World Wide Web (2020) 23:2979–3000
https://doi.org/10.1007/s11280-020-00809-8

A general strategy for researches on Chinese "的(de)" structure based on neural network

Bingqing Shi[1] · Weiguang Qu[1,3] · Rubing Dai[2] · Bin Li[2] · Xiaohui Yan[4] · Junsheng Zhou[1] · Yanhui Gu[1] · Ge Xu[3]

Received: 10 May 2019 / Revised: 7 January 2020 / Accepted: 6 March 2020 /
Published online: 13 September 2020

Abstract
Noun phrases reflect people's understanding of the world entities and play an important role in people's language system, conceptual system and application system.
With the Chinese "的(de)" structure, attributive noun phrases of the combined type can accommodate more words and syntactic structures, resulting in rich levels and complex semantic structures in Chinese sentences. Moreover, the Chinese elliptical "的(de)" structure is also of vital importance to the overall semantic understanding of the sentence. Many researches focus on rule-based models and semantic complement of "verb+ (de)" structure. To tackle these issues, we propose a general three-stage strategy utilizing neural network for the researches on all "的(de)" structure. Experimental results demonstrate that the proposed strategy is effective in boundary definition, elliptical recognition and semantic complement of "的(de)" structure.

Keywords De structure · neural network · ellipsis · Chinese language processing

1 Introduction

With the rapid development of Internet, the amount of the Chinese text data is increasing in an incredibly rapid speed, which has brought new requirements and challenges to natural

✉ Weiguang Qu
wgqu@njnu.edu.cn

✉ Yanhui Gu
gu@njnu.edu.cn

[1] School of Computer Science and Technology, Nanjing Normal University, Nanjing, China

[2] School of Chinese Language and Literature, Nanjing Normal University, Nanjing, China

[3] Fujian Provincial Key Laboratory of Information, Processing and Intelligent Control, Minjiang University, Fuzhou, China

[4] Poisson Lab, Huawei Technologies, Beijing, China

language processing (NLP). In modern Chinese, the Chinese word “的(de)” is one of the words with strong word-formation ability which can express various grammatical relations [1, 2]. The present academic circles to the usages of the Chinese word “的(de)” have not achieved the mutual recognition. There are two main viewpoints. Firstly, as a structural auxiliary word, the Chinese word “的(de)” is used between attributives and centers [3]. The “的 (de)” structure(hereinafter referred to as the DE) is the structure “attributive +‘的(de)’ +center”, one of the basic syntactic structures in Chinese. The elliptical DE is a variation [2]. It omits the center, does not depend on any component, independently acts as a nominal component in the sentence, and undertakes the role of the omitted center in the sentence. In Figure1 (a), the Chinese sentence, “他们*(they)* 要*(want)* 找*(find)* 的*(de)* 是*(is)* 失散*(lost)* 多*(many)* 年*(year)* 的*(de)* 女儿*(daughter)*”, has two DEs. The first DE, “他们 *(they)* 要*(want)* 找*(find)* 的*(de)*”, omits the center “人 *(person)*” , and becomes the subject in the sentence. The second DE, “失散*(lost)* 多*(many)* 年*(year)* 的 *(de)* 女儿*(daughter)*” , is an integral structure, so it is not elliptical. Secondly, as a non-structural auxiliary word, the Chinese word “的(de)” usually comes at the end of the sentence [4]. In this case, it can be further subdivided into tense-aspect auxiliary and modal particle. This paper mainly studies the elliptical DE.

With the DE, attributive noun phrases of the combined type can accommodate more words and syntactic structures, resulting in rich levels and complex semantic structures in Chinese sentences. Moreover, the Chinese elliptical DE is also of vital importance to the overall semantic understanding of the sentence. Therefore, in processing of massive Chinese texts, the correct analysis of the DE is the basis of syntactic analysis and semantic understanding, which is an unavoidable problem. Based on previous related literature reports, we divide tasks of the DE into two types, boundary definition and usage recognition. Boundary definition aims at finding all DEs in sentences. In usage recognition, researchers can study the DE from different perspectives, such as syntactic and semantic ellipsis. Moreover, the follow-up work of recognizing elliptical DE is to find the omitted center. These two types of tasks have their own characteristics, but have some connections. We can either accomplish them independently, or take into consideration of their connections while coming up with solutions. Here, we choose the former.

In this paper, by referring to the theoretical researches, literature reports based on rules and statistical methods and the practical experience of deep learning related models and methods that have been employed for numerous NLP tasks, we propose a three-stage strategy to solve tasks on the background of DE for the first time.

For the task of usage identification, researchers describe the usage and characteristics of DE from syntactic perspective. They cannot dig deeper into the semantics of the central word that is omitted due to the constraints of context and language economic principles, which is significant for sentence-level researches. For the semantic completion task of elliptical DE, current researches only focus on “verb+ (de)” structure. The rule-based approaches in elliptical DE research limit the performance of boundary definition, recognition and semantic complement of elliptical DE. They have two disadvantages. First, the design rules need to rely on previous theoretical knowledge and it is difficult to properly cover all the research domain. Second, it suffers from feature engineering. In statistical-based methods, experimental performance depends on the setting of feature templates, which is limited by the research background of the researcher.

To tackle these issues, we propose a general three-stage strategy for boundary definition of the DE, recognition of the elliptical DE, and semantic complement for the elliptical DE. The strategy contains three parts: semantic representation learning of word and sentence, task-oriented cooperation of linguistic knowledge and output construction.

In addition, motivated by many classic NLP tasks, including event extraction [5], natural language modeling [6], text summarization [7], machine translation [8] and question answering [9], that benefit from abstract meaning representation (AMR) [10], we consider to automatically build the experimental dataset based on the Chinese abstract meaning representation(CAMR) corpus [11] to solve the problem that the existing machine-oriented researches on DE use different corpus and require manual labeling.

Our main contributions are:

- A three-stage strategy targeted at solve tasks on the background of DE by neural network methods with experimental datasets which are automatically build based on the CAMR corpus and CTB corpus.
- According to the proposed strategy, we accomplish three tasks on the background of DE, namely, boundary definition of the DE, recognition of the elliptical DE, and semantic complement for the elliptical DE.
- The experimental results show that based on proposed strategy, our methods can effectively find the boundaries of DE, recognize DE with semantic ellipsis and complement the center word which is omitted due to constraints of the context and linguistic economic principles.

We survey related work in Section 2. This is followed by proposed strategy in Sections 3. Sections 4 and 5 and 6 detail solutions to our three research works respectively. This followed by our empirical evaluations in Section 7 and conclusions in Section 8.

2 Related work

In this section, we first describe some previous works on DE, then introduce some researches on CAMR, and finally briefly discuss deep learning in NLP.

2.1 De

Boundary definition of DE is to extract all DEs automatically from sentences by locating the left and right boundaries. Qian [12, 13] first studied the boundary definition of Maximal Noun Phrase that contains “的(de)”(deMNP), a special structure of DE, and formulated the corresponding corpus annotation standards. Qian comprehensively analyzed the different features of the right boundaries as well as the left boundaries of deMNP, such as internal structures, syntactic features, and linear distributions, advanced a strategy of “Identify the right boundary first, and then identify the left one”, and recognized the two boundaries by the method of “Boundary Distribution Probability”. For recognizing the right boundary of the DE, Xiao [14] proposed a bottom-up analysis method supplemented by top-down analysis method. The algorithm first uses a bottom-up method to find the candidate center and determine other components according to the verb frame. Then, a top-down method is used to determine whether it is a DE.

In tasks of usage recognition, some researchers studied the DE from the syntactic perspective. With nominalization as the guideline, Yang [15] divided the “verb+ (de)” structure, a special structure of DE, into different types, and proposed the classification criteria for computer understanding based on the analysis of syntax, semantics and pragmatics. Based on the triune Chinese function word usage knowledge base (CFKB) [16, 17], Han [18] and Liu [19] studied the automatic recognition of usages of auxiliary “的(de)” successively. In their works, the usage is divided into 39 types. Han used the rule-based method and pointed out that, automatic annotation of auxiliary “的(de)” is helpful to improve the quality of Chinese corpus and reduce the artificial work. In addition to using the rule-based method and conditional random field(CRF) model, Liu also adopted gated recurrent neural network to automatically recognize the usages of auxiliary “的” for the first time. Liu found that, compared with rule-based method and CRF model that need either laborious feature engineering work or massive extra linguistic resources, gated recurrent unit can be used to automatically acquire long distance features and improve the performance of automatic recognition.

Wang [20] studied the DE from the perspective of semantic ellipsis. Wang theoretically analyzed the ellipsis of three kinds of “的(de)” structures, namely, “verb+ (de)”, “adjective + (de)+ noun” and “noun 1 + (de)+ noun 2”, and then, based on the verb information and verb-noun collocations, worked out a rule-based method for automatically complementing the omitting center of the elliptical “verb+ (de)” structure.

In summary, researchers have carried out various machine-oriented studies on DE. However, existing researches use different corpus and require labor-intensive and time-consuming manual annotation. Some of them even do not have concrete results on the large-scale dataset. Moreover, in boundary definition, the depth and extent of studying is insufficient. In usage of recognition, most previous works on DE only focus on the usage and characteristics of auxiliary “的(de)” from the syntactic perspective, and only a small number of people paid attention to DE from the perspective of omission of semantic components, which is essential for complementing the center which is omitted due to constraints of the context and linguistic economic principles. In this paper, we propose a three-stage strategy to solve tasks on the background of DE with experimental datasets, which are automatically build based on the CAMR corpus and CTB corpus. According to the proposed strategy, we will introduce three research works on the background of DE, namely, boundary definition of the DE, recognition of the elliptical DE, and semantic complement for the elliptical DE.

2.2 CAMR

Abstract meaning representation (AMR) is a new sentence-level semantic representation in the world, which is designed to describe and reveal the complete and deep semantic information contained in sentences in order to solve various NLP problems [21]. Unlike traditional syntactic and semantic representations, AMR is a brand new domain-independent sentence-level semantic representation in English with the following characteristics [22]. First, the single directed acyclic graph is used to represent the semantics of sentences, which effectively solves the problem of argument sharing. Second, in the abstraction process in which words are converted to concepts, conceptual nodes can be added to represent implied semantics and restore the complete semantics of sentences, which is conducive to ellipsis and intra-word analysis that traditional syntactic representations cannot cope with. Third, for sentences with different surface syntactic structures and the same underlying semantics, their corresponding AMR representations are semantically identical. At present, AMR has become a hotspot in language resource construction and sentence semantic parsing. However, compared with the studies on English AMR, the development of CAMR started relatively late. According to the characteristics of Chinese, Li et al. [11, 22] made an in-depth analysis of the original framework of English AMR, inherited the tagging system applicable to Chinese, deleted or modified the inapplicable norms, added some special semantic representation methods, established the CAMR and published corresponding annotated corpus consisting of AMRs of some sentences selected from the Chinese Treebank (CTB) which are illustrated in Fig. 1.

Song [23] made a statistical analysis of elliptical DEs in annotated corpus, and found that it is comparable to other sentence semantic representation methods, semantic role labeling framework of Chinese predicates adopted by CAMR can well complement the elliptical core argument information of predicate in Chinese special structure and fully express the semantic information of predicate. That is to say, CAMR is suitable for DE. In Fig. 2, the elliptical DE, "跳舞*(dance)* 的*(de)*", omits the agent, "人*(person)*", of the predicate "跳舞*(dance)*". Chinese AMR annotators can add the node "person" to represent the omitted agent, i.e. the omitted semantic concept.

CAMR parsing is still in its infancy. By analyzing various AMR parsing algorithms, Gu [24] conducted an experimental study on CAMR automatic parsing and proposed a CAMR parsing model by utilizing transition-based neural network for the first time. However, the design of this parser has not involved the processing of DE yet. Thus, machine-oriented researches on DE can provide supports for CAMR parsing and other upper applications, such as dependency parsing and machine translation.

2.3 Deep learning in NLP

Deep learning architecture and algorithms have achieved remarkable results in computer vision. In this trend, a variety of model designs and methods have blossomed in the context of NLP. Traditional machine learning methods often rely on manual feature extraction, which is not only time-consuming and laborious, but also incomplete. In contrast, deep learning offers a way to harness large amount of computation and data [25].

At present, the achievements of deep learning in NLP applications are not as significant as those of in computer vision. A possible reason is that compared with image, natural language is abstract, high-level information [26]. Therefore, how to make full use of the advantages of neural network models in representation learning and other aspects, and how to improve the effectiveness and efficiency of NLP tasks based on neural network are still an urgent problem to be solved.

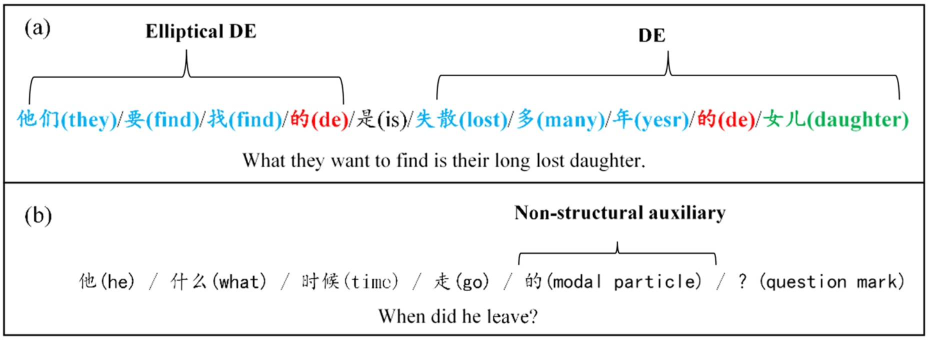

Fig. 1 Examples of different usages of the Chinese word "的(de)". DE is marked in bold. All attributives are marked in blue. Those "的(de)" that dominate DE are marked in red. Center is marked in green

跳舞(dance) / 的(de) / 走(go) / 了(modal particle)
The dancer has gone.

x0/走-01(go)
  :arg0 x1/ person
    :arg0-of x2/ 跳舞-01(dance)
  :aspect x3/ 了(modal particle)

Fig. 2 An example of concept adding of an elliptical DE. The dominating-dominated relationship between concepts is illustrated by text-indent. "的(de)" dominates an elliptical DE is marked in red. The concept that can represent the omitted center is marked in green

3 Solution strategy

We propose a three-stage strategy for each task. The framework of each task is shown in Fig. 3.

In this section, the general introduction of the strategy for each task is as follows:

For **boundary definition** task, the three-stage strategy is:

(a) **Semantic Representation learning of global information for boundary prediction**. Word embedding are obtained based on external corpuses. Global information in a sentence is extracted by a semantic representation model for the prediction.
(b) **Local information extraction based on DC-CNN.** The boundary definition task needs to locate the starting position of the modifier accurately, so it is necessary to consider the internal correlation between the n-element models in semantic modeling. We utilizes the Densely Connected CNN (DC-CNN) [27] proposed by Wang [20] to incorporate local information by introducing linguistic knowledge.
(c) **Boundary Prediction**. Identify the left and right boundary of DE based on the information learned in the former two models.

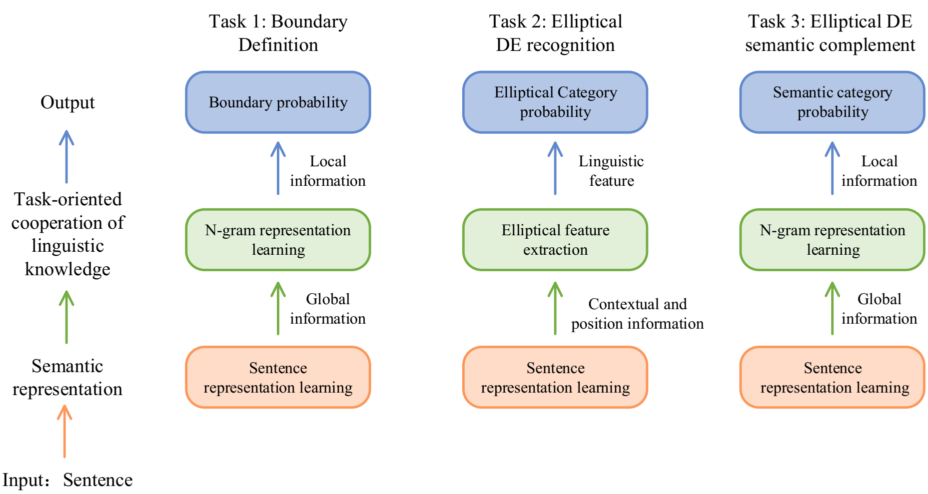

Fig. 3 The framework of three tasks

For **elliptical DE recognition** task, the three-stage strategy is:

(a) **Semantic Representation learning of context information for elliptical DE**. Learn the information from DE and its context utilizing word order information and relative position information between words.
(b) **Elliptical feature extraction**. Capture the elliptical features of DE by simulating the situation where people actually recognize elliptical DE.
(c) **Category prediction of elliptical DE**. Predict the type of elliptical DE based on the omitted features and context representation.

For **elliptical DE semantic complement** task, the three-stage strategy is:

(a) **Semantic Representation learning of context information for semantic complement**. Learn the information from DE and its context utilizing word order information and relative position information between words.
(b) **Local information learning for semantic complement.** The context produced by the verb and the semantic collocation relationship between the verb and the noun can provide effective guidance for the completion of the head word. Considering the role of semantic collocation, we apply CNN to extract local information.
(c) **Category prediction of omitting center**. Based on the contextual global semantic information and local information, the category of omitting center is predicted.

Generally, each three-stage strategy can be organized as: semantic representation learning of word and sentence, task-oriented cooperation of linguistic knowledge and output construction. We present each task in detail in the following sections.

4 Boundary definition

Boundary definition of DE is to extract all DEs automatically from sentences by locating the left and right boundaries. As shown in Fig. 3, we propose a three-stage strategy for boundary prediction. First, we learn the hierarchical information of sentences. Then we employ DC-CNN to learn variable n-gram features flexibly which can facilitate the identification of the core components in the sentence. With global information of whole sentence and local information in a word, we can predict where the boundary of DE is in a sentence.

Let's assume a sentence to be considered as: $c = \{c_1, c_2, \ldots, c_{de}, \ldots, c_n\}$, where $c_i = \{word_i, pos_i\}$, $1 \le i \le n$. Here, $word_i$ and pos_i represent thei^{th} word and part of speech respectively. de is the index of the Chinese word "的(de)" that dominates the DE, and n is the length of the sentence. The goal of this task is to extract all DEs automatically from sentences by locating the left and right boundaries.

4.1 Semantic representation learning of global information for boundary prediction

Like other deep learning models, each word or part of speech is represented as a dense vector extracted from an embedding matrix. Then the distributed representation of each word can be formulated as follow:

$$\boldsymbol{v}_i = [\boldsymbol{w}_i; \boldsymbol{p}_i], 1 \leq i \leq n, \tag{1}$$

where $\boldsymbol{w}_i \in \mathbb{R}^{d_w}$ and $\boldsymbol{p}_i \in \mathbb{R}^{d_p}$ represent the embedding of i^{th} word and part of speech respectively, $[\boldsymbol{w}_i; \boldsymbol{p}_i] \in \mathbb{R}^{d_w + d_p}$ is the concatenation of two vectors.

Considering that the semantic information of a word is not only related to the information before the word, but also to the information after the word, a bi-directional Long Short-Term Memory (Bi-LSTM) [28] which combines a forward-traversal LSTM and a backward-traversal LSTM is used to model the representation of a word in the sentence.

$$\boldsymbol{h}_t^l = \left[\overrightarrow{\boldsymbol{h}_t^l}; \boldsymbol{h}_t^l\right], 1 \leq t \leq n. \tag{2}$$

$$\overrightarrow{\boldsymbol{h}_t} = lstm\left(\overrightarrow{\boldsymbol{h}_{t-1}}, \boldsymbol{v}_t\right), 1 \leq t \leq n, \tag{3}$$

$$\boldsymbol{h}_t = lstm\left(\boldsymbol{h}_{t-1}, \boldsymbol{v}_t\right), 1 \leq t \leq n, \tag{4}$$

Syntactic structures are hierarchical, while semantic structures are more complex. It is not enough to rely solely on the information obtained by sequence traversal. Therefore, stack LSTM is used to improve the effect of semantic modeling further. It's known to us that a deeply layered model is more capable of getting more abstract features [29].

$$\boldsymbol{h}_t^l = \left[\overrightarrow{\boldsymbol{h}_t^l}; \boldsymbol{h}_t^l\right], \boldsymbol{h}_t^0 = \boldsymbol{v}_t, 1 \leq t \leq n, \tag{5}$$

$$\overrightarrow{\boldsymbol{h}_t^l} = lstm\left(\overrightarrow{\boldsymbol{h}_{t-1}^l}, \boldsymbol{h}_t^{l-1}\right), 1 \leq t \leq n, , \tag{6}$$

$$\boldsymbol{h}_t^l = lstm\left(\boldsymbol{h}_{t-1}^l, \boldsymbol{h}_t^{l-1}\right), 1 \leq t \leq n, . \tag{7}$$

With the increment of layers, there are gradient explosion and gradient vanishing in the training of stack LSTM. However, DEs appear mostly in nested forms. For example, in "我们*(we)* 的*(de)* 邻居*(neighbor)* 的*(de)* 小孩*(child)*", "我们*(we)*" is the modifier of "邻居*(neighbor)*" rather than "小孩*(child)*". Another example is "我*(my)* 的*(de)* 红色*(red)* 的*(de)* 书包*(schoolbag)*", in which "我*(my)*" modifies "书包*(schoolbag)*" and "红色*(red)*" also modifies "书包*(schoolbag)*". Thus, in this task, we need a model with more layers, which can be successfully trained, to get hierarchical information.

To solve this problem, many solutions have been proposed, such as Highway LSTM [30] which extends stacked LSTM by introducing gated direct connections between memory cells in adjacent layers, and densely connected Bi-LSTM (DC-Bi-LSTM) [31] in which the upstream layer can directly access the outputs of all the downstream layers. Here, we use DC-Bi-LSTM to integrate syntax and semantic information into the representation of words. The model structure is shown in Fig. 4.

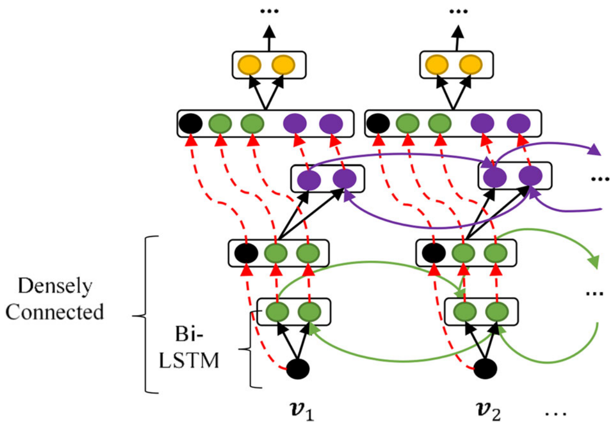

Fig. 4 Illustration of DC-Bi-LSTM. Each black node denotes an input. Green, Purple, and yellow nodes denote hidden units

As shown in Fig. 4, DC-Bi-LSTM essentially represents each layer by the concatenation of its hidden state and all preceding layers' hidden states, followed by recursively passing each layer's representation to all subsequent layers. DC-Bi-LSTM can alleviate the problems of vanishing-gradient and overfitting and can be successfully trained when the networks are as deep as dozens of layers. For the l^{th} layer, the above process is formulated as follows:

$$\overrightarrow{\boldsymbol{h}_t^l} = lstm\left(\overrightarrow{\boldsymbol{h}_{t-1}^l}, \boldsymbol{M}_t^{l-1}\right), \tag{8}$$

$$\boldsymbol{h}_t^l = lstm\left(\boldsymbol{h}_{t-1}^l, \boldsymbol{M}_t^{l-1}\right),, \tag{9}$$

$$\boldsymbol{M}_t^{l-1} = \left[\boldsymbol{h}_t^1; \boldsymbol{h}_t^2; \ldots; \boldsymbol{h}_t^{l-1}\right]. \tag{10}$$

For a L layer DC-Bi-LSTM, the output is $\boldsymbol{h}^L = \left\{\boldsymbol{h}_1^L, \boldsymbol{h}_2^L, \ldots, \boldsymbol{h}_n^L\right\}$.

4.2 Local information extraction based on DC-CNN

Considering DEs appear mostly in various nested forms, it is necessary to consider the internal relationship between n-grams of various granularities in order to accurately locate the beginning of modifiers and centers of Des. For example, trigram features can be equivalently constructed on the basis of several bigram features or the combination of unigram features and bigram features [27]. Convolutional neural network (CNN) is an effective model for extracting semantic representations and capturing salient features in a flat structure [32]. Here, before using DC-Bi-LSTM to model the hierarchical information of the sentence, we employ Densely Connected CNN (DC-CNN) [27] to learn variable n-gram features flexibly, to facilitate the identification of the core components in the sentence. We briefly introduce the model below.

In Figure 5, w is the window-size of convolutional filter, $f(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is the feature from the second layer, $f(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ is the feature from the third layer, and so on. By adding dense

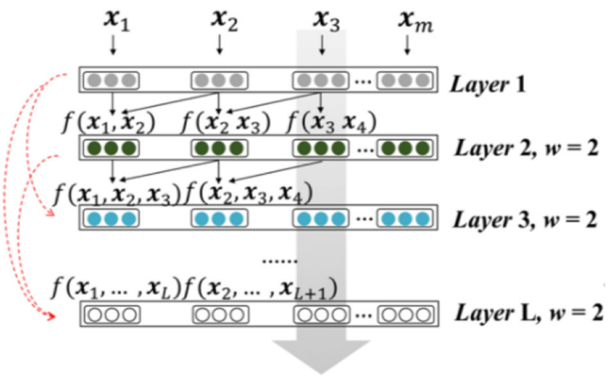

Fig. 5 Intuitive illustration of how to generate multi-scale features in DC-CNN

connections between layers, feature maps at downstream layers can be effectively re-used at upstream layers. Consequently, such a deep model offers more flexibility to construct multi-scale features than a wide model. DC-CNN takes the outputs of all downstream layers, $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, …, $\boldsymbol{X}_{l-1}$, as input and generates feature maps for the l^{th} layer, as formulated below:

$$\boldsymbol{X}_l = f'\prime(\boldsymbol{W}_l, [\boldsymbol{X}_1; \boldsymbol{X}_2; \ldots; \boldsymbol{X}_{l-1}]). \tag{11}$$

Here, $\boldsymbol{W}_l$ is a set of convolutional filters. Each convolutional filter is a list of linear layers with shared parameters [33]. For a L layer DC-CNN, the output is $\boldsymbol{X} = \{\boldsymbol{X}_1; \boldsymbol{X}_2; \ldots; \boldsymbol{X}_L\}$, where $\boldsymbol{X}_l = [\boldsymbol{x}_l^1; \boldsymbol{x}_l^2; \ldots; \boldsymbol{x}_l^n] \in \mathbb{R}^{n \times k}$, $1 \le l \le L$.

Moreover, because of the diversity of nested forms, the contradiction between granularity and semantics should also be considered [34]. Here, we use a multi-scale feature attention to select task-friendly features adaptively based on results of different granularities [27]. The process is formulated as follows:

$$\boldsymbol{s}_l^i = F_{ensem}(\boldsymbol{x}_l^i) = \sum_{j=1}^{k} \boldsymbol{x}_l^i(j), 1 \le l \le L , \tag{12}$$

$$\boldsymbol{a}^i = softmax(MLP(\boldsymbol{s}^i)), 1 \le i \le n , \tag{13}$$

$$\boldsymbol{x}_{atten}^i = \sum_{l=1}^{L} a_l^i \boldsymbol{x}_l^i, 1 \le i \le n . \tag{14}$$

Here, $\boldsymbol{s}_l^i$ is a scalar can be used as a descriptor of the feature vector, $\boldsymbol{x}_l^i$, $\boldsymbol{x}_{atten}^i \in \mathbb{R}^{\mathrm{k}}$, and $\boldsymbol{a}^i$ is the attention weight. And the output of DC-CNN, $\{\boldsymbol{x}_{atten}^1, \boldsymbol{x}_{atten}^2, \ldots \boldsymbol{x}_{atten}^n\}$, is the input of DC-Bi-LSTM.

4.3 Boundary prediction

The prediction probabilities is computed by

$$\widehat{\boldsymbol{y}} = softmax(\boldsymbol{W}_d \boldsymbol{h}^L + \boldsymbol{b}_d). \tag{15}$$

And we use the negative log likelihood of correct boundaries as training loss.

5 Recognition of elliptical DE

Recognition of elliptical DE aims at figure out whether a DE is elliptical. We establish three-stage strategy for recognition. First of all, we utilize Bi-LSTM to obtain semantic and syntactic contextual information. After that we capture features of ellipsis by which the network can recognize the DE containing the usage of semantic ellipsis. By applying the semantic, syntactic contextual information and the features, the probability of whether a DE is elliptical DE is predicted.

Let's assume a sentence to be considered as $c = \{c_1, c_2, \ldots, c_{de}, \ldots, c_n\}$, where $c_i = \{word_i, pos_i\}$, $1 \leq i \leq n$. The goal of this task is to automatically recognize the DE containing the usage of semantic ellipsis. That is to say, it's a binary classification task. Figure 6 depicts the model architecture for recognition of elliptical DE.

5.1 Semantic representation learning of context information for elliptical DE

According to previous studies, the recognition of the elliptical DE mainly depends on the syntax information. Since dimension of the word vector is much larger than that of the part-of-speech vector, we separately deal with words and part-of-speech information to avoid the possibility that the model is insensitive to part-of-speech information. In semantic modeling, same operations are performed on the words and parts of speech of the DE separately, and the results of them are merged in the output module. Considering that the recognition of the elliptical DE relies more on shallow syntactic information than deep semantic information, in this task, we only utilize Bi-LSTM to obtain semantic and syntactic contextual information. Specifically, we choose Bi-LSTM for word and the layer of stack Bi-LSTM for part-of-speech is 2.

Besides contextual information obtained by sequence traversal, we also take the position information into consideration. The calculation of word representation of i^{th} word $\boldsymbol{m}_i$ is as follows:

$$u_i = \frac{i-de}{n}, 1 \leq i \leq n, \tag{16}$$

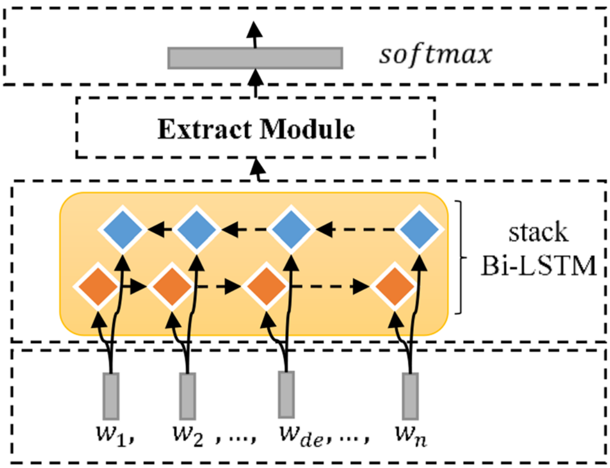

Fig. 6 The illustration of model architecture for recognition of elliptical DE

$$weight_i = 1 - |u_i|, 1 \le i \le n, , \tag{17}$$

$$\boldsymbol{m}_i = \left[weight_i \odot \boldsymbol{m}_i^*; u_i\right], 1 \le i \le n. \tag{18}$$

Here, m_i^* is the word representation of i^{th}word calculated by Bi-LSTM, the distance formula is proposed by Chen [35], ⊙ means element-wise multiplication, *de* is the index of the Chinese word "的(de)" that dominates the DE, and u_i is the distance between the i^{th} word and the word "的(de)".

5.2 Elliptical feature extraction

In order to recognize the elliptical DE, a critical task is to extract appropriate elliptical features. In some simple cases, the elliptical DE can be recognized by the word next to "的(de)". For example, in the Chinese sentence,"分区定价法*(zone pricing)* 力图*(strive to)* 做*(do)* 的*(de)* 远远不止*(far more than)* 反映*(reflect)* 不同*(different)* 税率*(tax rate)*", has one elliptical DE, "分区定价法*(zone pricing)* 力图*(strive to)* 做*(do)* 的*(de)*", of which the semantic ellipsis is the omitted center "事情*(thing)*". According to the Chinese word "做(do)", nearest to "的(de)", we can determine that this structure is elliptical. At the same time, there are many cases in which several words can be the evidences for judgment. In Chinese sentence, "他们*(they)* 要*(want)* 找*(find)* 的*(de)* 是*(is)* 失散*(lost)* 多*(many)* 年*(year)* 的*(de)* 女儿*(daughter)*", the DE, "他们*(they)* 要*(want)* 找*(find)* 的*(de)*" is elliptical. In this example, according to several Chinese words "找*(find)*", "是*(is)*" and "女儿*(daughter)*", we can determine that it is elliptical. These motivate us to develop a extract module, which is capable of capture features of ellipsis by which the network can recognize the DE containing the usage of semantic ellipsis.

In the extract module, we use two methods. One is Max-Pooling operation, and the other is GRU based multiple attention mechanism.

Max-Pooling operation has outstanding performance in tasks of sentence classification [36, 37]. The detailed computation is described as follows:

$$\boldsymbol{single\_feature} = max_{\boldsymbol{j=1}}^{\boldsymbol{n}} \boldsymbol{m}_j , \tag{19}$$

$$\boldsymbol{e} = MLP(\boldsymbol{single\_feature}) . \tag{20}$$

Here, *MLP* is a feedforward layer.

In GRU based multiple attention mechanism, we first extract features of ellipsis, and then combine them in a reasonable way. The detailed computation is described as follows:

$$\boldsymbol{g}_j^t = \boldsymbol{W}_{att}\left[\boldsymbol{m}_j; \boldsymbol{e}_{t-1}; \boldsymbol{v}_{de}; u_j\right] + \boldsymbol{b}_{att}, 1 \le j \le n, \tag{21}$$

$$a_j^t = \frac{\boldsymbol{exp}\left(\boldsymbol{g}_j^t\right)}{\sum_{\boldsymbol{k=1}}^{\boldsymbol{n}} \boldsymbol{g}_k^t}, 1 \le j \le n, \tag{22}$$

$$\boldsymbol{i}_t = \sum_{\boldsymbol{j}=\boldsymbol{1}}^{\boldsymbol{n}} a_j^t \boldsymbol{m}_j, 1 \le j \le n. \tag{23}$$

a_j^t is the attention score, $\boldsymbol{i}_t$ is the output of the t^{th} attention layer. Then a GRU unit to combine the output of each attention layer in a nonlinear way. Thus, different locations of the input can be noticed in different attention layers [38]. GRU unit is calculated as follows.

$$\boldsymbol{r} = \boldsymbol{\sigma}(\boldsymbol{W}_r \boldsymbol{i}_t + \boldsymbol{U}_r \boldsymbol{e}_{t-1}), 1 \le t \le L, \tag{24}$$

$$\boldsymbol{z} = \boldsymbol{\sigma}(\boldsymbol{W}_z \boldsymbol{i}_t + \boldsymbol{U}_z \boldsymbol{e}_{t-1}), 1 \le t \le L, \tag{25}$$

$$\boldsymbol{e}_t' = \boldsymbol{tanh}\left(\boldsymbol{W}_i \boldsymbol{i}_t + \boldsymbol{W}_g (\boldsymbol{r} \odot \boldsymbol{e}_{t-1})\right), 1 \le t \le L, \tag{26}$$

$$\boldsymbol{e}_t = (\boldsymbol{1} - \boldsymbol{z}) \odot \boldsymbol{e}_{t-1} + \boldsymbol{z} \odot \boldsymbol{e}_t', 1 \le t \le L. \tag{27}$$

Here, we set the layer number L of GRU based multiple attention mechanism is 3.

5.3 Category prediction of elliptical DE

Motivated by TD-LSTM, an extension on LSTM proposed by Tang [39], in output module, we feed the elliptical features into a *softmax*classifier and obtain predicted probability distribution over all labels. A reasonable training objective to be minimized is the categorical cross-entropy loss.

6 Semantic complement for elliptical DE

AMR has a complete set of named entity concepts. *Thing* and *person* are the two most commonly used concepts. In the abstraction process in which words are converted to concepts, concepts in the set can be added to represent implied semantics. Motivated by this approach of AMR, we define the problem as a multi-class classification task. Let's assume a sentence to be considered as$c = \{c_1, c_2, \ldots, c_{de}, \ldots, c_n\}$, where$c_i = \{word_i, pos_i\}$, $1 \le i \le n$. The goal of this task is to choose a concept in the set to complement the semantic ellipsis of an elliptical DE automatically. The approach for this task is shown in Figure 7.

Firstly, we incorporate deep, abstract contextual syntax and semantic information into the representation of words. Then, based on the partial language phenomenon of elliptical DE, the network extract local features. Based on the contextual global semantic information and local features, the category of the default omitted center of DE can be predicted.

6.1 Semantic representation learning of context information for semantic complement

The distributed representation of each word is defined by Eq. (1).

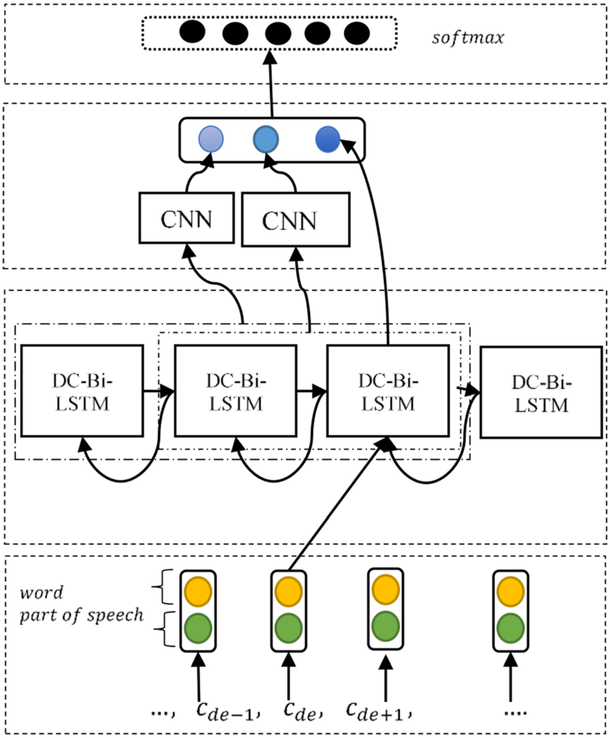

Fig. 7 The illustration of model architecture for semantic complement for elliptical DE

Complement of the center depends on the comprehensive analysis of semantics and context. We use the DC-Bi-LSTM mentioned in Section 4.2 to effectively integrate rich and abstract contextual syntactic and semantic information into the representation of words. The output is$\boldsymbol{h} = \{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_n\}$, $\boldsymbol{h}_i \in \mathbb{R}^d$ represents a d-dimensional embedding for the i^{th} word.

6.2 Local information learning for semantic complement

A human asked to do this task will selectively focus on parts of the structure and acquire information. In some cases, based on the information offered by words just before the dominant word "的(de)", the appropriate concept can be chosen to representing the omitted content. This linguistic phenomenon inspires us to pay attention to semantics of n-grams of various granularities in sentences.

We enrich the network model with multiple convolutional filters with different widths to generate local context representations. Convolutional filters with widths of 2 and 3 encode the semantics of bigrams and trigrams. Local information can help our model fix some errors due to lexical ambiguity [40].

$$\boldsymbol{phrase}_i^l = Relu\left(\boldsymbol{W}_p^l \cdot [\boldsymbol{h}_{i-l}; \ldots; \boldsymbol{h}_i] + \boldsymbol{b}_p^l\right), \tag{28}$$

Furthermore, in order to preserve the context information of the elliptical DE, we concatenate the phrase vectors of i^{th} word and the corresponding hidden state h_i produced in DC-Bi-LSTM. For i^{th} word, the representation is denoted as $e_i = \left[h_i; phrase_i^2; phrase_i^3\right]$.

6.3 Category prediction of omitting center

In output module, we directly extracts the word representation e_{de} of the word "的(de)" that dominates the elliptical DE, and then feeds it into a *sofmax* classifier, which takes e_{de} as features and generates predicted probability distribution over all labels. And the training objective is the categorical cross-entropy loss.

7 Experiments

7.1 Dataset

CAMR can describe and reveal the complete and deep semantic information contained in sentences [21]. In graph of CAMR, structural auxiliary "的(de)" will be labeled in the edge and non-structural auxiliary "的(de)" will not appear in the edge, or in the graph. We automatically extract the experimental datasets based on CAMR corpus and the number the word "的(de)" is 11,829. As shown in Table 1, by synthesizing Eight Hundred Words in Modern Chinese, Semantic Knowledge—base of Contemporary Chinese [16, 17] and CAMR annotated corpus, and analyzing the contextual features of "的(de)", we divide the meaning of "的(de)" into four meaning items.

In this paper, we deal with three research works on the background of DE. The specific research object of each task is different, and the experimental dataset is also different. Below is a brief introduction.

For boundary definition of DE, the dataset includes those DEs which are extract from CAMR corpus. Considering the strong learning ability of neural networks, we also add those non-structural auxiliary words into the dataset. Unlike the DE that dominated by structural auxiliary "的(de)", non-structural auxiliary "的(de)" does not have its left and right boundaries. In this paper, we define the left and right boundaries of non-structural auxiliary "的(de)" is "的(de)" itself. Moreover, the right boundary of elliptical DE is also the "的(de)" itself. The number of structural auxiliary "的(de)" and non-structural auxiliary "的(de)" is 11,116 and 713 respectively.

In Table 1, all the Des of item 3 and some Des of item 4 are elliptical. The rest do not belong to the usage of ellipsis. The number of elliptical DE is 664. Obviously, the dataset has unbalanced classes. Thus, besides elliptical DEs mentioned above, we also manually extract elliptical DEs in some sentences of CTB corpus without the annotation of CAMR. Moreover, we also halve the amount of samples without the usage of ellipsis. For the recognition of

Table 1 Interpretation the meaning items of "的(de)"

Item ID	Interpretation	Usage
1	The DE is a noun phrase in sentence.	Noun | verb | adjective | adverb | clause | prepositional phrase + "的(de)" + noun
2	Express a certain tone.	At the end of a declarative or interrogative sentence
3	It's elliptical DE, a variation of DE	Noun | verb | adjective | clause + "的(de)"
4	The DE is a predicate.	Noun | pronoun | verb | adjective | clause | idiom + "的(de)"

elliptical DE, the number of “的(de)” without the usage of ellipsis and elliptical DEs is 6191 and 1830 respectively.

For semantic complement for elliptical DE, the research object is the elliptical DE. Besides, elliptical DEs mentioned above, we continue to expand data scale. The number of elliptical DEs increases to 2265. There are 23 different concepts in the corpus. The number of concept *thing* and *person* accounts for 94% of the entire dataset. The number of most of the remaining concepts does not exceed five. Inspired by concept classification of several Chinese language knowledge resources, e.g., Semantic Knowledge—base of Contemporary Chinese (SKCC) [41] and HowNet [42], we merge all concepts into 5 categories, *thing*, *person*, *organization*, *location*, and *animal*. The number of samples in these categories is 1639, 515, 51, 41, and 19, respectively.

8 Results

8.1 Boundary definition

For boundary definition of DE, the evaluation metrics is the accuracy of boundary definition. Because of the inconsistency between the research objects, the experimental results of the relevant working methods are not given in this paper. We focused on the effect of several improvements mentioned in Section 4.2 on performance.

There are several hyper parameters. We use 100-dimension word vectors pre-trained by Word2vec with the CTB corpus, and randomly initialize 20-dimension part of speech vectors. We only tune the embedding of part of speech during training. In DC-CNN, the initial filter window is 2, the number of filters is 200, and the network layers is 2. The layer number of DC-Bi-LSTM is 5. The number of hidden units of top Bi-LSTM (the last Bi-LSTM layer in DC-Bi-LSTM) is 100. For the rest layer of DC-Bi-LSTM, the number is 50. Adam [43] is used to optimize the neural network, and the learning rate is 0.001. For regularization, in DC-Bi-LSTM module and DC-CNN module, dropout operation [44] is performed for each output layer to prevent over-fitting, and the ratio is set to 0.5.

Qian [12] recognized right boundary and left boundary successively, and utilized the features of right boundary to the identification of left boundary. Motivated by this approach, we decide to recognize right boundary and left boundary simultaneously by parameter sharing.

As shown in Table 2, with task-related improvements, LSTM+CNN * outperforms other methods. On the one hand, it is reasonable and effective to adopt the appropriate approach to model the semantics of the sentence according to characteristics of task. In other words, effective use of context information determines better experimental results. One the other hands, in the aid of joint learning, the performance of our proposed approach is improved. It is mainly because joint learning can be used to learn rich features that might not be easy to learn just using the original task. LSTM+CNN is worse than LSTM. A potential reason might be that DC-CNN has larger number of parameters, which cannot be easily optimized with the small number of corpus. The results of recognition of left boundary are worse than that of the right boundary. It main because those DEs appear mostly in nested forms.

Although intuitively, in deeper layer, the convolution filters can cover more words, even the whole sentence, so that the global feature representations of the sentence can be obtained. However, as shown in Table 2, the method of CNN still needs to work with RNN to achieve better experimental results. This is consistent with Yin’s viewpoint that the experimental

Table 2 Experimental results of boundary definition of DE. LSTM: DC-Bi-LSTM. CNN: DC-CNN. *: Joint learning. ACC-L: the accuracy of left boundary identification. ACC-R: the accuracy of right boundary identification. ACC-LR: the accuracy of boundary identification. The best method in each setting is in bold

Method	*ACC-L*	*ACC-R*	*ACC-LR*
LSTM	0.8351	0.9150	0.7634
LSTM+CNN	0.8123	0.9052	0.7323
LSTM+CNN *	**0.8469**	**0.9170**	**0.7772**

results often depend on the specific need of the task for global semantics, rather than the methods which have been proposed in [45].

8.2 Recognition of elliptical DE

For recognition of elliptical DE, the main evaluation metrics are the accuracy of classification of DE and F1 value of recognition elliptical DE. Referring to related works, we use two baseline methods: CRF_1 and GRU_1. CRF_1 is a traditional conditional random field based model with the feature template proposed by Liu. In GRU_1. Because Liu did not specify the hyper-parameter settings of the network model (such as the dimension of word vectors etc.), the hyper-parameter settings are the same as those in our work and the hidden state corresponding to the position of the word "的(de)" is used as the output. We conduct comparative experiments to prove the effectiveness of proposed methods.

There are several hyper parameters. The dimension of word and part of speech representation are 100 and 20, respectively, and only the part of speech vectors are updated with iteration during the training process. The number of hidden units of Bi-LSTM of word and part of speech is 64 and 32 respectively. The dimension of output in Max-Pooling layer is 16. The dimension of output of word and part of speech in GRU based multiple attention layers are 16 and 20 respectively. We use Adam [43] to optimize the neural network and the learning rate is 0.001.

Experimental results of recognition of elliptical DE are shown in Table 3. It is clear that the recognition performances of proposed models are better than that of CRF_1 and GRU_1, which indicates that the mode and method of human problem solving is useful for the construction of neural network model.

From Table 4, we can find that None and Att perform worse than Max-pooling and GRU + Att which demonstrates that how to extract features of ellipsis from the structure is directly connected to the overall performance. Besides, Comparing between GRU + Att and Att, we

Table 3 Experimental results of recognition of elliptical DE. ACC: the accuracy of classification of DE. *F1-nonE:* the f1 value of the recognition of elliptical DE. *F1-E* the f1 value of the recognition of DE without ellipsis. CRF and GRU are baseline methods. GRU + ATT: GRU based multiple attention layers mentioned in Section 5.2. Max-pooling: Max-Pooling layer mentioned in Section 5.2

Method	**ACC**	**F1-nonE**	**F1-E**
CRF_1	0.9451	0.9649	0.8739
GRU_1	0.8737	0.9209	0.6873
GRU + ATT	0.9838	0.9895	0.9642
Max-pooling	**0.9850**	**0.9903**	**0.9667**

Table 4 Experimental results of different extract methods. None: without extract module. Att: the setting of extract module is a soft-attention mechanism [46]

Method	*P*	*R*	F1
None	0.9407	0.9641	0.9523
Att	**0.9641**	0.9432	0.9536
Max-pooling	0.9613	**0.9720**	**0.9667**
GRU + Att	0.9615	0.9669	0.9642

find that it is necessary to combine the elliptical features in a reasonable way, some related contributions are also described in [47, 48].

8.3 Semantic complement for elliptical DE

We employ classification accuracy and F1 score along with precision and recall metrics to measure the performance of our proposed method. F1 score gives equal weight to each class label that is suitable for classification tasks with unbalanced categories.

There are several hyper parameters. The dimension of word and part of speech representation are 100 and 20, respectively. The number of each filter in CNN is 20. The layer number of DC-Bi-LSTM is 5. The number of hidden units of top Bi-LSTM (the last Bi-LSTM layer in DC-Bi-LSTM) is 100. For the rest layer of DC-Bi- LSTM, the number is 40. The mini-batch size is 10 and the learning rater of Adam [43] is 0.001.

As mentioned above, most previous works on DE only focus on the usage and characteristics of auxiliary "的(de)" from the syntactic perspective, thus we do not compare our proposed method against their methods directly. Since contextual information is the decision fundament, we design a baseline model, CRF_2, a CRF model with the feature template proposed by Liu [19], instead of comparing with rule-based methods.

Table 5 depicts the performance of six models. From row 1, we can see that feature-driven CRF is an extremely strong performer. Comparing between row 1 and 6, we find the effectiveness of our proposed method, which benefits from rich automatic features generated from the well-designed network. Comparing between row 2 and 6, we can see that even if we just integrate a simple linguistic phenomenon of the elliptical DE into the network construction, the performance of the model can be improved. Thus, building networks with the linguistic knowledge of elliptical DE is a worthwhile option to try in the future work. Row 3 and 6 show the effectiveness of densely connected Bi-LSTM. In Stack, the layer number is 3, which is differ from DC. Although we can learn abstract representations by increasing the

Table 5 Experimental results of semantic complement for elliptical DE. CNN: the method mentioned in Section 6.2. Stack: stack Bi-LSTM. DC: DC-Bi-LSTM. ATT_2: a soft- attention mechanism [46]. $MaxPooling_2$: an outstanding performance in tasks of sentence classification [36, 37]

#	Method	Accuracy
1	CRF_2	0.8145
2	DC + DE	0.8244
3	Stack+CNN + DE	0.8156
4	DC + CNN + ATT_2	0.7911
5	DC + CNN + $MaxPooling_2$	0.8067
6	DC + CNN + DE	**0.8400**

Table 6 Experiment results of different concepts

#	thing			person			organization			location			animal		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
1	0.8245	0.9538	0.8844	0.7536	0.5049	0.6047	0.5000	0.0833	0.1428	1.0000	0.2222	0.3636	1.0000	1.0000	1.0000
2	0.8457	0.9534	0.8963	0.7215	0.5534	0.6263	0.7500	0.2500	0.3750	1.0000	0.5000	0.6666	1.0000	0.2222	0.3636
3	0.8361	0.9348	0.8826	0.5825	0.7229	0.6451	0.6667	0.1667	0.2667	1.0000	0.3333	0.4999	1.0000	0.2500	0.4000
4	0.8079	0.9534	0.8746	0.6949	0.3981	0.5062	0.8571	0.5000	0.6315	0.5000	0.1111	0.1818	0.5000	0.2500	0.3333
5	0.8301	0.9410	0.8820	0.7353	0.4854	0.5847	0.5000	0.3333	0.3999	0.5000	0.3999	0.3999	1.0000	0.7500	0.8571
6	0.8772	0.9317	0.9036	0.7174	0.6408	0.6769	0.6667	0.5000	0.5714	1.0000	0.3333	0.4615	1.0000	0.7500	0.8571

depth of network layers, we also need to prevent side effects of this choice, such as gradient vanishing. We also conduct experiments with different operations in output module. The elliptical DE is a complex structure, so that how to extract information from the structure to complement semantic ellipsis is directly connected to the overall performance. From row 4–6, compared with ATT_2 and $MaxPooling_2$, De has better performance than the other two methods. A potential reason might be that the attention-based method has larger number of parameters, which cannot be easily optimized with the small number of corpus. What's more, we think the extraction method is still worth exploring in output module.

Because the dataset has unbalanced classes, F-measure is also reported. In Table 6, the corresponding model settings for each serial number are the same as Table 5. The analyses are as follows.

First, due to the amount of data of each concept is unbalanced, the model must have the ability to distinguish concept thing and other concepts first. From row 2,3, and 6,considering other concepts and their number, we find if a model has good performance on recognizing concept thing and person at the same time, it may be competent for this task.

Second, one can notice that the poor performances of concept organization, location and animal do not affect the overall performance obviously, due to the amount of data of these concepts is small. Regardless of the module settings, these results of concept organization, location are not good. Thus, the way we merge concepts may also be improved in the future work. Moreover, we believe that increasing the size of the corpus of this task is a worthwhile option to try in the future work.

9 Conclusions and future work

By referring to the theoretical researches, literature reports based on rules and statistical methods and the practical experience of deep learning related models and methods that have been employed for numerous NLP tasks, we propose a three-stage strategy, to solve tasks on the background of DE. Based on the strategy, we accomplish three research works on the background of DE, namely, boundary definition of DE, the recognition of elliptical DE, and semantic complement for elliptical DE. Experimental results show that the proposed models can effectively define the boundaries of DE, recognize elliptical DE and complement the center word that is omitted due to the constraints of the context and linguistic economic principles. We hope that the proposed strategy can provide inspiration and help for researchers in the follow-up work.

As future work, we will complete more tasks on the background of DE with proposed strategy to provide strong supports for upper applications, such as dependency parsing, CAMR parsing, and machine translation. Meanwhile, boundary definition and usage identification have their own characteristics, but they are not completely independent. Making good use of the relationship between them is also the direction in the future.

Acknowledgements This work is supported by National Natural Science Foundation of China under Grant 61772278, 61472191, and 41571382, National Social Science Fund of China under Grant 18BYY127, the project for Jiangsu Higher Institution's Excellent Innovative Team for Philosophy and Social Science 2017STD006, Fujian Provincial Fund under Grant MJUKF201705, and Science & Technology Planning Project of Fuzhou City No.2017-G-106.

References

1. Zhang, X.: A survey of the studies on "de" constructions in modern Chinese language. Journal of Kaifeng Institute of Education. **32**(4), 50–52 (2012)
2. Xu, Y.: A Study on Form Word "de" and its Related Matters. Fudan University, Doctor (2003)
3. Liu, Y.: Error Analysis and Teaching Strategies of Structural Auxiliary Word "de" of International Students. Nanchang University, Master (2018)
4. Lai, Y.: A Study of the Word "de" in the End of Sentence. Nanchang University, Master (2018)
5. Rao, S., Marcu, D., Knight, K., Daumé III, H.: Biomedical event extraction using abstract meaning representation. In: BioNLP. pp. 126–135 (2017)
6. Garg, S., Steeg, G. V., Galstyan, A.: Stochastic Learning of Nonstationary Kernels for Natural Language Modeling. arXiv preprint arXiv: 1801.03911 (2018)
7. Dohare S., Karnick, H.: Text Summarization using Abstract Meaning Representation. arXiv preprint arXiv: 1706.01678 (2017)
8. Jones, B. K., Andreas, J., Bauer, D., Hermann, K. M., Knight K.: Semantics-based machine translation with hyperedge replacement grammars. In: COLING. pp. 1359–1376 (2012)
9. Mitra, A., Baral, C.: Addressing a question answering challenge by combining statistical methods with inductive rule learning and reasoning. In: AAAI. pp. 2779–2785 (2016)
10. Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., Schneider, N.: Abstract meaning representation for sembanking. In: LAW@ACL. pp. 178–186 (2013)
11. Li, B., Wen, Y., Song, L., Bu, L., Qu, W., Xue, N.: Constriction of a Chinese abstract meaning representation corpus with concept-to-word alignment. Journal of Chinese Information Processing. **31**(6), 93–102 (2017)
12. Qian, X.: Automatic Identification of Maximal Noun Phrase that Contains "de". Nanjing Normal University, Master (2007)
13. Qian, X.: Recognition of MNP with "de-phrase". Computer Engineering and Applications. **46**(18), 138–141 (2010)
14. Xiao, C., Zhou, J.: Method of identifying the border of de phrases in Chinese. Natural Science Journal of Xiangtan University. **02**, 140–144 (1999)
15. Yang, X., Xu, W., Zhao, K., Zhang, N.: Analysis of "Vde" structure in Chinese. Electronic Science and Technology. 21(06), 54–57+69 (2008)
16. Zan, H., Zhang, K., Chai, M., Yu, S.: Studies on the functional word knowledge base of modern Chinese. Journal of Chinese Information Processing. **21**(5), 107–111 (2007)
17. Zhang, K., Zan, H., Chai, M., Han, Y., Zhao, D.: Survey of the Chinese function word usage knowledge base. Journal of Chinese Information Processing. 29(3), 1–8+15 (2015)
18. Han, Y., Zan, H., Zhang, K., Chai, M.: Automatic annotation of auxiliary words usage in rule-based Chinese language. Journal of Computer Applications. **31**(12), 3271–3274 (2011)
19. Liu, Q., Zhang, K., Xu, H., Yu, S., Zan, H.: Research on automatic recognition of auxiliary "DE". Acta Sci. Nat. Univ. Pekin. **54**(3), 466–474 (2018)
20. Wang, L.: The Recovery of Omissible Noun in Natural Language Understanding and its Application on Product Design. Xidian University, Master (2010)
21. Qu, W., Zhou, J., Wu, X., Dai, R., Gu, M., Gu, Y.: Survey on abstract meaning representation. Journal of Data Acquisition & Processing. **32**(1), 26–36 (2017)
22. Li, B., Wen, Y., Bu, L., Qu, W., Xue, N.: A comparative analysis of the AMR graphs between English and Chinese corpus of the little prince. Journal of Chinese Information Processing. 31(1), 50–57+74(2017)
23. Song, L., Wen, Y., Ge, S., Li, B., Zhou, J., Qu, W.: Semantic role labeling of Chinese predicates based on the AMR corpus. Corpus Linguistics. 5(01), 45–58+116 (2018)
24. Gu, M.: Research on Chinese AMR Parsing Using Transition-Based Neural Network. Nanjing Normal University, Master (2018)
25. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature. **521**(7553), 436–444 (2015)
26. Shi, Y.: Research on Text Categorization Technology Based on Deep Learning. Beijing University of Posts and Telecommunications, Master (2018)
27. Wang, S., Huang, M., Deng, Z.: Densely connected CNN with multi-scale feature attention for text classification. In IJCAI. pp. 4468–4474 (2018)
28. Schuster, M., Paliwal, K.K.: Bidirectional recurrent neural networks. IEEE Trans. Signal Process. **45**(11), 2673–2681 (1997)
29. Sukhbaatar, S., Szlam, A., Weston, J., Fergus, R.: End-to-end memory networks. In NIPS. pp. 2440–2448 (2015)

30. Zhang, Y., Chen, G., Yu, D., Yao, K., Khudanpur, S., Glass, J. R.: Highway long short-term memory RNNS for distant speech recognition. In ICASSP. pp. 5755–5759 (2016)
31. Ding, Z., Xia, R., Yu, J., Li, X., Yang, J.: Densely connected bidirectional LSTM with applications to sentence classification. In NLPCC. pp. 278–287 (2018)
32. Liu, Y., Wei, F., Li, S., Ji, H., Zhou, M., Wang, H.: A dependency- based neural network for relation classification. In: ACL. pp. 285–290 (2015)
33. Zhang, Y., Wallace, B. C.: A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. In: IJCNLP. pp. 253–263 (2017)
34. Zhang, T., Huang, M., Zhao, L.: Learning structured representation for text classification via reinforcement learning. In: AAAI. pp. 6053–6060 (2018)
35. Chen, P., Sun, Z., Bing, L., Yang, W.: Recurrent attention network on memory for aspect sentiment analysis. In: EMNLP. pp. 452–461 (2017)
36. Conneau, A., Kiela, D., Schwenk, H., Barrault, L., Bordes, A.: Supervised learning of universal sentence representations from natural language inference data. In: EMNLP. pp. 670–680 (2017)
37. Collobert, R., Weston, J.: A unified architecture for natural language processing: deep neural networks with multitask learning. In: ICML. pp. 160–167 (2008)
38. Kumar, A., Irsoy, O., Ondruska, P., Iyyer, M., Bradbury, J., Gulrajani, I., Zhong, V., Paulus, R., Socher, R.: Ask me anything: dynamic memory networks for natural language processing. In: ICML. pp. 1378–1387 (2016)
39. Tang, D., Qin, B., Feng, X., Liu, T.: Effective LSTMs for target-dependent sentiment classification. In: COLING. pp. 3298–3307 (2016)
40. Feng, X., Huang, L., Tang, D., Ji, H., Qin, B., Liu, T.: A language-independent neural network for event detection. In: ACL. pp. 66–71 (2016)
41. Bai, M., Hsieh, Y., Chen, K., Chang, J. S.: Translating Chinese unknown words by automatically acquired templates. In: IJCNLP. pp. 839–843 (2013)
42. Dong, Z., Dong, Q., Hownet - a hybrid language and knowledge resource. In: NLP-KE. pp. 820–824(2003)
43. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv: 1412.6980v8 (2014)
44. Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
45. Yin ,W., Kann, K., Yu, M., Schutze, H.: Comparative study of cnn and rnn for natural language processing. arXiv preprint arXiv:1702.01923 (2017)
46. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)
47. Wei, T., Qu, W., Song, L., Dai, R.: A survey on the study of compound sentences with Chinese abstract meaning representation. Journal of Xiamen University (Natural Science). **57**(06), 849–858 (2018)
48. Young, T., Hazarika. D., Poria, S., Cambria, E.: Recent trends in deep learning based natural language processing. arXiv preprint arXiv: 1708.02709 (2017)