2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Research on Confusing Entity Linking Method Based on Graph Neural Network

Zhen Zhang[1, 2], Jiaxing Fan[1, 2], Jing Wang[1, 2], Yemao Zhang[1, 2], Lingnan Bai[1, 2], Zhe Xu[1, 2], Ruiyao Han[1, 2] and Jun Yu[1, 2, *]

[1] Shanxi Taihang Laboratory Co., Ltd., Taiyuan Shanxi 030024, China
[2] Taihang Laboratory in Shanxi Province (Advanced Computing Laboratory in Shanxi Province), Taiyuan Shanxi 030024, China
18435996998@163.com; c1111111c@163.com; 273569698@qq.com; 867296114@qq.com; bailingnan@vip.qq.com; xz96424@163.com; hanruiyao163@163.com; *y.j@sxu.edu.cn

Abstract. Entity Linking (EL), as a knowledge-driven semantic understanding technique, focuses on resolving the many-to-one mapping problem between name mentions in open-domain text and entities in a knowledge base. It plays a pivotal role in various applications such as knowledge graph construction, question answering systems. Building upon the rapid development of deep learning, EL is gradually moving away from traditional rule-based systems and evolving into intelligent solutions centered around semantic embedding. However, these methods still face challenges when dealing with ambiguous candidate entities, such as homonyms, synonyms, or cases with insufficient contextual information, which often cause inaccurate entity resolution. To overcome this challenge, this paper proposes **C**onfusing **E**ntity **L**inking model based on **G**raph **N**eural **N**etworks (CEL-GNN), which leverages graph structures to capture subtle differences between candidate entity descriptions, thereby improving the precision and robustness. The proposed model first employs a BERT-based encoding layer to generate representations for both short texts and candidate entity descriptions. It then applies the TF-IDF method to extract keywords and construct a knowledge graph. Subsequently, a Graph Distillation Operator (GDO) is introduced to capture discriminative characteristics, further improving the disambiguation performance. Experimental results indicate that the proposed approach achieves outstanding outcomes on the CCKS2020 benchmark for entity linking in Chinese short texts. As opposed to the baseline BERT model, our method achieves an F1 score of 88.9, significantly improving entity linking effectiveness.

Keywords: Entity Linking, Graph Neural Network, BERT.

1 Introduction

EL as one of the core technologies in Natural Language Processing (NLP), has widespread applications in various fields such as information extraction [1], knowledge graph construction [2], and question-answering systems [3]. With the rapid advancement of information technology, the scale of textual data has grown exponentially. Extracting and associating key information from massive texts efficiently has become an urgent issue for intelligent information processing systems. The primary goal of entity linking is to match named entities (such as names of persons, places, entities, produc, etc.) in texts with predefined entities in a knowledge base, thereby constructing a more comprehensive knowledge system and enhancing information retrieval and reasoning capabilities [4, 5]. As a logographic language, Chinese has a flexible grammatical structure and a high degree of lexical ambiguity, posing numerous challenges for entity linking tasks. For instance, "Support" can refer to both heavy artillery fire support and individual soldier equipment supply, while "The Great Wall" may denote a historical site or a brand/company. This polysemy and ambiguity require entity linking techniques to fully understand the contextual semantics to make accurate linking decisions. Additionally, issues such as synonym expressions, abbreviations, and ambiguity resolution further increase the complexity of EL tasks.

To achieve more accurate of EL, scholars have explored various methods to enhance the contextual understanding of models. Traditional rule-based and statistical methods have achieved certain effectiveness in specific scenarios. However, due to the limitations of rule scalability and data sparsity, these approaches struggle to adapt to large-scale and dynamic textual environments. Stimulated by the growth of AI, deep learning-based entity linking methods have become a research hotspot [6,7]. Foundation language models [8-12] have demonstrated powerful capabilities in textual semantic understanding, enabling entity linking tasks to capture contextual information more effectively and enhance disambiguation performance. These models compare each entity in the text with candidate entities in the knowledge base one by one, leveraging contextual information to make precise linking decisions (see Fig. 1. a).

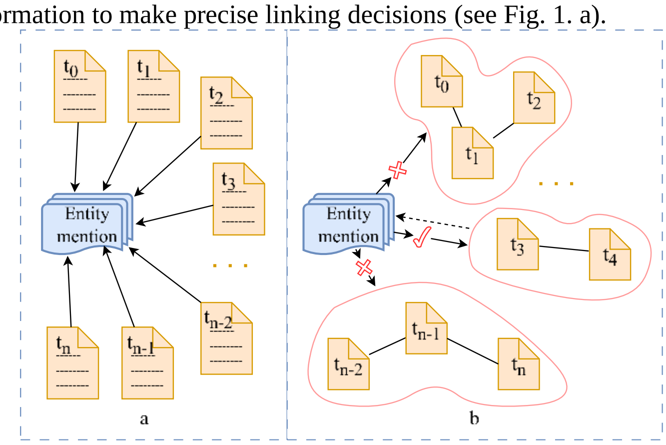

Fig. 1. a. Traditional entity matching using similarity scoring. b. Entity matching via relationship graph construction.

Despite the significant progress made by deep learning techniques in entity linking, current mainstream methods still primarily rely on empirical conclusions drawn from large-scale data training. In certain scenarios, coarse-grained features fail to capture subtle differences between candidate entities accurately, limiting disambiguation effectiveness. This is particularly evident when dealing with semantically similar but contextually distinct entities, where existing methods still exhibit limitations. To address this issue, this paper proposes an improved approach based on BERT-based entity linking. By constructing a relationship graph among candidate entities (see Fig. 1. b), this approach refines entity descriptions, extracts distinguishable key terms from candidate entity descriptions, and analyzes them at a finer granularity. This method not only enhances the accuracy of entity linking but also effectively reduces ambiguity in complex contexts, thereby improving the precision of knowledge association.

2 Related Work

2.1 Pretrained Language Model BERT

BERT is a multi-layer Transformer-based architecture that employs self-supervised learning for pretraining on large-scale corpora [13]. Self-supervised learning in BERT is primarily achieved through two core objectives: masked token prediction and sentence-level coherence modeling. The MLM mechanism learns word or character representations in bidirectional contexts. Specifically, during training, BERT randomly replaces certain words or characters in the input using a custom token [MASK] and predicts the original tokens. The NSP mechanism focuses on sentence-level coherence. BERT processes sentence pairs by distinguishing them with learnable token type embeddings and inserting a special separator token [SEP] between them. Additionally, The BERT model embeds predefined control symbols ([CLS] and [SEP]) at the beginning and end of the text sequence. The [CLS] token is primarily used to capture sentence-level information, indicating whether the given sentence pair is semantically coherent.

2.2 Entity Linking

The research in EL has undergone a transformation from traditional feature engineering approaches to automatic feature learning based on deep learning models. In recent years, studies on entity linking have primarily focused on leveraging deep learning models to enhance semantic understanding of entities, improve disambiguation accuracy, and reduce computational complexity.

In the early stages of deep learning, entity linking primarily relied on traditional classic neural models, exemplified by feedforward and convolutional neural networks [14]. The core idea behind these methods was to automatically learn the correlation features between input text and candidate entities through neural networks. Early deep learning-based entity linking methods predominantly relied on vectorizing entity descriptions and performing matching and disambiguation using neural networks. However, these methods had significant limitations, as they struggled to handle complex contextual information and capture long-range dependencies, leaving considerable room for improvement in disambiguation accuracy. With the emergence of RNN and

LSTM, researchers began leveraging these models to better process the sequential characteristics of text. RNNs and LSTMs effectively capture sequential dependencies in text, mitigating the limitations of early neural network models in handling long-range dependencies [15]. However, despite their ability to model contextual information, LSTMs still suffer from inefficiencies in processing long texts due to their sequential nature of information propagation.

In 2018, the introduction of BERT marked a significant breakthrough in natural language processing [13]. By leveraging the self-attention mechanism, BERT enables bidirectional context modeling, capturing complex semantic relationships on a global scale and significantly improving performance across various NLP tasks. The emergence of BERT represented a major advancement in entity linking technology. Yamada et al. [16] proposed a BERT-based entity linking method, in which BERT serves as the core model to encode the context of entities and compute similarity with description in a predefined knowledge base, ultimately achieving precise EL. The key advantage of this approach lies in BERT's bidirectional modeling capability, which substantially enhances its ability to disambiguate polysemous and synonymous entities. Additionally, Wang et al. [17] introduced an end-to-end BERT-based entity linking model that formulates entity linking as a joint optimization problem, utilizing BERT to perform entity recognition, candidate entity generation, and disambiguation simultaneously. Unlike traditional methods that rely on handcrafted features, this approach directly learns entity representations from raw text through an end-to-end framework, leading to significant improvements in entity linking performance. Furthermore, Sebbaq [18] proposed a multi-task learning architecture built on BERT, in which entity recognition, mention-knowledge base alignment, and entity disambiguation are treated as three interconnected subtasks trained jointly. By sharing hidden layers in BERT, the model facilitates information exchange between different tasks, ultimately enhancing overall entity linking performance.

With the growing maturity of graph neural networks (GNNs), researchers have started integrating graph structures into entity linking tasks, leveraging relational graphs to further optimize entity linking accuracy. GNNs effectively capture intricate relationships among entities and propagate information through graph structures, thereby improving entity matching precision [19,20].

3 The Basic Principles of the Model

We introduce an entity linking framework built upon Graph Neural Networks (GNNs), aiming to improve the accuracy of entity disambiguation, particularly when handling knowledge bases containing ambiguous data. The model integrates graph-based ideas by leveraging GNN methods, allowing it to better capture the interactions between short texts and entities. This enables the model to focus on subtle differences between candidate entities, effectively extracting distinguishable features from them, thus enhancing entity linking performance in complex text environments. In Fig. 2, the framework adopts a hierarchical architecture design, including a basic encoding layer, a distinguishable feature encoding layer, and a fully connected layer. Through this architecture, the model establishes strong associations between short texts and candidate entity descriptions and, using GNN techniques, further mines and retains discriminative features from the candidate entity descriptions, ultimately matching the most suitable entity description for each short text.

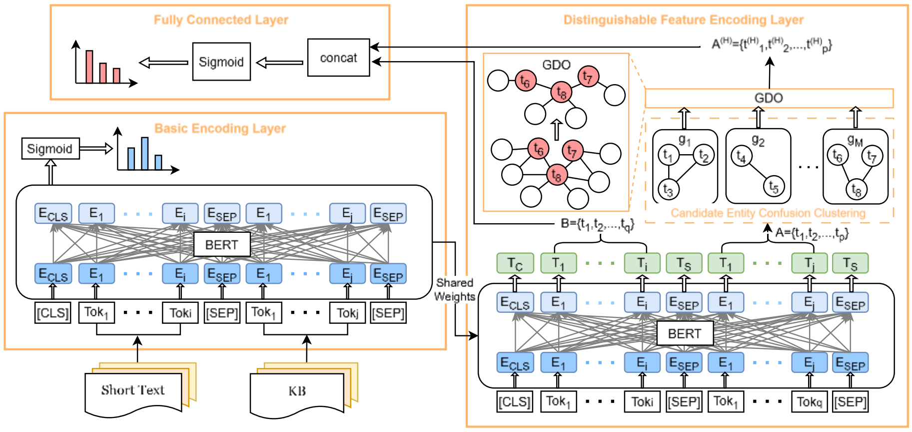

Fig. 2. CEL-GNN model includes a basic encoding layer, a distinguishable feature encoding layer, and a fully connected layer, capturing entity interactions and discriminative features to achieve accurate entity linking.

3.1 Basic Encoding Layer

The rapid advancement of pretrained language models has established them as fundamental cornerstones in NLP applications. In collaboration with iFlytek, Harbin Institute of Technology (HIT) has enhanced the BERT model by incorporating the Whole Word Masking (WWM) technique. Trained on the Chinese Wikipedia corpus, this improved model exhibits superior Chinese text encoding capabilities, providing richer semantic representations that are particularly beneficial for short-text entity disambiguation (ED). Therefore, this paper adopts Chinese-BERT-wwm-ext as the fundamental encoding layer to compute the correlation between short texts and candidate entities, extracting the corresponding feature vectors. Specifically, for each entity in the CCKS2020 knowledge base, we concatenate its triple set into a single string, referred to as the entity description text, where each individual triple is transformed into a triple text through element concatenation.

As illustrated in Fig. 2, during training, for each entity mention, we concatenate its contextual text with the description text of each candidate entity and formulate entity disambiguation into a framework of binary classification. The input to the BERT encoder follows a structured format:

- A [CLS] token is inserted at the beginning of the sentence pair.
- A [SEP] token is used to separate the short text from the entity description text.
- Another [SEP] token is appended at the end of the sentence pair.

The final-layer output of the BERT model, specifically the [CLS] token vector, is leveraged to capture the global semantic differences between the short text and knowledge base text. Furthermore, to retain local entity-specific information, we record the beginning ($begin$) and ending (end) positions of the entity mention in the text and concatenate the corresponding encoded position vectors to construct the local entity features $Entity(begin, end)$. The formal computation process is provided in Equation (1), (2),

$$text = [CLS]t_1, t_2, \ldots, t_n[SEP]d_1, d_2, \ldots, d_m[SEP] \tag{1}$$

$$s_{link} = BERT(text) + Entity(begin, end) \tag{2}$$

Let $t_1, t_2, \ldots, t_n$ represent the context of the entity mention, while $d_1, d_2, \ldots, d_m$ denote the description text of a candidate entity. The feature vector obtained from the CLS position is concatenated with the entity position vectors, and the resulting representation is passed through a fully connected layer that uses a Sigmoid function to perform classification. To mitigate overfitting, a Dropout layer with a dropout rate of 0.15 is introduced. Finally, the model is used to predict the matching degree between the ambiguous entity and the descriptive texts of all its potential candidates, generating a matching score that quantifies the likelihood of each candidate entity being the correct disambiguation target.

3.2 Distinguishable Feature Encoding Layer

In the experiment, although the entity disambiguation method based on the BERT-encoded basic encoding layer can achieve entity disambiguation using coarse-grained relevance features, this approach primarily relies on empirical results learned from large-scale training data. However, in certain special cases, such coarse-grained empirical information may fail to capture the subtle differences between candidate entities. To address this limitation, as illustrated in Fig. 2: Distinguishable Feature Encoding Layer, this study leverages graph-based modeling to identify and analyze fine-grained distinctions between key descriptive terms in the textual representations of candidate entities.

For each entity mention, multiple candidate entity descriptions may exist. For the candidate entity set L, the fully connected graph topology G^* is initialized based on the principle of maximizing mutual information, where the nodes represent the semantic features of the entities. The adjacency weight between entity pair $L_i, L_j \in L$ is modeled as the angular proximity metric in vector space, denoted as $\mathrm{tf\_idf}_i$ and $\mathrm{tf\_idf}_j$, derived from their respective term-weighting vectors. These vectors are generated using the TF-IDF statistical model, which quantifies lexical significance through frequency analysis based on the corpus. Since ambiguous candidate entity descriptions tend to be semantically similar while still containing sufficient distinguishing information, we refine the graph G^* structure by removing edges with weights below a predefined threshold. Through the application of a suitable threshold, the association strength between entities is determined, we construct a new graph $G = \{g_i\}_{i=1}^{M}$, which consists of multiple disconnected subgraphs $g_1, \ldots, g_M$. Each subgraph contains at least one candidate entity description, ensuring that the remaining graph structure effectively captures semantically relevant yet distinguishable candidate entities for the given mention.

Next, to extract distinguishable information from each subgraph g_i, a straightforward approach would be to remove repeated words and sentences from candidate entity descriptions within the subgraph. However, this method cannot be directly integrated into an end-to-end neural architecture. To overcome this limitation, our model initializes a new BERT model by transferring the pretrained weights and bias matrices from the basic encoding layer BERT model. The fused entity mention representation, incorporating all candidate entity descriptions, is denoted as $\mathrm{A} = \{t_1, t_2, \dots, t_p\}$, where $a \in [1, p]$ represents the feature vector t_i of words in the candidate entity description L_i.

To effectively extract distinguishable features, we introduce an improved Graph Distillation Operator (GDO) based on the Graph Convolution Operator (GCO) [21]. Unlike GCO, which propagates and aggregates information among neighboring nodes to enrich their representations, GDO is designed to learn discriminative features by removing redundant information shared between nodes. Specifically, for any candidate entity description L_i, GDO utilizes a learnable weight matrix to find the similarity between L_i and the adjacent nodes in graph G. Subsequently, a linear transformation matrix is applied to map the representation of L_i into a new space, thereby extracting discriminative semantic features. Each network layer ($l \geq 0$) performs similarity-aware feature decoupling by explicitly subtracting the association features between L_i and its neighboring nodes from its embedding representation, ensuring that only distinctive features are retained. As shown in Equation 3,

$$t_i^{(l+1)} = \Phi^{(l)} t_i^{(l)} - \sum_{j \in N_i} \frac{\Psi^{(l)}\left[t_i^{(l)}, t_j^{(l)}\right]}{|N_i|} + b^{(l)} \tag{3}$$

Let $t_i^{(l)}$ denote context-aware encoding of the candidate description text L_i in the l-th graph distillation layer. N_i represents the group of neighbors of the candidate entity within the graph G, while $b^{(l)}$ is the bias term. $\Phi^{(l)}$ and $\Psi^{(l)}$ are trainable self-weighting matrices.

By employing a GDO with H layers, we obtain the final representation $t_i^{(H)}$ for the candidate entity description L_i at the last layer. The representation $t_i^{(H)}$ encapsulates rich and distinctive features, enabling effective differentiation of L_i from other candidate entities within the same subgraph.

3.3 Fully Connected Layer

Extracting the vectors of all candidate entities in the subgraph g from the embedded node feature matrix, we denote them as $\mathrm{A}^{(H)} = \left\{t_1^{(H)}, t_2^{(H)}, \dots, t_p^{(H)}\right\}$, where $a \in [1, p]$. The vector representation of an individual candidate entity is denoted as $t_a^{(H)}$. Similarly, the entity mentions obtained by integrating all short-text features are denoted as $\mathrm{B} = \{t_1, t_2, \dots, t_q\}$, where $b \in [1, q]$, and the representation of a single entity mention is denoted as t_b. Finally, we concatenate $\mathrm{t}_a^{(H)}$ and t_b to obtain T.

$$T = \text{concat}\left(t_b, t_a^{(H)}\right) \tag{4}$$

$$Score = \text{Sigmoid}(T) \tag{5}$$

In the final stage, a fully connected layer coupled with a Sigmoid activation function is used to derive the normalized similarity scores of the candidate entity set, determining and selecting the target entity with the highest confidence.

4 Experimental Results and Analysis

4.1 Dataset and Data Preprocessing

In this study, we validate our model using the dataset provided by the CCKS 2020 Chinese Short Text Entity Linking Task. The dataset consists of short texts with an average length of 21.73 Chinese characters, primarily sourced from online platforms such as Weibo, encyclopedia entries, news video titles, article headlines, and user conversations. The corpus is divided into a training set and a validation set, where the training set contains 70,000 annotated short texts, and the validation set consists of 10,000 annotated short texts. The knowledge base is semi-structured, with each entry containing both triples and structured long texts. The triples encapsulate entities and their relationships, providing essential contextual information.

To achieve EL model, we conduct dataset analysis and preprocessing. First, since the official test set of this dataset is not publicly available, we use the CCKS2020-dev set as our test set. Additionally, the training dataset (CCKS2020-train) is partitioned into training and validation subsets using seven-fold cross-validation. During training, we save 14 models based on both loss and F1 score criteria. Second, to mitigate the impact of erroneous characters on downstream disambiguation results, we perform a structured alignment operation to map the entity mentions in the raw data to the corresponding canonical entries in the knowledge base. This prevents mismatches caused by punctuation differences, letter case variations, or misspellings, thereby improving the robustness of entity linking.

4.2 Evaluation Criteria

The experiment evaluates disambiguation results using F1-score.

$$R = \frac{TP}{TP+FN} \tag{6}$$

R represents recall, where TP denotes the number of correctly predicted entity links, and FN denotes the number of missed entity links.

$$P = \frac{TP}{TP+FP} \tag{7}$$

P represents precision, where FP denotes the number of incorrectly predicted entity links.

$$F_1 = \frac{2*P*R}{P+R} \tag{8}$$

Finally, the model's performance is evaluated by calculating the F_1 score, with higher values indicating better model performance.

4.3 Experimental Setup

The fundamental encoding layer utilizes the official Chinese Bert_wwm_ext model as the factual encoder, with accelerated computation powered by eight NVIDIA A100 40GB GPUs. Given that the length of each triple's text varies and that BERT imposes a strict 512-token limit, entity descriptions exceeding this length are truncated, retaining only the first 512 tokens while discarding the excess. Since the default truncation strategy may result in information loss, we assume that the length of a triple's text is proportional to the amount of information it contains. To minimize the loss of essential information, we sort all triples in descending order based on their text length before truncation. This ensures that tokens removed due to truncation primarily belong to shorter triples, which inherently contain less information, thereby maximizing the retention of meaningful content in entity descriptions. Additionally, when the input text length is shorter than the maximum allowable length, padding is applied to extend it to 512 tokens. Padding is typically performed using a special [PAD] token. Finally, all triple texts are concatenated to serve as the descriptive text for the candidate entity.

In the experiment, the Adam optimizer was employed with a learning rate set to $2*10^{-5}$, the number of epochs set to 3, and a batch size of 32. The BERT model was fine-tuned based on the characteristics of the short-text entity linking dataset to enhance experimental performance. For the distinguishable feature encoding layer of the GNN, the learning rate was set to 10^{-4}, and the similarity threshold for removing redundant features was set to 0.35.

4.4 Experimental Result

To validate the effectiveness of the proposed entity linking model based on Graph Neural Networks (GNN), extensive experiments were conducted on the CCKS2020 Chinese short-text entity linking dataset. The F1-score was used as the evaluation metric, and the performance of different models on the entity linking task was compared. Experimental results demonstrated that the proposed model significantly outperformed other baseline models across multiple evaluation metrics. Table 1 shows the experimental results of the proposed model and the comparison models. Notably, it achieved an F1-score of 88.9, surpassing ERNIE [22] (86.7), BERT-base (87.4), BERT-large (87.8), Bert_wwm_ext (88.1) and RoBERTa-wwm-ext [23,24] (87.5). This indicates that the proposed GNN-based entity linking model substantially improves disambiguation accuracy, particularly in distinguishing and correctly matching highly similar candidate entities.

In traditional pre-trained models, while BERT-base and BERT-wwm-ext exhibit strong performance in disambiguation tasks, they often struggle to capture subtle differences when dealing with multiple candidate entities with semantically similar descriptions, leading to decreased matching accuracy. Although RoBERTa-wwm-ext has advantages in processing long texts, it does not incorporate further optimizations tailored to the fine-grained details of candidate entity descriptions. In contrast, the proposed model leverages GNNs to better explore both local and global relationships among candidate entity descriptions, enabling the extraction of more fine-grained information. Specifically, the graph structure among candidate entity descriptions enhances the model's ability to capture inter-entity relationships and helps extract more discriminative features, even when candidate entities exhibit high semantic similarity. This, in turn, improves the accuracy of entity matching.

In summary, by integrating Graph Neural Networks, the proposed model significantly enhances entity disambiguation performance. It demonstrates superior capability in distinguishing similar candidate entities compared to traditional models, providing a more reliable solution for precise entity matching in entity linking tasks.

Table 1. Experimental results (%).

Model	Recall	Precision	F1
ERNIE	86.7	86.7	86.7
Bert-base	87.4	87.4	87.4
Bert-large	87.8	87.8	87.8
Bert_wwm_ext	87.9	88.3	88.1
Roberta_wwm_ext	87.2	87.7	87.5
CEL-GNN	88.7	89.1	**88.9**

4.5 Case Analysis

To verify the effectiveness of CEL-GNN in extracting discriminative features, we conducted a visualization analysis of its encoder's attention mechanism. Fig. 3 illustrates the contextual focus of the encoder on both short texts and the knowledge base, where darker word colors indicate higher attention weights. Taking the phrase "The Moonlight Before My Bed" as an example, there are five candidate entities associated with it in the knowledge base, spanning multiple musical works and cultural references. Therefore, a deeper contextual analysis is required to achieve accurate entity disambiguation.

When using only the base BERT encoder for entity prediction (Fig. 3. a), the model primarily focuses on the keywords "Anita Mui" and "The Moonlight Before My Bed". Since candidate entity 51069 contains frequent mentions of "Anita Mui", the model assigns it a higher attention weight, mistakenly linking to entity ID 51069. In contrast, when CEL-GNN is employed (Fig. 3. b), the model first utilizes the strong prior information of "Anita Mui" in the short text to filter out unrelated candidate entities for "The Moonlight Before My Bed", narrowing down the options to entity IDs 331515 and 51069. Subsequently, the discriminative encoding layer distinguishes 331515 as a song and 51069 as an album. By incorporating the short-text context, CEL-GNN successfully identifies 331515 as the correct entity in the knowledge base. This case study highlights that ambiguous entity names may correspond to multiple references, and precise entity linking requires fine-grained contextual matching to ensure accurate entity recognition.

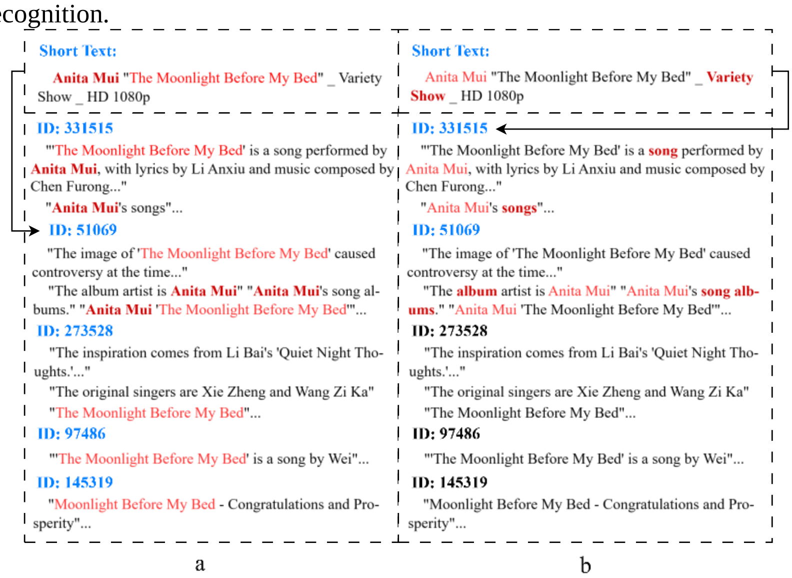

Fig. 3. a. Only the base BERT encoder for entity prediction. b. CEL-GNN for entity prediction.

5 Conclusion

The **C**onfusing **E**ntity **L**inking Model based on **G**raph **N**eural **N**etworks (**CEL-GNN**) proposed in this paper effectively addresses the limitations of traditional BERT-based methods in disambiguation tasks, particularly demonstrating strong capabilities in handling subtle semantic differences between candidate entities. By incorporating GDO, the model not only improves the precision of candidate entity disambiguation but also enhances the understanding of complex entity relationships. In the CCKS2020 Chinese short-text entity linking task, the proposed model significantly outperforms traditional methods in terms of F1 score, validating its potential application in the field of entity linking. However, the model still faces certain limitations when dealing with long texts and noisy data. Future research could further optimize the structure of the Graph Neural Network and introduce additional semantic information and contextual constraints to enhance the model's disambiguation ability. Furthermore, integrating more knowledge graphs and multimodal information is expected to further improve the model's generalization ability and practical applicability.

Acknowledgments. This study was funded by the National Key R&D Program of China (2023YFB4502904).

References

1. Tsujimura T, Yamada K, Ida R, et al.: Contextualized medication event extraction with striding NER and multi-turn QA[J]. Journal of Biomedical Informatics **144**, 104416 (2023)
2. Al-Moslmi T, Ocaña M G, Opdahl A L, et al.: Named entity extraction for knowledge graphs: A literature overview[J]. IEEE Access **8**, 32862-32881 (2020)
3. Li L, Zhang M, Chao Z, et al.: Using context information to enhance simple question answering[J]. World Wide Web **24**, 249-277 (2021)
4. Barlaug N, Gulla J A.: Neural networks for entity matching: A survey[J]. ACM Transactions on Knowledge Discovery from Data (TKDD) **15**(3), 1-37 (2021)
5. Möller C, Lehmann J, Usbeck R.: Survey on english entity linking on wikidata: Datasets and approaches[J]. Semantic Web **13**(6), 925-966 (2022)
6. Fang, Z., Cao, Y., Li, Q., Zhang, D., Zhang, Z., & Liu, Y.: Joint entity linking with deep reinforcement learning. In The world wide web conference, pp. 438-447. Association for Computing Machinery, New York (2019)
7. Chen, S., Wang, J., Jiang, F., & Lin, C.-Y.: Improving Entity Linking by Modeling Latent Entity Type Information. Proceedings of the AAAI Conference on Artificial Intelligence, **34**(05), 7529-7537 (2020)
8. Yin X, Huang Y, Zhou B, et al.: Deep entity linking via eliminating semantic ambiguity with BERT[J]. IEEE Access **7**, 169434-169445 (2019)
9. Shi Y, Yang R, Yin C, et al.: Entity linking method for chinese short texts with multiple embedded representations[J]. Electronics **12**(12), 2692 (2023)
10. Chen L, Varoquaux G, Suchanek F M.: A lightweight neural model for biomedical entity linking[C]. Proceedings of the AAAI conference on artificial intelligence **35**(14), 12657-12665 (2021)
11. Zhang Y, Liu J, Huang B, et al.: Entity Linking Method for Chinese Short Text Based on Siamese-Like Network[J]. Information **13**(8), 397 (2022)
12. Shi Y, Yang R, Yin C, et al.: Entity linking method for chinese short texts with multiple embedded representations[J]. Electronics **12**(12), 2692 (2023)
13. Aftan, Sulaiman, and Habib Shah.: A survey on bert and its applications. 20th Learning and Technology Conference (L&T) 2023, pp. 161-166. IEEE (2023)
14. Li H, Chen Q, Tang B, et al.: CNN-based ranking for biomedical entity normalization[J]. BMC bioinformatics **18**, 79-86 (2017)
15. Luo A, Gao S, Xu Y.: Deep semantic match model for entity linking using knowledge graph and text[J]. Procedia Computer Science **129**, 110-114 (2018)
16. Yamada, K. Washio, H. Shindo: Matsumoto, Global entity disambiguation with pretrained contextualized embeddings of words and entities, https://arxiv.org/abs/1909.00426v3
17. Wang Y.: End-to-End Entity Linking Combined with Bert-based Siamese and Interaction Network[C]. Proceedings of the 2022 10th International Conference on Information Technology: IoT and Smart City, pp. 47-53. (2022)
18. Sebbaq H.: Mtbert-attention: An explainable bert model based on multi-task learning for cognitive text classification[J]. Scientific African **21**, e01799 (2023)
19. De Bonis M, Falchi F, Manghi P.: Graph-based methods for Author Name Disambiguation: a survey[J]. PeerJ Computer Science **9**, e1536 (2023)
20. Chen Y, Yuan H, Liu T, et al.: Name disambiguation based on graph convolutional network[J]. Scientific Programming **2021**(1), 5577692 (2021)
21. Xu N, Wang P, Chen L, et al.: Distinguish confusing law articles for legal judgment prediction[J], https://arxiv.org/abs/2004.02557

22. Sun Y, Wang S, Feng S, et al.: Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation[J]. arXiv preprint arXiv:2107.02137 (2021)
23. Cui Y, Che W, Liu T, et al.: Pre-training with whole word masking for chinese bert[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing **29**, 3504-3514 (2021)
24. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al.: HuggingFace's Transformers: State-of-the-art Natural Language Processing, https://arxiv.org/abs/1910.03771