Satellite Data[J]. Photogrammetric Engineering and Remote Sensing, 2015, 70(11):1241-1248.

[3] Shirahama K, Grzegorzek M. Towards large-scale multimedia retrieval enriched by knowledge about human interpretation[J]. Multimedia Tools and Applications, 2016, 75(1):297-331.

[4] Paulheim H, Cimiano P. Knowledge graph refinement: A survey of approaches and evaluation methods[J]. Semantic Web, 2017, 8(3):489-508.

[5] Uyar A, Aliyu F M. Evaluating search features of Google Knowledge Graph and Bing Satori[J]. Online Information Review, 2015, 39(2):197-213.

[6] Ernst P, Siu A, Weikum G. KnowLife: a versatile approach for constructing a large knowledge graph for biomedical sciences[J]. Bmc Bioinformatics, 2015, 16(1):1-13.

[7] Shengtian S, Zhihao Y, Lei W, et al. SemaTyP: a knowledge graph based literature mining method for drug discovery[J]. BMC Bioinformatics, 2018, 19(1):1-11.

[8] Sun Z, Wang C, Hu W, et al. Knowledge Graph Alignment Network with Gated Multi-Hop Neighborhood Aggregation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1):222-229.

[9] Wang T, Shi D, Wang Z, et al. MRP2Rec: Exploring Multiple-Step Relation Path Semantics for Knowledge Graph-Based Recommendations[J]. IEEE Access, 2020, PP(99):1-1.

[10] Wu Y, Liu Q, Chen R, et al. A Group Recommendation System of Network Document Resource Based on Knowledge Graph and LSTM in Edge Computing[J]. Security and Communication Networks, 2020, 2020(1):1-11.