Article

Knowledge Graph Completion Algorithm Based on Probabilistic Fuzzy Information Aggregation and Natural Language Processing Technology

Canlin Zhang [1] **and Kai Lu** [2,*]

1   Sorenson Communications, Salt Lake City, UT 84123, USA
2   Department of Public Safety Technology, Hainan Vocational College of Political Science and Law, Haikou 571100, China
*   Correspondence: lklgx316@hainnu.edu.cn

Abstract: The knowledge graph was first used in the information search of the Internet as a way to improve the quality of the search because it contains a huge amount of structured knowledge data. In this paper, the knowledge map algorithm is studied through natural language processing technology and probabilistic fuzzy information aggregation, and the knowledge map completion algorithm is cognitive-fitted. NLP is natural language processing. Based on the experiments in this paper, it can be seen that, after combining the algorithm, the behavior data set of 1000 Amazon users was analyzed, and it can be found that the accuracy of the algorithm improves as the proportion of data in the experiment increases. Among them, the 10% dataset has a correct rate of 0.66; the 30% dataset has a final accuracy rate of 0.68; and the 50% dataset has a final accuracy rate of 0.70. The experimental results of this paper show that using probabilistic fuzzy information aggregation and natural language processing technology as a way to complete the knowledge graph can improve the accuracy of the operation. It plays an important role in the development of intelligent cognition and search engines.

Keywords: probabilistic fuzzy sets; information aggregation; natural language processing technology; knowledge graph completion

MSC: 65C20

1. Introduction

Knowledge graph is an important branch of AI. It was proposed by Google in 2012. It is a structured semantic knowledge base, which is used to describe concepts and their relationships in the physical world in symbolic form. Its basic unit is the "entity relationship entity" triplet and entities and their related attributes value pairs. Entities are connected through relationships to form a network of knowledge structures. A knowledge graph is a collection of human cognition and cognitive processes. In the age of intelligence, it is more closely related to computer technology, which, to a certain extent, facilitates the public to have a more unified cognition of the world. Knowledge graphs are always flawed because people's cognitive ability is constantly improving, and computer technology is constantly developing, so knowledge graphs need to be supplemented. Since this is related to public perception, the research on knowledge graph completion algorithms is very important. There are many scholars who have studied and had discussion based on the above purposes, but they mainly study the related concepts and application scope of knowledge graphs, including artificial neural networks and other methods. Probabilistic fuzzy information aggregation and natural language processing techniques are rarely analyzed because knowledge graphs involve fuzzy concepts. This paper uses probability fuzzy sets and natural language processing technology as the basic methods of research

and discusses the information aggregation of knowledge graph completion algorithms so as to promote the development of technologies in related fields.

Natural language processing is an important direction in the field of computer science and artificial intelligence. It studies various theories and methods that can realize effective communication between people and computers using natural language. Natural language processing is a science that integrates linguistics, computer science, and mathematics.

There are many scholars who want to complete the knowledge graph. Reyes O believed that in real life, a large amount of unlabeled data was generated, which had become a relatively serious problem, and active learning was a relatively important area of discussion, which can visually compare learning curves. Therefore, two comparison methods were proposed in his research [1]. Ji K's research aimed to illustrate that graph neural networks were very effective in the learning process of knowledge graphs and can bring advantages to the completion of knowledge graphs. In this form, he proposed a graph attention network that can perform neighborhood aggregation strategies and can represent complex semantics [2]. Bai L believed that the temporal knowledge graph was a very useful artificial intelligence resource, but it is not complete, so the completion of the knowledge graph was the focus of his research, and a trend-guided prediction model was proposed to guide the aggregation process [3]. In the experiment, Che F mainly used the prediction missing link model in the learning of knowledge graph and used the translation characteristic, while the convolutional network model has a strong fitting space, but it lacks the above advantages. Therefore, in his specific research, he proposed a new knowledge graph embedding model [4]. Zhang Z believed that the knowledge graph was rapidly popularizing, but there were still missing values in the triples in the knowledge, so based on the method of completing the knowledge graph, a relational graph neural network with hierarchical attention was proposed [5]. It can be clearly found that the above-mentioned scholars' research on the completion of knowledge graphs has progressive significance, and their research methods are also constantly innovating but are less often combined with probabilistic fuzzy information aggregation and natural language processing techniques.

Fuzzy information and natural language processing technology are more commonly used research methods. Yu T believed that traditional Chinese medicine, as an important cultural heritage, is of great significance to the inheritance of human civilization. Therefore, it was very important to protect and inherit TCM(traditional Chinese medicine) knowledge by means of knowledge graph [6]. Natthawut K believed that the knowledge graph can be constructed by analyzing Java bytecode, which is a general method belonging to natural language text processing and can be defined at compilation time [7]. Lin Z Q said that the development of software intelligence is a contemporary development trend, and in the specific research, he proposed the concept of intelligent development environment and software knowledge graph and studied the architecture and discussed it [8]. Jia Y believed that knowledge graphs can predict entities in knowledge and can express them in a structured way through knowledge graph embedding. However, due to the multi-layer relationships in the graph, the current methods are still insufficient [9]. Wang C believed that open data in current publications were an important part of literature, but they were challenging to analyze. Therefore, semantic analysis was carried out in combination with knowledge graphs to establish connections [10]. These scholars' research on probabilistic fuzzy information aggregation and natural language processing technology was relatively in-depth and has reference significance, but it was not fully integrated with the specific content of knowledge graph completion.

In this paper, probabilistic fuzzy information aggregation and natural language processing technology are used as related methods for knowledge graph completion, and experimental results are obtained by using related models, and conclusions with reference significance are obtained from them. The innovation of this paper is to use the probabilistic fuzzy information aggregation and natural language processing technology as the main research methods to analyze the knowledge graph completion in the intelligence era. The innovation of this paper is the combination of probabilistic fuzzy information aggregation

and natural language processing technology to study the knowledge map, which ensures the accuracy of data and makes the data more efficient in intelligent cognition and search engine use. It has been supplemented in the text.

2. Method of Knowledge Graph Completion Based on Probabilistic Fuzzy Information Aggregation and Natural Language Processing Technology

2.1. Knowledge Graph Completion

The knowledge map is a classification system of relevant knowledge provided for an organization's knowledge management system. It is convenient for organization members to read and use, acquire knowledge, share knowledge, and create knowledge. With the support of modern information technology, its forms include a basic layered display of three-dimensional dynamics. The function is to provide a comprehensive introduction to all the knowledge of the organization and to focus on a certain field. The knowledge graph is mainly used to describe related concepts, entities, and events and the relationship between them in the real world. In this part, concept refers to the conceptual representation of objective things, including but not limited to people, animals, etc. Entity refers to specific things in the objective world, such as desks, computers, etc. Events are activities that exist in the objective world, including earthquakes, trading relationships, etc. [11]. Relationship is used to describe the objective relationship between the three. The knowledge graph technology is a means used in the process of establishing a knowledge graph, such as the fusion of computational knowledge, information retrieval and extraction, etc. It is mainly used to solve the problem of how computer algorithms acquire knowledge from the objective world and is applied in intelligent service systems [12,13]. The existing knowledge picture resources are mainly knowledge graphs constructed by manual construction, swarm intelligence, Internet link data construction, and machine learning [14]. Figure 1 shows the main research directions of knowledge graphs:
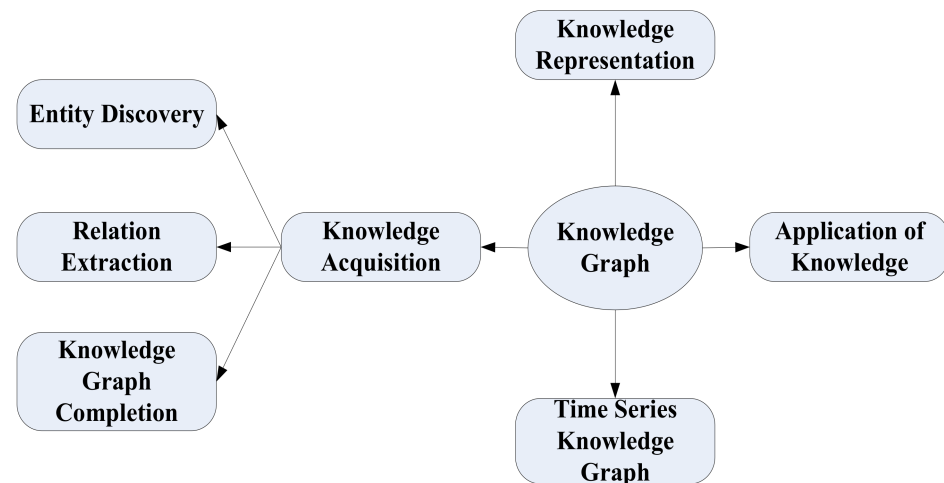

Figure 1. Main research directions of knowledge graph.

In Figure 1 the knowledge map includes the application of knowledge, knowledge representation, time series knowledge map, and knowledge acquisition. Knowledge acquisition mainly includes entity discovery, relationship extraction, and knowledge graph completion.

2.2. Probabilistic Fuzzy Information Aggregation

2.2.1. Probabilistic Fuzzy Algorithm

In real life, it is easy to encounter vague language such as "higher". In order to describe the fuzzy variables existing in this language more clearly, some scholars put forward the concept of fuzzy set and choose to use the membership function to describe it [15]. The triangle is a commonly used function to discuss fuzzy set theory, as shown in Figure 2.
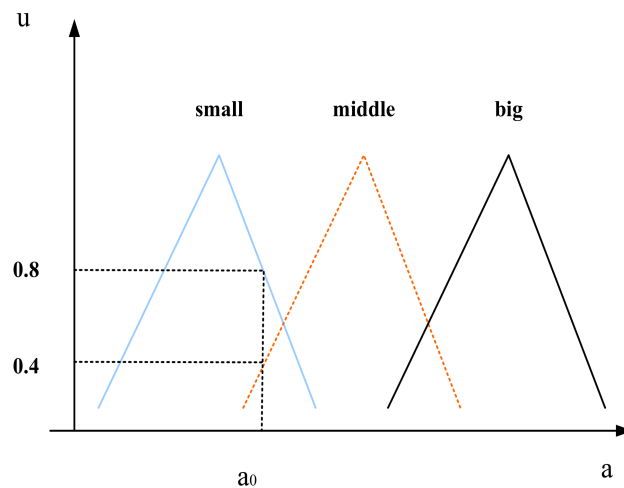
Figure 2. Typical fuzzy set.

It can be seen from Figure 2 that there are three fuzzy sets of short, medium, and high on the domain U = [0, 30]. This membership function can be used to describe any point on the field, and there are memberships belonging to the three fuzzy sets mentioned above. The value corresponding to the point is called the degree. In Figure 2, the element a_0 is included in the fuzzy set of "short", and its degree is 0.8, and the degree belonging to "medium" is 0.4. On the basis of fuzzy sets, in order to better adapt to the research, some scholars put forward the fuzzy logic system. It can simulate the thinking process of the human brain and use fuzzy sets to describe the characteristics of the knowledge graph so as to deal with the fuzzy uncertainty in the whole system [16,17]. The basic fuzzy inference framework is shown in Figure 3.
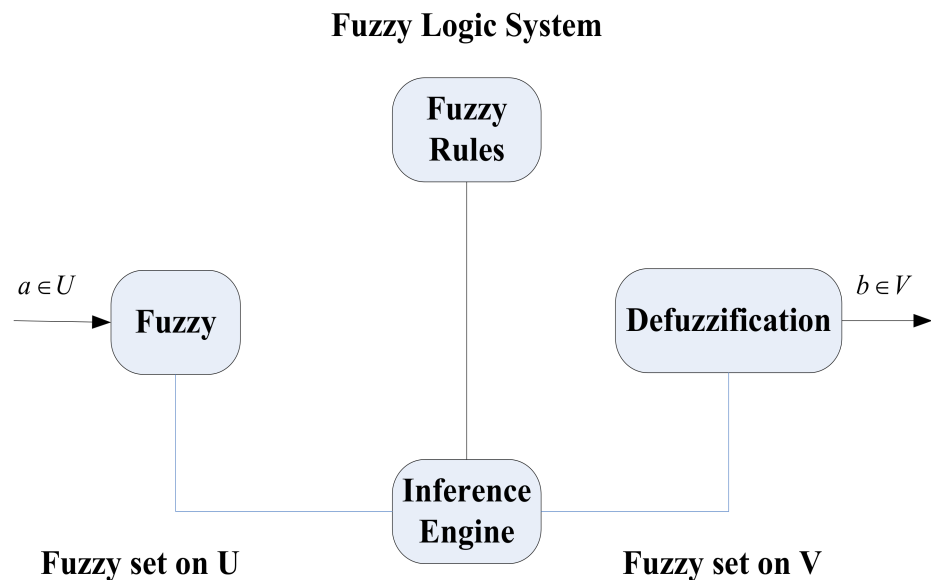

Figure 3. Basic fuzzy inference framework.

Therefore, in view of the uncertain characteristics of fuzzy sets, they need to be processed. Common processing methods include probability fuzzy sets and random sets. All of them retain the characteristics of fuzzy sets and combine probability theory and fuzzy theory in different directions, thereby enhancing their ability to deal with uncertainty in random environments [18]. Among them, the random set is derived from the concept of geometry, which belongs to the expansion of the traditional random variable, which describes the random element and belongs to the set mapping function. The generation

of this model provides a good mathematical basis for completing the knowledge graph and can express vague concepts in clear mathematical expressions. It provides favorable conditions for specific analysis [19,20].

Fuzzy random variable is a kind of random phenomenon that occurs easily and does not have a clear meaning. It is not only a further extension of traditional random variables but also an extension of the concept of random sets. In terms of definition, it can be expressed that it defines a function that can map all possible results of random experiments into fuzzy sets so as to simulate the uncertainty of the coupling between fuzzy and random phases [21]. The specific definition can be expressed as the following: there is a probability space (A, B, C); let $a_1, a_2, \cdots a_I$ represent I fuzzy variables; then, $\xi(m)$ represents a fuzzy random variable, and $\xi(m) = a_n, n = 1, 2, \cdots I$, as specifically shown in Figure 4.
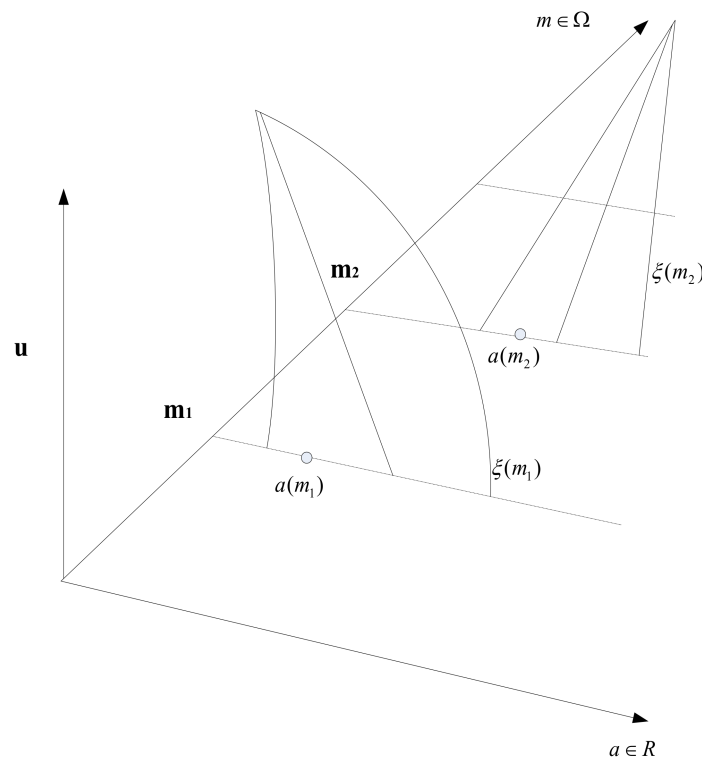

Figure 4. Typical fuzzy random variable plot.

This representation is intended to illustrate specific events in the event space set. At this time, the specific value of the fuzzy random variable becomes a fuzzy set rather than a single value. The probability set appears to deal with the problems arising from pattern recognition and decision making. It can directly assume the membership degree as a random variable and integrate the random theory into the fuzzy set framework, thereby improving the processing ability of the fuzzy set. Because the probability set only proposes related concepts without establishing a system, it cannot be directly applied in data modeling.

Unstable fuzzy set is a viewpoint that has been produced in recent years, which is mainly reflected in the uncertainty of two aspects. One part is the different opinions of a group of people when making decisions; the other part is that the opinions of any one person change over time. In view of this situation, parameters related to time are added to the traditional fuzzy membership function so as to describe the difference caused by time changes in the actual inference. Thus far, an unstable fuzzy set is formed, as shown in Table 1. M is different opinions when making decisions, F is that opinions will change over time, and the middle point indicates the fuzzy relationship between them.

Table 1. A person's perception of old age in 4 weeks.

Week	M.F.	M.F.	M.F.
1	X_1	70	5
2	X_2	68	5
3	X_3	70	6
4	X_4	72	5

Table 1 shows an individual's perception of old age over time over a 4-week period. In the specific operation, the Gaussian membership function is used to represent it, where c represents the center of the function, and ε represents the variance of the function. The Gaussian membership function can be combined with the actual parameters and time to obtain an unstable fuzzy set:

$$G = \int_{t \in T} \int_{a \in A} e^{-\frac{(a-c(t))^2}{2\varepsilon(t)^2}} \, da \, dt \tag{1}$$

Among them, four different situations correspond to 4 weeks:

$$G_1 = X_1, G_2 = X_2, G_3 = X_3, G_4 = X_4, T = \{1,2,3,4\} \tag{2}$$

The basis function $u_G(a)$ can be obtained:

$$u_G(a) = e^{-\frac{(a-70)^2}{2(4^2)}} \tag{3}$$

The perturbation function $f_c(t)$ is defined as follows:

$$f_c(t) = \begin{cases} 0 & if \, t \in \{1,3\} \\ -2 & if \, t \in 2 \\ 2 & if \, t \in 4 \end{cases} \tag{4}$$

Let $h_c = 1$, then the function

$$c(t) = 70 + f_c(t) \tag{5}$$

Then, the perturbation function is defined as follows:

$$f_\varepsilon(t) = \begin{cases} -1 & if \, t \in \{1,2,3\} \\ 0 & if \, t \in 2 \end{cases} \tag{6}$$

Let $h_\varepsilon = 1$; then, the function

$$\varepsilon(t) = 5 + f_\varepsilon(t) \tag{7}$$

Therefore, on this basis, the specific algorithm concepts of probabilistic fuzzy sets and probabilistic fuzzy logic systems are proposed. The former is inspired by the TYPE2 fuzzy set, which can simultaneously deal with the fuzzy and random uncertainties generated in more complex systems. The discrete probability fuzzy set is the earliest model of the probability fuzzy set, which is defined as follows:

The fuzzy membership degree corresponding to the input variable $a \in A$ is v, so the probability fuzzy set X can be described by the probability space (V_a, σ, P). The content represented by V_a is mainly the set of all possible events $\{v \in [0,1]\}$ in the probability space, σ represents the ε domain, and P represents the probability on the domain σ. Therefore all events G_i in V_a satisfy

$$P(G_i) \geq 0, \quad P(\textstyle\sum G_i) = \textstyle\sum P(G_i) \quad P(V_a) = 1 \tag{8}$$

For the event $v = v_i \in [0,1], v_1(i = 1, 2, \cdots M)$ corresponding to G_i represents a specific value of a fuzzy membership degree, $P(G_i)$ represents the occurrence probability of event G_i, and M represents the number of events in G_i. Therefore, a probabilistic fuzzy set X can be shown as a union in a finite sub-probability space:

$$\overline{X} = \bigcup_{a \in A} (V_a, \sigma, p) \tag{9}$$

In the specific three-dimensional membership function, $v_X(a)$ represents that the fuzzy domain variable is membership of the fuzzy set X corresponding to the input a. $P(a, v_X(a))$ represents the probability distribution function indicated in the probability domain, which means the probability distribution characteristic is used to represent the degree of membership. For example, if the input value of $a = 1$ is input, the corresponding membership degree would generate three values, and the corresponding probability is:

$$v_1(a) = 0.8186, P(a, v_1(a)) = 0.3 \tag{10}$$

$$v_2(a) = 0.9047, P(a, v_2(a)) = 0.5 \tag{11}$$

$$v_3(a) = 0.9444, P(a, v_3(a)) = 0.2 \tag{12}$$

According to the relevant definitions of discrete probability fuzzy sets, the continuous form of probability fuzzy sets is generated, and the continuous probability density is proposed to further describe the continuous random variable of membership degree. The basic definition is as follows:

Input variable $a \in A$, and the corresponding fuzzy membership degree is $v \in [0,1]$. A probabilistic fuzzy set X can be described using a probability space (V_a, σ, P), where $V_a = [0,1]$. σ is just one of the ε domains. Define $p(v)$ as the probability density function on σ that can describe the random characteristics of membership degree. Therefore, for any v in the domain V_a, it satisfies

$$p(v) \geq 0, \quad \int_0^1 p(v) dv = 1 \tag{13}$$

Then, the probability fuzzy set can be expressed as

$$\overline{X} = \bigcup_{a \in A} (V_a, \sigma, p) \tag{14}$$

However, due to the complexity of probabilistic fuzzy sets, the research on them is not deep enough, and further research is needed. The probabilistic fuzzy set system contains a main part, which is more consistent with the traditional fuzzy logic system in this respect. Usually, the rules of these probabilistic fuzzy sets can be expressed in the form of IF–THEN, such as

$$rulei : if\ a_1 \in \widetilde{X}_{1,i}, and \cdots a_n \in \widetilde{X}_{n,i}, then\ b \in \widetilde{Y}_i \tag{15}$$

However, the probabilistic fuzzy logic system mainly uses fuzzy sets with three-dimensional membership functions. Therefore, it is necessary to further express continuous probability fuzzy sets.

2.2.2. Information Aggregation

Information aggregation can be referred to as sensor information fusion. It mainly refers to the multi-level processing process, and multi-source data are detected and combined to estimate so as to achieve accurate identity estimation and timely and effective situation and threat assessment. The completion of the knowledge graph is the fusion of multi-source data, and there are many advantages in data fusion. It has strong fault tolerance, fast information processing speed, good robustness, and low cost of information acquisition. Common information fusion methods are mainly divided into different

combination levels of data layer, feature layer, and decision layer for fusion, as shown in Figure 5.
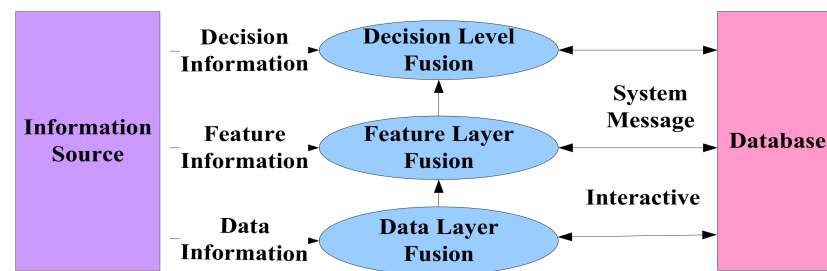
Figure 5. Information fusion model.

In the specific information fusion process, uncertain factors are generally represented by fuzzy language, so the probabilistic fuzzy information aggregation method can be better used to deal with knowledge graph completion. In Figure 5, the decision layer, feature layer, and data layer are integrated through the database to obtain information sources for decision information, function information, and data information. The sources of information acquisition mainly include information interaction and system information.

2.3. Natural Language Processing Technology

Natural language processing actually explores language problems that arise in human interactions and in human–computer communication. It is currently defined as a collection of theories and means of how to more efficiently realize human–computer communication in human–computer interaction, covering a wide range. Natural language has a wide range of applications. It can not only deal with the intelligent development of the search field but also be used in the military field. In addition, it is also widely used in the field of medical treatment. It also has certain limitations, the most prominent of which is the ambiguity of semantics, which is more related to complex human activities. Therefore, taking Chinese as an example, one of the disambiguation tasks is word segmentation, and the second is the acquisition of contextual content, which is also a big challenge for machine translation. However, as a new technology, natural language processing technology plays an increasingly prominent role in the computer field and occupies an important position.

3. Experiment of Knowledge Graph Completion

3.1. Background Introduction

The knowledge map can deal with this kind of multi-step reasoning problem by using the path based search method. The traditional path lookup method is mainly PRA (path ranking algorithm); however, this method will lead to a sharp expansion of the feature space due to the explosive growth of the number of paths for knowledge maps with a large scale. To solve this problem, we can try to express the relationship by embedding, generalizing the relationship, and modeling the completion of knowledge based on this so as to alleviate the problem of feature space expansion caused by too many paths. As a technology widely used in the Internet field, knowledge graph plays an important role in recommendation systems and intelligent dialogue. Since it is an important research method in natural language processing, but it lacks the specific description of entity relationships, it would also cause many communication barriers and troubles in practical applications, which also increases the demand for knowledge graph completion. Since the knowledge graph itself is prone to incomplete missing relationships, it is necessary to infer the missing relationships from the existing knowledge graph through analysis and then complete the entity relationships within it. With the development of technology, the current completion technology can be expressed as the following: knowledge representation (as the most direct representation method, it mainly uses the low-dimensional embedded method of the knowledge graph to illustrate the entities and relationships in the knowledge graph. Generally, the TransE

method is used to represent the core rule, which can predict the missing triples); path finding (because the previous method cannot handle multi-step knowledge reasoning. Therefore, the path finding method is successfully generated. During the inference process, the path ranking algorithm method is used to find the path, and the graph is completed by the correlation between the path vector and the relationship vector to be predicted); inference rules (mainly focus on modeling the logic rules themselves, so they are usually combined with embedding or traditional inference models with neural network models); reinforcement learning (a learning method that has difficulty in clearly evaluating similar entities and relationships in the knowledge graph but easily affects path finding due to its own imperfection); and meta-learning (mainly for the completion of the long-tail data of the knowledge graph, mainly including two methods of measurement and optimization). Based on the above knowledge completion methods, although the applied method is relatively simple, it can have better completion performance through continuous method improvement. Therefore, this paper combines probability fuzzy information aggregation with natural language processing technology to complete the knowledge graph, which has certain reference significance.

3.2. Experimental Analysis

Combined with the relevant completion methods in the background introduction, in order to verify the feasibility of the above methods, this paper uses the relevant behavior records of real users of the Amazon shopping network as the experimental data set of this paper. It involved 1000 users, 201,526 items, and 562,135 user records. According to the entity association graph of the triplet of "entity-association-entity", namely entity association graph (EAG), its efficiency, speedup ratio, parallel efficiency, and feasibility of graph completion were tested.

3.2.1. The Efficiency of Constructing EAG

In order to visually demonstrate the efficiency of EAG, the relevant data of 2.5×105, 5.0×105, and 7.5×105 customers in the Amazon shopping network dataset were randomly selected. The execution time and speedup ratio of related algorithms were tested when the number of workers is two, four, and eight. The specific situation is shown in Figure 6.
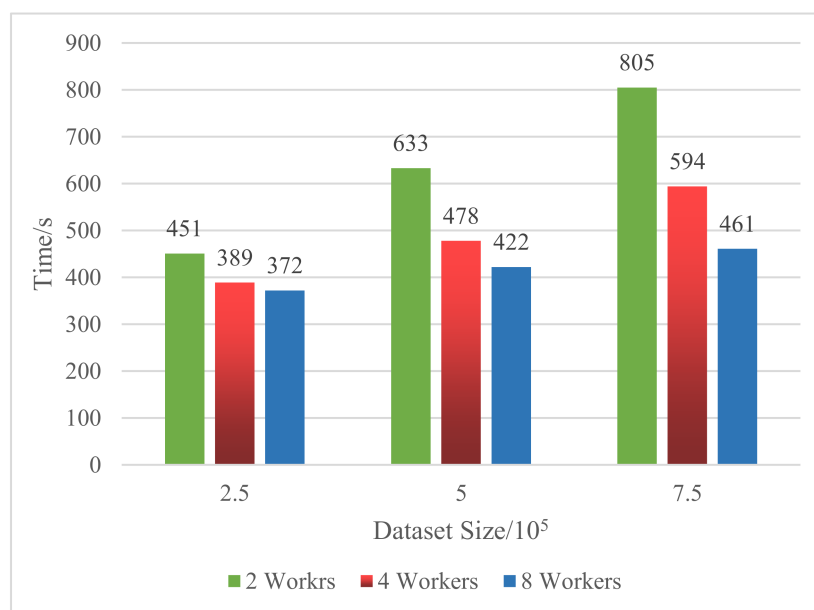

Figure 6. Build time of EAG.

It can be seen from Figure 6 that the data scale increases, and the running time also increases with different numbers of workers. Taking the running time of two workers as an example, the running time of the 2.5×10^5 dataset is 451 s, the running time of the 5×10^5 dataset is 633 s, and the running time of the 7.5×10^5 dataset is 805 s. Taking the running time of four workers as an example, the running time of 2.5×10^5 dataset is 389 s, the running time of 5×10^5 dataset is 478 s, and the running time of 7.5×10^5 dataset is 594 s. Taking the running time of eight workers as an example, the running time of the 2.5×10^5 dataset is 372 s, the running time of the 5×10^5 dataset is 422 s, and the running time of the 7.5×10^5 dataset is 461 s.

3.2.2. Speedup Ratio of EAG

Since the speedup ratio of the parallel algorithm represents the ratio of the execution time in the case of a single node and the case of multiple nodes, the speedup ratio for different scales of Amazon data sets is shown in Figure 7.
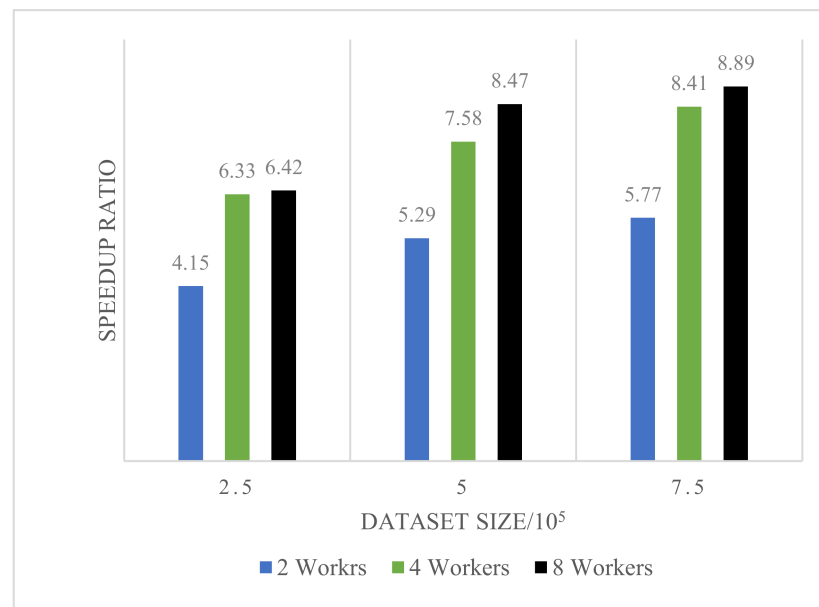
Figure 7. Algorithm speedup for different dataset sizes.

According to Figure 7, under different data scales, the speedup ratio of the algorithm is not the same, but the overall increase. From the data scale set of 2.5×10^5, the speedup ratio of two workers is 4.15, the speedup ratio of four workers is 6.33, and the speedup ratio of eight workers is 6.42. From the data scale set of 5×10^5, the speedup ratio of two workers is 5.29, the speedup ratio of four workers is 7.58, and the speedup ratio of eight workers is 8.47. In terms of the data scale set of 7.5×10^5, the speedup ratio of two workers is 5.77, the speedup ratio of four workers is 8.41, and the speedup ratio of eight workers is 8.89.

In order to better explore the speedup ratio, this paper combines the relevant Amazon data sets to calculate the parallel efficiency of the algorithm. The specific situation is shown in Figure 8.
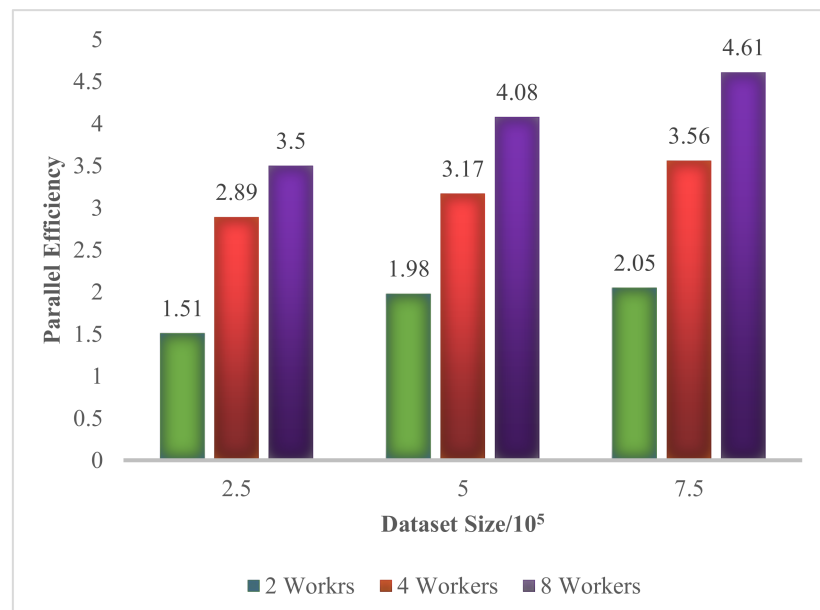
Figure 8. Algorithm parallel efficiency under different dataset sizes.

It can be seen from Figure 8 that under different Amazon data scale sets, the parallel efficiency of the algorithm improves. Taking the running efficiency of two workers as an example, the parallel efficiency of the 2.5×10^5 dataset is 1.51, the parallel efficiency of the 5×10^5 dataset is 1.98, and the parallel efficiency of the 7.5×10^5 dataset is 2.05. Although the parallel efficiency of the algorithm of two workers is not particularly high, with the increase of data sets, the parallel efficiency gradually improves. Taking the running efficiency of four workers as an example, the parallel efficiency of the 2.5×10^5 dataset is 2.89, the parallel efficiency of the 5×10^5 dataset is 3.17, and the parallel efficiency of the 7.5×10^5 dataset is 3.56. This parallel efficiency also gradually improves with the increase of the data set. By comparing the parallel efficiency of two workers and four workers, it can be clearly seen that the increase of the number of workers and the growth of the data set improve the parallel efficiency of the algorithm to a certain extent.

3.2.3. Effectiveness of Knowledge Graph Completion

In order to better compare and test the effectiveness of the relationship between entity nodes for completing the knowledge graph, this paper selects some data from the experimental data set to verify the correct rate of the algorithm. The details are shown in Table 2.

Table 2. Correct rate of the algorithm.

Data Set Proportion	Threshold Value	Number of Entity Nodes	The Number of Entity Nodes in the Algorithm	Correct Rate
10%	0.56	102,576	67,701	0.66
20%	0.54	140,218	93,946	0.67
30%	0.56	185,146	125,899	0.68
40%	0.51	224,514	154,915	0.69
50%	0.52	272,103	190,473	0.70

According to Table 2, in order to better complete the knowledge graph, it is necessary to verify the effectiveness of the algorithm in this paper. Considering that the thresholds of different datasets would be different, this paper calculates the thresholds from the proportions of 10%, 20%, 30%, 40%, and 50% of the datasets. It can be clearly seen from the data in Table 2 that as the proportion of data increases, the accuracy of the algorithm also gradually improves.

Among them, the threshold of 10% of the Amazon dataset is 0.56, and the number of entity nodes is 102,576. The number of entity nodes calculated in the algorithm is 67,701, and the final correct rate is 0.66. The threshold of 30% dataset is 0.56, and the number of entity nodes is 185,146. The number of entity nodes calculated in the algorithm is 125,899, and the final correct rate is 0.68. The threshold of 50% of the dataset is 0.52, the number of entity nodes is 272,103, the number of entity nodes that can be obtained by the algorithm is 190,473, and the final accuracy rate is 0.70. This gradually increasing accuracy also verifies the effectiveness of the algorithm proposed in this paper to a certain extent.

3.3. Improvement Suggestions

Based on the above analysis, it can be seen that knowledge graphs generally have limitations and deficiencies, but due to their wide range of use, it is necessary to complement them to better assist humans in their intelligent life. The algorithm proposed in the experiment only constructs EAG for analysis by calculating the relationship between entity nodes in the graph. This is a relatively innovative superposition method, which can calculate the potential relationship between non-adjacent entity nodes. Using this as a complementing method can effectively deal with the situation that the relationship is incorrectly represented due to sparse nodes. In the specific experiment process, it can also be found that nodes with strong correlation can more easily calculate their mutual relations, but nodes with weak correlation are easily ignored. Therefore, in future experiments, we should consider removing weakly related entity nodes to enhance the effectiveness of knowledge graph completion.

4. Discussion

This paper combines the research methods of probabilistic fuzzy information aggregation and natural language processing technology with the relevant analysis of the completion of knowledge graphs, which is a method to further expand the completion of knowledge graphs. Based on the research purpose of this paper, two specific methods, namely probabilistic fuzzy information aggregation and natural language processing technology, were combined to analyze 1000 Amazon users and related data sets. The effective performance of the algorithm was used to mine the development potential of the two methods in the application of knowledge graph. Finally, the final conclusion of this paper was obtained through the in-depth study of the experimental part.

5. Conclusions

Combined with the analysis of this paper, the following conclusions can be drawn. By describing the node relationship of the entity, the data of the stronger and weaker relationships in the nodes can be fed back. In specific experiments, selective deletions can be made to obtain knowledge graph completion algorithms with better accuracy. There are many methods that can be used to complete the knowledge graph. The related methods proposed in this article can also be analyzed in combination with the description information of the entity. It is necessary to pay attention to the quality and value of multi-source information to prevent incorrect or invalid information from being displayed in the knowledge graph. Of course, when completing the completion, if the relationship between some entities is relatively simple, some precise relationships can be introduced, and the analysis can be carried out by combining probabilistic fuzzy information aggregation and natural language processing technology. With the development of science and technology, natural language processing technology will gradually be used in more places to process more convenient information for people. This paper tests the relationship between entity nodes to complete the effectiveness of the knowledge map. However, this paper only used five groups of data from the experimental dataset to verify the correctness of the algorithm. It is hoped that more data can be collected in the subsequent work to verify the experiment and make the results more representative.

References

1. Reyes, O.; Altalhi, A.H.; Ventura, S. Statistical Comparisons of Active Learning Strategies over Multiple Datasets. *Knowl.-Based Syst.* **2018**, *145*, 274–288. [CrossRef]
2. Ji, K.; Hui, B.; Luo, G. Graph Attention Networks with Local Structure Awareness for Knowledge Graph Completion. *IEEE Access* **2020**, *8*, 224860–224870. [CrossRef]
3. Bai, L.; Ma, X.; Zhang, M.; Yu, W. TPmod: A Tendency-Guided Prediction Model for Temporal Knowledge Graph Completion. *ACM Trans. Knowl. Discov. Data* **2021**, *15*, 1–17. [CrossRef]
4. Che, F.; Zhang, D.; Tao, J.; Niu, M.; Zhao, B. ParamE: Regarding Neural Network Parameters as Relation Embeddings for Knowledge Graph Completion. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 2774–2781.
5. Zhang, Z.; Zhuang, F.; Zhu, H.; Shi, Z.; He, Q. Relational Graph Neural Network with Hierarchical Attention for Knowledge Graph Completion. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 9612–9619.
6. Yu, T.; Li, J.; Yu, Q.; Tian, Y.; Shun, X.; Xu, L.; Zhu, L.; Gao, H. Knowledge graph for TCM health preservation: Design, construction, and applications. *Artif. Intell. Med.* **2017**, *77*, 48–52. [CrossRef] [PubMed]
7. Natthawut, K.; Ryutaro, I. An Automatic Knowledge Graph Creation Framework from Natural Language Text. *IEICE Trans. Inf. Syst.* **2018**, *101*, 90–98.
8. Lin, Z.Q.; Xie, B.; Zou, Y.Z.; Zhao, J.F.; Li, X.D.; Wei, J.; Sun, H.-L.; Yin, G. Intelligent Development Environment and Software Knowledge Graph. *J. Comput. Sci. Technol.* **2017**, *32*, 242–249. [CrossRef]
9. Jia, Y.; Wang, Y.; Jin, X.; Lin, H.; Cheng, X. Knowledge Graph Embedding: A Locally and Temporally Adaptive Translation-Based Approach. *ACM Trans. Web* **2017**, *12*, 1–33. [CrossRef]
10. Wang, C.; Ma, X.; Chen, J.; Chen, J. Information extraction and knowledge graph construction from geoscience literature. *Comput. Geosci.* **2018**, *112*, 112–120. [CrossRef]
11. Roh, J.S.; Jagvaral, B.; Lee, W.G.; Park, Y.T. Approach for Managing Multiple Class Membership in Knowledge Graph Completion Using Bi-LSTM. *J. KIISE* **2020**, *47*, 559–567. [CrossRef]
12. Jagvaral, B.; Kim, M.S.; Park, Y.T. Path Embedding-Based Knowledge Graph Completion Approach. *J. KIISE* **2020**, *47*, 722–729. [CrossRef]
13. Wang, S.; Zhijuan, D.U.; Meng, X. Research progress of large-scale knowledge graph completion technology. *Sci. Sin. Inf.* **2020**, *50*, 551–575.
14. Wu, Y.; Zhu, D.; Liao, X.; Zhang, D.; Lin, K. Knowledge Graph Reasoning Based on Paths of Tensor Factorization. *Moshi Shibie Yu Rengong Zhineng/Pattern Recognit. Artif. Intell.* **2017**, *30*, 473–480.
15. Zhang, C.; Miao, Z.; Xiao, H.; Zheng, H.; Yang, J. Knowledge graph embedding for hyper-relational data. *Tsinghua Sci. Technol.* **2017**, *22*, 185–197. [CrossRef]
16. Danyang, L. Knowledge Graph Analysis of Mathematics Anxiety Research Based on CNKI and Web of Science Database. *Adv. Psychol.* **2021**, *11*, 352–359.
17. Fan, T.; Song, S.; Li, H.; Bai, Y.; Chen, Y.; Cheng, B. Knowledge graph characteristics of sepsis research based on scientometric study. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue* **2021**, *33*, 433–437. [PubMed]
18. Duan, Y.; Hou, L.; Leng, S. A novel cutting tool selection approach based on a metal cutting process knowledge graph. *Int. J. Adv. Manuf. Technol.* **2021**, *112*, 3201–3214. [CrossRef]
19. Peng, Z.; Yu, H.; Jia, X. Path-based reasoning with K-nearest neighbor and position embedding for knowledge graph completion. *J. Intell. Inf. Syst.* **2022**, *58*, 513–533. [CrossRef]
20. He, Y.; Xu, Z.; Jiang, W. Probabilistic Interval Reference Ordering Sets in Multi-Criteria Group Decision Making. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **2017**, *25*, 189–212. [CrossRef]
21. Nandan Challapalli, S.S.; Jaiswal, S.; Bahadur, P.S. Latest Advances of Natural Language Processing and their Applications in Everyday life. *Int. J. Mod. Trends Sci. Technol.* **2020**, *6*, 31–35.