



Research Progress of Automatic Question Answering System Based on Deep Learning

Shiyao Zhao, Zhezhi Jin*

Yanbian University, Yanji, China

Email: *1780012422@qq.com

How to cite this paper: Zhao, S.Y. and Jin, Z.Z. (2020) Research Progress of Automatic Question Answering System Based on Deep Learning. *Open Access Library Journal*, 7: e6046.

<https://doi.org/10.4236/oalib.1106046>

Received: January 2, 2020

Accepted: June 14, 2020

Published: June 17, 2020

Copyright © 2020 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

With the rapid development of deep learning, a large number of machine reading comprehension models based on deep learning have emerged. Firstly, the paper points out the shortcomings of traditional search engines and explains the advantages of automatic question answering systems compared with them. Secondly, it summarizes the development process of the deep learning-based machine reading comprehension model, and expounds the overall framework and operation principle of the model, as well as the advantages and application scope of the model. Finally, it points out where the development trend lies, and lays the foundation for follow-up researchers.

Subject Areas

Artificial Intelligence, Computer Engineering, Education

Keywords

Natural Language Processing, Deep Learning, Machine Reading Comprehension, Automatic Question Answering, Attention Mechanism

1. Introduction

With the vigorous development of the Internet and the sharp increase in data, we have entered the era of big data. The increasing in the amount of data is a challenge for how to quickly retrieve key information from large amounts of data. However, traditional search engines have the following disadvantages: 1) Search engines usually use keywords to guess the user's intentions, but they cannot fully understand them. 2) It takes a lot of time to return a large number of webpages related to keywords. It is inefficient to read and extract key information manually. Therefore, it can no longer meet people's needs for accurate information.

Automatic question answering system is one of the most popular directions in the field of natural language processing, it can input questions in the form of natural language, and the system analyzes the semantics, so that it can more accurately understand the needs of users and filter invalid information. Finally, it is sorted by relevance to return, which is more accurate and efficient. Therefore, the intelligent question answering system has become a hot topic in the field of natural language processing.

2. Deep Semantic Matching Model

Semantic matching technology plays an important role in information retrieval and search engines, and plays an important role in the recall of results and precise ranking. The semantic matching technology in the traditional sense pays more attention to the degree of semantic matching at the text level, which we will temporarily call the semantic matching at the language level.

DSSM is an abbreviation of Deep Semantic Similarity Model, which is established by Huang [1] and others on the basis of deep neural networks. Its main principle is to represent the problem and the document as a low-dimensional semantic vector through DNN, and calculate the distance of the semantic vector through the cosine similarity. Although the DSSM model has supervisory capabilities, it loses contextual information. To make up for this shortcoming, the CLSM model proposed by Shen [2] *et al.* is improved on the basis of the DSSM model, replacing the DNN with CNN to learn the semantic vector, and retains some context information. Because CLSM has some difficulties in saving long-distance semantic dependencies, Palangi [3] constructed the LSTM-DSSM model. Later, Elkahky [4] and others also proposed the MV-DSSM model based on the DSSM model.

3. Machine Reading Comprehension Model

3.1. Definition of Machine Reading Comprehension

Machine reading comprehension is to make machines imitate humans to read text and answer questions based on understanding. Machine reading comprehension is a research hotspot of scholars at home and abroad today, involving many fields such as information retrieval and natural language processing. It can help users find the answers they need in a large amount of text through a computer, reducing the cost of searching for information.

3.2. Machine Reading Comprehension Model Based on Deep Learning

Deep learning is famous for its powerful learning capabilities, and it has been widely used in the field of natural language processing and has achieved great results. Attention mechanism was first applied to image processing. Later, people began to use it in natural language processing and achieved good results.

Hermann [5] and others built the CNN/Daily corpus and proposed the con-

cept of Attentive Reader (**Figure 1**). The model encodes the problem and the document using bidirectional LSTMs, and then concatenating the hidden state of the problem in both directions to form the problem representation Q . Second, calculating the attention weight of the words in the document, and perform weighted average to obtain document D . Next, the non-linear Tanh function is used to integrate the representation of the problem and the document. Finally, use softmax to predict the answer.

BiLSTM is called Bi-directional Long Short-Term Memory, and it consists of two LSTMs, forward and backward. Compared with LSTM, it can better capture the information of sentence context, but it is limited in the hidden layer when calculating the long-distance dependencies. Therefore, Tan [6] and others combined the attention mechanism and proposed the QA_LSTM-Attention model. The model enhances the weight of more relevant words and reduces the weight of less relevant words, which can effectively alleviate the disadvantages of BiLSTM. The specific steps are as follows: First, BiLSTM modeling is performed on the input, and features are extracted to obtain a representation feature map. Before pooling, the BiLSTM output is multiplied by the softmax weight, and finally the cosine similarity is used to calculate the matching degree.

The attention mechanism in QA_LSTM_Attention only considers the impact of the question on the answer, and ignores the impact of the answer on the question. Cicero [7] and others simultaneously applied attention to questions and answers, and proposed the AP_BiLSTM model (**Figure 2**), which is more comprehensive than QA_LSTM_Attention. The innovation of the model lies in mapping the input ends Q and A into a high-dimensional space through the parameter matrix U , and then constructing the matrix G through the features of Q and A , where G represents the interaction between the answer and the question. Next, max pooling is performed on the columns and rows of G , which respectively indicate the importance scores of the answer to the question and the question to the answer. Finally, after getting the attention vector of the question and answer, do the matching work.

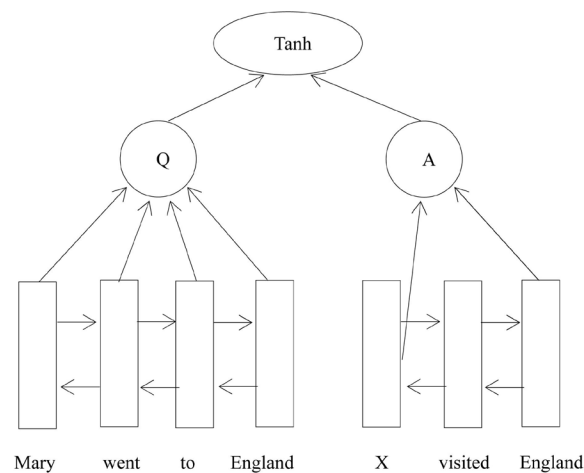


Figure 1. Architecture of attentive reader.

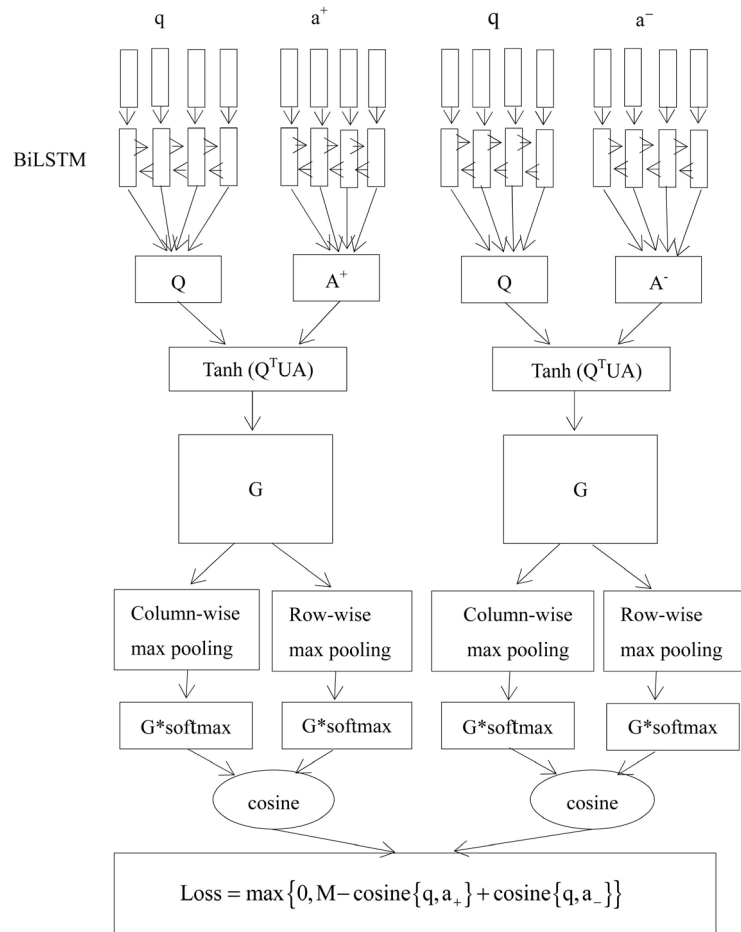


Figure 2. Architecture of AP_BiLSTM.

Wang [8] *et al.* proposed a new model with reading comprehension ability-Match-LSTM, which uses two different recurrent neural networks to encode questions and documents. And combined Tanh function and LSTM to calculate the attention weight of each word in the document in the problem and weight it. Finally, the boundary model is used to predict the start and end positions of the answer. The disadvantage of the model is that it takes too long to train and reason, which is not suitable for daily use.

Compared with match-lstm, it only predicts the location of the answer once through pointer network. Xiong C [9] *et al.* proposed a dynamic coattention network using multiple iterations at the prediction layer. The model is mainly divided into two parts: the coating encoder and the dynamic pointer decoder. The answer position is continuously updated through multiple iterations. It can solve the problem of trapping in the local optimal solution.

Seo [10] and others proposed the Bi-Directional Attention Flow (BiDAF) model. It also uses Char-CNN and GloVe to map the vocabulary to a high-dimensional vector space, stitches it, and then uses two layers of Highway Network processing to obtain the problem and document matrix. Through the dual-end attention mechanism, the attention weights of the problem to the document side and the

document to the problem side are calculated, normalized processing is performed using softmax, and weighted summing is performed. It fully considers the interaction between the problem and the document, and reduces the loss of effective information. Finally, use Pointer Network to predict the answer position. BiDAF is faster than Match-LSTM in training time.

Wang W [11] first proposed the R-Net model. The model uses two interaction layers, which are used to capture the interaction information between the problem and the document and the interaction information between the words in the document. The model introduces self-attention and gating mechanisms, and expands the receptive field of each vocabulary to the full text, which enhances the ability to understand documents.

4. Summary and Outlook

With the rapid development of deep learning technology, machine reading comprehension technology based on deep learning has been widely improved. This paper introduces and sorts out the development history of the machine reading comprehension model of deep learning, and briefly describes the overall framework, operation principle, the scope and advantages of the model. The current reading comprehension model only calculates text similarity based on shallow semantics, which cannot model the deep semantic, so it does not have the ability to reason. Therefore, the automatic question answering system needs further development, and the future development trend is the study of machine reading comprehension models with inference ability.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Huang, P., He, X., Gao, J., *et al.* (2013) Learning Deep Structured Semantic Models for Web Search Using Click through Data. *ACM International Conference on Conference on Information & Knowledge Management*, San Francisco, 27 October-1 November 2013, 2333-2338.
- [2] Shen, Y., *et al.* (2014) A Latent Semantic Model with Convolutional-Pooling Structure for Information Retrieval. *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, Shanghai, 3-7 November 2014, 101-110. <https://doi.org/10.1145/2661829.2661935>
- [3] Palangi, H., *et al.* (2014) Semantic Modelling with Long-Short-Term Memory for Information Retrieval. arXiv:1412.6629
- [4] Elkahky, A.M., Song, Y. and He, X. (2015) A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems. *Proceedings of the 24th International Conference on World Wide Web*, International World Wide Web Conferences Steering Committee, Florence, 18-22 May 2015, 278-288. <https://doi.org/10.1145/2736277.2741667>
- [5] Hermann, K.M., Kocisky Hermann, K.M., Kocisky, T., *et al.* (2015) Teaching Ma-

chines to Read and Comprehend. *Advances in Neural Information Processing Systems*, **28**, 1693-1701.

<http://papers.nips.cc/paper/5945-teaching-machines-to-read-and-comprehend>

- [6] Tan, M., *et al.* (2015) LSTM-Based Deep Learning Models for Non-Factoid Answer Selection. arXiv:1511.04108
- [7] dos Santos, C., Tan, M., *et al.* (2016) Attentive Pooling Networks. arXiv:1602.03609
- [8] Wang, S. and Jiang, J. (2016) Machine Comprehension Using Match-Lstm and Answer Pointer. arXiv:1608.07905
- [9] Xiong, C., Zhong, V. and Socher, R. (2016) Dynamic Coattention Networks for Question Answering. arXiv:1611.01604
- [10] Seo, M., Kembhavi, A., Farhadi, A., *et al.* (2016) Bidirectional Attention Flow for Machine Comprehension. arXiv:1611.01603
- [11] Wang, W., Yang, N., Wei, F., *et al.* (2017) Gated Self-Matching Networks for Reading Comprehension and Question Answering. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Stroudsburg, July 2017, 189-198. <https://doi.org/10.18653/v1/p17-1018>