An Improved TransE Algorithm

Haoyu Chang\(^1\), Xin Luo\(^2\), Youqun Shi\(^3\), Xinjie Li\(^4\), and Tao Yang\(^5\)\(^,\ast\)

\(^1\)School of Computer Science & Technology, Donghua University, Shanghai, China
\(^2\)School of Computer Science & Technology, Donghua University, Shanghai, China
\(^3\)School of Computer Science & Technology, Donghua University, Shanghai, China
\(^4\)School of Information Science & Technology, Donghua University, Shanghai, China
\(^5\)Continuing Education College, Donghua University, Shanghai, China

Abstract. This paper proposes the TransE-CBA (TransE Based On Improved Bernoulli And Adam) model, an improved TransE model. We used an improved Bernoulli distribution sampling method to improve the accuracy of negative samples, and used the Adam algorithm to update the gradient to improve the performance of the algorithm. Experiment result shows that, on the FB13 and WN18 datasets, the TransE-CBA model has a lower average ranking score (MeanRank) than TransE as a whole, and the proportion of the top ten triples (HITS@10) Higher, better sorting effect.

1 Overview

Internet content has the characteristics of large-scale and loose organizational structure. These characteristics pose challenges for people to obtain information and knowledge efficiently. In order to improve the capabilities of search engines and improve users’ search experience, Google proposed the concept of Knowledge Graph (KG) on May 17, 2012. With the development and application of artificial intelligence technology, KGs are widely used in many fields, such as Intelligent-searching, Intelligent question and answering, and Personalized-recommendations. At the same time, many large-scale Knowledge Bases have been constructed, such as the Knowledge Base for language named WordNet ([1]) and the Knowledge Base of the world named Freebase ([2]). The essence of the KG is a semantic network that reveals the relationship between entities, and it can formally describe things and their relationships in the real world. As a structured semantic Knowledge Base, Knowledge Graphs usually use the form of triples \((h, r, t)\) to represent knowledge, \(h\) and \(t\) represent head and tail entities, and \(r\) represents relationship. The construction of Knowledge Graph mainly includes knowledge acquisition, knowledge fusion, knowledge verification, knowledge reasoning and application. Knowledge representation is the foundation of the construction and application of Knowledge Graphs. Traditional knowledge representation methods mainly use the triplet which as a standard of the Resource Description Framework (RDF) ([3]) to symbolically describe the relationship between those entities. This method of representation is widely recognized because of its general simplicity, but it faces many problems in terms of computational efficiency and data sparsity.

Representative models for knowledge representation learning include Translation-based models, Structured Embedding (SE) ([4]) models, Semantic Matching Energy (SME) ([5])
models, and tensor Neural Network Models (NTN) ([6]) etc. Among the translation-based models, the most typical one is the TransE ([7]) model, which aims to express entities and relationships into the same embedded vector space. When the triplet \((h, r, t)\) is established in this space, the corresponding embedding vector satisfies \(h + r \approx t\), that is, the tail entity embedding vector \(t\) is closest to \(h + r\) as for the head entity \(h\) translated through the relationship \(r\) ([8]). Literature ([9]) obtained the distributed vectors of words in a certain area through training of knowledge representation models such as TransE. Literature ([10]) added transformation vectors on the basis of TransE model to extend TransE and enhanced the ability of knowledge representation. Reference ([11]) proposes a collaborative filtering recommendation algorithm based on the TransE model, which improves the accuracy of recommendation effectively. The TransE model has fewer parameters and lower computational complexity. It also has good performance and scalability on a large-scale sparse knowledge base, but it is easy to introduce wrong negative examples into the training process with the method of Original Uniform Sampling ([7]). Therefore, the score of the negative example in the loss function has a low value, contrary to expectations. Accordingly, our paper uses an improved Bernoulli sampling method to improve the accuracy of negative examples. On the other hand, in the TransE model, a single learning rate is used to update the weights in the gradient update stage. A single learning rate that is too large or too small will affect convergence. The same learning rate cannot be well adapted to data sets with large differences in sparseness. In response to this, our method uses the Adam (adaptive moment estimation) method ([12]) to update the gradient. This method provides an adaptive learning rate, with a higher convergence rate and a better learning effect. Based on the above two improvements, the algorithm in this paper has a lower average ranking score (MeanRank) as a whole, and the proportion of the top ten triples (HITS@10) is higher, and the ranking effect is better. We named the model as TransE-CBA.

2 Related work

Knowledge graph-oriented representation learning is a new method that supports the calculation and reasoning of knowledge graphs. While retaining the specific attributes of the original graph, the knowledge graph is mapped into a continuous vector space. The low-dimensional vector of the knowledge graph represents a distributed representation ([13]), that is, it expresses a latent feature that has no clear corresponding meaning (called "Feature dimension") when looking at each dimension in the representation vector in isolation, but the integration of each dimension to form a vector can represent the semantic information of the object ([14]).

Among the knowledge representation models, the translation-based model believes that for each triple \((h, r, t)\), the relationship \(r\) is a translation operation from the head entity vector \(h\) to the tail entity vector \(t\) ([14]), and a distance-based scoring function is used to estimate the probability of a triple. There are many translation-based models, such as TransE, TransH ([15]), TransR ([16]), CTransR ([17]), TransD ([18]), etc.

Bordes et al. Proposed the TransE model in 2013, which measures the semantic similarity between computing entities based on the offset in Euclidean distance ([19]). The TransE model believes that if the triple \((h, r, t)\) holds, the addition of the head entity embedding and the relation embedding is approximately equal to the embedding of the tail entity, \(h + r \approx t\), and its scoring function is as follows ([7]), where \(L_1\) is Manhattan distance, \(L_2\) is European distance:

\[ f_c(h, t) = ||h + r - t||_{L_1}/||h + r - t||_{L_2} \]  (1)
For the correct triplet, there should be a lower score, and for the wrong triplet, the score should be higher. For a triple \((h, r, t)\) in the knowledge graph \(S\), there is the following loss formula:

\[
L = \sum_{(h, r, t) \in S} \sum_{(h', r, t') \in S'} \left[ \gamma + \|h + r - t\| - \|h' + r - t'\| \right],
\]

where \(\gamma\) is the interval distance parameter with a value greater than 0, which is a hyperparameter, and the general value is 1. \([\cdot]\) indicates a positive value function:

\[
[x]_+ = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases}
\]

\(S\) is the training set of triples in the knowledge base. \(S'\) is a negatively sampled triplet, obtained by the uniform sampling method [7], that is, one of the correct triplet's head entity, tail entity, or relationship is randomly replaced with other entities or relationships.

\[
S'_{(h, r, t)} = \{ (h', r, t) \in E \} \cup \{ (h, r, t') \in E \}
\]

where \(h'\) and \(t'\) are randomly selected entities in the entity set used to replace the \(h\) and \(t\) entities in the original triple.

In uniform sampling, it is easy to introduce negative examples into the training process. For example, for positive examples (Tang Dynasty, Emperor, Li Shimin), random replacement of Li Shimin for Li Longji (Tang Dynasty, Emperor, Li Longji) as a negative example, but this is not a negative example, but also the correct triplet. Therefore, this method is easy to introduce wrong negative examples into the training set, which leads to a lower negative example score and the opposite of the expected value, reducing the effectiveness of training. According to this, the Bernoulli sampling method is introduced.

In the Bernoulli distribution sampling method ([14]), for the 1-N relationship, the head entity is replaced with a greater probability, and for the N-1 relationship, the tail entity is replaced with a greater probability. For all triples containing the relationship \(r\), we define two statistics: \(tph\) averages how many tail entities each head entity corresponds to; \(hpt\) averages how many head entities each tail entity corresponds to. The parameters of the Bernoulli distribution is as follow ([14]):

\[
P = \frac{tph}{tph + hpt}
\]

Replace the head entity of the triple with probability \(p\), and replace the tail entity of the triple with probability \(1-p\) ([14]). \(tph\) is divided by the number of tail entities corresponding to the current \(r\) by the total number of head entities corresponding to these tail entities, \(hpt\) is divided by the number of head entities corresponding to the current \(r\) entities by the number of tail entities corresponding to these head entities, the method ignores the weight problem in the many-to-one relationship, which leads to the inaccuracy of the parameter \(p\) because it is easy to ignore the sparsity of the data set. This paper proposes a changed Bernoulli sampling method to improve the TransE model to enhance the accuracy of negative examples and increase the performance of the algorithm.

In the TransE model, for each positive triplet, when the loss function is greater than zero, Stochastic Gradient Descent (SGD) ([20]) is used to update the triplet. A single learning rate is used throughout the update process. The basic idea is to use the gradient calculated on a random, small batch of subsets to approximate the true gradient calculated on the entire data set, and iteratively update the weights with small batch samples at each step ([20]):

\[
\theta_{t+1} = \theta_t + \Delta \theta_t
\]

\[
\Delta \theta_t = -\eta \nabla \theta E(\theta_t)
\]

where \(\eta\) is the learning rate, \(\theta_{t+1}\) is the weight value at time \(t+1\), \(\theta_t\) is the weight value at time \(t\), \(E(\theta_t)\) is the loss function about the weight \(\theta_t\) at iteration \(t\), \(\nabla \theta E(\theta_t)\) is the gradient of the weight \(\theta\) with respect to the loss function at time \(t\) and \(\Delta \theta_t\) is the gradient operator. The single learning rate of SGD is less effective on data with sparse features, and is easily
limited to the local best ([20]). Based on this, Adam is used in this paper to replace the traditional stochastic gradient descent process to improve the efficiency of the algorithm.

A series of extended models have also been produced after the TransE model to solve the limitations in dealing with complex relationships. The TransH model allows each entity to have a different representation under different relationships by assuming a hyperplane for each relationship, the entity is projected onto this hyperplane for translation. The projection vector in the TransD model is not only related to the relationship, but also to the entity ([17]). The model sets projection matrices for the head entity and the tail entity separately, and reduces the calculation complexity, it is more suitable for large-scale knowledge graph calculation.

3 TransE-CBA representation method

The TransE model has fewer parameters and high calculation efficiency, and has better performance and scalability on a large-scale sparse knowledge base. Aiming at the problem of easily introducing wrong negative examples into the training set in uniform sampling and the convergence problem caused by the single learning rate of the stochastic gradient descent method, our paper proposes a knowledge graph representation based on the improved Bernoulli sampling method under Adam gradient update, this model is called TransE-CBA model.

3.1 Improved Bernoulli sampling method

Two statistics are defined for each relationship $r$: $tph'$ is the average number of tail entities corresponding to each head entity; $hpt'$ is the average number of head entities corresponding to each tail entity. The sampling method proposed in this paper considers the proportion of tail entities corresponding to head entities in all tail entities when calculating $tph'$, and considers the proportion of head entities corresponding to tail entities in all head entities when calculating $hpt'$, to increase the accuracy of the sampling parameter $p$.

Let $hn$ be the number of head entities corresponding to the current $r$, and let $tn_i$ be the number of tail entities corresponding to these head entities, where $i=\{1,2,3…hn\}$, $tph'$ is calculated as follows:

$$tph' = \frac{m_1}{\sum_{i=1}^{hn} tn_i} \times tn_1 + \frac{m_2}{\sum_{i=1}^{hn} tn_i} \times tn_2 + \ldots + \frac{m_{hn}}{\sum_{i=1}^{hn} tn_i} \times tn_{hn}$$

which is

$$tph' = \frac{\sum_{i=1}^{hn} tn_i^2}{\sum_{i=1}^{hn} tn_i}$$

(8)

Let $tn$ be the number of head entities corresponding to the current $r$, and let $hn_i$ be the number of tail entities corresponding to these head entities, where $i=\{1,2,3…tn\}$, $hpt'$ is calculated as follows:

$$hpt' = \frac{hn_1}{\sum_{i=1}^{tn} hn_i} \times hn_1 + \frac{hn_2}{\sum_{i=1}^{tn} hn_i} \times hn_2 + \ldots + \frac{hn_{tn}}{\sum_{i=1}^{tn} hn_i} \times hn_{tn}$$

which is

$$hpt' = \frac{\sum_{i=1}^{tn} hn_i^2}{\sum_{i=1}^{tn} hn_i}$$

(10)

The parameters of the improved Bernoulli distribution are as follows:

$$p' = \frac{tph'}{tph' + hpt'}$$

(12)

Replace the head entity of the triple with probability $p'$, and replace the tail entity of the triple with probability $1-p'$. 

44
3.2 Adam gradient update

The Adam algorithm is a first-order optimization algorithm that can replace the traditional stochastic gradient descent (SGD) process. It can update the neural network weights iteratively based on the training data, and it is different by calculating the first-order moment estimation and second-order moment estimation of the gradient. Parameter design independent adaptive learning rate. This method reserves a learning rate for each parameter to improve the performance on the sparse gradient, and the average value based on the latest magnitude of the weight gradient adaptively retains the learning rate for each parameter. It can be seen from Fig. 1 ([21]) that compared with the base SGD, Adam is harder to fall into the local best, and the update speed is faster ([21]).

\[ \text{Figure 1. Update process.} \]

The algorithm update parameter method is as follows ([21]):

While \( \theta_t \) not converge do:

\[ t = t + 1 \]
\[ g_t = \nabla f_t(\theta_{t-1}) \]
\[ m_t = \beta_1 m_{t-1} + (1-\beta_1)g_t \]
\[ v_t = \beta_2 v_{t-1} + (1-\beta_2)g_t^2 \]
\[ \theta_t = \theta_{t-1} - \alpha \frac{m_t}{\sqrt{v_t} + \epsilon} \]

End while

where \( \theta_{t-1} \) is the parameter to be updated; \( \alpha \) is the learning rate; \( g_t \) is the gradient of the random objective function; \( m_t \) is the partial first-order moment estimation, \( m_0 = 0 \); \( v_t \) is the partial second-order moment estimation, \( v_0 = 0 \); \( \beta_1 \) and \( \beta_2 \) are the moment estimation Exponential decay rate; \( \epsilon \) is a small positive number ([12]).

3.3 TransE-CBA algorithm

This model regards the relationship in the knowledge base as some kind of translation vector between entities. For each triple \((h, r, t)\), the vector \(l_r\) of the relationship \(r\) is used as the translation between the head entity vector \(l_h\) and the tail entity vector \(l_t\), we can also think of \(l_r\) as a translation from \(l_h\) to \(l_t\) ([19]). The main goal is that if the triple \((h, r, t)\) holds, the head entity embedding and the relation embedding add up approximately equal to
the tail entity embedding, which is \( h+r=t \), and its scoring function is as follows, where \( L_1 \) is Manhattan distance, \( L_2 \) is European distance:

\[
f'_f(h, t) = \|h + r - t\|_{L_1/L_2}
\]

For the correct triplet, there should be a lower score. For a triplet \((h, r, t)\) in the knowledge graph \( S \), we can use the loss formula to train:

\[
L = \sum_{(h, r, t) \in S} \sum_{(h', r', t') \in S'(h, r, t)} [\gamma + \|h + r - t\| - \|h' + r' - t'\|]_+
\]

\( S \) is the training set of triples in the knowledge base. \( S' \) is a negatively sampled triplet. In our algorithm, one of the correct triplet's head entity, tail entity, and relationship is randomly replaced with other entities or relationships, where for all triplets containing the relationship \( r \), the triplet is replaced by the probability \( p' \) with head entity, replacing the tail entity of the triple with probability \( 1 - p' \). among them:

\[
p' = \frac{\text{typ}}{\text{typ} + \text{hp'}}
\]

\( \text{tp}h \) is the average number of tail entities corresponding to each head entity; \( \text{hp}' \) is the average number of head entities corresponding to each tail entity. For calculation formula see section 3.1.

where \( \gamma \) is the interval distance parameter with a value greater than 0, which is a hyperparameter, and the general value is 1. \([\cdot]_+ \) indicates a positive value function:

\[
[x]_+ = \begin{cases} 
  x, & x > 0 \\
  0, & x \leq 0 
\end{cases}
\]

First, select a positive triple \((h, r, t)\), and then sample from \( S \) to get a negative triple \((h', r', t')\). Then calculate the positive score \( f'_f(h, t) \) and the negative score \( f_r(h', t') \), if \( f_r(h, t) - f_s(h', t') + \gamma > 0 \), then use the Adam algorithm to update the gradient \( h, r, t, h', r', t' \). For the specific steps, see section 3.2. The algorithm flow is shown in Table 1.

**Table 1.** Algorithm flow of TransE-CBA.

| Algorithm | TransE-CBA |
|-----------|-----------|
| **Input** | Training set \( S = \{(h, r, t)\} \), entities and rel. Sets \( \alpha, E \) and \( L \), margin \( \gamma \) embeddings dim. \( k \) |
| **1:** initialize \( l \leftarrow \text{uniform}(\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}}) \) for each \( l \in L \) |
| **2:** \( l \leftarrow 1/\|l\| \) for each \( l \in L \) |
| **3:** \( e \leftarrow \text{uniform}(\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}}) \) for each \( e \in E \) |
| **4:** loop |
| **5:** \( e \leftarrow e/\|e\| \) for each \( e \in E \) |
| **6:** \( S_{\text{batch}} \leftarrow S \text{ sample (}S, b\text{) } // \text{sample a minibatch of size } b \) |
| **7:** \( T_{\text{batch}} \leftarrow \emptyset // \text{initialize the set of pairs of triplets} \) |
| **8:** for \((h, r, t) \in S_{\text{batch}}\) do |
| **9:** \( p' = \frac{\text{tp}h'}{\text{tp}h + \text{hp'}} \) |
| **10:** \((h', r', t') \leftarrow \text{sample } (S'(h, r, t)) \text{ use } p' // \text{Sample a corrupted triplet} \) |
| **11:** Calculate loss use Adam Update gradient |
| **12:** \( T_{\text{batch}} \leftarrow T_{\text{batch}} \cup \{(h, r, t), (h', r', t')\} \) |
| **13:** end for |
| **14:** Update embeddings w.r.t. \( \sum_{(h, r, t), (h', r', t') \in T_{\text{batch}}} \nabla[\gamma + d(h + r - t) - d(h' + r' - t')]_+ \) |
| **15:** end loop |
4 Analysis of experimental results

4.1 Experimental data set

In this paper, two data sets FB13 ([22]) and WN18 ([23]) are used to evaluate the effectiveness of the TransE-CBA model. Among them FB13 is a subset of Freebase ([24]), WN18 is a subset based on WordNet ([25]). The experimental data set is shown in Table 2.

| Data  | FB13     | WN18     |
|-------|----------|----------|
| #Relation | 13       | 18       |
| #Entity   | 75043    | 40943    |
| #Train    | 316232   | 141442   |
| #Valid    | 5908     | 5000     |
| #Test     | 23733    | 5000     |

4.2 Experimental content

The experiment mainly completes the link prediction task. The given head entity and the relationship prediction tail entity or the given head and tail entity predict the relationship between the two entity keys, that is, given $h$, $r$ predicts $t$ or given $h$, $t$ predicts $r$. For the original triple $(h, r, t)$ in the test set, the probability $p'$ calculated by the improved Bernoulli sampling method is used to randomly replace the entity $h$, and the tail entity $t$ is replaced by $1-p'$ to obtain the mutated triple, using $f_r(h,t)$ to calculate the scores of the original and mutant triples, and sorts the score results. The average ranking score ($\text{MeanRank}$) and the proportion of triples ranked within top10 ($HITS@10$) are used to measure the ranking results of the original triples. The erroneous mutation that will be generated when generating a mutant triplet, that is, the new triplet generated after replacing the entity happens to be an original triplet. Perform filtering to ensure that the generated mutation triplets belong to the test set, training set, and verification set, this process is called Filter. The unfiltered case is called Raw. Compared to Raw, Filter has better test results, with higher $HITS@10$ and lower $\text{MeanRank}$.

4.2 Experimental results

In our experiment, TransE-CBA, TransE, TransH, and TransD are tested on two data sets, FB13 and WN18, respectively. On the FB13 data set, the initial learning rate is $\alpha=0.01$, $\gamma=1$, $k=60$, and $L_1$ is used as Distance function. On the WN18 data set, $\alpha=0.01$, $\gamma=1$, and $k=100$. In addition to TransE-CBA using Adam algorithm for gradient update on the two data sets, other models use SGD method.
**Figure 2.** Comparison of MeanRank results in FB13.

**Figure 3.** Comparison of MeanRank results in WN18.

**Fig. 2** and **Fig. 3** are the comparison of the effects of several models on MeanRank under the FB13 and WN18 data sets. From the figure, it can be seen that the TransE-CBA model has a better average ranking score compared to TransE, of which FB13 The ranking score is reduced by 7.8% (Raw) and 1.5% (Filter), and it is reduced by 19% (Raw) and 27.3% (Filter) on WN18. Since TransH and TransD introduce the concept of hyperplane, the timetable is better when dealing with one-to-many, many-to-one, and many-to-many relationships, so there are better test results overall.

**Figure 4.** Comparison of HITS@10 results in FB13.
Fig. 4 and Fig. 5 are the comparison of the effects of several models on HITS@10 under the FB13 and WN18 data sets. From the figure, it can be seen that the TransE-CBA model has better ranking results compared to TransE, of which in FB13 the proportion of the top ten triples increased by 17.5% (Raw) and 3% (Filter), and compared to other models HITS@10 results are better. Increased by 2.45% on WN18. Since TransH and TransD introduce the concept of hyperplane, the timetable is better when dealing with one-to-many, many-to-one, and many-to-many relationships, so there are better test results overall.

Conclusion

In this paper, we proposed the TransE-CBA model. This model uses the improved Bernoulli distribution sampling method to generate negative examples for the 1-N and N-1 relations during the generation of negative examples. The weights are introduced on the basis of the original sampling method. Entities with more head (tail) entities account for a larger percentage when calculating the average. On the other hand, for the defect that the traditional stochastic gradient descent method has a single learning rate and is easy to fall into the local best advantage, the Adam algorithm is used to replace the original gradient update method, and the adaptive learning rate is used.

Experimental results show that this algorithm has a certain improvement in link prediction compared to the TransE model, has a lower average ranking score (MeanRank), and the proportion of the top ten triples (HITS@10) Higher, that is, there is a better sorting result but the improvement is not obvious. The next step is to optimize the improved Bernoulli sampling algorithm.

The authors would like to thank Guangdong Province collaborative innovation and platform Environmental Science build of special funds (2014B090908004); Dongguan City professional town innovation service platform construction project "Dongguan City Humen garment Collaborative Innovation Center", whose constructive comments and suggestions helps us to improve the quality of this paper.

References

1. Miler G A. WordNet: A lexical database for English[J].Communications of the ACM, 1995, 38(11):39-41.
2. Bolacker K, Evans C, Paritosh P, et al. Freebase: A collaboratively created graph database for structuring human knowledge[C]//Proc of KDD. New York: ACM, 2008:1247-1250.
3. Miler E. An introduction to the resource description framework [J]. Bulletin of the American Society for Information Science and Technology, 1998, 25(1):15-19.
4. Bordes A, Weston J, Collobert R, et al. Learning structured embeddings of knowledge bases[C] //Proc of AAAI. Menlo Park, CA: AAAI, 2011:301-306.
5. Bordes A, Glorot X, Weston J, et al. A semantic matching energy function for learning with multi-relational data [J]. Machine Learning, 2014, 94(2):233-259.
6. Socher R, Chen D, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion[C]//Proc of NIPS. Cambridge, MA: MIT Press, 2013:926-934.
7. Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data [C] // Proc of NIPS. Cambridge, MA: MIT Press, 2013:2787-2795.
8. Han Xiao. Research on distributed representation of knowledge graph [D]. Beijing University of Posts and Telecommunications, 2019.
9. Wang Fei, Tan Xin, Yi Mianzhu, Chen Yongsheng. Research on knowledge discovery in Tibet based on TransE model [J]. Journal of Information Engineering University, 2018, 19 (04): 441-446.
10. García-Durán A, Bordes A, Usunier N. Composing relationships with translations [C] Proceedings of the 2015 conference on EMNLP. 2015:17-21.
11. Wu Xiyu, Chen Qimai, Liu Hai, He Chaobo. Collaborative filtering recommendation algorithm based on knowledge graph representation learning [J] .Computer Engineering, 2018, 44 (02): 226-232 + 263.
12. Kingma D P, Ba J. Adam: A Method for Stochastic Optimization [J]. Computer Science, 2014.
13. Turian J, Ratinov L, Bengio Y. Word representations: A simple and general method for semi-supervised learning[C] Proc of ACL. Stroudsburg, PA: ACL, 2010:384-394.
14. Fang Yang, Zhao Xiang, Tan Zhen, Yang Shiyu, Xiao Weidong. An improved translation-based knowledge graph representation method [J]. Computer Research and Development, 2018, 55(01): 139-150.
15. Wang Zhen, Zhang Jianwen, Feng Jianlin, et al. Knowledge graph embedding by translating on hyperplanes[C]//Proc of AAAI 2014.Menlo Park, CA: AAAI, 2014:1112-1119.
16. Lin Yankai, Liu Zhiyuan, Sun Maosong, et al. Learning entity and relation embeddings for knowledge graph completion [C]//Proc of AAAI 2015. Menlo Park, CA: AAAI, 2015.
17. Ji Guoliang, He Shizhu, Xu Liheng, et al. Knowledge graph embedding via dynamic mapping matrix [C]//Proc of ACL 2015.Stroudsburg PA: ACL, 2015:687-696.
18. Wang Q, Mao Z, Wang B, et al. Knowledge Graph Embedding: A Survey of Approaches and Applications[J]. IEEE Transactions on Knowledge & Data Engineering, 2017, 29(12):2724-2743.
19. Liu Zhiyuan, Sun Maosong, Lin Yankai, Xie Ruobing. Research progress in knowledge representation learning [J]. Computer Research and Development, 2016, 53(02): 247-261.
20. Zhang Jinjing. Neural network weight optimization algorithm based on stochastic gradient descent [D]. Southwest University, 2018.
21. Tang Ye, Lu Weizhong, Chen Cheng, Huang Hongmei. Illumination calculation method based on Adam algorithm and neural network [J]. Journal of Lighting Engineering, 2019, 30 (02): 50-54.
22. Freebase[EB/OL].(2011-07-18)[2011-07-15].http://wiki.freebase.com/wiki/What_is_Freebase%3F.

23. Yao Tianshun, Zhang Li, Gao Zhu. Summary of WordNet [J]. Application of Language and Characters, 2001 (01): 27-32.

24. Li Jun. Freebase Research on Semantic Database [J]. Modern Library and Information Technology, 2011 (10): 18 -23.

25. Miller, G.A, Beckwith, R., Fellbaum, C., Gross, D., and Miller, K., “Introduction to WordNet: An on-Line Lexical Database”, in Five Papers on WordNet, CSL report, Cognitive Science Laboratory, Princeton University, 1993.