information

MDPI

Article

Entity Linking Method for Chinese Short Text Based on Siamese-Like Network

Yang Zhang, Jin Liu *, Bo Huang and Bei Chen

School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China
* Correspondence: liujin@sues.edu.cn; Tel.: +86-137-6193-6139

Abstract: Entity linking plays a fundamental role in knowledge engineering and data mining and is the basis of various downstream applications such as content analysis, relationship extraction, question and answer. Most existing entity linking models rely on sufficient context for disambiguation but do not work well for concise and sparse short texts. In addition, most of the methods use pre-training models to directly calculate the similarity between the entity text to be disambiguated and the candidate entity text, and do not dig deeper into the relationship between them. This article proposes an entity linking method for Chinese short texts based on Siamese-like networks to address the above shortcomings. In the entity disambiguation task, the features of the Siamese-like network are used to deeply parse the semantic relationships in the text and make full use of the feature information of the entity text to be disambiguated, capturing the interdependent features within the sentences through an attention mechanism, aiming to find out the most critical elements in the entity text description. The experimental demonstration on the CCKS2019 dataset shows that the *F*1 value of the method reaches 87.29%, increase of 11.02% compared to the *F*1 value(that) of the baseline method, fully validating the superiority of the model.

Keywords: Chinese short text; entity linking; named entity recognition; entity disambiguation

check for updates

Citation: Zhang, Y.; Liu, J.; Huang, B.; Chen, B. Entity Linking Method for Chinese Short Text Based on Siamese-Like Network. *Information* **2022**, *13*, 397. https://doi.org/10.3390/info13080397

Academic Editor: Ralf Krestel

Received: 10 June 2022
Accepted: 10 August 2022
Published: 22 August 2022

1. Introduction

Entity Linking (EL) is considered an integral part of natural language processing and is often used as pre-processing for missions such as information extraction [1] and intelligent question and answer [2]; it also plays a critical role in the construction and updating of knowledge bases by linking different texts with structured information.

At present, the reason why entity linking technology has been significantly developed in English is that the language is not limited by word segmentation, and there are many knowledge bases of entity linking. In the early stages, the shortcomings in the development of Chinese entity linking were mainly as follows: (1) Chinese word segmentation and entity recognition have a more significant impact on entity linking; (2) Chinese entity linking started late and lacked a more mature knowledge base. OpenKG integrates 120 Chinese knowledge base projects, such as the graph database of Peking University, the CN-DBpedia, and CN-Probase projects of Fudan University, and the three major Chinese encyclopaedia projects, forming a relatively developed Chinese semantic knowledge base. Chinese entity linking has developed by leaps and bounds. However, the effect of short text entity linking is still not ideal.

Nowadays, Chinese short texts, such as questions and answers, queries, and headlines, are essential in the Internet corpus. For short texts, it is difficult to determine the specific meaning of a word without contextualizing it, so accurately linking entities in short texts remains a significant challenge.

With the rise of pre-training language models, entity linking methods based on pre-training language models have become a focus of research in recent years. Word2Vec and GloVe were widely used until 2019. Kolitsas et al. added contextual information encoding on top of character-level embedding in order to make word embeddings aware of the importance of local context, resulting in better experimental results. Gupta et al. [3] obtained the meaning of entities in different contexts from introducing external data sources to enrich the semantic and contextual information of the entities. The method utilises GloVe as a source of word embedding, and uses an LSTM-based encoder to capture lexical and syntactic local contextual information of mentions. The word embedding produced by Word2Vec is static and independent of the sentence in which the word is located. For example, in the sentences "I want to access my bank account" and "We went to the river bank", the vector of "bank" given by Word2Vec is fixed. Often the same entity is represented in different sentences with various meanings, requiring consideration of contextual semantic information. The advantage of BERT lies in the fact that the generated word embedding is dynamic and able to deal with ambiguity.

Entity linking can be seen as a problem of semantic matching between the text where the entity is mentioned and the text describing the candidate entity. Two methods are currently used to calculate the semantic similarity of sentence pairs: Cross-Encoder and Bi-Encoder. Cross-Encoder: two sentences are passed simultaneously to the BERT model, which produces an output value between 0 and 1, indicating the similarity of the input sentence pair. Cheng et al. [4] stitched the entity text to be disambiguated and the candidate entity description text, and inputted them into the BERT model. The probability of the candidate entity was then obtained by concatenating the vector output at the [CLS] position with the feature vector of the candidate entity. Logeswaran et al. [5] combined the mention of context and entity description text, using [SEP] to separate the text as input to the BERT, and subsequently scored the output. For many applications, Cross-Encoder is not practical (time-consuming), does not produce sentence embedding, and does not allow comparison using cosine similarity. Whereas Bi-Encoders can generate separate sentence embedding, Siamese networks are those that define two network structures to characterise the sentences in a sentence pair separately. Reimers et al. [6] modified BERT to use Siamese and triple network structures to derive semantically meaningful sentence embedding that can be compared using cosine similarity. This is improved by considering that in Siamese network, encoding only sentence pairs does not take full advantage of the information in the text of the entities to be disambiguated.

With the advent of pre-training models, the ability of machines to understand text content has become increasingly powerful. However, for disambiguation aspect, the similarity between the entity description text to be disambiguated and the candidate entity description text, is calculated directly through a pre-training model, which has the disadvantage of not making full use of the text information of the entity to be disambiguated. This paper proposes a short Chinese text entity linking method based on a Siamese-like network to solve the above drawbacks. RoBERTa-BiLSTM is used to generate sentence-level embeddings, extract key features in the text through an attention mechanism, and use the properties of the Siamese-like network for information transfer to achieve the purpose of text information reuse. By deeply parsing the semantic relationship of the text, mining the correlation between the entity text to be disambiguated and the candidate entity description text, and leveraging feature information of the text to be disambiguated to solve the ambiguity problem, the accuracy of entity linking in short texts can be improved. Taking kb_data in CCKS2019 as the basic knowledge base, the experimental verification of this method shows that the $F1$ value reaches 87.29%.

The rest of the paper is structured as follows. The latest techniques and various approaches to entity linking are discussed in Section 2. Section 3 introduces the entity linking model, explains the named entity recognition model, and details the entity disambiguation model in this paper. Section 4 presents the pre-processing and the results of the experiments. In Section 5, the conclusions and prospects are presented.

2. Related Work

Entity linking is a crucial technique to resolve the problem of words with multiple meanings and multiple words with the same meaning. Entity linking plays a crucial role in question and answer [7], text analysis [8], information retrieval [9], and knowledge base augmentation [10]. For question and answer, entity linking is an immediate need for KBQA, as the graph database can only be queried for subsequent work after the entity has been accurately linked. Entity linking works in two ways: end-to-end [11] and linking-only [12]. In the former, the entities in the text (i.e., NER) are extracted first and then the extracted entities are disambiguated. In the latter, on the other hand, the text and the entities are mentioned directly as input (the default entities are already identified), and candidate entities are found and disambiguated. The ultimate aim of both ways of working is to accurately match ambiguous entities. Most of the time, it is considered that entity disambiguation is equivalent to entity linking, except that the step of named entity recognition is missing. Many scholars [13,14] view entity linking as a binary classification problem. For example, given <entity referent term, candidate entity>, a dichotomous model is used to determine whether they match or not. Nowadays, scholars have proposed a series of effective methods to disambiguate entities, divided into two approaches: (1) methods based on graph models; (2) methods based on deep learning.

Alhelbawy et al. [15] used a graph-based approach to perform joint entity disambiguation. All entities in the text are represented as nodes in the graph, which are ranked based on their PageRank values, and entity disambiguation is performed according to the magnitude of the numerical value. Zhang et al. [16] designed a unified framework LinKG to merge different types of knowledge graphs, and proposed a Heterogeneous Graph Attention Network (HGAN) to match entities with ambiguous moulds correctly (e.g., human names). Hu et al. [17] used a graph neural network model to solve the entity disambiguation problem. An entity-word heterogeneous graph is constructed for each document to build global semantic relationships among disambiguated entities in the text, obtain entity graph embeddings that encode global semantic features, and transfer the embedding representation to a statistical model to disambiguate entities. However, graph-based models treat all candidate entities the same, which leads to increased difficulty in differentiation, Fang et al. [18] designed a Sequence Graph Attention Network (SeqGAT), which takes full advantage of graph and sequence methods to dynamically encode before and after entities, exploiting local consistency and reducing noise interference. In the context-limited short text with insufficient number of entities and relationships, the graph model-based approach ignores the entity-to-entity relationship, and therefore cannot obtain the global semantic features of entities. To address this problem, Zhang et al. [19] first predicted the classification of entities to be disambiguated, which yielded the set of candidate entities based on category, and then found the association with the candidate entities for disambiguation through a knowledge graph. The knowledge graph itself is a graph-based data structure that not only contains information about entities, but also holds the relationship between entities, which can display the global semantic features of entities. Zhu et al. [20] designed a Category2Vec embedding model based on joint learning of word and category embeddings to calculate the semantic similarity between contextual and informational words of entities in the knowledge graph, and improve the performance of disambiguation methods based on context similarity.

In 2018, Google proposed the BERT pre-training model [21]. Since then, the large-scale pre-training language model has become a mainstream approach, allowing the field of natural language processing to develop by leaps and bounds. A large number of researchers have improved and fine-tuned the model and achieved better performance on numerous sub-tasks of natural language processing (e.g., entity linking). Zhan et al. [22] used BERT as a basic framework to extract the connections between the context of entities to be disambiguated and candidate entities and refine entity associations by analysing entity interrelationships. Later, [23] introduced a multi-task learning approach into the short textual entity linking process to build a multi-task learning model for entity linking and classification, to mitigate the problem of insufficient information in the short textual entity linking process. Zhang et al. [24] stitched the context of the entity mention with the description text of the candidate entities and introduced a local attention mechanism to alleviate the long-distance problem and enhance local context information. Zhao et al. [25] presented a BERT-Random Walk with Restart (BERT-RWR) entity linking model by combining deep neural networks and graph neural networks, aiming to compute the semantic similarity between entity mentions and candidate entities, with better results. Gu et al. [26] designed a multi-round reading framework to perform the entity linking of short texts. Generating a separate question for each entity to be disambiguated, forming a candidate set through the entities in the knowledge base, and using the linked entities to update the current interrogative is equivalent to adding a contextual description of the entities to be disambiguated, thus improving the accuracy of entity linking. Barbara et al. [27] defined the entity disambiguation task as a text extraction task, where the text to be disambiguated and all candidate entity texts are stitched together. The output of the model is then the index of words starting and ending with the target option, and the method achieves optimality on several datasets.

In this paper, the disambiguation entities in the text are identified using the RoBERTa-WWM-BiLSTM-CRF (Conditional Random Field) network framework, and the information of the text to be disambiguated is fully utilized for entity disambiguation using the properties of a Siamese-like Network.

3. Entity Linking

For a given Chinese short text, the entity recognition model is used to identify the entity designation items contained in the text to form the set of entities to be disambiguated; then, the entities to be disambiguated are matched by name in the knowledge base to construct the candidate entity set; finally, the best entity is selected through the analysis of text similarity.

A piece of text is represented as: $s = \{c_1, c_2, \ldots, c_n\}$, where c_i represents the i-th character. The purpose of the entity recognition model is to use the BIOES marking method to mark each character. The named entity recognition model can be used to retrieve entity mentions in the text. The entity mention set is represented as $M = \{ m_1, m_2, \ldots, m_j \}$, where j is the number of entity referents. Each identified entity is an object for disambiguation (entity mention), and a candidate entity set is constructed by matching entities with the same name from the knowledge base, e.g., the candidate entity set of m_j is $D = \{ d_{j1}, d_{j2}, \ldots, d_{jn} \}$, where n is the number of entities with the same name in the knowledge base. Since the number of entities in the knowledge base is vast and it is time-consuming to compare each entity in the knowledge base, the best method is to filter out the entity objects that are not likely to be the target entities by generating the candidate entity set. The details of the procedure for entity linking are shown in Figure 1.

3.1. Named Entity Recognition

This paper uses RoBERTa-WWM-BiLSTM-CRF as a network architecture for named entity recognition. RoBERTa-WWM [28] is an improvement on BERT by using the Chinese whole word masking technique in the pre-training phase, and using HIT LTP (Language Technology Platform) as a word separation tool. This makes the semantic representation generated by the model possess word information, which is more suitable for Chinese natural language processing tasks to some extent. The pre-training language model RoBERTa-wwm is used to generate a semantic representation of the text of the entity to be disambiguated; the resulting semantic representation is fed into a BiLSTM network to capture contextual information; and the sequence relationships between the labels are restricted using conditional random fields to obtain the best results to ensure the legitimacy of the prediction. The Entity Recognition (ER) model is shown in Figure 2.

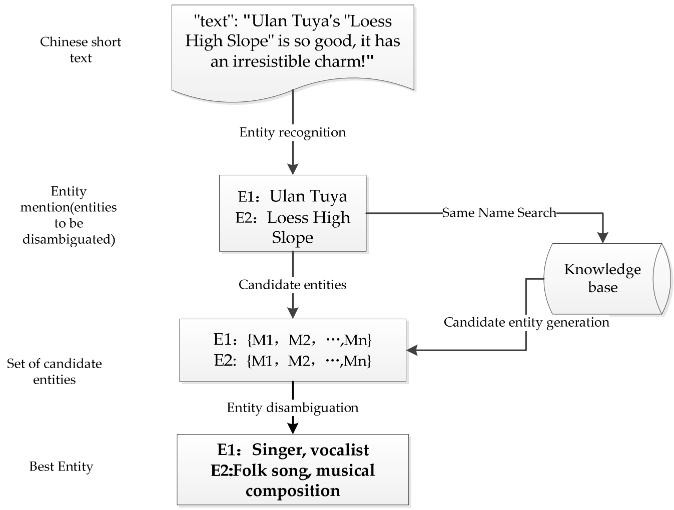

Figure 1. Entity linking flow chart.

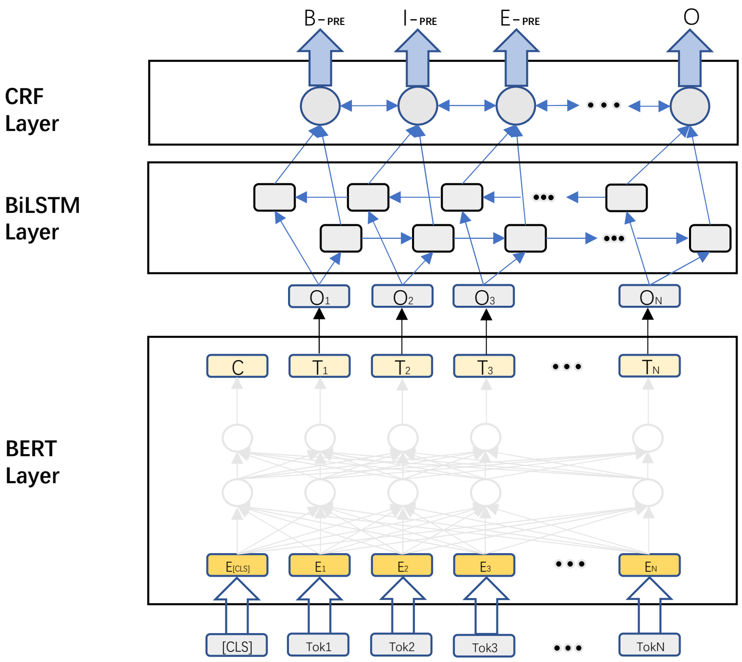

Figure 2. Entity Recognition Model (ER).

First, the input of RoBERTa-WWM is the cumulative sum of the character vector, text vector, and position vector. The input text needs to be processed, and the input text is defined as $s = \{c_1, c_2, \ldots, c_n\}$. The initial character vector $\{E_{CLS}, E_1, \ldots, E_n, E_{SEP}\}$, sentence s is obtained by adding $[CLS]$ tags at the beginning of the text and $[SEP]$ labels at the end of the text, which the RoBERTa-WWM model processes. The text vector is divided into 0 and 1, where 0 indicates the first sentence and 1 indicates the second sentence, corresponding to the named entity recognition task for which there is only one sentence. Hence, the type of the text vector is 0. To learn the time series features, BERT uses location vectors to add time-series information by varying the sine and cosine so that the location encoding contains information about the relative position of the context.

$$PE(pos, 2i) = sin\left(pos/1000^{2i/d_{model}}\right) \tag{1}$$

$$PE(pos, 2i+1) = cos\left(pos/1000^{2i/d_{model}}\right) \tag{2}$$

BERT Layer: RoBERTa-WWM is a structure containing 12 layers of Transformer encoder, which uses a bidirectional transformer encoding structure to analyse the weight of text through a self-attention mechanism, to evaluate the relationship between each one of the words in a piece of text and all the terms in the whole part of the text, to obtain the association of each word with all the text, which allows better learning of contextual linguistic information, where Q, K and are the feature vectors for calculating the weights, V is the vector representing the input features, and $\sqrt{d_k}$ is the deflation factor.

$$(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

To capture richer features in the text and calculate the attention of all words in a sentence, Transformer uses a "multi-headed attention" mechanism.

The Transformer coding unit incorporates residual networks and layer normalization to address the problem of gradient disappearance and explosion in deep learning.

$$LN(x_i) = \alpha \times \frac{x_i - \mu_L}{\sqrt{\sigma_L^2 + \varepsilon}} + \beta \tag{4}$$

$$FFN = max(0, xW_1 + b_1)W_2 + b_2 \tag{5}$$

BiLSTM Layer [29]: In the classical RNN (Recurrent Neural Network) algorithm, the LSTM [30] (Long Short-Term Memory) is used for the long-term dependency problem. In-text, the meaning of words changes with the context in that the words are related to the text above the words and related to the text below. The core of an LSTM is made up of the following components: input character x_t at the moment t, unit state C_t, temporary unit state $\widetilde{C}_t$, hidden layer state h_t, forgetting gate f_t, memory gate i_t, and output gate o_t. The following equation represents the structure.

$$f_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_f\right) \tag{6}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{7}$$

$$\widetilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{8}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{9}$$

$$o_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_0) \tag{10}$$

$$h_t = o_t * tanh(C_t) \tag{11}$$

The BiLSTM model results from the combination of the forward and backward LSTM models. The output of the BERT model is used as the input of the BiLSTM model, and then the output of the BiLSTM model's forward and backward dissemination is spliced so that it can extract the context information in the text.

CRF Layer [31]: Conditional Random Fields (CRF) is a popular conditional probability model in natural language processing that can be used to derive conditions for maximizing the probability of a sequence. BiLSTM only considers information about the context in the text and does not consider the dependencies between the tags. In this paper, the tag sequence is modelled using the CRF layer to ensure the legitimacy of the predicted titles by adding constraints to get the best results.

3.2. Entity Disambiguation

This paper only considers disambiguation using entity mention context and description text of candidate entities, and proposes a short Chinese textual entity linking method based on Siamese-like networks for this process. This paper uses RoBERTa [32]-BiLSTM as the central network architecture for text encoding, by connecting the Attention layer to capture the interdependent features between sentences. By focusing on the hidden states of candidate entity texts through the semantic representation of the text to be disambiguated, the aim is to identify the most critical features in the entity descriptions, evaluate the semantic associations between them using cosine similarity, and select the best entities. The Entity Disambiguation (ED) model is shown in Figure 3.

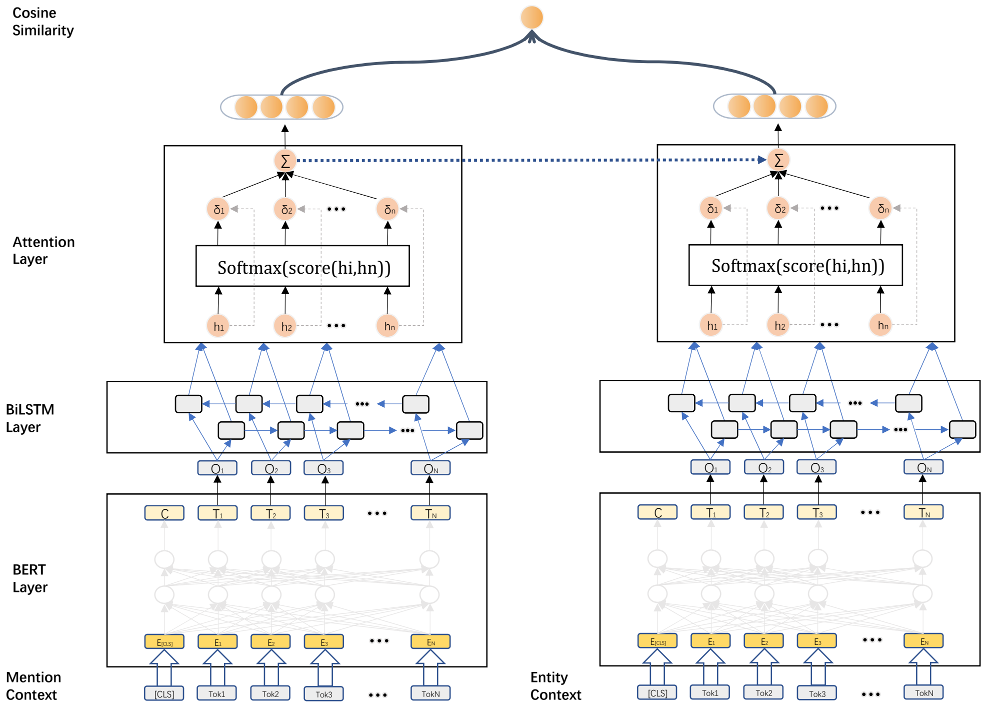

Figure 3. Entity disambiguation model (ED).

Compared with other models, BERT has the advantage of being able to observe text sequences dynamically. As an embedding layer in the similarity calculation model proposed in this paper, word sequences are received via BERT, and feature vectors are generated simultaneously by BERT. The generated feature vectors are passed through the BiLSTM layer to gain the semantic-level connections in the clause and form a whole sentence-level feature representation. Each BiLSTM encoder will be multiplied by the respective weight of a word to highlight the vital messages in the sentence.

The input on the left side of Figure 3 is the text of the entity description to be disambiguated, and the hidden state sequence of its BERT-BiLSTM output is $H_m = \{h_1^m, h_2^m, \ldots, h_n^m\}$. The forward hidden state output $\overrightarrow{h_i}$ and the backward hidden state output $\overleftarrow{h_i}$ are stitched together, and their metrics are shown as follows:

$$h_i^m = \left[\overrightarrow{h_i^m}; \overleftarrow{h_i^m}\right] \tag{12}$$

Stitching the forward and reverse hidden states with the results of going through the Attention layer,

$$g_m = [h_i^m; h_{Attention}^m] \tag{13}$$

where the Attention hidden layer is calculated as follows:

$$\alpha_i^m = \omega_\alpha^T h_i^m \tag{14}$$

$$\alpha_i^m = \frac{\exp(\alpha_i^m)}{\sum_{k=b}^{e} \alpha_k^m} \tag{15}$$

$$h^m_{Attention} = \sum_{k=b}^{e} \alpha^m_k h^m_i \tag{16}$$

The resulting g^m is then passed through a superficial fully connected layer $(Dense)$, which is used to obtain the final entity mention representation:

$$M^r = Dense(g_m) \tag{17}$$

The input on the right side of Figure 3 is the candidate entity description text, whose BERT-BiLSTM hidden state sequence output is $H_e = \{h^e_1, h^e_2, \ldots, h^e_n\}$, and the stitching result of the hidden state output is h^e_i. This paper uses the attention mechanism to generate the entity representation, which is slightly different from the attention mechanism used on the left side. Here, the entity mention representation M^r is used to focus on the hidden state of the candidate entity, aiming to identify the most critical feature points in the entity description. The entity representation E^r based on the attention mechanism is shown as follows:

$$\alpha^e_i = score(M^r, h^e_i) \tag{18}$$

$$\alpha^e_i = \frac{\exp(\alpha^e_i)}{\sum_{t=1}^{n} \alpha^e_t} \tag{19}$$

$$E^r = \sum_{t=1}^{n} \alpha^e_t h^e_t \tag{20}$$

where α^e_i is the attention score M^r and the hidden state of the candidate entity h^e_i. The calculation is performed using a splicing approach, and the formula is as follows:

$$score(M^r, h^e_i) = v^T_a tanh(W_a[M^r; h^e_i]) \tag{21}$$

The use of a Siamese-like network model allows for the full exploitation of the semantic features of the text to be disambiguated and the candidate entity description text. Given the semantic text representation M^r of the entity to be disambiguated, and the candidate entity description text semantic representation E^r, where i is the number of candidate entities mentioned by the entity, using cosine similarity between these two vectors represents their semantic relevance, namely:

$$P(M^r, E^r_i) = consine(M^r, E^r_i) \tag{22}$$

In the prediction process, the cosine metric is used to calculate the similarity between the contextual reference pairs and each candidate entity, and the one with the most resemblance is selected as the final result.

4. Experiments

4.1. Dataset

The data used for the experiments is the CCKS2019 dataset. The CCKS2019 training set provides 90,000 tagged sentences, and the test set provides 30,000 untagged sentences, most of which have an average length of 22; the knowledge base includes 400,000 entities. The training data includes a piece of text and a mention_data data field, which includes the span of entity mention and the corresponding ID of the knowledge base. The knowledge base contains fields of "subject_id", "subject", "alias", "data", etc. The "data" fields contain multiple predicate and object fields.

4.2. Data Pre-Processing

The task of entity linking is to match the entities in the text to be disambiguated with the knowledge base, so the contents of the text and the knowledge base need to be processed.

Entity Description Dictionary: concatenates all tuples (prediction, object) from the "data" field into a single sentence to represent the entity description.

Alias Dictionary: Since the number of entities in the knowledge base is enormous, it is time-consuming to pair up each entity in the knowledge base. The best way is to filter out the entities that are not likely to become the target entities is by generating the candidate entity set. This paper constructs an "Alias Dictionary" to screen the candidate entities by searching the "Alias Dictionary" fields to generate the candidate entities. The Alias Dictionary is shown in Table 1.

Table 1. Alias Dictionary.

Subject	Alias
Bridge to the Stars	Singbridge, Bridge to the Stars, ...
Robin Williams	Robin Williams, Robin William, ...
Scud missiles	Scud, Scud missiles, ...

EntityName Dictionary: The "EntityName Dictionary" is constructed by analysing the entity names in the knowledge base. After construction, each entity name corresponds to one or more entity IDs, e.g., "LiQing" corresponds to multiple entity IDs. The EntityName Dictionary is shown in Table 2.

Table 2. EntityName Dictionary.

Entity (Subject | Alias)	Entity ID
LiQing	10015, 23433, 27209, 31339, 42267, ..., 408084

4.3. Evaluation Indicators

There must be standard criteria in considering the merits of the various methods of entity disambiguation. In this paper, precision (P), recall (R), and $F1$-score ($F1$) are used to discriminate the accuracy of entity linking. The $F1$-score ($F1$) is the balance of precision and chances of completeness, which integrates the overall effect.

There are N entity mention in Text: $M_n = \{m_1, m_2, \ldots, m_n\}$ each mention is linked to the entity of the knowledge base: $E_n = \{e_1, e_2, \ldots, e_n\}$, the output of the model is: $E'_n = \{e'_1, e'_2, \ldots, e'_n\}$:

Precision (P):

$$P = \frac{\sum_{n\in N} |E_n \cap E'_n|}{\sum_{n\in N} |E'_n|} \times 100\% \tag{23}$$

Recall (R):

$$R = \frac{\sum_{n\in N} |E_n \cap E'_n|}{\sum_{n\in N} |E_n|} \times 100\% \tag{24}$$

$F1$:

$$F1 = \frac{2PR}{P+R} \times 100\% \tag{25}$$

4.4. Experimental Results

The results of the named entity recognition model tested on the CCKS2019 dataset are shown in Table 3, which mainly compare the effects of using different BERT models on the same dataset.

Table 3. Named entity recognition experimental results.

Model	*P*/%	*R*/%	*F1*/%
BERT-BiLSTM-CRF	86.38	85.24	85.81
BERT-WWM-BiLSTM-CRF	87.63	87.55	87.59
Our Model	88.15	88.38	88.26

To test the validity of the entity disambiguation methods in this paper, separate experiments were conducted on the CCKS2019 dataset in different ways. Table 4 shows the comparative experimental results. As can be gleaned from Table 4, the RoBERTa performs slightly better as the embedding layer of the entity disambiguation model.

Table 4. Entity disambiguation experiment results.

Model	*P*/%	*R*/%	*F*1/%
Word2Vec-BiLSTM	68.74	66.98	67.85
BERT-BiLSTM	78.68	78.61	78.64
BERT-CNN	80.24	79.36	79.80
bert-base-chinese	84.66	83.80	84.23
RoBERTa-WWM	84.85	85.25	85.05
Our Model	87.07	87.51	87.29

The above table shows that BERT has obvious advantages compared with Word2Vec. This is because Word2Vec produces static word representations that do not take into account the semantic information of the context, whereas for entity linking, the problem of multiple meanings of a word and multiple words with the same meaning requires contextual resolution. The advantage of BERT is that the resulting word representations are dynamic and capable of handling textual ambiguity.

In the above table, it can be seen that the BERT-CNN framework obtains better results than BERT-BiLSTM. The primary reason is that the text processed in this paper is a short Chinese text, and the advantage of BiLSTM in processing long text is not reflected, so the performance of CNN is slightly more effective than BiLSTM.

In order to find the most suitable BERT model for processing Chinese text matching, this paper conducts experiments using several popular BERT models as Siamese-like network embedding layers simultaneously, such as bert-base-chinese, RoBERTa-wwm, and RoBERTa (both the experiments conducted for this paper use the 12-layer model BERT-base). From the results in Table 4, we can learn that RoBERTa performs better in the task of text matching.

For the correlation degree between the Chinese short text and the knowledge base entity description text, this paper obtains the sentence-level feature representation through the RoBERTa-BiLSTM model; it captures the interdependent features between sentences through the Attention layer; it focuses the information obtained from the entity description text to be disambiguated on the candidate entity description text information; and it finds out the most critical feature points in the candidate entity description text, so as to filter out the candidate entity with the best matching degree. Therefore, this model outperforms other models in the short Chinese text entity linking task.

5. Conclusions and Feature Work

This paper proposes an entity linking method based on a Siamese-like network, which is different from the traditional method of simply calculating the similarity between entity mentions and entities. This paper profoundly analyses the semantic relationship in the text, not only by feature representation of the disambiguated entity text features and the candidate entity text features, but also by making full use of the text features of the entity to be disambiguated in order to extract the most critical information. The experimental performance shows that the method proposed in this paper achieves better accuracy in Chinese short textual entity linking compared with the baseline model.

Future work will examine how to adapt the current model to other languages for entity linking issues and multimodal entity linking. Data in the real world is not all structured or semi-structured data, but more of a "picture + sound + video" format, involving visual and auditory modalities. Taking only textual information into account can result in inaccurate entity linking due to missing data. In addition, using information such as picture, sound, and video to enrich the context will absolutely improve the accuracy of the entity linking.

Author Contributions: Conceptualization, Y.Z.; Data curation, B.C.; Investigation, Y.Z. and B.H.; Methodology, J.L.; Project administration, J.L.; Supervision, B.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Scientific and technological innovation 2030—major project of new generation artificial intelligence (2020AAA0109300); National Natural Science Foundation (U1831133); Shanghai Science and Technology Commission Science and Technology Innovation Action Plan (22S31903700, 21S31904200).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wu, F.; Weld, D.S. Open information extraction using Wikipedia. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden, 11–16 July 2010; pp. 118–127.
2. Fleischman, M.; Hovy, E.; Echihabi, A. Offline strategies for online question answering: Answering questions before they are asked. In Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, Sapporo, Japan, 7–12 July 2003; pp. 1–7.
3. Gupta, N.; Singh, S.; Roth, D. Entity linking via joint encoding of types, descriptions, and context. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 9–11 September 2017; pp. 2681–2690.
4. Cheng, J.; Pan, C.; Dang, J.; Yang, Z.; Guo, X.; Zhang, L.; Zhang, F. Entity linking for Chinese short texts based on BERT and entity name embeddings. In Proceedings of the China Conference on Knowledge Graph and Semantic Computing, Hangzhou, China, 24–27 August 2019.
5. Logeswaran, L.; Chang, M.; Lee, K.; Toutanova, K.; Devlin, J.; Lee, H. Zero-shot entity linking by reading entity descriptions. *arXiv* **2019**, arXiv:1906.07348.
6. Reimers, N.; Gurevych, I. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv* **2019**, arXiv:1908.10084.
7. Zhang, Y.; He, S.; Liu, K.; Zhao, J. A joint model for question answering over multiple knowledge bases. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; p. 30.
8. Michelson, M.; Macskassy, S.A. Discovering users' topics of interest on twitter: A first look. In Proceedings of the Fourth Workshop on Analytics for Noisy Unstructured Text Data, Toronto, ON, Canada, 26–30 October 2010; pp. 73–80.
9. Lin, Y.; Shen, S.; Liu, Z.; Luan, H.; Sun, M. Neural relation extraction with selective attention over instances. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, 7–12 August 2016; pp. 2124–2133.
10. Ji, H.; Grishman, R. Knowledge base population: Successful approaches and challenges. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA, 19–24 June 2011; pp. 1148–1158.
11. Broscheit, S. Investigating entity knowledge in BERT with simple neural end-to-end entity linking. *arXiv* **2020**, arXiv:2003.05473, 2020.
12. Nie, F.; Cao, Y.; Wang, J.; Lin, C.-Y.; Pan, R. Mention and entity description co-attention for entity disambiguation. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; p. 32.
13. Davletov, A.; Arefyev, N.; Gordeev, D.; Rey, A. LIORI at SemEval-2021 Task 2: Span Prediction and Binary Classification approaches to Word-in-Context Disambiguation. In Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021), Online, 5–6 August 2021; pp. 780–786.
14. Yunfang, W.; Miao, W.; Peng, J.; Shiwen, Y. Ensembles of Classifiers for Chinese Word Sense Disambiguation. *Comput. Res. Dev.* **2008**, *45*, 1354–1361.
15. Alhelbawy, A.; Gaizauskas, R. Graph ranking for collective named entity disambiguation. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Baltimore, MD, USA, 22–27 June 2014; pp. 75–80.
16. Zhang, F.; Liu, X.; Tang, J.; Dong, Y.; Yao, P.; Zhang, J.; Gu, X.; Wang, Y.; Shao, B.; Li, R.; et al. OAG: Toward Linking Large-scale Heterogeneous Entity Graphs. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2585–2595.
17. Hu, L.; Ding, J.; Shi, C.; Shao, C.; Li, S. Graph neural entity disambiguation. *Knowl. Based Syst.* **2020**, *195*, 105620. [CrossRef]
18. Fang, Z.; Cao, Y.; Li, R.; Zhang, Z.; Liu, Y.; Wang, S. High quality candidate generation and sequential graph attention network for entity linking. In Proceedings of the Web Conference, Taipei, Taiwan, 20–24 April 2020; pp. 640–650.

19. Zhang, K.; Zhu, Y.; Gao, W.; Xing, Y.; Zhou, J. An approach for named entity disambiguation with knowledge graph. In Proceedings of the International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, China, 16–17 July 2018; pp. 138–143.
20. Zhu, G.; Iglesias, C.A. Exploiting semantic similarity for named entity disambiguation in knowledge graphs. *Expert Syst. Appl.* **2018**, *101*, 8–24. [CrossRef]
21. Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805, 2018.
22. Fei, Z.; Yanhui, Z.; Wentong, L.; Xiangbing, J. Entity linking methods based on BERT and TextRank keyword extraction. *J. Hunan Univ. Technol.* **2020**, *34*, 63–70.
23. Fei, Z.; Yanhui, Z.; Wentong, L.; Xu, Z.; Kang, O.; Lingwei, K. Short text entity linking method based on multi-task learning. *Comput. Eng.* **2022**, *48*, 6.
24. Shengqi, Z.; Yuanlong, W.; Ru, L.; Xiaoyue, W.; Xiaohui, W.; Zhichao, Y. Entity Linking of Chinese Short Texts Based on Local Attention Mechanism. *Comput. Eng.* **2021**, *47*, 8.
25. Zhao, Y.; Wang, Y.; Yang, N. Chinese Short Text Entity Linking Based On Semantic Similarity and Entity Correlation. In Proceedings of the 32nd International Conference on Tools with Artificial Intelligence (ICTAI), Baltimore, MD, USA, 9–11 November 2020; pp. 426–431.
26. Gu, Y.; Qu, X.; Wang, Z.; Huai, B.; Yuan, N.J.; Gui, X. Read, retrospect, select: An mrc framework to short text entity linking. *arXiv* **2021**, arXiv:2101.02394.
27. Barba, E.; Procopio, L.; Navigli, R. ExtEnD: Extractive Entity Disambiguation. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Dublin, Ireland, 22–27 May 2022; pp. 2478–2488.
28. Cui, Y.; Che, W.; Liu, T.; Qin, B.; Yang, Z. Pre-Training with Whole Word Masking for Chinese BERT. *arXiv* **2019**, arXiv:1906.08101. [CrossRef]
29. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
30. Cornegruta, S.; Bakewell, R.; Withey, S.; Montana, G. Modelling radiological language with bidirectional long short-term memory networks. *arXiv* **2016**, arXiv:1609.08409.
31. Lafferty, J.; McCallum, A.; Pereira, F.C.N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the 18th International Conference on Machine Learning, Morgan Kaufmann, New York, NY, USA, 16–21 July 2022; pp. 282–289.
32. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.