FOCUS

An attention mechanism and multi-granularity-based Bi-LSTM model for Chinese Q&A system

Xiao-mei Yu[1,2] · Wen-zhi Feng[1] · Hong Wang[1,2] · Qian Chu[1] · Qi Chen[1]

Abstract

Natural language processing (NLP) is one of the key techniques in intelligent question-answering (Q&A) systems. Although recurrent neural networks and long short-term memory (LSTM) networks exhibit obvious advantages on well-known English Q&A datasets, they still suffer from several defects including indeterminateness, polysemy and the lack of changing morphology in Chinese, which results in complex NLP on large and diverse Chinese Q&A datasets. In this paper, we first analyze limitations of applying LSTM and bidirectional LSTM (Bi-LSTM) models to noisy Chinese Q&A datasets. Then, we focus on integrating attention mechanisms and multi-granularity word segmentation into Bi-LSTM and propose an attention mechanism and multi-granularity-based Bi-LSTM model (AM–Bi-LSTM) which combines the improved attention mechanism with a novel processing of multi-granularity word segmentation to handle the complex NLP in Chinese Q&A datasets. Furthermore, similarity of questions and answers is formulated to implement the quantitative computation which helps to achieve better performance in Chinese Q&A systems. Finally, we verify the proposed model on a noisy Chinese Q&A dataset. The experimental results demonstrate that the novel AM–Bi-LSTM model achieves significant improvement on evaluation metrics of accuracy, mean average precision and so on. Moreover, the experimental results indicate that the novel AM–Bi-LSTM model outperforms baseline methods and other LSTM-based models.

Keywords NLP · Artificial intelligence · Long short-term memory · Question-answering system

1 Introduction

The research on natural language processing (NLP) plays an important role in many fields (Almomani et al. 2018;

✉ Xiao-mei Yu
    yxm0708@126.com

    Wen-zhi Feng
    13695486206@163.com

    Hong Wang
    teacher_sdnu@sina.com

    Qian Chu
    cq19940413@163.com

    Qi Chen
    906689923@qq.com

1   School of Information Science and Engineering, Shandong Normal University, Jinan 250014, People's Republic of China

2   Shandong Provincial Key Laboratory for Distributed Computer Software Novel Technology, Jinan 250014, People's Republic of China

Negi et al. 2013; Dkaich et al. 2017; Demin et al. 2019). In industry, NLP is involved in human–computer interaction, business information analysis and web software development (Zheng et al. 2016; Cheng et al. 2019). In academia, NLP is always known by the name of "Computational Linguistics" and is essential in fields from humanities computing and corpus linguistics to computer science and artificial intelligence (Chang et al. 2016; Li et al. 2017). As one of the core domains in artificial intelligence, the research on natural language processing has been rapidly progressed, and the related theories and methods have been deployed in a variety of new language applications (Bird et al. 2009; Wang et al. 2015; Ji et al. 2018). These applications range from the syntactic, such as part-of-speech tagging, to the semantic, such as keyword extraction (Erhan et al. 2010; Hu et al. 2017; Liu et al. 2018).

As a complex natural language processing application, the question-answering (Q&A) system has attracted great attention and its popularity attributes to the people's rising needs for fast and accurate access to information (Zheng et al. 2018; Wang et al. 2019). For different applications, the Q&A systems in different forms are built, in which the corpus and

techniques required are not universal (Liu et al. 2015; Yu et al. 2016; Wu et al. 2018). In a limited domain Q&A system, the problems to be solved are limited to a certain field and it is easy to generate an answer if all the domain knowledge required is expressed in an internal structured form. However, some common sense and general knowledge are needed in an open domain Q&A system, and a semantic dictionary is indispensable for answering questions in the open field. For example, the English WordNet is widely applied in English open domain Q&A systems, while the Chinese WordNet and "synonym word forest" are usually used in Chinese open domain Q&A systems. Generally, the Q&A system is an advanced information retrieval system that answers questions in accurate and concise natural language, which requires a thorough understanding of text and the ability to reason over relevant facts.

Compared with the English Q&A systems, there are many challenges in the research of Chinese Q&A systems which result from the characteristics of Chinese that are outlined as follows.

– *Link writing* Chinese writing is continuous with no spaces between the words; thus, a Chinese Q&A system is also a sentence-level information retrieval system. Therefore, it is necessary to analyze the sentences and perform word segmentation before building the Chinese Q&A system.
– *Form* Chinese lacks a narrow sense of morphological change, such as the active and passive voice in English. Therefore, it is more difficult in building a Chinese Q&A system than an English one, for the reason that the changes of forms in English can be marked that contribute to the processing on emotions and semantics, while it is impossible in Chinese.
– *Grammar* Chinese grammar is flexible. The order of words and the ideas in combined functional words all play key roles in keeping relationships between the various components in a sentence. The flexible grammar makes Chinese more difficult to perform semantic analysis in a Chinese Q&A system. Furthermore, relative few studies have been conducted on Chinese grammar in data processing, and research materials are insufficient for the formalization of Chinese sentence patterns or the transitions between different sentence patterns, which are essential in a Chinese Q&A system.
– *Semantics* Polysemy, homonyms, synonyms, as well as rich means of expression, high contextual dependency, omissions, etc., are all difficult problems in Chinese semantic analysis.
– *Related resources* There is a lack of Chinese linguistic resources such as semantic dictionaries and related mature corpus, which are necessary in a Chinese Q&A system.

In order to tackle some of these limitations in building a Chinese Q&A system, there is a demand to improve the existing Chinese information processing technology on the basis of the comprehensive analysis of the characteristics of both questions and answers. In this paper, we firstly analyze potential difficulties of applying LSTM and bidirectional LSTM (Bi-LSTM) to noisy Chinese Q&A datasets. Then, we focus on the characteristics of LSTM networks and the mechanisms of attention-based LSTM model and propose an attention mechanism and multi-granularity-based Bi-LSTM model (AM–Bi-LSTM) which combines the improved attention-based Bi-LSTM (A-Bi-LSTM) with a novel processing of multi-granularity word segmentation to handle the complex natural language processing in Chinese Q&A datasets. To alleviate the missing of key information as well as settle the difficulty of indeterminateness, the proposed model performs multi-granularity word segmentation at the Chinese character level, the word level and the phrase level to enhance the generalization ability of training accuracy of the model. Moreover, to tackle the limitations of indeterminateness and polysemy in Chinese, an improved attention mechanism is applied to achieve the most important final features. Finally, the cosine similarity is used to calculate the matching degree of the question and answers to get the best possible answer. With the coordination of the improved attention mechanism and a novel processing of multi-granularity word segmentation, the proposed AM–Bi-LSTM model fully demonstrates its superiority.

The main contributions of the paper are outlined as follows.

– LSTM, Bi-LSTM and A-Bi-LSTM models are analyzed to resolve some complex issues in Chinese Q&A applications, such as the problems of indeterminateness, polysemy and the lack of changing morphology. And possible solutions are acquired in the Chinese Q&A systems.
– AM–Bi-LSTM model which combines the improved attention mechanism with a novel multi-granularity word segmentation is proposed. Also, similarity of the question and answers is formulated to implement the quantitative computation.
– The experiments are conducted to verify the performance of the proposed AM–Bi-LSTM model. In comparison with several baseline models, AM–Bi-LSTM model outperforms the state-of-the-art methods on the three widely used metrics as accuracy (ACC), $F1$ score ($F1$) and mean average precision (mAP).

The remainder of the paper is organized as follows. In Sect. 2, related work on Q&A systems is briefly reviewed. And then three models of LSTM, Bi-LSTM and A-Bi-LSTM are introduced in Sects. 3–5, respectively. Based on the analysis of the advantages and disadvantages of the three models,

Sect. 6 presents the proposed AM–Bi-LSTM model. Section 7 describes the experimental setup and provides the experimental results. Finally, the work is concluded and the future research direction is indicated in Sect. 8.

2 Related work

Historically, the well-known early Q&A system includes BASEBALL (Green et al. 1961), LUNAR (Woods 1973), PLANES (Paris 1985), etc. In 1996, Suppes et al. (1996) conducted research to understand the words, grammar and semantics for ten languages, which provided the initial research on a English Q&A system. And then, the Text Retrieval Evaluation Organization (TREC), which initiated the question-answering track in 1999, has been contributing greatly to multilingual and cross-lingual Q&A tracks and has received intense attention (Sakre et al. 2009). The previous work on Q&A applications normally used machine learning (ML) approaches, and Allam and Haggag (2012) provided an overview of Q&A systems and gave an analytical discussion on the work of the top-publishing and top-cited authors in the Q&A field. Hermjakob (2001) described machine learning-based parsing and question classification for question-answering. Zhang and Lee (2003) compared various machine learning approaches for the task of automatic question classifications, with the result that the SVM outperforms the other four methods. Chang et al. (2016) proposed a new nearly isotonic SVM classifier to exploit the constructed semantic ordering information in Q&A systems with untrimmed video data. Consequently, as the ML approaches are designed and trained on annotated corpuses composed of labeled questions, they might suffer from the unavailability of additional resources, the effort of feature engineering and the systematic complexity by introducing linguistic tools.

Recently, deep learning models have obtained a significant success on Q&A applications. Iyyer et al. (2014) introduced a recursive neural network (RNN) for factoid question answering; Feng et al. (2015) applied a convolutional neural network (CNN) to address the non-factoid question answering tasks; Tan et al. (2015) built LSTM-based deep learning models on both questions and answers, respectively, for the answer selection task. Compared with several ML approaches, the general deep learning frameworks did not depend on manually defined features or linguistic tools, and demonstrated superior performance. Then, Seo et al. (2016) introduced the bidirectional attention flow (BIDAF) network for English Q&A applications. The BIDAF model used a multi-stage hierarchical process to represent the context at different granularity levels and used the bidirectional attention flow mechanism to obtain a query-aware contextual representation. The model achieved good

experimental results on the Stanford Question Answering Data set (SQuAD) and the CNN/Daily Mail cloze test. Liu et al. (2017) proposed an attention-based, bilinear function-based Bi-LSTM model, which has good performance in extracting the semantics of questions, candidates and articles. In addition, the accuracy was improved obviously after adding the inference structure with multilayer attention. Generally speaking, these models are successfully used in English Q&A applications with natural language words and obtain satisfied effect. Furthermore, in English Q&A applications with both images and texts, Socher et al. (2011) introduced a recursive neural network which successfully merges image segments or natural language words based on deep learned semantic transformations; Cheng et al. (2019) applied a multi-modal aspect-aware topic model in Q&A applications with both textual reviews and item images. These two models achieve excellent performance in English Q&A applications with both texts and images, while they have little consideration for Q&A applications with Chinese texts.

The initial research on Chinese Q&A systems mostly adopted hybrid approaches. Peng et al. (2005) explored a hybrid approach for Chinese definitional question answering by combining deep linguistic analysis with surface pattern learning. It is the first formal study about linguistic analysis and pattern learning issues in Chinese definitional Q&A applications. Day et al. (2007) proposed an integrated genetic algorithm (GA) and ML approach for question classification in English–Chinese cross-language question answering. Lee et al. (2008) proposed two lightweight methods for a Chinese factoid Q&A system. The Sum of Co-occurrences of Question and Answer Terms (SCO-QAT) calculates co-occurrence scores on basis of the passage retrieval results, and alignment-based surface patterns (ABSPs) are syntactic patterns trained from question-answer pairs with a multiple alignment algorithm. However, both of the methods suffered from the word canonicalization problem. Therefore, rules with taxonomy or ontology resources should be applied to solve most canonicalization problems in both methods.

Compared with the significant achievements made in English Q&A applications, there were few studies on Chinese Q&A systems, and little effective progress has been made on the Chinese corpus. In recent years, deep learning models have obtained a significant success on various natural language processing tasks, which promotes the Chinese Q&A applications in different domains. Qiu and Huang (2015) proposed convolutional neural tensor network architecture to model the complicated interactions between question and answer in Chinese Q&A systems. However, the performances on Chinese dataset are slightly worse than that on English one for the reason that there are errors in Chinese word segmentation. Yin et al. (2015) presented an end-to-

end neural network model for generative question answering, based on the facts in a knowledge base. The experiment on Chinese Q&A datasets demonstrated that the proposed model outperformed an embedding-based Q&A model as well as a neural dialog model with the test accuracy of not more than 52%. Lai et al. (2017) proposed a CNN-based method to rank the entity-predicate pairs generated on shallow features and then introduced a system to answer single-relation factoid questions in Chinese. The whole system achieved the $F1$-score of 47.23% on test data and obtained the first place in the NLPCC KBQA evaluation task. Yao and Huang (2016) held that the bidirectional LSTM network (Bi-LSTM) performs considerably well in word segmentation on both traditional Chinese datasets and simplified Chinese datasets. Using a large Chinese knowledge base, Zhou et al. (2018) combined two different attention mechanisms with the Bi-LSTM for question-property mapping. In the NLPCC-ICCPOL 2016 KBQA task, the system achieved more competitive results than the best existing one with the average $F1$ value of 0.81. On Chinese Q&A applications with both images and texts, many methods are inspired by the m-RNN model (Mao et al. 2014) for the image captioning and image-sentence retrieval tasks. It contains a deep CNN for vision part and a RNN for language model part. For example, Gao et al. (2015) presented the mQA model to answer questions about the content of an image and addressed the task of visual question answering in Chinese Q&A applications. In the mQA model, LSTM is used to extract the question representations and store the linguistic context in an answer, while CNN is applied to extract the visual representation. Finally, a fusing component is designed to combine the information and to generate the answer.

In the Q&A applications, traditional machine learning approaches can explicitly interpret the results in question answering tasks. However, as the ML approaches normally use feature engineering, linguistic tools, or external resources, they suffer from the choice of features and classifiers. In contrast, most of the deep learning models do not rely on any linguistic tools and can be applied in different languages or domains. However, there is not a universal deep learning model that is suitable for any language or each domain. As we have seen, CNN is effective in visualization part of a Q&A system, and RNN as well as LSTM-based models play an important role in language part of a Q&A system. However, there is not enough intensive achievements of applying Bi-LSTM-based models in Chinese Q&A applications, though the up-to-date studies indicate that the Bi-LSTM network and attention mechanism achieve impressive performance in a Chinese Q&A system. Inspired by the related research, we will analyze LSTM, Bi-LSTM and A-Bi-LSTM in the following sections and then propose the AM–Bi-LSTM model.
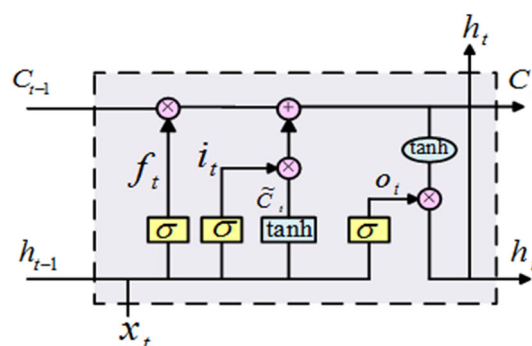

Fig. 1 A LSTM memory unit

3 Long short-term memory network

Long short-term memory network is a special RNN that efficiently avoids the problem of vanishing gradient suffered in the traditional RNN. The LSTM network is composed of specialized memory storage units that continuously respond to an input sequence or time series data. Correspondingly, the previous information is commonly propagated backwards, and the control weights are optimized based on the resulting loss gained in the forward stage (Sutskever et al. 2014).

A LSTM memory unit is controlled by three carefully designed gates(as shown in Fig. 1) which modulate the information that flows into or out of the unit. The input gate i_t controls incoming information to flow into the memory unit, while the output gate o_t allows the state of the memory unit to have an effect on the network's output at the present time step. Besides, the forget gate f_t makes decisions on how long an existing piece of information is preserved. At every time step t, the value of the memory unit C_t is updated with the status after filtering the previous information of ($f_t * C_{t-1}$) and obtaining the candidate information of ($i_t * \tilde{C}_t$), as described in the equations below.

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

where

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{2}$$
$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{3}$$
$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{4}$$

Here, σ is a function of the corresponding elemental multiplication, such as the sigmoid function or tanh function. Once the memory unit is updated, according to the result of the current output gate o_t, the current hidden layer h_t is calculated as follows.

$$o_t = \sigma(W_o x_t + U_o h_{t-1}) \tag{5}$$
$$h_t = o_t * \tanh(c_t) \tag{6}$$

Different units in the LSTM play their roles in forgetting, remembering and displaying information, respectively, and adaptively. Take the incoming sequence of $S : t_1, t_2, \ldots, t_n$ as an example. The previous information of element t_i is $\overrightarrow{S} : t_1, t_2, \ldots, t_{i-1}$ in sequence S, while the following information is $\overleftarrow{S} : t_{i+1}, t_{i+2}, \ldots, t_n$. Once the sequence of S is received, each unit in the network reacts on the information mentioned above and passes it to the subsequent unit, while it has no relevance on the information in the following sequence. Consequently, such a model is inadequate for addressing the long-term memory problem in Chinese Q&A systems.

The LSTM network encounters great challenges in Chinese Q&A applications. For example, ellipses and semantic references are common in Chinese, and they may result in semantic problems for the omissions of important contextual relations in a sentence. Furthermore, the poor continuity in the sequence would lead to an extremely low accuracy of the model. Inspired by the idea in LSTM, a reverse LSTM network can be introduced to remember the information of $\overleftarrow{S} : t_{i+1}, t_{i+2}, \ldots, t_n$ in the context of a sentence. Similarly, the following sequence $\overleftarrow{S}$ can be trained using the reverse LSTM network, and each unit only reacts on the following information in sequence S or responds to the previous information in the reverse sequence $\overleftarrow{S}$.

Generally, in the research area of Chinese Q&A applications, the LSTM network has achieved significant advantages compared with other models such as the N-GRAM and RNN. However, the LSTM network still has insurmountable difficulties in addressing long-term memory problems. Considering the complexity in Chinese Q&A systems, some researchers (Zhou et al. 2018) have proposed the Bi-LSTM model, which combines the forward LSTM and reverse LSTM networks to achieve a full deep learning process.

4 Bi-LSTM model

The Bi-LSTM model is constructed on a unidirectional LSTM network. In each epoch, the forward LSTM network and reverse LSTM network are trained separately, and both networks are connected to the same output layer, which completely captures the contextual information about both the past and the future in the input sequence.

As shown in Fig. 2, the intermediate encoding vectors of S_i and S'_i, are both used to retain the contextual information. In a Bi-LSTM model, the hidden state H_t, which serves as the LSTM memory blocks, contains the forward pass $\overrightarrow{h_t}$ and the reverse pass $\overleftarrow{h_t}$ at time t to perfectly compensate for the inadequacy in the unidirectional LSTM model, where the

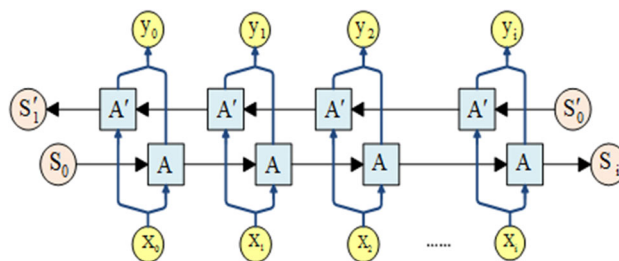**Fig. 2** The Bi-LSTM architecture

latter is incapable of exploiting the subsequent information in the future (Wang et al. 2015).

$$\overrightarrow{h_t} = \overrightarrow{\mathrm{LSTM}}(h_{t-1}, w_t, c_{t-1}), t \in [1, T] \qquad (7)$$

$$\overleftarrow{h_t} = \overleftarrow{\mathrm{LSTM}}(h_{t+1}, w_t, c_{t+1}), \quad t \in [T, 1] \qquad (8)$$

$$H_t = \left[\overrightarrow{h_t}, \ \overleftarrow{h_t}\right] \qquad (9)$$

As an extension of the traditional LSTM network, the Bi-LSTM model achieves better performance than LSTM does in many research fields including sequence classification. When all of the time steps in the input sequence are available, it is provided as the input sequence of the Bi-LSTM model in the opposite direction, too. In other words, the original sequence is input in the forward direction, while the inverted copy of the input sequence is available in the opposite direction. Finally, the output is determined by both unidirectional LSTM models. Consequently, such a Bi-LSTM model is capable of capturing the complete information in features and achieving effective performance in Chinese Q&A applications.

However, there are still some shortcomings in Bi-LSTM model that makes difficulties in addressing Chinese Q&A tasks. For example, as the input sequence is encoded into a representation with a fixed-length vector, it imposes restrictions on the decoding of the sequence. Once the input sequence is too long, it is difficult for the Bi-LSTM model to acquire a reasonable vector representation. Since all the contextual inputs are limited to a fixed length, the learning ability of the Bi-LSTM model is also limited. Therefore, when confronted with long input sequence in a Chinese Q&A system, the Bi-LSTM model demonstrates relatively poor performance in deep learning .

5 A-Bi-LSTM model

In the attention mechanism, more attention is allocated to the important content, while less attention is paid to the remaining content. At present, the general attention-based model performs well in understanding the semantics of English arti-
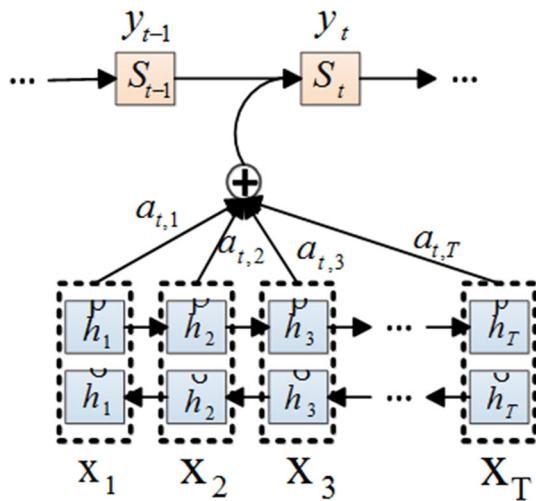
Fig. 3 The attention mechanism

Fig. 4 The A-Bi-LSTM model

cles and discovers a relatively reasonable answer. However, the attention-based model still encounters great challenges in tackling some complex semantic comprehension problems in noisy Chinese Q&A datasets. Therefore, we make some improvements on general attention-based model firstly and then combine the improved attention mechanism to Bi-LSTM (A-Bi-LSTM) model for better settling the problems of semantic comprehension in Chinese. The implementation of the A-Bi-LSTM model in Chinese Q&A applications is close to the one proposed by Cho et al. (2014) (shown in Fig. 3).

In the A-Bi-LSTM model, the conditional probability is defined as follows:

$$p\left(y_t|y_1,\ldots,y_{t-1},x\right) = g\left(y_{t-1}, s_t, c_t\right) \tag{10}$$

$$S_t = f\left(S_{t-1}, y_{t-1}, c_t\right) \tag{11}$$

$$c_t = \sum_{j=1}^{T_x} \partial_{tj} h_j \tag{12}$$

where S_t is the hidden state in the decoder at time t, and c_t is a weighted value in which h_j is the hidden vector of the jth word on the encoder side. ∂_{tj} is the probability between the jth word on the encoder side and the tth word on the decoder side, which denotes the influence of the jth word on the tth word of the target. The value of ∂_{tj} can be calculated as follows:

$$\partial_{tj} = \frac{\exp\left(e_{tj}\right)}{\sum_{k=1}^{T_x} exp\left(e_{tk}\right)} \tag{13}$$

$$e_{tj} = a\left(S_{t-1}, h_j\right) \tag{14}$$

Here, ∂_{tj} is the output of the Softmax model, and the sum of the probabilities is 1. e_{tj} is an alignment model that measures the influence of the jth word (at the encoder side) on the tth
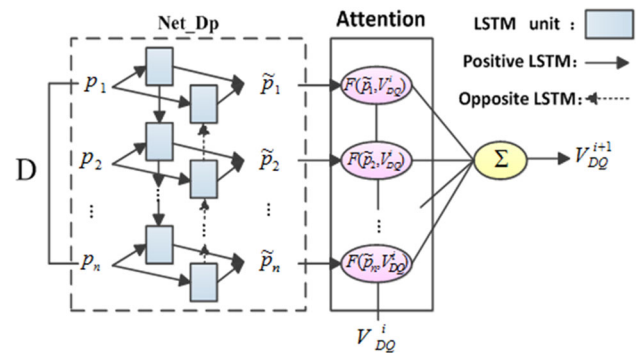
word (at the decoder side). In an alignment model of the dot product matrix, the hidden state h_t at the target end is matrix-multiplied with the hidden state $\bar{h}_s$ at the source end. The alignment calculation is shown as follows:

$$\text{score}\left(h_t, \bar{h}_s\right) = \begin{cases} h_t^T \bar{h}_s & \text{dot} \\ h_t^T W_a \bar{h}_s & \text{general} \\ W_a\left[h_t, \bar{h}_s\right] & \text{concat} \end{cases} \tag{15}$$

The A-Bi-LSTM model concentrates on different parts of a sentence as various features are input. Therefore, the output of the Bi-LSTM model is taken as a text-encoded feature vector, and the similarity between the questions and answers is calculated. The overall architecture of the model is shown in Fig. 4. As we can see, the layer of Net_Dp extracts the semantic vector representation of text D, and the attention layer calculates the attention value of the answering text about the question. In this way, the A-Bi-LSTM model contributes to solving the problem of poor learning in long sequences that suffers the Bi-LSTM model.

6 The proposed AM–Bi-LSTM model

At present, the technology of deep learning promotes the development of Q&A applications, while there are still challenges in the research area of Chinese Q&A applications.

6.1 Multi-granularity processing

The characteristics of Chinese vocabularies, such as the indeterminateness, polysemy and flexibility, cause problems for semantic analysis in Chinese Q&A applications. Moreover, high contextual dependency and the lack of changing morphology in texts pose enormous challenges in semantic comprehension in Chinese Q&A systems. In some situations, it is difficult to determine the most suitable answer that is required for a question, thereby increasing the difficulty in choosing the polarity of granularity in opinion calculation.

On the one hand, the emotion in language expression is often embarrassing, and the boundaries of the opinions that are expressed are always vague. For example, in the sentence of "the weather may be fine tomorrow", it is difficult to classify the sentence as positive or neutral. After all, the concepts of positive and neutral are inherently ambiguous, and this distinction relies heavily on subjective judgements. Another example is in the Chinese idioms and allusions such as "Actionless Governance." (无为而治) It is difficult to accurately classify such an expression using traditional word segmentation, thus resulting in the loss of the overall characteristics in the whole sentence. Therefore, the essential characteristics of the sentences should be analyzed, and the effective methods and technologies need to be carefully chosen in dealing with the complex NLP problems in Chinese. On the other hand, when the context is incomplete in a sentence, the traditional methods perform poorly in word segmentation. For instance, in the Chinese sentence of "what happened next will be analyzed and explained below (预知后事如何，且听下回分解)", there is no perfect solution using Chinese word segmentation. Therefore, the word segmentation should be robust enough to avoid the impacts of potential errors on Chinese Q&A applications.

To avoid the adverse effect of inadequate Chinese word segmentation on the performance of AM–Bi-LSTM model, we improve the multi-granularity processing to increase the robustness of the novel model. More precisely, the text is processed at different levels, such as the Chinese character level, the word level, and the phrase level, respectively. In this way, the novel semantic comprehension is constructed with a hierarchical multi-granularity processing. The following excerpt is a toy example which briefly shows the multi-granularity processing in texts.

describing some basic problems on a professional website, Guan et al. (2006) presented the basic theory on system similarity, and theoretically reorganized the related technologies in a Chinese Q&A system. Zhen et al. (2012) addressed a matching algorithm for Chinese sentence similarity and completed a Chinese Q&A system based on common problems, which achieved better performance than the popular SVM method in Chinese sentence similarity matching. Xu et al. (2015) addressed a hybrid framework where the text features are preliminary extracted for semantic comprehension and optimized using a novel information gain method. Moreover, the features are expressed in the vector space model and integrated using the traditional machine learning algorithm. The experimental results show its improvement with respect to efficiency and accuracy. Li et al. (2017) proposed a new criterion to maximize the weighted harmonic mean of trace ratios. Furthermore, the novel algorithm updates the transformation matrix just using eigenvalue decomposition, which gives us inspiration in feature selection.

Generally, the up-to-date studies indicate that the largest eigenvalue represents the strongest feature which play important role in matching results between the question and answers, while the weak features play negligible roles in the results (Zhu et al. 2018; Shi et al. 2018). Therefore, in proposed AM–Bi-LSTM model, we extract a number of eigenvalues from a filter and take the highest eigenvalue as the retention value in the pooling layer. In this way, the attention vector of the question and the alternative answers are obtained separately, which indicates the semantic relationship between the texts and can be used to calculate the similarity of the question and answers using formula 16.

$$\text{Cosine} - \text{Similarity}\left(V_Q, V_A\right) = \frac{V_Q \cdot V_A}{\|V_Q\| \times \|V_A\|} \qquad (16)$$

"下回分解下一句" (The next sentence will be explained later)
['下回分解','下','一句'] ['下回分解','下一句'] ['下','回','分','解','下','一','句']

"梁山泊位于山东省西南部梁山县境内"
(Liangshanpo is located in Liangshan County, southwest of Shandong province.)
['梁','山','泊','位','于','山','东','省','西','南','部','梁','山','县','境','内']"

6.2 Similarity computations on the question and answers

With respect to the similarity computation in Chinese Q&A applications, Li et al. (2002) presented the approach of computing the similarity and relevancy between words to get the relationship between two sentences and then obtain the optimal answer in a Chinese Q&A system. This approach avoids some difficulties in conventional NLP techniques. By

It shows that the matching answers of the question have slightly higher similarities than the mismatched ones, although all of them have a high degree of similarity with the question. Since some answers without keywords are also considered as the correct answers to the question, we increase the distance by setting the loss function to evaluate whether the answer matches the question well in the word vector. In the loss function, we use a variant of the SVM objective function (Zhang et al. 2018). The AM–Bi-LSTM model trains

the similarity between the two samples, where the following formula is used:

$$LOSS = \max \left\{ 0, m - \cos\left(V_Q, V_{A+}\right) + \cos\left(V_Q, V_{A-}\right) \right\} \tag{17}$$

Here, m is the parameter margin. V_Q, V_{A+} and V_{A-} are the vectors that provide semantic representation of the question, the positive answer and the negative answer, respectively. In the loss function, the cosine value of the vectors between the positive answer and question should be greater than that of the negative answer and question, which is defined by the margin parameter. The larger the cosine value is, the closer the relationship of the two vectors is. In other words, the loss function is set with the purpose of making the relationship between the positive answer and the question increasingly more similar and making the relationship between the negative answer and the question increasingly more dissimilar. In the objective function of the proposed model, it is the similarity of the maximum feature between the answer and question that plays a decisive role. Therefore, it can be concluded that the effect of the model depends heavily on the similarities.

6.3 The procedure of AM–Bi-LSTM

In order to gain significant performance, the method of multi-granular fusion is preferred in building a Chinese Q&A system (Xu 2017). Including the improved attention mechanism in A-Bi-LSTM models and a novel processing of multi-granularity word segmentation, the proposed AM–Bi-LSTM model is mainly composed of the following steps (shown in Fig. 5).

1. *Data extension* When the imbalanced dataset of texts with both questions and answers is obtained, data extension methods are needed to balance the texts with right answers and wrong answers.
2. *Chinese word segmentation* The text is conducted with the novel processing of multi-granularity word segmentation. Differences in the granularity are considered for words with several Chinese characters, except for non-compositional words such as 'Peihuai' (徘徊) in Chinese.
3. *Word vector computation* With the open source tool of word2vector, the words vectors are trained by Wikipedia, and each word can be transformed into a dense vector to fully explore the contextual relationships between words.
4. *Attention modeling* In building the A-Bi-LSTM model, max-pooling is applied to calculate the maximum features of the problem and the candidate answers (Wang et al. 2016). Since the largest eigenvalue reflects the strongest feature and the weak features play negligible roles in the results, we extract a number of eigenvalues
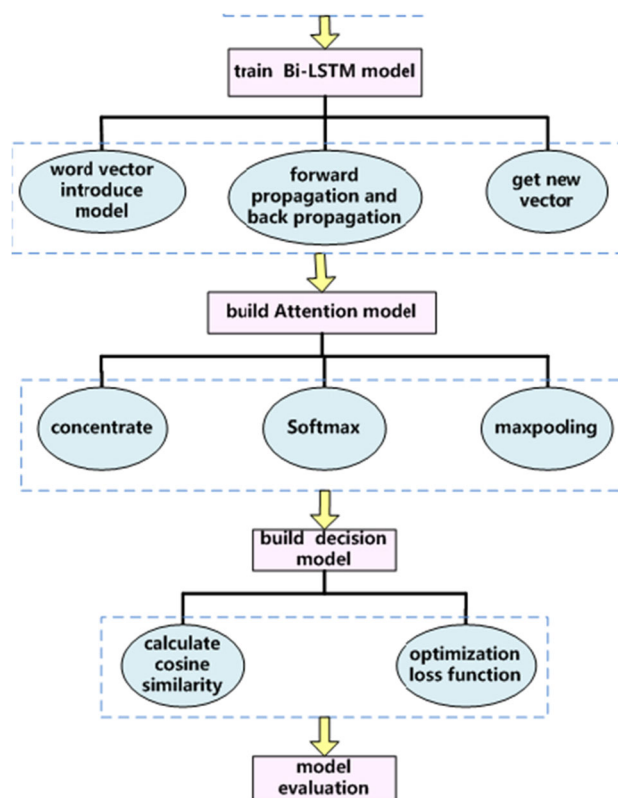

Fig. 5 The proposed AM–Bi-LSTM model

from a filter and take the maximum eigenvalue as the retention value in the pooling layer. In this way, the attention vector of the question and the alternative answer is obtained separately, which denotes the semantic relationship between the texts and can be used to calculate the similarity of the question and answer using formula 16.

7 Experiments

In this section, we evaluate the performance of these models on an noisy Chinese Q&A dataset with the three most widely used metrics.

7.1 Evaluation criterion

Accuracy The accuracy of a system reflects the ability to correctly classify the entire samples. That is, in the set of experimental samples, the accuracy describes the closeness of a predicted result to the true value. The accuracy focuses on the predictive capability of a model and is calculated as follows:

$$ACC = \frac{TP + TN}{TP + TN + FN + FP} \tag{18}$$

Table 1 Symbolic representation

Total number of samples	TP + FP + FN + TN	
Real	True positive (TP)	False negative (FN)
	False positive (FP)	True negative (TN)
Prediction	To be positive(TP + FP)	To be negative (FN + TN)

F1 score The *F*1 score is widely used in the field of information retrieval to measure the performance of classification. It is defined as the harmonic average of the precision and recall.

$$F1 = \frac{2 \times \text{TP}}{\text{Total number of samples} + \text{TP} - \text{TN}} \quad (19)$$

The meanings of the abbreviations in formula 18 are shown in Table 1, where 1 represents a positive value and 0 represents a negative value.

Mean average precision The accuracy only considers the number of related documents in the results without considering the order between the documents. However, the output results in a Q&A system should be ordered, and the more relevant that the document is ranked, the better the evaluation results are.

The mAP is a single-valued metric that indicates the performance of a system on all relevant documents. The more relevance that the documents have, the higher the value of the mAP would be.

$$\text{mAP} = \frac{\sum_{q=1}^{Q} \text{AveP}(q)}{Q} \quad (20)$$

Here, Q is the number of queries under consideration.

7.2 Data sets

We conduct experiments to evaluate the performance of the proposed AM–Bi-LSTM model on a noisy Chinese Q&A dataset. The experimental data are divided for training and testing. The training set consists of 30,000 questions and 477,019 answers. A sample in training set is shown in Table 2.

Some characteristic in the Chinese training samples is addressed before data preprocessing, which are important in building a suitable model.

– There are redundant statements in the samples. Since repeated sentences occur in different contents, it is difficult to distinguish the best matched answer with a single granularity method.
– Incomplete sentences appear in many samples in the dataset. Since missing words and sentence fragments frequently appear in the Chinese samples, significant challenges occur when handling contextual or temporal information with a relatively long duration.
– Ancient texts and modern texts are alternately presented in the contexts. The mixture of ancient and modern texts increases the difficulties of syntactic analysis and semantic comprehension.
– Spoken language occurs in the samples and is accompanied by some acronyms. As we know, spoken language and acronyms are always constructed with abnormal grammatical structures, which are difficult to understand for an open Chinese Q&A system.

Confronted with all the difficulties, we endeavor to solve some problems in the process of data preprocessing and then to effectively settle the challenges in the subsequent proposed model. To address the seriously imbalanced training sam-
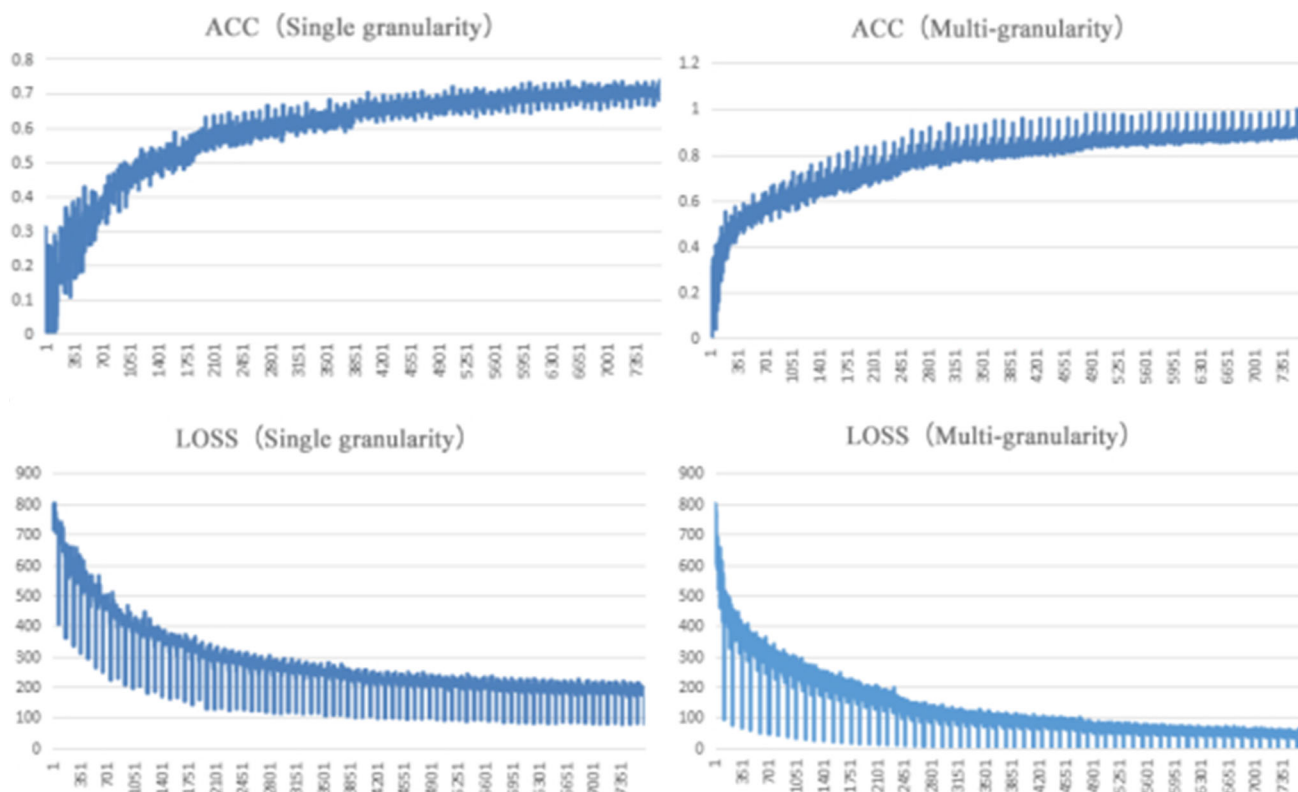
Table 2 The format of a sample in the dataset

QUESTION
高速占用应急车道扣多少分？
(How many points does the high-speed emergency lane take?)
CONTENT 1 with LABEL 0
高速公路的应急车道是"生命车道"，很多朋友产生疑惑，换驾驶员、困了想临时休息、"方便"、打电话
(The emergency lane of the expressway is a passage for rescuing a patient. Many friends have doubts, take turns driving, sleep as temporary rest, go to the toilet, and make a phone call)
CONTENT 2 with LABEL 1
6分 (6 points)
CONTENT 3 with LABEL 1
违法占用应急车道扣几分？核心内容：车辆在高速公路上行驶时候，违法占用应急车道要不要扣分？要扣分。那么 要扣几分？按照法律规定，违法占用应急车道要记6 分，罚款200 元。法律小编为你提示...
(How many points does it take to occupy the emergency lane? Core content: When the vehicle is driving on the expressway, do you need to deduct points for illegally occupying the emergency lane? Then, how many points should be deducted? According to the law, it is necessary to record 6 points for illegal lanes and a fine of 200 yuan. The Legal Express reminds ...)

Fig. 6 A comparison of multi-granularity and single granularity processing

ples, some measures such as data expansion are conducted on sparse answers in the dataset. For example, a question in the training set corresponds to a group of n answers, of which there is only one correct answer with a label of 1 and $n - 1$ unsuitable answers with a label of 0. Therefore, the data expansion has to be performed on the correct answer to obtain the same number of correct answers as incorrect answers.

7.3 Experimental setup

The models in this work are implemented with Tensorflow from scratch, and all experiments are conducted on an Intel Corporation Xeon E7 CPU. The texts in Chinese are converted into sequences of words using the Jieba for Chinese word segmentation. The word embedding is trained by word2vec, and the word vector size is 50. The word embedding is also parameters and is optimized as well during the training. The dimensions of the hidden states of encoder and decoder are both set to 200. We train our models with the mini-batch size of 256, and the maximum length of questions and answers is 2000. Stochastic Gradient Descent is the optimization strategy. We tried different margin values, such as 0.1 and 0.2, and finally fixed the margin as 0.2. The training of each model takes about 3 days.

7.4 Performance evaluation of multi-granularity processing

We carry out experiments to compare the performance of multi-granularity processing with single granularity processing. The experimental results are shown in Fig. 6, from which we observe that the multi-granularity processing significantly outperforms single granularity processing in terms of both accuracy and the number of loss.

As we can see, the accuracies in both multi-granularity and single granularity processings rise with the increasing of training epochs. However, the accuracy in multi-granularity processing rises more quickly and more dramatically, and the trend continues until the ACC reach its peak. Finally, the final accuracy of the model with multi-granularity processing is nearly 10% higher than that in the model with single granularity processing. In contrast, at the beginning of the experimental evaluation, the model with single granularity processing suffers more information loss than the model with multi-granularity processing. As the training goes on, the loss metric descends in both multi-granularity and single granularity processing, though the downward trend in multi-granularity processing is faster and sharper than that in single granularity processing. Nevertheless, the multi-granularity processing keeps less loss throughout the entire training procedure than the single granularity processing.
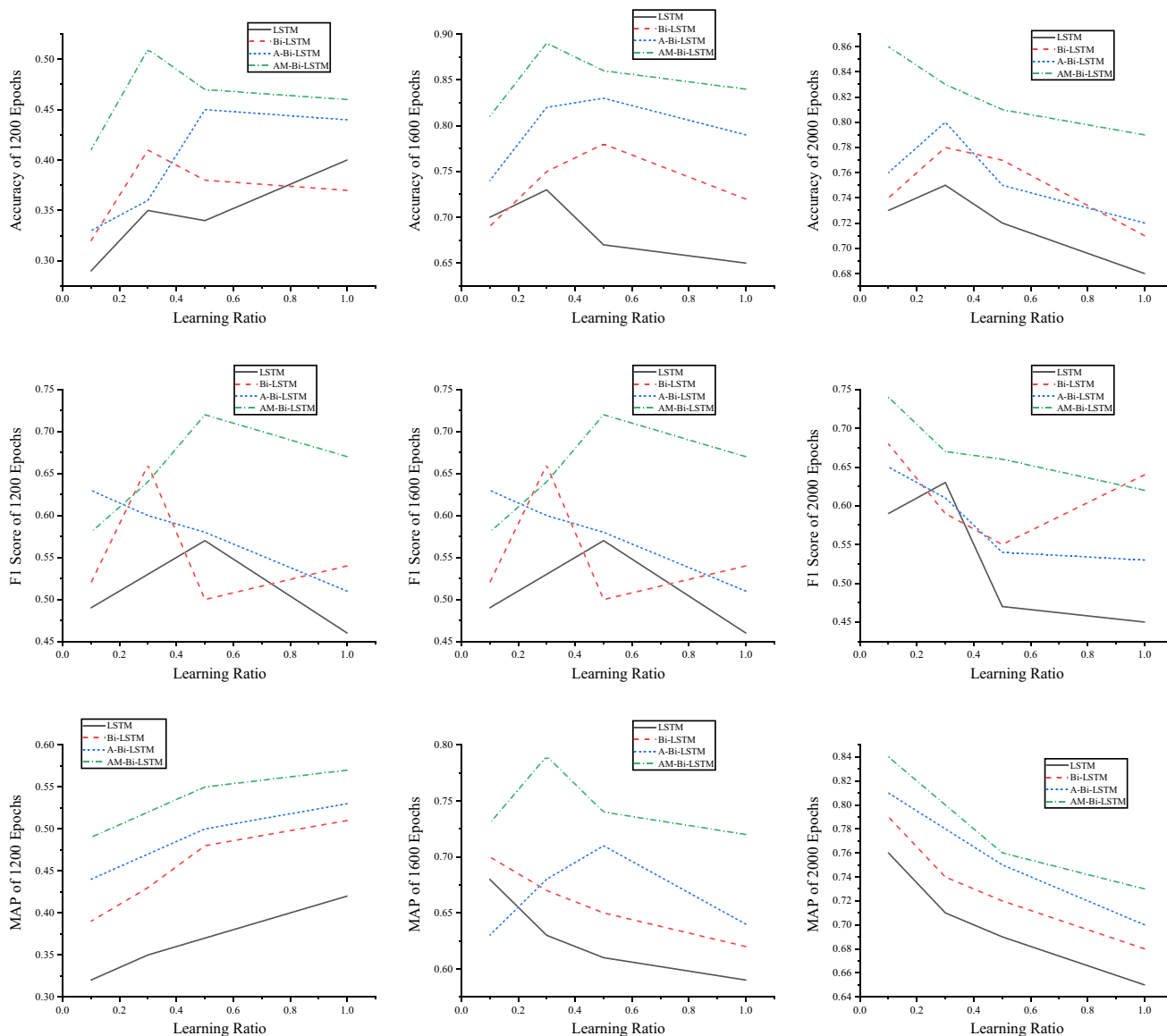
Fig. 7 Performance of the models with the varying of the learning ratio

7.5 Contrastive experiments

Assume that the known question and answer library is as follows:

$$QASet = (\langle Q_i, A_i \rangle)\,(1 \le i \le n) \tag{21}$$

Here, Q_i is the question and A_i is the answer in the Q&A dataset. For the proposed problem, the mapping function is as follows:

$$
F(Q_i, A_i)
= \begin{cases}
1, & \text{if } (Q_i \text{ and } A_i \text{ are semantically related}) \\
0, & \text{if } (Q_i \text{ and } A_i \text{ are semantically independent})
\end{cases} \tag{22}
$$

In the experiments, we use the noisy Chinese Q&A dataset to train and test the LSTM model, the Bi-LSTM model, the A-Bi-LSTM model and the AM–Bi-LSTM model, respectively. All experiments are conducted with the same main parameters, and the experimental results of each model are shown in Fig. 7.

First, the learning ratio is set to the default value of 0.5 since no true value is available in the previous experiments. Then, the learning ratio is adjusted with the fluctuation of the performance. We attempt to double the learning ratio with the value of 1 and conduct the training. However, the poor performance discourages us from raising the learning ratio. It seems that a larger learning ratio leads to oscillations or overshoots of the training in the gradient descent process. Therefore, a much smaller learning ratio is adopted in the
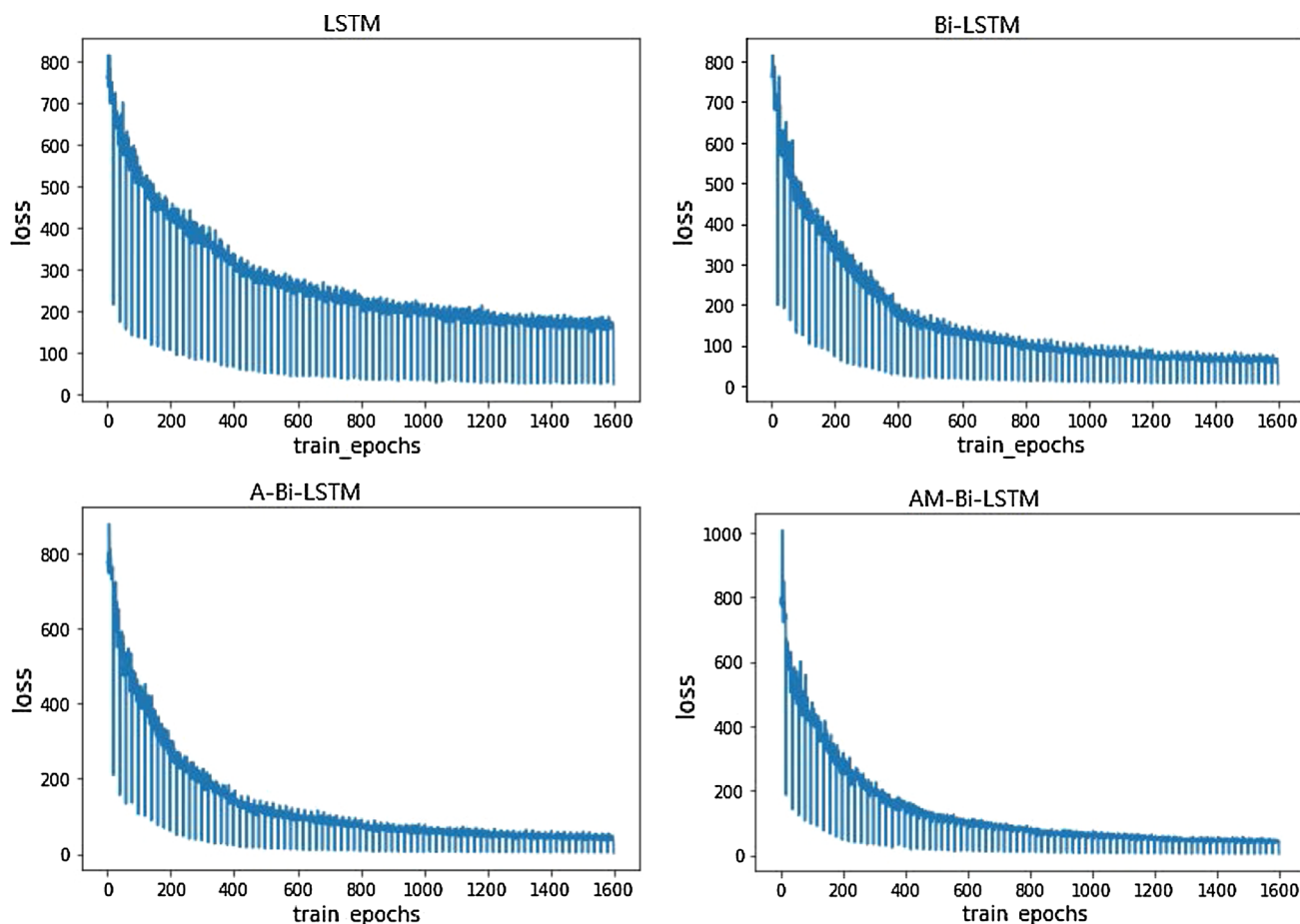
Fig. 8 A comparison of the LOSS of the four models

following training, and some improved performance occurs, which encourages us to set the value of the learning ratio even smaller, such as 0.1. With the extremely slow training speed that is experienced, the performance of the model is even worse. Obviously, a too small learning rate will harm the performance of the models. With the experimental results shown in Fig. 7, we conclude that the AM–Bi-LSTM model may acquire the best performance at a learning ratio around 0.3, and it is almost the same as the other three models.

Continually, we run our training with varying epochs. In this way, we look forward to increasing the opportunities of following the decreasing slopes and updating the weights in the gradient descent process. As shown in Fig. 8, more training results in better performance. Combined with the performance comparison in Fig. 8, we can conclude that a smaller learning ratio is truly helpful in promoting performance under the circumstances of more epochs being provided.

In general, the experimental results on the noisy Chinese Q&A dataset manifest the similar distinct advantages on the main metrics and imply that the proposed AM–Bi-LSTM

model performs better than the other three models, which can be summarized as follows.

– The LSTM model has a certain effect on processing the temporal data in the noisy Chinese Q&A dataset. With the experimental results shown in Fig. 7, LSTM exhibits its ability of utilizing the previous states in the self-recurrent connections of its memory cells. In other words, the LSTM model captures the context in the previous part of a sequence and participates in the present calculation. Therefore, the LSTM model achieves acceptable performance with an ACC and $F1$ of 78.02% and 73.62%, respectively.
– Compared with LSTM, the Bi-LSTM model has a slight advantage of approximately 1.85–4.17% with respect to the three metrics. The context in the previous parts of the sequences and the context in the following parts of the sequences are both considered in order to make full use of contextual information in the Bi-LSTM model.
– As to the A-Bi-LSTM model with the attention mechanism, its ACC, $F1$ and mAP are all higher than those of the Bi-LSTM model without the attention mechanism,

though the former is better by a narrow margin of 0.33–3.83%. The advantage possibly benefits from the ability of the attention layer at the top level of the model to learn the potential dependencies in labels. It shows that the introduction of the improved attention mechanism captures the importance of words in contexts.

– As shown in Fig. 7, the proposed AM–Bi-LSTM model, which combines the improved attention mechanism with a novel processing of multi-granularity word segmentation, obtains better performance in terms of ACC, $F1$ and mAP. Moreover, the AM–Bi-LSTM model achieves the least information loss which is shown in Fig. 8. Consequently, the significant performance of AM–Bi-LSTM model reflects its effectiveness in building a Chinese Q&A system.

The results of the experimental evaluation indicate that the Bi-LSTM model considers the contextual information in the noisy Chinese Q&A dataset and achieves better performance than the LSTM model, the proposed AM–Bi-LSTM model with improved attention mechanism and multi-granularity processing is beneficial in improving the main performance metrics of the ACC, $F1$ and MAP in Chinese Q&A applications.

8 Conclusion

The Q&A system is a research focus in the fields of artificial intelligence and natural language processing. Compared with an English Q&A system, a Chinese Q&A system is confronted with many difficulties and deficiencies. On the basis of the in-depth analysis of the characteristics of the Chinese Q&A system, the paper proposes an AM–Bi-LSTM model to overcome the difficulties in Chinese Q&A applications which are resulted from the grammar, semantics and morphology limitations in Chinese. The experimental evaluations verify that the AM–Bi-LSTM model achieves the best performance in the four LSTM-based models.

Due to the different application scenarios and requirements of Q&A systems, the modeling process is also constantly changing. The future work will be conducted as follows:

– By adding more hidden layers and/or using optimized dropout, we will improve the consistency on the architecture of the model, so as to effectively promote the model's ability in inference and achieve better performance in Chinese Q&A applications.
– As data with various types are input, it is difficult for the present model to deal with the multi-modal data efficiently. A hybrid framework using the deep learning model and the multi-source data fusion method are

preferred in previous studies, and they achieve better performance in Chinese Q&A applications with multi-modal data. It is the future work interested us.
– In an intelligent Chinese Q&A system, not only the correct answers should be provided, but also the underlying reasons for the answering should be explainable and can be understood. However, the deep learning model alone is not competent for such a task. In the future, we plan to study on a hybrid framework which incorporates machine learning methods and deep learning models with knowledge base, aiming to gain good interpretability for intelligent Chinese Q&A applications.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

References

Allam AMN, Haggag MH (2012) The question answering systems: a survey. Int J Res Rev Inf Sci 2(3):211–221

Almomani A, Alauthman M, Albalas F et al (2018) An online intrusion detection system to cloud computing based on NeuCube algorithms. Int J Cloud Appl Comput 8(2):96–112

Bird S, Klein E, Loper E (2009) Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Beijing

Chang X, Yu YL, Yang Y et al (2016) Semantic pooling for complex event analysis in untrimmed videos. IEEE Trans Pattern Anal Mach Intell 39(8):1617–1632

Cheng Z, Chang X, Zhu L et al (2019) MMALFM: explainable recommendation by leveraging reviews and images. ACM Trans Inf Syst 37(2):16

Cho K, van Merrienboer B, Bahdanau D, Bengio Y (2014) On the properties of neural machine translation: encoder–decoder approaches. In: Eighth workshop on syntax, semantics and structure in statistical translation, 10

Day MY, Ong CS, Hsu WL (2007) Question classification in English-Chinese cross-language question answering: an integrated genetic algorithm and machine learning approach. In: IEEE international conference on information reuse and integration, pp 203–208

Demin B, Parlati S, Spinnato PF et al (2019) U-LITE, a private cloud approach for particle physics computing. Int J Cloud Appl Comput 9(1):1–15

Dkaich R, El Azami I, Mouloudi A (2017) XML OLAP cube in the cloud towards the DWaaS. Int J Cloud Appl Comput 7(1):47–56

Erhan D, Bengio Y, Courville A et al (2010) Why does unsupervised pre-training help deep learning? J Mach Learn Res 11(3):625–660

Feng M, Xiang B, Glass MR et al (2015) Applying deep learning to answer selection: a study and an open task. In: IEEE workshop on

automatic speech recognition and understanding (ASRU). IEEE, pp 813–820

Gao H, Mao J, Zhou J et al (2015) Are you talking to a machine? Dataset and methods for multilingual image question. In: Advances in neural information processing systems, pp 2296–2304

Green Jr BF, Wolf AK, Chomsky C et al (1961) Baseball: an automatic question-answerer. In: Proceedings of western joint IRE-AIEE-ACM computing conference, Los Angeles, 9–11 May, pp 219–224

Guan Y, Wang XL, Zhao J (2006) The research on professional website oriented Chinese question answering system. Nat Immunol 8(1):92–100

Hermjakob U (2001) Parsing and question classification for question answering. In: Proceedings of the ACL 2001 workshop on open-domain question answering

Hu B, Wang H, Yu X et al (2017) Sparse network embedding for community detection and sign prediction in signed social networks. J Ambient Intell Humaniz Comput 1:1–12

Iyyer M, Boyd-Graber J, Claudino L et al (2014) A neural network for factoid question answering over paragraphs. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp 633–644

Ji C, Liu S, Yang C et al (2018) A shapelet selection algorithm for time series classification: new directions. Procedia Comput Sci 129:461–467

Lai Y, Jia Y, Lin Y et al (2017) A Chinese question answering system for single-relation factoid questions. In: National CCF conference on natural language processing and Chinese computing. Springer, Cham, pp 124–135

Lee CW, Day MY, Sung CL et al (2008) Boosting Chinese question answering with two lightweight methods: ABSPs and SCO-QAT. ACM Trans Asian Lang Inf Process 7(4):12

Li S, Zhang J, Huang X et al (2002) Semantic computation in a Chinese question-answering system. J Comput Sci Technol 17(6):933–939

Li Z, Nie F, Chang X et al (2017) Beyond trace ratio: weighted harmonic mean of trace ratios for multiclass discriminant analysis. IEEE Trans Knowl Data Eng 29(10):2100–2110

Liu H, Hu B, Moore P (2015) HCI model with learning mechanism for cooperative design in pervasive computing environment. J Internet Technol 16(2):201–210

Liu FL, Hao WN et al (2017) Attention of bilinear function based Bi-LSTM model for machine reading comprehension. Comput Sci 44(s1):92–96

Liu R, Wang H, Yu XM (2018) Shared-nearest-neighbor-based clustering by fast search and find of density peaks. Inf Sci 450:200–226

Mao J, Xu W, Yang Y et al (2014) Deep captioning with multimodal recurrent neural networks (m-RNN). arXiv preprint arXiv:1412.6632

Negi P, Mishra A, Gupta BB (2013) Enhanced CBF packet filtering method to detect DDoS attack in cloud computing environment. arXiv preprint arXiv:1304.7073

Paris CL (1985) Towards more graceful interaction: a survey of question-answering programs. Columbia University Computer Science Technical Reports

Peng F, Weischedel R, Licuanan A et al (2005) Combining deep linguistics analysis and surface pattern learning: a hybrid approach to Chinese definitional question answering. In: Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics, pp 307–314

Qiu X, Huang X (2015) Convolutional neural tensor network architecture for community-based question answering. In: Twenty-Fourth international joint conference on artificial intelligence

Sakre MM, Kouta MM, Allam AMN (2009) Automated construction of Arabic-English parallel corpus. J Adv Comput Sci 3:1–8. https://doi.org/10.13140/RG.2.1.2135.0880

Shi D, Zhu L, Cheng Z et al (2018) Unsupervised multi-view feature extraction with dynamic graph learning. J Vis Commun Image Represent 56:256–264

Seo M, Kembhavi A, Farhadi A et al (2016) Bidirectional attention flow for machine comprehension. arXiv preprint arXiv:1611.01603

Socher R, Lin CC, Manning C et al (2011) Parsing natural scenes and natural language with recursive neural networks. In: Proceedings of the 28th international conference on machine learning (ICML-11), pp 129–136

Suppes P, Liang L, Bottner M (1996) Machine learning comprehension grammars for ten languages. Comput Linguist 22(3):329–350

Sutskever I, Vinyals O, Le Q (2014) Sequence to sequence learning with neural networks. In: Advances in neural information processing systems

Tan M, Santos C, Xiang B et al (2015) LSTM-based deep learning models for non-factoid answer selection. arXiv preprint arXiv:1511.04108

Wang YL, Chuan Z, Qiuliang X et al (2015) Fair secure computation with reputation assumptions in the mobile social networks. Mobile Inf Syst. https://doi.org/10.1155/2015/637458

Wang Y, Huang M, Zhu X et al (2016) Attention-based LSTM for aspect-level sentiment classification. In: Conference on empirical methods in natural language processing, pp 606–615

Wang YL, Zhao M, Hu YM, Gao YJ, Cui XC (2019) Secure computation protocols under asymmetric scenarios in enterprise information system. Enterp Inf Syst. https://doi.org/10.1080/17517575.2019.1597387

Woods WA (1973) Progress in natural language understanding: an application to lunar geology. In: Proceedings of the national conference of the American Federation of Information Processing Societies, 4–8 June, pp 441–450

Wu H, Zhang H, Cui L et al (2018) A heuristic model for supporting users' decision-making in privacy disclosure for recommendation. Secur Commun. https://doi.org/10.1155/2018/2790373

Xu C (2017) Research on multi-granularity analysis and processing method of time series signal based on convolution-long-term memory neural network, Harbin Institute of Technology

Xu J, Xu Y, Zhang Y et al (2015) Combining semantic comprehension and machine learning for Chinese sentiment classification. Open Autom Control Syst J 7(1):1660–1666

Yao Y, Huang Z (2016) Bi-directional LSTM recurrent neural network for Chinese word segmentation. In: International conference on neural information processing. Springer, Cham, pp 345–353

Yin J, Jiang X, Lu Z et al (2015) Neural generative question answering. arXiv preprint arXiv:1512.01337

Yu XM, Wang H, Zheng X et al (2016) Effective algorithms for vertical mining probabilistic frequent patterns in uncertain mobile environments. Int J Ad Hoc Ubiquitous Comput 23(3/4):137

Zhang D, Lee WS (2003) Question classification using support vector machines. In: Proceedings of the 26th annual international ACM SIGIR conference on research and development in information retrieval. ACM, pp 26–32

Zhang B, Zhu L, Sun J, Zhang H (2018) Cross-media retrieval with collective deep semantic learning. Multimed Tools Appl 77(17):22247–22266

Zhen L, Wenxian X, Wenlong W et al (2012) The research of Chinese Q&A system based on similarity algorithm. Computer, informatics, cybernetics and applications. Springer, Dordrecht, pp 981–990

Zheng XW, Li Y, Liu H et al (2016) A study on a cooperative character modeling based on an improved NSGA II. Multimed Tools Appl 75(8):4305–4320

Zheng X, Tian J, Xiao X et al (2018) A heuristic survivable virtual network mapping algorithm. Soft Comput 23:1453. https://doi.org/10.1007/s00500-018-3152-7

Zhou B, Sun C, Lin L et al (2018) LSTM based question answering for large scale knowledge base. Beijing Da Xue Xue Bao 54(2):286–292

Zhu L, Huang Z, Li Z et al (2018) Exploring auxiliary context: discrete semantic transfer hashing for scalable image retrieval. IEEE Trans Neural Netw Learn syst 29(11):5264–5276