Question Classification Based on MAC-LSTM (Postprint)

Authors: Yu Bengong, Xu Qingtang, Zhang Peixing

Date: 2018-11-29T00:00:00+00:00

Abstract

Question classification constitutes a fundamental component of question analysis research in question answering systems, and its accuracy directly influences the efficacy of natural language understanding. To address challenges such as the typically short length of question texts, insufficient semantic information, and inadequate word co-occurrence statistics, we propose a question classification method based on a Multi-level Attention Convolutional LSTM neural network (MAC-LSTM). In contrast to deep learning models based on word embeddings, this method employs a question-word attention mechanism to specifically emphasize interrogative word features within questions. Simultaneously, by leveraging the attention mechanism to combine the distinct text modeling strengths of convolutional neural networks and long short-term memory models, it can extract lexical-level features in a parallel fashion while also learning higher-level long-distance dependency features. Experimental results demonstrate that this method yields significant performance improvements over traditional machine learning methods, as well as standard convolutional neural networks and long short-term memory models.

Full Text

Question Classification Based on MAC-LSTM

Yu Bengonga,b, Xu Qingtanga, Zhang Peihanga

aSchool of Management, bKey Laboratory of Process Optimization & Intelligent Decision-making of Ministry of Education, Hefei University of Technology, Hefei 230009, China

Abstract

Question classification is a fundamental component of research on question answering systems. Its accuracy directly affects the quality of natural language understanding. Most existing question classification methods are based on supervised learning algorithms that require word embeddings and do not consider interrogative features. However, question text is usually short, and semantic information and word co-occurrence information are insufficient. To address these problems, this paper proposes a multi-level attention convolutional LSTM neural network (MAC-LSTM) for question classification. This approach uses an interrogative word attention mechanism to focus on interrogative features in question contexts. Simultaneously, by combining the attention mechanism with the advantages of convolutional neural networks and long short-term memory recurrent neural networks (LSTM), MAC-LSTM is able to capture both local phrase features as well as global and temporal sentence semantics. Experiments show that our approach achieves better performance than traditional machine learning methods, ordinary convolutional neural networks, and traditional LSTM models on question classification tasks without requiring prior knowledge.

Keywords: question answering; question classification; attention mechanism; interrogative attention mechanism; convolutional neural networks; LSTM

0 Introduction

Question answering (QA) systems are a hot research topic in the field of information retrieval. They can provide concise and accurate answers to natural language questions posed by users, effectively satisfying the need for rapid and precise information acquisition. The processing pipeline of QA systems mainly consists of four steps: question classification, semantic understanding, text retrieval, and answer extraction. Among these, question classification determines the target answer type of a question, providing semantic constraints for subsequent information retrieval and answer extraction, thereby narrowing the search scope for candidate answers. For example, the question "Where is the most famous Hui cuisine restaurant?" is a location-based question, so during answer extraction, only location-type entities in candidate answers need to be matched, which effectively improves QA accuracy. Meanwhile, question classification can also provide a basis for selecting answer generation strategies. For instance, "What are the characteristics of authentic Hui cuisine?" is a description question, so when generating an answer, the focus should be on describing and introducing entity features to provide users with the knowledge they most want to obtain.

Unlike general text classification, question classification corpora exhibit the following characteristics: (a) interrogative words in questions have strong correlations with question categories, and incorporating interrogative word features can make classification results more accurate; (b) question texts are typically short, with insufficient feature co-occurrence information, and they exhibit colloquial and ambiguous characteristics. Existing question classification methods are primarily based on rule-based approaches and machine learning methods. These methods do not consider the particularities of question text corpora, thus facing issues of data feature sparsity and semantic sensitivity at the semantic level, and they struggle to extract interrogative word features effectively.

To address the above problems, this paper combines deep learning models with attention mechanisms and proposes a multi-level attention convolutional long short-term memory model. This method uses an interrogative word attention mechanism to focus on interrogative features in questions, enhancing the model's ability to extract interrogative word features. Simultaneously, by inserting an attention mechanism between the convolutional layer and LSTM layer, the model can extract lexical-level features in parallel while learning higher-level long-distance dependency features.

1 Related Work

Early question answering systems were domain-specific and emphasized particular domain knowledge, so question classification rules needed to be formulated by experts with relevant backgrounds. For example, Biswas et al. extracted fixed grammatical patterns for medical data and then used these patterns to determine question categories. This manual feature design requires substantial human effort and is only applicable to specific datasets, resulting in poor transferability. Machine learning-based question classification methods, by contrast, have stronger applicability. Research on these methods has primarily focused on model selection and feature extraction—namely, how to extract features from short question texts and obtain shallow semantic information. For instance, Li et al. used bilingual information (English-Chinese), sentence length, and syllable count as extended features for a support vector machine model, achieving excellent results in English sentence classification. Considering the importance of interrogative words in questions, Hu et al. proposed the concept of interrogative semantic units and combined them with HOWNET for semantic expansion, achieving high classification accuracy in open-domain question classification experiments. However, machine learning-based question classification algorithms face the problem of data feature sparsity and cannot effectively extract semantic information embedded in text, leaving considerable room for improvement in classifier performance.

Currently, deep learning has achieved excellent performance in natural language processing and other fields, and many scholars have utilized deep learning techniques to solve question classification tasks in QA systems. Li et al. applied convolutional neural networks to question classification in knowledge base QA systems and extended features such as answer text, answer type, and answer paths to learn question classification rules. For the same problem, Feng et al. used shared convolutional neural networks to train question-answer pairs and performed semantic similarity calculations on this basis. These methods used convolutional neural networks for question classification tasks and could effectively capture complex mappings from raw data to high-level semantics, demonstrating far greater expressive power than traditional machine learning models. However, these methods have complex parameters and do not consider the temporal features of text corpora. To address this issue, some scholars began using long short-term memory models to process sequential data like questions. LSTM is a special improved structure of recurrent neural networks that can solve the problems of gradient vanishing and long-distance dependency in the original model. Wang et al. trained joint feature vectors of question-answer pairs based on a multi-layer LSTM network, converting the question-answer matching problem into a classification or ranking problem. Additionally, considering the complementary advantages of CNN and LSTM in text modeling, Zhou et al. combined the two models to propose a novel combined model (C-LSTM) and demonstrated that CNN and LSTM provide complementary information, achieving excellent performance on both sentiment classification and question classification tasks.

The principle of attention mechanisms is to imitate the attention mechanism of the human brain, achieving significant accuracy improvements in deep learning models by automatically weighting to focus locally on overall information. In 2014, Bahdanau et al. used attention-based recurrent neural networks to solve machine translation tasks, initiating the application of attention mechanisms in natural language processing. Zhou et al. proposed a hierarchical attention mechanism for bilingual LSTM to solve cross-lingual sentiment analysis tasks, where word-level attention patterns can identify which words in each sentence are decisive, while sentence-level attention patterns learn which sentences are more important for determining the overall document sentiment. Additionally, considering the excellent performance of convolutional neural networks in short text classification scenarios, scholars have explored the combination of attention mechanisms with CNNs. Among them, reference [19] proposed a hierarchical convolutional neural network by combining user and product information with text semantic information through attention mechanisms. This method can reduce model parameters to a certain extent and enrich semantic information in text.

These research achievements have inspired the introduction of attention mechanisms into question classification to measure feature importance, and combining them with relevant models can achieve effective feature extraction. The multi-level attention convolutional long short-term memory model proposed in this paper uses convolutional neural networks to extract local features from the interrogative word attention matrix and combines the attention matrix to filter the most relevant feature mappings for question classification to input into the bidirectional long short-term memory model. Furthermore, in question classification tasks within the QA domain, interrogative words are important influential features. To effectively extract interrogative word features, this paper proposes an interrogative word-based attention mechanism that strengthens the model' s focused attention on interrogative word features by constructing an interrogative word attention matrix. Thus, MAC-LSTM simultaneously possesses the advantages of LSTM in capturing global context and CNN in capturing local question features, as well as the ability to extract key features such as interrogative words.

2 Question Classification Based on MAC-LSTM

The MAC-LSTM model proposed in this paper learns interrogative features from question text through two levels of attention mechanisms, effectively improving question classification accuracy. First, at the model input layer, an interrogative word attention mechanism is introduced to construct an interrogative word attention matrix through an interrogative word dictionary, strengthening the interrogative word semantic information in question representations, and inputting the interrogative word attention matrix into the convolutional layer to extract local features. Second, the model uses an attention matrix to filter the most useful convolutional features for question classification and feeds them to the bidirectional long short-term memory model layer for high-level temporal feature extraction. Finally, the fused feature vector is fed into a Softmax classifier to complete the question classification task from a global perspective. The model structure is shown in Figure 1 [Figure 1: see original paper].

2.1 Input Representation Based on Interrogative Word Attention Mechanism

Unlike ordinary text classification, question type discrimination relies more heavily on interrogative words as classification features. This is because question texts are short and have insufficient semantic information and word co-occurrence information, so interrogative words in questions have a greater impact on classification results. However, word vectors generally used in deep learning models do not pay special attention to interrogative word information. To address this problem, this paper proposes an interrogative word-based attention mechanism that combines word vectors with an interrogative word diagonal attention matrix by constructing an interrogative word dictionary, enabling the model to determine which parts of a sentence are most influential for interrogative words. The specific method is: collect interrogative words from Chinese questions to establish an interrogative word dictionary, where words like "谁" (who), "哪里" (where), and "多少" (how many) are all interrogative words; then perform word segmentation on questions and search the interrogative word dictionary to find interrogative words in the questions; subsequently train word vectors on a word-by-word basis; finally introduce a diagonal attention matrix to represent the contextual relevance and connection strength between ordinary words and interrogative words in the sentence.

The calculation formula for its elements is as follows:

$$a_i = e \cdot w_i$$

where: interrogative word e is obtained by scanning the question segmentation sequence through the interrogative word dictionary constructed in this paper; the function is the inner product operation of word vectors; β is a parameter vector that is updated during the training process through backpropagation.

Simultaneously, for each diagonal attention matrix element, the relative importance between the i-th ordinary word w_i and the interrogative word e is represented, and based on this, an interrogative word vector based on the attention mechanism is constructed:

$$x'_i = a_i \cdot w_i$$

The final result of the model input representation layer is the interrogative word vector arrangement matrix based on the attention mechanism:

$$X' = [x'_1, x'_2, \dots, x'_n]$$

where: the matrix dimension is $l \times n$, l is the dimension of the word vector, and n is the length of the question; the comma represents row vector concatenation. Figure 2 [Figure 2: see original paper] describes the detailed process of constructing the interrogative word vector attention matrix representation.

2.2 Feature Extraction Through Convolution Operations

Convolutional neural networks can extract local features from raw data through convolution operations. This paper uses d one-dimensional filters of the same size sliding on the interrogative word vector sequence to extract text features at different positions, obtaining interrogative word semantic dependency relationship feature mappings. The one-dimensional filter is represented as $f \in \mathbb{R}^{k \times l}$, where k is the width of the filter. At this point, the window matrix $W_i \in \mathbb{R}^{k \times l}$ corresponding to the i-th word in the sentence is composed of k one-dimensional vectors:

$$W_i = [x_i, x_{i+1}, \dots, x_{i+k-1}]$$

where: the comma represents row vector concatenation. The convolution operation between the convolution filter f and the window matrix (k-gram) at each position generates a feature mapping $c \in \mathbb{R}^{n-k+1}$, where each element c_j of the feature mapping is defined by the following formula:

$$c_j = f(W_i * f) + b$$

where: $*$ represents element-wise multiplication; b is the bias term. Relevant studies have shown that ReLU has unilateral inhibition and relatively broad excitation boundaries, and by introducing network sparsity, it can obtain faster feature learning rates. The nonlinear activation function f in this paper selects ReLU, with the formula shown in Equation (7):

$$f(x) = \max(0, x)$$

To preserve the temporal features of the original data, for d feature mappings c of the same length, they are transposed and rearranged into feature representations $E \in \mathbb{R}^{d\times(n-k+1)}$ corresponding to each window matrix W_i:

$$E = [c_1; c_2; \dots; c_d]$$

where: the semicolon represents column vector concatenation; c_i is the feature mapping generated using the i-th filter. Each column E_j of E is the new feature representation form of the window vector at position j in the original question text.

2.3 Temporal Feature Extraction Based on Attention Mechanism

For question classification tasks, classification rules depend on forward and backward context information. Therefore, this paper inputs the window mappings extracted by the convolutional layer into a bidirectional long short-term memory model to further extract text temporal features. The structure of the bidirectional long short-term memory model layer is shown in Figure 3 [Figure 3: see original paper].

In Figure 3, the boxes represent LSTM units: $\overrightarrow{h_t}$ is the output of the forward LSTM at time t; $\overleftarrow{h_t}$ is the output of the backward LSTM at time t. By concatenating the forward and backward sequence outputs, the contextual semantic representation of the text can be obtained.

To enable the overall model to automatically identify the most relevant parts of a sentence for question classification at the sentence level, this paper proposes an attention-based connection pattern to link the convolutional layer with the long short-term memory model. The specific approach is: construct a weighted attention matrix G for convolutional features to compare the importance of convolutional features. Different weight values reflect the magnitude of semantic importance, focusing on more important parts of the sentence while reducing information loss and redundancy in the feature extraction process.

The attention matrix is calculated as follows:

$$A = \tanh(U \cdot E + b)$$

where: U is the weighted parameter matrix learned by the neural network; b represents the linear bias of the attention mechanism. A is the importance function of E for question classification, which is further normalized:

$$G_{ij} = \frac{\exp(A_{ij})}{\sum_{m=1}^{n-k+1} \exp(A_{im})}$$

Finally, this attention matrix is multiplied with the output E of the convolutional layer, weighting the vector representation of important parts for classification results and feeding it as input to the bidirectional long short-term memory model. After temporal semantic modeling by the bidirectional long short-term memory model, the output vector of the hidden layer is used as input to the Softmax classifier, which analyzes features from a global perspective to complete the question classification task.

3 Experiments

3.1 Dataset Selection and Processing

The datasets selected in this paper are divided into three parts: 6,205 questions from the Baidu Lab dataset, 9,604 questions from the NLPCC 2016 International Conference on Natural Language Processing and Chinese Computing QA evaluation set, and 9,518 questions from the NLPCC 2017 QA evaluation set, totaling 25,327 questions. Among them, the public dataset provided by NLPCC QA has a clear format and high quality, but only contains question-answer pairs, so this paper performed manual annotation on this data. In this task, each question was independently annotated by three people, and data with disagreements were stored in a database for final collaborative annotation. Additionally, to exclude interference from accidental errors in the dataset, this paper randomly divided each dataset into training and test sets at approximately a 20% ratio.

The construction of a question classification system is a prerequisite for question classification. For English question classification, various institutions often adopt the UIUC classification system for the TREC QA standard dataset, dividing questions into six major categories: ABBR, DESC, ENTY, HUM, LOC, and NUM. However, there is no unified classification system for Chinese question classification. The system proposed by the Information Retrieval and Social Computing Center of Harbin Institute of Technology is adopted by most scholars. Based on the characteristics of Chinese, it proposes six major categories including description (DES), person (HUM), location (LOC), number (NUM), time (TIME), and entity (OBJ), with examples shown in Table 2 .

Table 2 Category System

Category	Example
Description (DES)	What business does Wanda Plaza mainly operate?
Person (HUM)	Who is the author of the book “Fundamentals of Mechanical Design” ?
Location (LOC)	Which country is Andre from?
Number (NUM)	What is the total investment of Hefei Metro Line 1?
Time (TIME)	When was Huayan Temple built?

Category	Example
Entity (OBJ)	What is the second largest ethnic group in China?

3.2 Comparison Experiment Settings

To verify the effectiveness of the proposed model in solving question classification tasks, this paper conducts comparative experiments on different Chinese short text datasets. The experimental environment and configuration are shown in Table 1 .

Table 1 Experimental Environment and Configuration

Component	Configuration
Deep Learning Framework	DeepLearning4J
Operating System	Windows 10 Enterprise Edition
CPU	Intel Core i5-6500 3.2GHz
GPU	NVIDIA GeForce GTX 1070
Word Segmentation Tool	ICTCLAS 2016

For comparison experiments, this paper adopts the same word vector training configuration as Zhou et al., using the skip-gram training mode of word2vec with a context window size of 5 and a word vector dimension of 100 to obtain the input representation of the original short text, i.e., the word group mapping matrix.

The following models are compared:

a) **SVM**: Based on the linear kernel SVM model proposed by Li et al., using the bag-of-words model for text representation and the TF-IDF algorithm for word weight calculation, which is an effective traditional classification model.

b) **CNN**: The basic convolutional neural network model proposed by Kim, consisting of convolutional layers, pooling layers, and fully connected layers.

c) **LSTM**: A bidirectional long short-term memory model proposed in the literature, suitable for processing and predicting text sequences with relatively long intervals and delays in time series.

d) **C-LSTM**: The combination of convolutional neural network and long short-term memory model proposed by Zhou et al., where features extracted by the convolutional layer are input to the long short-term memory model, adopting a novel vector rearrangement pattern.

e) **MAC-LSTM**: The model proposed in this paper, which adds interrogative word matrix input based on the attention mechanism and connection layer attention matrix to C-LSTM, enabling adaptive identification of the most important parts of a sentence.

3.3 Experimental Results and Analysis

The MAC-LSTM model proposed in this paper combines convolutional neural networks with long short-term memory models through a multi-level attention mechanism, focusing on interrogative word features for question classification tasks. To verify the model' s effectiveness, this paper sets up multiple models for comparative experiments on two datasets. The experimental results are shown in Table 3 .

Table 3 Comparison of Evaluation Scores

Model	Accuracy
SVM	76.69%
CNN	89.14%
LSTM	90.82%
C-LSTM	92.65%
MAC-LSTM	94.31%

As shown in Table 3, due to different quantity distributions of major categories across datasets, the accuracy of each model fluctuates to varying degrees, but this does not affect the comparative reference between models. This paper draws a histogram to visually demonstrate the accuracy comparison of several models on different datasets, and MAC-LSTM shows better advantages on all three datasets.

Figure 4 [Figure 4: see original paper] shows that the results of the proposed model are superior to the traditional SVM model with highly manually designed features. Manual feature design requires substantial human labor and cannot be well generalized to other datasets and tasks. The MAC-LSTM model proposed in this paper has the ability to automatically learn semantic sentence representations, requiring no manual feature extraction and offering better scalability. Comparing the results of single convolutional neural networks and long short-term memory models, convolutional neural networks achieve higher accuracy on the NLPCC dataset, indicating that convolution operations can more effectively extract text features when texts are short and data is sufficient. Compared with single convolutional neural networks and long short-term memory models, C-LSTM and the proposed MAC-LSTM achieve better classification results, demonstrating that model combination can extract local text features while further deeply mining sequence features, which is more conducive to understanding questions. Comparing the MAC-LSTM model with the C-LSTM method shows that the multi-level attention mechanism proposed in this paper enables the model to highly focus on target feature information during training, thereby better identifying interrogative word-related features and verifying the effectiveness of the attention mechanism in question classification tasks.

Based on the C-LSTM research, this paper proposes an interrogative word attention mechanism to strengthen the model' s extraction effect on question interrogative word features by constructing interrogative word vectors. The feature representation effect of interrogative word vectors is greatly related to the dimension of word vectors. This paper conducts experiments using word vectors of different dimensions to construct interrogative word vectors, verifying the effectiveness and stability of the interrogative word attention mechanism. The experimental results are shown in Figure 5 [Figure 5: see original paper]. The figure shows that when the vector dimension size changes, the model accuracy increases initially and becomes stable after the dimension size exceeds 50. This paper finally selects 100 as the vector dimension, which can achieve relatively good experimental results while maintaining computational simplicity.

4 Conclusion

This paper analyzes the shortcomings of traditional methods and the relevant applications of deep learning in this field from the perspective of question classification, and improves deep learning models by combining them with attention mechanisms. A multi-level attention deep learning model is proposed. First, the attention mechanism is applied to enhance the model' s focus on interrogative word features. Then, convolutional neural networks and long short-term memory models are effectively combined through an attention connection pattern, leveraging the advantages of both models for question classification. The MAC-LSTM model can learn phrase-level features of interrogative word vectors through convolution operations, and then feed the feature representations to the long short-term memory model to enhance its ability to capture text temporal dependency features. This paper evaluates the question type classification task using the MAC-LSTM model and achieves satisfactory results.

Current question classification tasks face the problem of lacking corpora. In addition to continuing to collect and annotate new data, utilizing large amounts of unlabeled corpora for semi-supervised learning is also a feasible direction. Furthermore, question classification is one of the fundamental studies of question answering systems, and the quality of question classification results affects the selection of answer extraction and answer generation strategies in QA systems. Therefore, research on combining answer extraction with question classification will also be a focus of future work.

References

[1] Dang H T, Kelly D, Lin J. Overview of the TREC 2007 question answering track [C]// Proc of the 16th Text Retrieval Conference. 2007: 115-123.

[2] Le-Hong P, Phan X H, Nguyen T D. Using dependency analysis to improve question classification [C]// Proc of Knowledge and Systems Engineering. 2015: 653-665.

[3] Hernault H, Prendinger H, Duverle D A, et al. HILDA: a discourse parser using support vector machine classification [J]. Dialogue & Discourse, 2010, 1(3).

[4] Li Xin, Roth D. Learning question classifiers [J]. Proc. COLING-2002, 2002, 12(24): 556-562.

[5] Biswas P, Sharan A, Kumar R. Question classification using syntactic and rule based approach [C]// Proc of International Conference on Advances in Computing, Communications and Informatics. 2014: 1033-1038.

[6] Li Kuan, Xiong Zhongyang, Zhang Yufang, et al. The spoken//written language classification of english sentences with bilingual information [M]// Natural Language Processing and Chinese Computing. Berlin: Springer, 2013: 370-377.

[7] Hu Yue, Liu Bo, Zhang Shouwei. Chinese question classify model based on interrogative semantic unit [C]// Proc of International Symposium on Computer Architecture. 2013, 92: 14-19.

[8] Kim Y. Convolutional neural networks for sentence classification [J]. arXiv preprint arXiv: 1408.5882, 2014.

[9] Li Dong, Wei Furu, Zhou Ming, et al. Question answering over freebase with multi-column convolutional neural networks [C]// Proc of Meeting of the Association for Computational Linguistics and the, International Joint Conference on Natural Language Processing. 2015: 260-269.

[10] Feng Minwei, Xiang Bing, Glass M R, et al. Applying deep learning to answer selection: a study and an open task [J]. Astrophysics Data System, 2015: 813-820.

[11] Ruder S, Ghaffari P, Breslin J G. A hierarchical model of reviews for aspect-based sentiment analysis [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2016: 999-1005.

[12] Wang Ding, Nyberg E. A long short-term memory model for answer sentence selection in question answering [C]// Proc of Meeting of the Association for Computational Linguistics and the, International Joint Conference on Natural Language Processing. 2015: 707-712.

[13] Zhou Chunting, Sun Chonglin, Liu Zhiyan, et al. A C-LSTM neural network for text classification [J]. arXiv preprint arXiv: 1511.08630, 2015.

[14] Yin Wenpeng, Schütze H, Xiang Bing, et al. Abcnn: attention-based convolutional neural network for modeling sentence pairs [J]. arXiv preprint arXiv: 1512.05193, 2015.

[15] Mnih V, Heess N, Graves A, et al. Recurrent models of visual attention [J]. Astrophysics Data System, 2014, 3: 2204-2212.

[16] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [J]. arXiv preprint arXiv: 1409.0473, 2014.

[17] Zhou Xinjie, Wan Xiaojun, Xiao Jianguo. Attention-based LSTM network for cross-Lingual sentiment classification [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2016: 247-256.

[18] Yang Zichao, Yang Diyi, Dyer C, et al. Hierarchical attention networks for document classification [C]// Proc of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2017: 1480-1489.

[19] Chen Huimin, Sun Maosong, Tu Cunchao, et al. Neural sentiment classification with user and product attention [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2016: 1650-1659.

[20] Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines [C]// Proc of International Conference on International Conference on Machine Learning. 2010: 807-814.

[21] Hao Tianyong, Xie Wenxiu, Wu Qingyao, et al. Leveraging question target word features through semantic relation expansion for answer type classification [J]. Knowledge-Based Systems, 2017.

Note: Figure translations are in progress. See original paper for figures.