ISSN: 2664-9640

Research on the Construction of Dinosaur Encyclopedia Knowledge Graph

Jinxuan Xie*

School of Automation and Electronic Information, Sichuan University of Science and Engineering; Yibin, Sichuan, China

* Corresponding Author

Abstract

The storage and inheritance of knowledge is the basis for the continuation of human civilization. With the development of natural language processing technology, the emergence of knowledge graphs makes the preservation and application of knowledge more efficient, and further promotes the development of youth science education. In order to improve the efficiency of information acquisition in the field of Dinosaur Encyclopedia, due to the particularity of Dinosaur Encyclopedia data, this article mainly focuses on manual collection, assisted by web crawler technology, to obtain Dinosaur Encyclopedia knowledge data. Use regularization, Chinese word segmentation and other technologies for data cleaning, use natural language technology to extract information from multi-source knowledge, name entity recognition, knowledge fusion, and sort out the entity and entity relationship obtained from entity extraction, and import it into the graph database Neo4j to construct knowledge Atlas to facilitate subsequent applications and research.

Keywords

Dinosaur Encyclopedia; Knowledge Graph; Web Crawler; Graph Database; Named Entity Recognition.

1. Introduction

Knowledge Graph [1] is an efficient knowledge expression model formally proposed by Google in 2012. Knowledge graph can help people to acquire the logical relationship between knowledge more quickly, which is conducive to the realization of intelligent reasoning between knowledge. It uses a series of string symbols to map to various entities or concepts that exist in the real world, and then uses the association relationship between these entities or concepts asconnectors to link different types of information together to form a huge Semantic Web. The essence of the knowledge graph is the semantic network, and the difference between it and the semantic network lies in its large-scale. The scale effect of the knowledge map has also brought about a qualitative change in utility, which is a product of the era of big data.

At present, with the continuous development of intelligent information service applications, knowledge graphs have been widely used in fields such as intelligent search, intelligent question and answer, and personalized recommendation [2]. Knowledge graphs can be divided into two categories: open domain knowledge graphs and vertical domain knowledge graphs based on their coverage. The open domain knowledge graph is usually not limited to a specific domain. It contains a lot of common sense knowledge, and it pursues the breadth of knowledge. The open domain knowledge graph is mainly used in search engines and intelligent question answering and other services. For example, YAGO[3], Freebase[4], DBpedia[5], Wikidata[6], these are relatively well-known English knowledge graphs; OpenKG is a Chinese open knowledge graph platform, which aims to promote the openness and interconnection of

ISSN: 2664-9640

Chinese knowledge graphs. Promote the popularization and application of knowledge graph technology in China, and contribute to the development of artificial intelligence and innovation and entrepreneurship in China. At present, 35 institutions have settled in. It has attracted the participation of the most famous knowledge graph resources in China, such as Zhishi.me[7], CN-DBPedia, PKUBase. It already contains 15 categories of open knowledge graphs from common sense, medical care, finance, urban, and travel.

The development of knowledge graph still has the following obstacles [8]. First of all, although the era of big data has produced a huge amount of data, the data release lacks standardization and the data quality is not high. Mining high-quality knowledge from these data needs to deal with the problem of data noise. Secondly, the construction of knowledge graphs in the vertical domain lacks resources in natural language processing, especially the lack of dictionaries, which makes the construction of knowledge graphs in the vertical domain very expensive, and it is difficult to implement a general knowledge graph construction platform.

The basic unit of the knowledge graph consists of the triples of head entity, relationship, and tail entity [9]. Although this method is very effective in representing structured data, there are still two problems: computational efficiency and data sparseness [10], In this regard, knowledge representation learning [11] and a better knowledge graph embedding (Knowledge Graph Embedding) [12],[13],[14] method are proposed.

In the ACL 2019 conference, Allen Lab proposed the Commonsense Transformers (COMET) [15] generative model. The main framework is the Transformer [16] language model. The seed knowledge training set is selected in the ATOMIC and ConceptNet knowledge bases for pre-training, so that the model can be automatically Build a knowledge base of common sense. The nature of this automatic construction also needs to be based on the existing triples. In terms of encyclopedic knowledge, the general knowledge graph, there are now enough open training sets, based on the industry knowledge graph of healthcare, finance, e-commerce, etc. In the early stage, a large amount of manual labeling of triples is also required. In particular, open data sets generally come from the accumulation of foreign institutions over the years, and there is still a big gap in the accumulation of Chinese knowledge bases.

The pre-trained model of natural language processing is currently the hottest topic. The BERT[17] model proposed by Google in the NAACL 2019 conference has achieved very good results by predicting the blocked words and using Transformer's multi-layer self-attention two-way modeling capabilities. However, the modeling object of the BERT model is mainly focused on the original language signal, and the semantic knowledge unit is less used for modeling. This problem is particularly obvious in Chinese. To this end, Baidu proposed the ERNIE[18] model in the ACL 2019 conference, which learns real-world semantic knowledge by modeling the words, entities, and entity relationships in massive data. Compared with BERT learning the semantic representation of local language co-occurrence, ERNIE directly models the semantic knowledge and enhances the semantic representation ability of the model.

Zigong is the most famous dinosaur producing area in the Sichuan Basin. A variety of dinosaurs such as sauropods, theropods, ornithopods, stegosaurus, etc. have been discovered in Zigong. The ages range from the early, middle and late Jurassic, filling the world where dinosaur fossils are scarce in the early and middle Jurassic. The missing ring, coupled with the large number of dinosaur fossils discovered in Zigong, and the complete preservation are rare in the world, so Zigong dinosaurs are famous in the world.

With the development of the Internet, web crawlers, and knowledge graph technology, it has greatly facilitated the collection, storage and display of knowledge. In the context of the country’s vigorous development of artificial intelligence, we should use the method of combining old industries with new technologies. To increase the influence of Zigong's dinosaur tourism industry in the new consumption and new network era, to enhance the fun and

ISSN: 2664-9640

technological nature of dinosaur visits, and to strengthen the popular science education for young people. This topic is based on Dinosaur Encyclopedia, combined with knowledge graph technology to collect relevant knowledge of various dinosaurs, and combine dinosaurs and dinosaur encyclopedia knowledge by means of knowledge links, so that users can learn about dinosaurs more conveniently and directly, and based on knowledge With Tupu's question-and-answer system, users can use natural language to exchange knowledge with intelligent customer service and learn more about dinosaur stories.

2. Methodology

2.1. Research Strategy

The purpose of this research is to establish a knowledge map of Dinosaur Encyclopedia, in order to improve the efficiency of data query and storage, to develop a question and answer system, to benefit the local tourism industry, and to enhance the science education of young people.

Due to the particularity of Dinosaur Encyclopedia data, the data in this article is mainly collected manually, supplemented by web crawler technology. The data sources are mainly Baidu Encyclopedia, Wikipedia, Dinosaur Encyclopedia, related documents and books in museum collections.

In terms of technology, this article mainly uses Chinese word segmentation technology and regularization to clean up the data, and then extracts entities and relationships from the data, and then stores the data in a graph database. The graph database uses Neo4j, which combines structured data with The form of graph storage, based on Java implementation (now also provides Python interface), is a high-performance data system with full transaction characteristics, with all the characteristics of a mature database. The query language it uses is cypher, and the creation of knowledge graph nodes and relationships (create command) and query (match command) can be realized through Neo4j.

2.2. Web Crawler Technology

The implementation principle and process of general web crawlers can be briefly summarized as follows:

(1) Get the initial URL. The initial URL address can be manually specified by the user, or it can be determined by one or several initial crawling webpages specified by the user.

(2) Crawl the page according to the initial URL and obtain a new URL. After obtaining the initial URL address, you first need to crawl the webpage in the corresponding URL address. After crawling the webpage in the corresponding URL address, store the webpage in the original database, and while crawling the webpage, discover the new URL At the same time, the crawled URL address is stored in a URL list for de-duplication and judging the crawling process.

(3) Put the new URL in the URL queue. In step 2, after obtaining the next new URL address, the new URL address will be placed in the URL queue.

(4) Read the new URL from the URL queue, crawl the web page according to the new URL, obtain the new URL from the new web page at the same time, and repeat the above-mentioned crawling process.

(5) When the stop condition set by the crawler system is met, the crawling is stopped. When writing a crawler, the corresponding stop conditions are generally set. If the stop condition is not set, the crawler will keep crawling until it cannot obtain a new URL address. If the stop condition is set, the crawler will stop crawling when the stop condition is met.

The process of crawling the data of the encyclopedia website this time is shown in Figure 2-1:

ISSN: 2664-9640

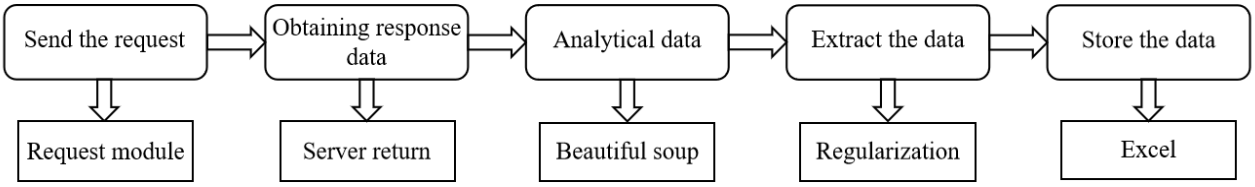

Figure 2-1 Flow chart of crawling encyclopedia website by crawler

After the crawler crawls the data, it needs to perform data cleaning and statistical analysis on the data, so that it is easy to store in the database.

2.3. Knowledge Graph Construction

Use the cleaned data to build a knowledge map, the overall technical block diagram is shown in Figure 2-2:

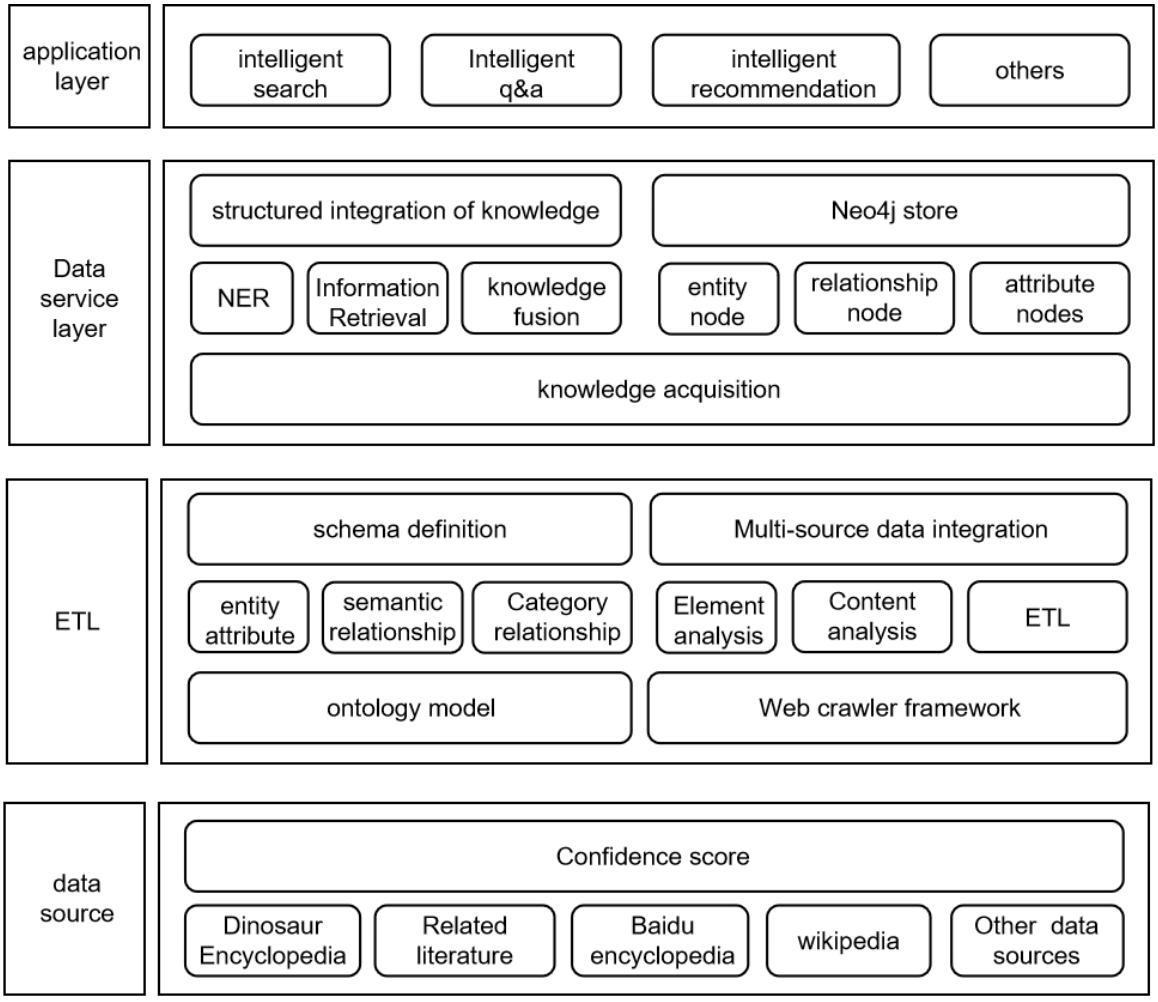

Figure 2-2 Overall technical block diagram

Through the analysis of the dinosaur encyclopedia data, the knowledge map knowledge is divided into static knowledge and dynamic knowledge. Among them, static product data sources are mainly Dinosaur Encyclopedia, Wikipedia, CNKI and other high-quality websites, and dynamic data are Dinosaur Forum and news information. Encyclopedia knowledge is stored in the form of triples, which are composed of three elements: head entity, relationship, and tail entity, and each element contains its own attributes. Examples of triples are shown in Table 2-3:

ISSN: 2664-9640

Table 2-1 Examples of dinosaur triads

Head entity	Describe relations	Tail entity
Liliensternus	Length of dinosaurs	2-5m
Liliensternus	The dinosaur weight	100-140kg
Liliensternus	S survival	Triassic

Ontology Construction

Construct the ontology structure of the knowledge map of Dinosaur Encyclopedia based on the defined knowledge map pattern layer, that is, the framework of the knowledge map. The process of ontology construction is quite cumbersome, and the construction process is often different depending on the respective fields and specific projects. Ontology construction usually has three construction methods: manual, automatic and semi-automatic. Due to the huge workload of manual construction, the coverage and accuracy of current automatic construction methods are still relatively low. This study chooses a semi-automatic construction method for ontology construction. Use statistical and unsupervised methods to obtain ontology knowledge, combine it with ontology knowledge of other knowledge graphs, and complete it through manual annotation.

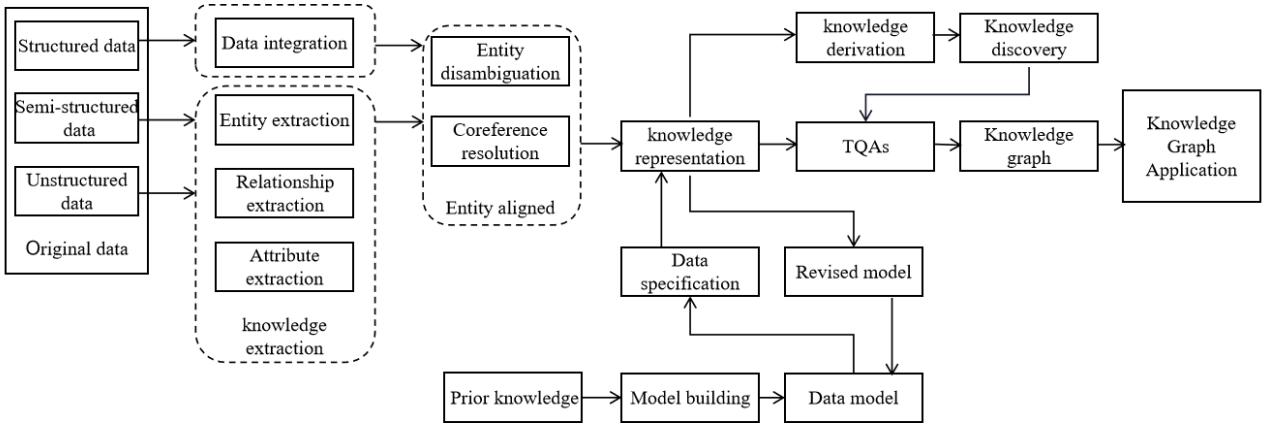

Figure 2-3 Flow chart of knowledge graph construction

Informatica Extraction

Entity extraction is also called named entity recognition, which is to automatically identify named entities from text. Plan to use the word vector method to use data to expand the entity set in the knowledge graph. Named entities refer to entity nouns with specific meanings, such as proper nouns such as names of persons, organizations, and places. Entities are the most basic elements in the knowledge graph, and their performance will directly affect the quality of the knowledge base. According to the characteristics of NER extraction technology, entity extraction technology can be divided into rule-based methods, statistical machine learning-based methods, and deep learning-based methods.

Relation extraction refers to extracting the relationship between entities and entities from the text, so that the scattered entities can be connected. Attribute extraction is to extract the attribute information of the entity from the text. Since the attribute of the entity can be regarded as a nominal relationship between the entity and the attribute, the attribute extraction problem can also be regarded as the relationship extraction problem. Three methods of unsupervised, supervised and semi-supervised can be used to extract relations.

The goal of relationship extraction is to extract the semantic relationship between two or more entities, so that the knowledge graph truly becomes a graph. The research on relation extraction is based on the MUC (Message Understanding Conference) evaluation conference and the ACE

ISSN: 2664-9640

(Automatic Content Extraction) evaluation conference that later replaced MUC. At present, the relationship extraction methods can be divided into two methods: template-based relationship extraction and relationship extraction based on supervised learning.

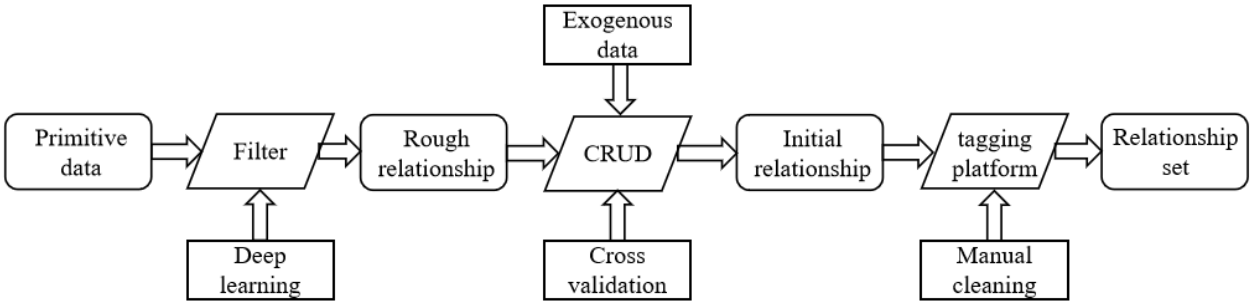

Figure 2-4 Relationship extraction process

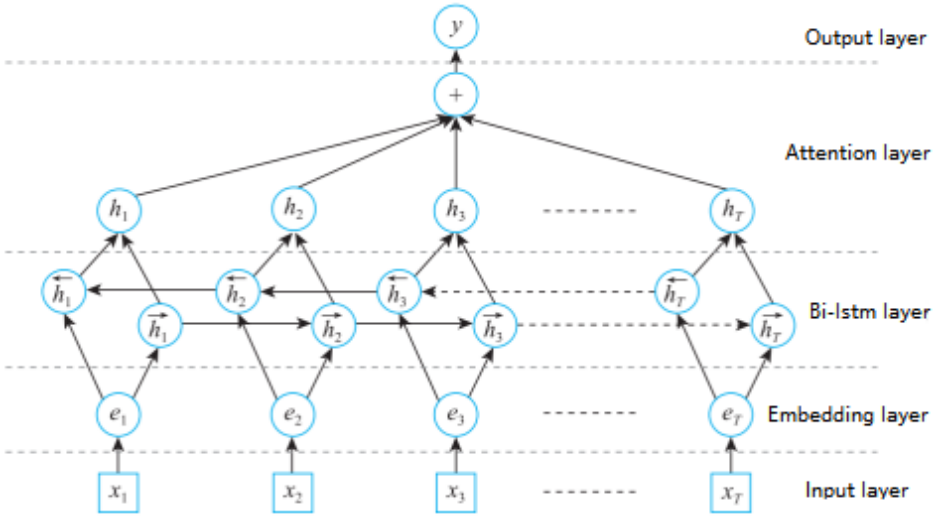

Figure 2-5 Classic deep learning relation extraction model architecture

Knowledge Fusion

Through knowledge extraction, we get a large number of entities (attributes) and relationships, but due to the difference in description and writing, there are a lot of redundancy and error information in the results. It is necessary to disambiguate, clean and integrate these data. As a key technology of knowledge fusion, the purpose of Entity Linking is to link the entity object extracted from the text to the uniquely determined entity object corresponding to it in the knowledge base to realize entity disambiguation and coreference resolution.

Entity Disambiguation is specifically used to solve the ambiguity problem of entities with the same name. The simplest method is to construct feature vectors through the attributes of the entities and surrounding words, and evaluate the similarity of two entities through the cosine similarity of the vectors. Based on this idea, we can have more semantic-based methods to characterize the target entity, so as to evaluate whether the two entities are the same.

Co-referential resolution (Entity Resolution) refers to solving the problem that multiple entities written in different ways point to the same entity. Generally, this kind of problem can be solved by referring to the entity disambiguation method, or it can be solved by specific analysis of specific problems and some rule methods.

Knowledge Storage

Classic relational databases are less flexible in dealing with complex entity relationships, and the effect is not satisfactory, while graph databases can achieve higher query efficiency when dealing with complex relationships and entity mapping, and are more humane in relational

ISSN: 2664-9640

storage. Therefore, it is planned to use the open source Neo4j graph database as a knowledge storage database and MySQL relational database as an auxiliary.

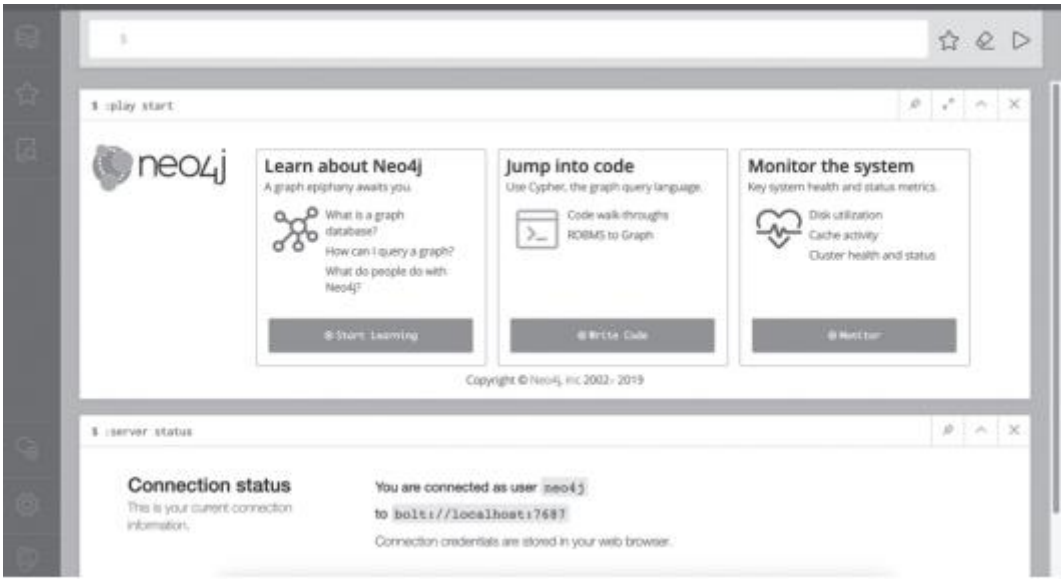

Figure 2-6 Neo4j Web visualization interface

Neo4j is an open source graph database. It stores structured data in the form of graphs. It is implemented based on Java (now also provides Python interface). It is a high-performance data system with full transaction characteristics and has all the characteristics of a mature database. The query language it uses is cypher, and the creation of knowledge graph nodes and relationships (create command) and query (match command) can be realized through Neo4j.

3. Algorithms and Research Results

3.1. Algorithms

Named Entity Recognition (NER) is a very basic task in NLP. NER is an important basic tool for many NLP tasks such as information extraction, question answering systems, syntax analysis, and machine translation. The accuracy of named entity recognition determines the effect of downstream tasks, which is a very important basic issue in NLP.

The current mainstream method for NER is to use LSTM as the feature extractor, and then connect a CRF layer as the output layer. Although CNN has weaknesses in feature extraction of long sequences, the CNN model has parallel capabilities and has the advantage of fast calculation speed. The introduction of dilated convolution enables CNN to take into account the calculation speed and feature extraction of long sequences in the NER task. BERT contains a lot of general knowledge. Using a pre-trained BERT model and using a small amount of labeled data for FINETUNE is a fast way to obtain a good NER, as shown in the figure below.

The Chinese Named Entity Recognition (NER) algorithm can be roughly divided into two types: Character-based (character) and Word-based (word) according to the type of input. Both of these two methods have some shortcomings. Character-based cannot use vocabulary and vocabulary sequence information; Word-based requires word segmentation, and word segmentation errors will have a greater impact on NER results. This paper uses a Lattice LSTM algorithm, which can use vocabulary information and avoid the impact of word segmentation errors on the model.

ISSN: 2664-9640

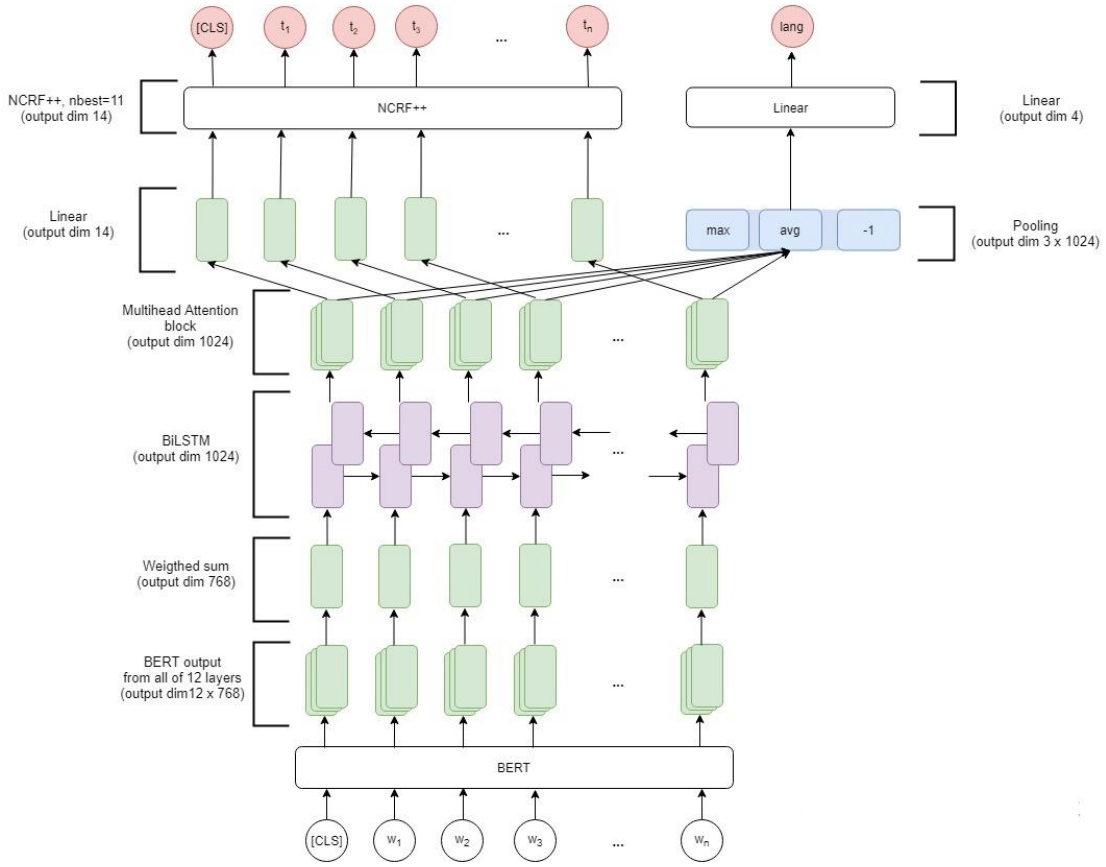

Figure 3-1 BERT+(LSTM)+CRF model

Lattice LSTM adds word-base cell and extra gate structure to Char-LSTM to control and select information flow. Some mathematical symbols used in Lattice LSTM are as follows: c_i represents the i-th character of the sentence, $w_{b,e}^{d}$ represents the word, b is the starting position of the word, e is the ending position, $x_j^c = e^c(c_j)$ represents the embedding of the j-th character, $x_{b,e}^{w} = e^{w}(w_{b,e}^{d})$ represents the word $w_{b,e}^{d}$ Embedding. The output of the traditional Char-LSTM calculation formula mainly includes cell state c and hidden state h. The calculation formula is as follows:

$$\begin{bmatrix} \boldsymbol{i}_j^c \\ \boldsymbol{o}_j^c \\ \boldsymbol{f}_j^c \\ \tilde{\boldsymbol{c}}_j^c \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{bmatrix} \left(W^{cT} \begin{bmatrix} x_j^c \\ \boldsymbol{h}_{j-1}^c \end{bmatrix} + b^c \right)$$

$$\boldsymbol{c}_j^c = \boldsymbol{f}_j^c \odot \boldsymbol{c}_{j-1}^c + \boldsymbol{i}_j^c \odot \tilde{\boldsymbol{c}}_j^c$$

$$\boldsymbol{h}_j^c = \boldsymbol{o}_j^c \odot \tanh\left(\boldsymbol{c}_j^c\right)$$

i_j^c represents the LSTM input gate, o_j^c represents the LSTM output gate, and f_j^c represents the LSTM forget gate. The above is the calculation formula of Char-LSTM. On this basis, Lattice LSTM adds word-base cell to calculate the cell state c of the word subsequence. The word-base cell calculation formula is as follows to generate a cell state containing word information:

ISSN: 2664-9640

$$\begin{bmatrix} \boldsymbol{i}_{b,e}^{w} \\ \boldsymbol{f}_{b,e}^{w} \\ \tilde{\boldsymbol{c}}_{b,e}^{w} \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \tanh \end{bmatrix} \left(W^{wT} \begin{bmatrix} x_{b,e}^{w} \\ \boldsymbol{h}_{b}^{c} \end{bmatrix} + b^{w} \right)$$

$$\boldsymbol{c}_{b,e}^{w} = \boldsymbol{f}_{b,e}^{w} \odot \boldsymbol{c}_{b}^{c} + \boldsymbol{i}_{b,e}^{w} \odot \tilde{\boldsymbol{c}}_{b,e}^{w}$$

In the above formula, the word-base cell does not include the output gate vector, because Lattice LSTM only outputs on Char-LSTM. After getting the cell state of word-base, Lattice needs to be fused to the cell state of Char-LSTM. This process requires adding an input gate vector and normalizing the input gate vector, as shown below:

$$\boldsymbol{i}_{b,e}^{c} = \sigma \left(W^{l^{T}} \begin{bmatrix} x_{e}^{c} \\ \boldsymbol{c}_{b,e}^{w} \end{bmatrix} + b^{l} \right)$$

$$\boldsymbol{\alpha}_{b,j}^{c} = \frac{\exp\left(\boldsymbol{i}_{b,j}^{c}\right)}{\exp\left(\boldsymbol{i}_{j}^{c}\right) + \sum_{b' \in \left\{ b'' | w_{b'',j}^{d} \in \boldsymbol{D} \right\}} \exp\left(\boldsymbol{i}_{b',j}^{c}\right)}$$

$$\boldsymbol{\alpha}_{j}^{c} = \frac{\exp\left(\boldsymbol{i}_{j}^{c}\right)}{\exp\left(\boldsymbol{i}_{j}^{c}\right) + \sum_{b' \in \left\{ b'' | w_{b'',j}^{d} \in D \right\}} \exp\left(\boldsymbol{i}_{b',j}^{c}\right)}$$

D is the dictionary, and finally Lattice LSTM gets the cell state of the j-th character as follows:

$$\boldsymbol{c}_{j}^{c} = \sum_{b \in \left\{ b' | w_{b',j}^{d} \in \boldsymbol{D} \right\}} \boldsymbol{\alpha}_{b,j}^{c} \odot \boldsymbol{c}_{b,j}^{w} + \boldsymbol{\alpha}_{j}^{c} \odot \tilde{\boldsymbol{c}}_{j}^{c}$$

The hidden state h calculation formula of Lattice LSTM is the same as Char-LSTM, and finally h is passed to the CRF layer for named entity recognition. Lattice LSTM can use the information of Chinese characters and words at the same time. By adding word-base cell and control gate, the information of characters and words can be selected to eliminate ambiguity. But because the number of added word nodes is different between characters, Lattice LSTM does not support batch training. If there are too many matching words in the sentence, the effect of Lattice LSTM will deteriorate (it may degenerate into word-based LSTM) and will be affected by word segmentation errors.

3.2. Research Results

A total of more than 700 kinds of dinosaur encyclopedia data were obtained through manual collection and crawler collection, and through data processing, information extraction, and data integration, the construction of the dinosaur encyclopedia knowledge graph was completed. At present, a total of 32,492 entities and 28,177 relationships have been constructed in the knowledge base. It includes more than ten attributes such as head height, hip height, body length, weight, distribution area, age of survival, main food, references, etc. In the graph database Neo4j, the dinosaur encyclopedia knowledge graph uses graph nodes and relationship edges. Store knowledge, the construction of the Dinosaur Encyclopedia knowledge graph transforms unstructured knowledge into structured knowledge, which lays the foundation for the later recommendation system based on the Dinosaur Encyclopedia knowledge graph, the question and answer system, and the research of knowledge mining and machine learning. It is a popular science education and humanities. The development of tourism provides ideas.

ISSN: 2664-9640

Figure 3-2 An example of the knowledge map of Dinosaur Encyclopedia

4. Conclusion

In this paper, a relatively complete dinosaur encyclopedia knowledge graph is constructed with the help of Neo4j graph database, and the named entity recognition algorithm is integrated with the help of domain ontology and vocabulary set, which lays the foundation for the application of knowledge graph.

This subject still has many shortcomings and parts to be implemented, such as algorithm improvement, applicability and accuracy enhancement, and the graph database can be further improved. In addition, the knowledge graph based on Dinosaur Encyclopedia can continue to develop a question and answer system, recommend The system, etc., will be further developed and improved in the later stage.

References

[1]. AMIT S. Introducing the knowledge graph[R]. America: Official Blog of Google, 2012.

[2]. Quan Wang, Zhendong Mao, Bin Wang, et al. Knowledge Graph Embedding: A Survey of Approaches and Applications[J]. IEEE Transactions on Knowledge & Data Engineering, 2017, PP(99):1-1.

[3]. F. M. Suchanek, G. Kasneci, G. Weikum. YAGO: A core of semantic knowledge[A]. Proceedings of the 16th international conference on World Wide Web[C]. Banff, Alberta, Canada, 2007, pp. 697–706.

[4]. K. Bollacker, C. Evans, et al. Freebase: A collaboratively created graph database for structuring human knowledge[A]. Proceedings of the 2008 ACM SIGMOD international conference on Management of data[C]. Vancouver, Canada, 2008, pp. 1247-1250.

ISSN: 2664-9640

[5]. J Lehmann, R Isele, M Jakob, et al. DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia[J]. Semantic Web, 2015, 6(2):167-195.

[6]. Vrande, Denny, M Tzsch. Wikidata: A Free Collaborative Knowledgebase[J]. Communications of the ACM, 2014, 57(10):78-85.

[7]. X Niu, X Sun, et al. Zhishi.me: weaving Chinese Linking Open Data[A]. International Conference on the Semantic Web[C]. Springer-Verlag, 2011, pp. 205-220.

[8]. Qi Guilin, Gao Huan, Wu Tianxing. Research Progress of Knowledge Graph[J]. Information Engineering, 2017, 3(01):4-25.

[9]. L. A. Galárraga, C Teflioudi, et al. AMIE: Association rule mining under incomplete evidence in ontological knowledge bases[A]. Proceedings of the 22nd international conference on World Wide Web[C]. ACM, 2013.

[10]. Liu Zhiyuan, Sun Maosong, Lin Yankai, et al. Research progress in knowledge representation learning[J]. Journal of Computer Research and Development, 2016, 53(02):247-261.

[11]. A Bordes, N Usunier, A García-Durán, et al. Translating Embeddings for Modeling Multi-relational Data[A]. Proc of NIPS. Cambridge[C]. MA: MIT Press, 2013, pp. 2787-2795.

[12]. Z. Wang, J. Zhang, J. Feng, and Z. Chen, Knowledge graph embedding by translating on hyperplanes[A]. 28th AAAI Conference on Artificial Intelligence[C]. 2014, pp. 1112–1119.

[13]. G Ji, S He, L Xu, et al. Knowledge Graph Embedding via Dynamic Mapping Matrix.[A]. Meeting of the Association for Computational Linguistics[C]. Beijing, China, 2015.

[14]. G Ji, K Liu, S He, et al. Knowledge Graph Completion with Adaptive Sparse Transfer Matrix[A]. 30th AAAI Conference on Artificial Intelligence[C]. AAAI Press, Phoenix, Arizona, 2016, pp. 985-991.

[15]. A Bosselut, H Rashkin, M Sap, et al. COMET: Commonsense Transformers for Automatic Knowledge Graph Construction[A]. 57th Annual Meeting of the ACL[C], Italy: Florence, 2019.

[16]. A Vaswani, N Shazeer, N Parmar, et al. Attention is All you Need[A]. Advances in Neural Information Processing Systems 30th[C]. Curran Associates, Inc., 2017: 5998–6008.

[17]. D Jacob, C Ming-Wei, L Kenton, T Kristina. BERT: Pre-training of deep bidirectional transformers for language understanding[A]. In Proceedings of NAACL-HLT[C]. America: Minneapolis, 2018.

[18]. Z Zhengyan, H Xu, L Zhiyuan, et al. ERNIE: Enhanced Language Representation with Informative Entities[A]. 57th Annual Meeting of the ACL[C], Italy: Florence, 2019.