

Article

Long Text QA Matching Model Based on BiGRU–DAttention–DSSM

Shihong Chen ^{1,†} and Tianjiao Xu ^{2,*,†}

¹ Laboratory of Language Engineering and Computing, School of Information Science and Technology, Guangdong University of Foreign Studies, Guangzhou 510000, China; 200911836@oamail.gdufs.edu.cn

² Faculty of Science and Technology, University of Macau, Macau 999078, China

* Correspondence: mc05402@um.edu.mo

† These authors contributed equally to this work.

Abstract: QA matching is a very important task in natural language processing, but current research on text matching focuses more on short text matching rather than long text matching. Compared with short text matching, long text matching is rich in information, but distracting information is frequent. This paper extracted question-and-answer pairs about psychological counseling to research long text QA-matching technology based on deep learning. We adjusted DSSM (Deep Structured Semantic Model) to make it suitable for the QA-matching task. Moreover, for better extraction of long text features, we also improved DSSM by enriching the text representation layer, using a bidirectional neural network and attention mechanism. The experimental results show that BiGRU–Dattention–DSSM performs better at matching questions and answers.

Keywords: QA matching; long text; DSSM; BiGRU–Dattention–DSSM



Citation: Chen, S.; Xu, T. Long Text QA Matching Model Based on BiGRU–DAttention–DSSM. *Mathematics* **2021**, *9*, 1129. <https://doi.org/10.3390/math9101129>

Academic Editor: Radir Romansky

Received: 19 April 2021

Accepted: 14 May 2021

Published: 17 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Text matching is an important task in natural language processing (NLP) and is widely used in information retrieval, QA systems, chatbots, machine translation, recommendation systems, machine reading and so on. Unlike textual similarity, text matching is more concerned with exploring the logical relationship between texts. For example, in an intelligent QA system, text matching comprises giving a question to find the matching answer in all candidate answers; in chatbots, it is a candidate response that matches the text of the previous conversation. Among them, the QA system is an application that allows users to ask questions in the form of natural language, and the system directly returns accurate and concise answers to the users. How to improve the efficiency of obtaining information, locate answers more intelligently and accurately, and provide a better experience for users have become the most popular research direction in the field of natural language processing. This project mainly focuses on the text of answering questions in the psychological counseling answering system. Since the description of a psychological problem is usually a long text, the matching of answering questions in a psychological counseling context can be summarized as the long text-matching task in the field of natural language processing [1]. Different tasks and objectives or different characteristics of the research text often need to be studied differently. Short-text QA matching has achieved fruitful results, but the matching method of long text should be distinguished from that of short text. In this project, we defined the length of long texts as between that of short texts and articles. Moreover, there are rare public question-and-answer datasets that are dominated by long text at present.

The QA system of psychological counseling can conveniently solve the superficial psychological problems of users in real time. Users may be bothered by such problems as, for example, “*what if your roommate borrows money and doesn't pay it back*” and so on. Nevertheless, these common problems are not necessarily urgent enough to seek the aid of a psychiatrist. Because of the growing threat of mental health problems in China

and the difficulties in the development of the psychological counseling industry, internet counseling platforms have become one of the most convenient channels for users. Users ask questions on the consulting platform and experts give corresponding answers. With the rapid development of internet technology and the wide application of big data technology, a large number of text data are constantly produced in the psychological counseling QA system, and accurately obtaining the real intention of users is becoming more and more important. However, feedback on the Internet consulting platform is not immediate and the quality of the reply is uneven. Some replies even induce consumption, fraud and other behaviors and are ineffective for convincing users. However, many similar answers to frequently asked questions (FAQs) are stored on psychological counseling platforms, for example, “how to do if a friend borrows money and never pay it back?” and so on. Their corresponding answers will be helpful to users. Therefore, it is of far-reaching significance and great application value to make full use of these question-answering data and construct a professional and efficient question-answering system for psychological counseling.

The process of psychological consultation is complex and has a large amount of information, which often requires long text for descriptions and replies. After a comparative analysis of the number of words in psychological counseling and community answering texts represented by Baidu Zhidao, as well as the Ant Financial text matching competition as shown in Figure 1, the lengths of psychological counseling QAs were found to be far more than those of the community QA and NLP competition task QA. In terms of the larger number of words in the psychological counseling QAs, these QAs are relatively rich in information. Compared with the common analysis objects of natural language processing—short texts and chapters—the long text of psychological counseling FAQ is often rich in important information and cannot be filtered optionally. As a result, adapting the commonly used text-matching model is difficult, so it is necessary to study the FAQ matching model of long text and apply it to the psychological counseling question-answering system.

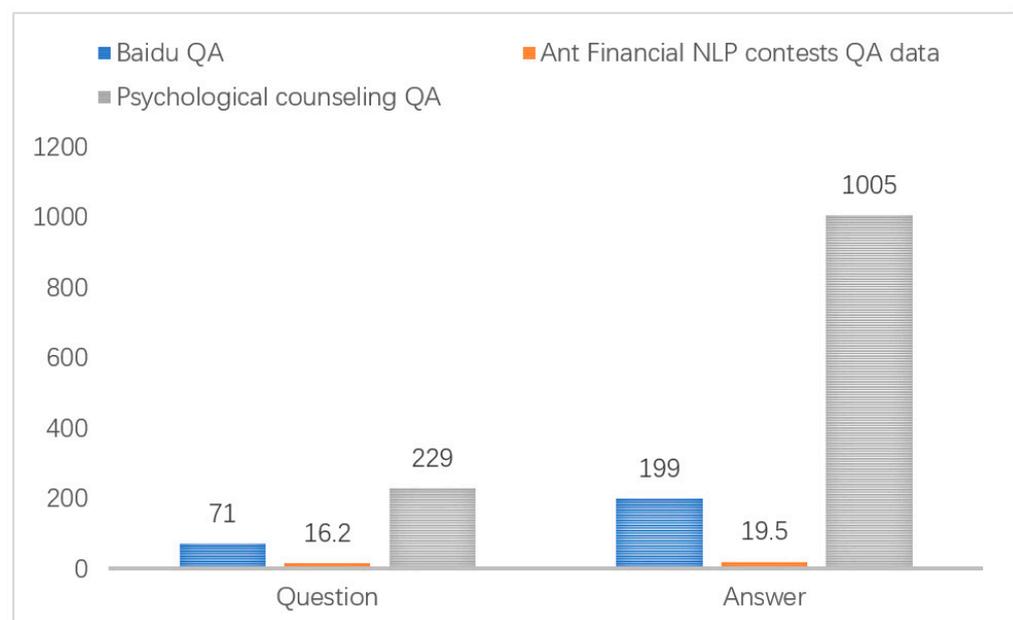


Figure 1. Comparison of the average number of words between psychological counseling QA, Ant Financial NLP contests and Baidu QA.

2. Related Work

Computation of text similarity is one of the traditional text-matching methods. The early idea was to extract the keywords from text objects and then calculate their cosine distance [2]. Later, subject models, such as LSI (Latent Semantic Indexing) and LDA (Latent Dirichlet Allocation), led the research trend of text matching [3,4] until deep learning

models emerged and became widely used. Because deep learning models can adapt to the multi-level natural language and its inherent logic, its rapid development has greatly promoted the development of natural language processing. CNN (Convolutional Neural Network) can scan the text and capture the characteristics of the text, a method that also works well for text matching [5–7]. RNN (Recurrent Neural Network) is a kind of memory neural network that can remember the previous characteristics and, according to these characteristics, infer subsequent results. Additionally, the overall network structure of RNN loops constantly, but it cannot very well capture dependencies for a long time. LSTM (Long and Short Memory Neural Network) is a special RNN designed to solve the problem of long-distance dependence, and it quickly has become a very famous research model in the NLP deep learning framework. LSTM and its variants have been widely used in text matching and have achieved convincing results [8,9].

In natural language processing, the text matching task uses training data to calculate weights to distinguish the boundary of the input source so that the learned structure will be able to distinguish the input source. Bai et al. [10] proposed a Supervised Semantic Indexing (SSI) model to calculate text. Huawei's Li et al. [11] proposed to complete text matching by performing regularized mapping to latent space (Regularized Mapping to Latent Space). Gao et al. [12] proposed a bilingual topic model (Bi-lingual Topic model) based on the topic model to model the probability of latent space to represent text.

Scholars such as Qiu of Fudan University [13] proposed a valuable CNTN model in the modern learning era based on previous research on deep learning in the academic world, and used tensor neural networks to systematically analyze the similarity of the text resource measure. Alibaba researchers [14], aiming at reducing the parameter amount of the matching model and obtaining faster inference speed, proposed three kinds of alignment features between sequences in the model, including the original point-by-point alignment feature, the previous alignment feature, and the context feature. The researchers of Ant Financial [15] proposed the Enhanced-RCNN model. By designing a combination of RNN- and CNN-encoding methods, and a special fusion layer to control the text feature representation, it has also achieved short text-matching tasks. Scholars such as Liu of Tencent [16], combined with the development needs of the big data era, focused on the idea of dividing and conquering and adopted a professional operation method to convert the long text information to be matched into multiple keyword matching applications. The information results are effectively merged to form an overall matching score, and propose two targeted Chinese long text-matching datasets and their effects.

Feng et al. [17] proposed a variety of text-matching models with CNN and LSTM structures to match questions and candidate answers. They can be applied to the insurance question answering system. ESIM (Enhanced LSTM for Natural Language Inference) is a model integrating BiLSTM and Attention mechanisms [18]. The original goal of the design is to make Natural Language Inferences based on short texts, but the effect is superior in the text-matching task after a slight modification. ESIM adds an interactive matching layer, changing the situation such that models have no interaction with the comparison objects in the past. Bi-MPM is a matching model that is combined with multiple perspectives, solving the previous interaction models' problem of insufficient matching, which has more advantages than ESIM in many aspects [19].

DSSM uses a deep neural network (DNN) to transform the text object into a low-latitude semantic vector, calculates the distance between two semantic vectors through cosine distance, and finally trains the semantic similarity model [20]. DSSM adopts a unified supervised training without doing unsupervised model mapping during the process, so the accuracy is relatively high. However, DSSM has the inherent disadvantage of losing contextual information. CDSSM was proposed by Microsoft in 2014, which is a transformed DSSM based on CNN. The difference is that DSSM's presentation layer DNN is replaced by CNN, which makes up for the problem of losing contextual information [21]. LSTM-DSSM [22] is also a variant of DSSM, which changes the presentation layer from DNN to LSTM. It can effectively retain contextual information and capture more remote contextual

features than CDSSM. This project mainly made relevant improvements based on the DSSM model, and the experimental results also proved that the improved two models performed much better than the original model. Research on the discourse level text mainly focuses on text classification, machine translation and machine reading, with little on QA matching.

To conclude, deep learning has become the mainstream approach for NLP text matching model research and has achieved fruitful results. We formerly tried to transplant a variety of insurance question-answering models from the essay [9] into the psychological question-answering system but found that the effect was very poor such that the MRR (Mean Reciprocal Rank) value was only about 0.0003. The text matching model based on long text needs new research, as it has important academic value and application value to carry out the work in this field.

3. Approaches

Psychology counseling questions and answers have far longer word lengths than community questions and are relatively more informative. According to the characteristics of question-and-answer pairs, we chose DSSM as the fundamental method, which has excellent performance, and modified it to adapt to the QA matching task. Later, we improved DSSM by changing the presentation layer from DNN to BiGRU and Double Attention through repeated trials to verify the effectiveness of the proposed model.

3.1. Modified DSSM for QA Matching

DSSM is a proven and commonly used text matching model that can be used to predict semantic similarity. We adjusted the DSSM model to fit the QA matching pattern, and its structure is shown in Figure 2, which consists of the input layer, presentation layer and matching layer. After calculating through these layers, the matching score is calculated finally.

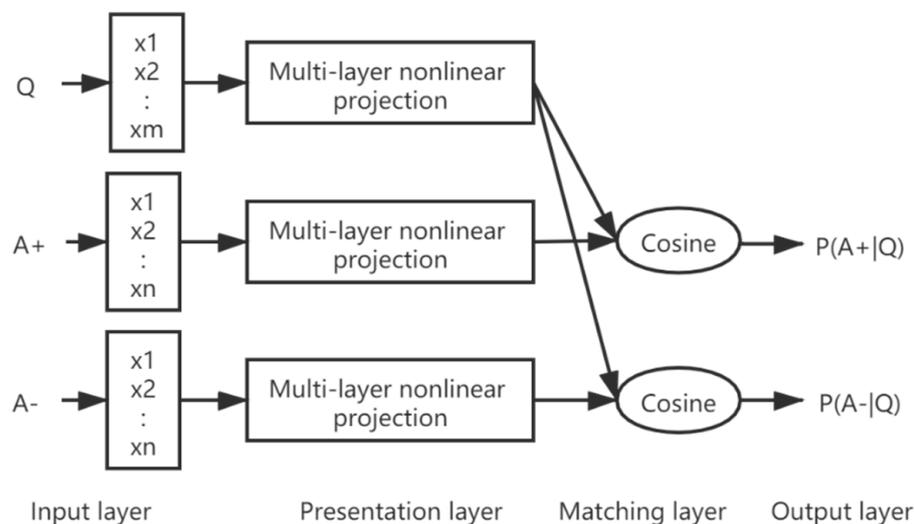


Figure 2. DSSM structure used in QA matching.

The main role of the input layer is to map text to a low dimensional vector space and convert it to vectors for the presentation layer. In the input layer of the DSSM model, the word vector, as the input, applies the one-hot method. The vector space is obviously a high-dimensional space so the initial dimensionality reduction is carried out through word hashing technology, and then the input is transferred to the presentation layer.

The presentation layer comprises multi-layer, nonlinear mapping and it usually adopts a multi-layer fully connected neural network or deep neural network, such as CNN, RNN, LSTM, etc., which is called deep DSSM. The presentation layer maps the features of the high-dimensional sparse text to the low-dimensional dense features in the semantic space.

Through sample training, feature extraction and mapping can be carried out to extract the feature information of the text. Furthermore, it can extract the characteristic information of the text as much as possible so that the output vector has semantic information.

The presentation layer of DSSM can be flexibly extended as, for example, RNN (Recurrent Neural Network). RNN can send previous output to the next hidden layer and train them together so that it can depict dynamic time behaviors. It also can be applied to sequence problems, which can be interpreted as sequence problems after vectorization of the QA input. BiRNN (Bidirectional recursive neural network) can use all available input information both in the past and future of a specific time frame for training [23]. When BiRNN is added to DSSM, the forward and backward directions work together on the results, making the output vector of the model consider both historical and future information, making the semantic information more comprehensive and richer.

The matching layer firstly calculates the distance $R(Q, A)$ between the question (Q) and the answer (A) through a cosine similarity calculation, which is represented as the matching score. The calculation formula is shown as Equation (1).

$$R(Q, A) = \cos \text{ine}(y_Q, y_A) = \frac{y_Q^T \cdot y_A}{\|y_Q\| \|y_A\|} \quad (1)$$

Secondly, the semantic correlation score is calculated by the Softmax function, which is used to calculate the posterior probability. In the calculation formula, γ is a smoothing factor in the Softmax function; A_+ represents the representation vector of the positive example answer, while A_- represents the representation vector of the negative example answer. The calculation formula is shown as Equation (2).

$$P(A_+|Q) = \frac{\exp(\gamma R(Q, A_+))}{\sum \exp(\gamma R(Q, A_-))} \quad (2)$$

During training process, the loss function formula is shown as Equation (3). It is modified based on the loss function of DSSM. The main goal is to maximize the probability of correct answers matching and minimize the probability of wrong answers matching.

$$L(A, Q) = \log P(A_-|Q) - \log P(A_+|Q) \quad (3)$$

The word vectors trained by the traditional unsupervised language model do not pay attention to the antagonism of words. Some emotional words, such as “happy” and “depressed”, have opposite meanings but are close in the vector representation. This is bad for certain NLP tasks, such as the sentiment analysis, QA, etc. In the process of psychological counseling, it is especially necessary to arouse the sensitivity of these emotional words.

In the DSSM model, questions and answers are input as pairs. From the perspective of the neural network trained in its presentation layer, it can be said that DSSM is a weakly supervised model. Therefore, the DSSM trained model can catch the implicit matching relationship between “too stupid”, “too stupid to understand”, “useless”, or “ass” and “inferiority”. This is very beneficial to the QA system of psychological counseling.

3.2. BiGRU-DAttention-DSSM

GRU (Gate Recurrent Unit) is a variant of LSTM (Long-Short-Term Memory), which has fewer parameters than LSTM due to the lack of an output gate. The performance of GRU is similar to that of LSTM, which is an improvement for the gradient disappearance problem of RNN. GRU and LSTM can both learn long-term dependent information and be more effective for long text. However, GRU can train and calculate faster than LSTM. Moreover, BiGRU (Bidirectional GRU) can remember the contextual information of sequences with a simple structure and high training speed [24]. The essence of the BiGRU-Dattention-DSSM structure is to replace the presentation layer in the DSSM structure with BiGRU-Dattention.

The complete structure of the BiGRU-Dattention-DSSM model is shown in Figure 3. With BiGRU and the Double Attention mechanism, the output vectors of the model can not only consider historical and future information, but also better remember important semantic information. After vectorization of the questions and answers, the matching scores are calculated by the cosine distance after one dense layer.

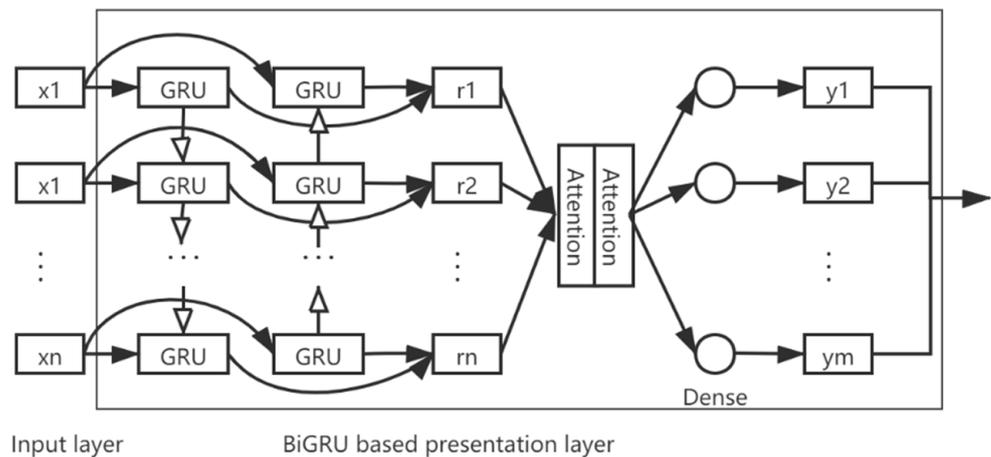


Figure 3. BiGRU-based presentation layer structure.

GRU model has only two gates, including an update gate and reset gate. The update gate is used to control the extent to which the state information of the previous moment is passed into the current state. The larger the value of the update gate, the more that the state information of the previous moment is passed into the current state. The reset gate is used to control the extent to which the status information of the previous moment is forgotten. The lower the reset gate value is, the more information is forgotten [25,26]. The calculation formula is shown below.

$$u_t = \sigma(W_u \cdot [h_{t-1}, x_t]) \tag{4}$$

In Equation (4), u_t is the calculated value of the update gate; x_t is the input data; h_{t-1} is the last output of the GRU unit; W_u is the weight matrix of the update gate; and σ is the sigmoid function. Additionally, u_t determines how much the last output h_{t-1} is passed to the next state h_t .

$$r_t = \sigma \cdot (W_r \cdot [h_{t-1}, x_t]) \tag{5}$$

In Equation (5), r_t is the value of the reset gate; W_r is the weight matrix of the reset gate; u_t and r_t control the value of the hidden state h_t .

According to the calculation results of the reset gate, update gate and hidden state, the output result h_t at the current moment can be obtained by Equations (6) and (7).

$$\tilde{h}_t = \tanh(W_h \cdot [r_t * h_{t-1}, x_t]) \tag{6}$$

$$h_t = (1 - u_t) \cdot h_{t-1} + u_t * \tilde{h}_t \tag{7}$$

The formula for calculating the attention mechanism is shown as Equation (8).

$$y = h_t \cdot \text{softmax}(w_a^T \cdot \tanh(h_t))^T \tag{8}$$

In Equation (8), y is the output of the attention layer, and w_a is the weight matrix of training and learning in the attention layer.

The structure of BiGRU-DAttention combined with DSSM is shown in Figure 4. First, the question-and-answer pairs are vectorized in a word-embedding process followed by sentence-embedding, composed with BiGRU, Double Attention, hidden layer, regulariza-

tion, and other structures. Finally, score calculation is performed by the DSSM matching layer to obtain the correctly matched score and incorrectly matched score.

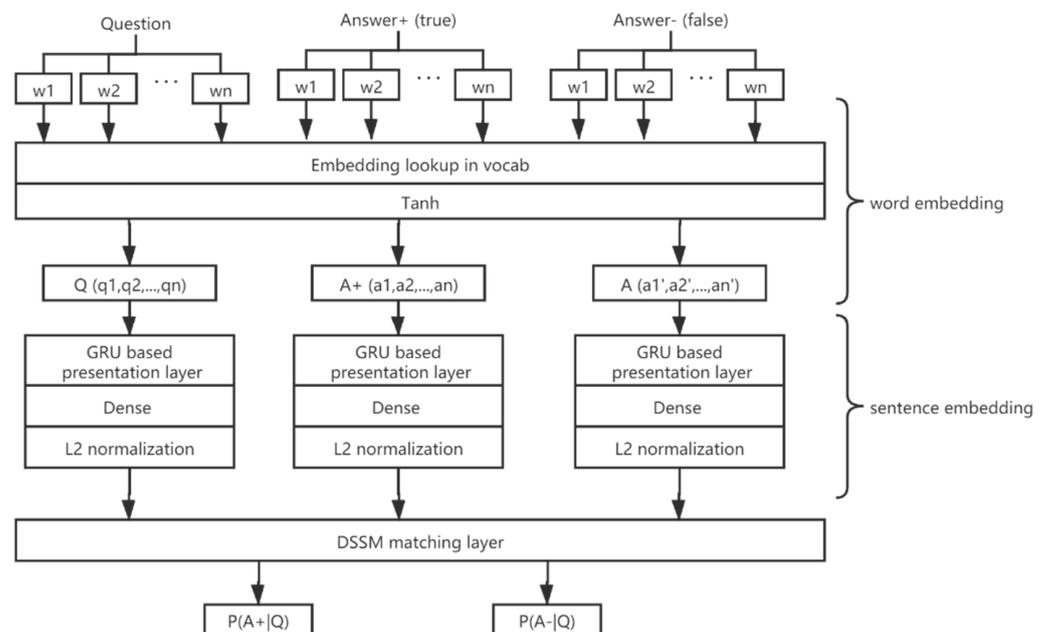


Figure 4. BiGRU–Dattention–DSSM structure.

4. Experiments

4.1. Data Pre-Processing

Firstly, we crawled the FAQ data from major community websites, such as “One Psychology” (<https://www.xinli001.com/> (accessed on 1 November 2019)), “Consulting China” (<http://www.psyzg.com/wd/> (accessed on 1 November 2019)) and other well-known psychological counseling platforms in China. The QA database was constructed with the structure of “question” and “answer” pairs, and a total of 14,204 high-quality question-and-answer pairs were selected. In the QA model training, each question should form question-and-answer pairs with matching and mismatching answers. The answers unrelated to the target questions were randomly selected from the question-answering library as the negative sample. Matching and mismatching QA pairs are represented as (Q, A_+) and (Q, A_-) , with the ratio of 1:1.

Secondly, the question-and-answer pairs were filtered by stop words. All questions and answers were segmented by Chinese text segmentation and the dictionary vocab files were established. Each word corresponds to a unique word number. For each pair with a question and matching answer, the word segmentation was mapped to the vocab file, and the sentence vector (Q, A) was synthesized as the input data of the model. The amount of training data and test data were divided into a ratio of 9:1.

4.2. Assessment Method

MRR (Mean Reciprocal Rank) was selected as our model performance evaluation reference, which is a commonly used indicator for rank. In fact, the QA-matching computation is a top N computation, as it finds the best N candidate answers from the library according to the given question. MRR takes the reciprocal of the matched answer rank given in the sorting results and averages all the calculated scores. The calculation formula is as follows.

$$MRR = \frac{1}{N_q} \sum_{i=1}^{N_q} \frac{1}{rank_i} \tag{9}$$

In Equation (9), N_q is the number of tested questions; $rank_i$ is the rank of the exactly matching answer among all candidate matching answers. For example, all matching answers to one certain question are ranked from high to low according to their matching score. If the matching answer ranks first in the candidate answers, then the MRR score of this question is 1; if it ranks second, then the MRR score of this question is 0.5; if it ranks n , then the MRR score of this question is $1/n$. The final score is the average of all the questions' MRR.

4.3. Experiment Details

In the input layer of the model, the numbers of dimensions are fixed. As shown in Figure 1, the average number of Chinese characters for psychological counseling questions and answers is 229 and 1005, respectively. The length of the answer is relatively long, but the analysis shows that answers with more words often contain multiple replies, or there are multiple rounds of questioning in the reply process. The content of about 600 words in front of the answer is often more important and representative. In Chinese, most words are composed of two Chinese characters. In other words, one word is generally equal to two characters. Therefore, the dimension of the input layer is fixed as 200 words for the question and 300 words for the answer. The extra words are truncated, and short sentences are filled in with zeros to a given length.

For the presentation layer of DSSM, we chose an attention layer followed by a full connection layer of 64 nodes, and got the final presentation layer vector after L2 normalization. The presentation layer of BiGRU-DAttention-DSSM consists of two same GRUs with 64 units connecting through RNN, the Double Attention layer, and then another 64 units of the hidden layer. The final representation layer vector is obtained after L2 normalization. The structure of BiRNN-DAttention-DSSM is similar to that of BiGRU-DAttention-DSSM.

Because of the different maximum output lengths of the questions and answers, the output values of these two types of text after passing through the presentation layer are denoted as Question_Encoder and Answer_Encoder, respectively. The cosine similarity is used to calculate the distance $R(\text{Question_Encoder}, \text{Answer_Encoder})$ as the matching score. Finally, the Softmax function is used to calculate the posterior probability according to the semantic correlation score.

4.4. Result and Analysis

By referring to the methods in the essay [17], we built a variety of QA-matching models based on CNN and LSTM and applied them to the QA system of psychological counseling. However, it was found that the effect was very poor. As shown in Table 1, the MRR value was very low, indicating that these model methods could not adapt to the question-answer matching of long text. In the proposed Attention-DSSM, the MRR of training and testing is close to 0.5, which is a qualitative improvement by comparison. The deep DSSM models we proposed significantly improved the matching effect. In particular, the MRR score of BiGRU-DAttention-DSSM in the test data reached 0.93363 and 0.93285 in the test data.

Table 1. MRR scores for different model structures.

ID	Architecture on QA Matching	MRR	
		Training Data	Test Data
1	Model structure in essay [9]	0.0003	0.00025
2	Attention-DSSM	0.499	0.449
3	BiRNN-Attention-DSSM	0.96315	0.93019
4	BiRNN-DAttention-DSSM	0.96436	0.93250
5	BiGRU-Attention-DSSM	0.96447	0.93285
6	BiGRU-DAttention-DSSM	0.96450	0.93285

According to the experimental results, traditional text-matching methods based on CNN or LSTM are not suitable for long text matching. However, with the improved model based on DSSM, we achieve excellent results by replacing the presentation layer with deep neural networks. Analyzing the reasons, the essence of DSSM is to use neural networks to express a pair of text objects to be matched as low-latitude semantic vectors, calculate the distance between these two semantic vectors through the cosine distance, and finally, train the semantic similarity model. DSSM is a weakly supervised model and has a strong semantic learning ability after training, which is helpful to capture implicit matching relationships in long text. Both BiRNN and BiGRU are neural networks that have obvious advantages in processing sequence data. The two-layer attention mechanism enables the model to pay more attention to important information and learn to absorb it, which is very important for long texts with rich information. The experimental results showed that BiGRU–Dattention–DSSM was superior to BiRNN–Dattention–DSSM, which also proved that BiGRU was better at solving the long-term dependence problem of sequence data.

5. Discussion

When commonly used text-matching methods are applied to psychological question answering, we find that the matching effect of these methods is not good. After analysis, it turns out that psychological question-and-answer pairs are long texts with rich information but with more distracting information than short text. In addition, the problem level of the questioner is not consistent and their language styles are different, which bring a great challenge to the processing of text information. Therefore, we need to study new ways to cope with these challenges. We improved the DSSM model to better adapt to the QA matching task and changed the DSSM presentation layer from DNN to BiGRU–Dattention to better extract the semantic information.

The psychological question-answering system that we study can meet the superficial psychological counseling needs of most users. Moreover, we find that our research objects have some special attributes, which are not present in the open research objects. The research objects of NLP are divided into many categories according to the length, such as characters, words, sentences, chapters, etc., and the processing methods of different research objects are very different. We hope to attract the attention of more scholars. In the following work, we also hope to introduce more research work, such as text classification and automatic labeling, to further optimize long text matching tasks.

6. Conclusions

Psychological consultation is a complex process, which requires more information to be described clearly, so the texts of psychological questions and answers are longer. The current QA is mainly based on short text; similar psychological QA studies based on longer text are not found. Long text research based on machine translation and reading comprehension tasks is common, but long text research based on QA matching is rare. In this paper, we found that traditional QA matching models based on CNN or LSTM are noneffective in long text QA matching for psychological counseling. The text-matching model based on short text is not suitable for long text. However, there are not many reference materials and cases for long text-matching models. Therefore, we modified DSSM to fit the QA matching task. The experimental results show that the matching effect of the DSSM-based matching model in long text is significantly improved.

DSSM has certain advantages in the text representation of weakly supervised training and an extensible presentation layer, which is why we changed DSSM's presentation layer with deep learning models. The experiment results shows that BiGRU–Dattention–DSSM reached the highest performance. With BiGRU and the Double Attention layer, we can better capture the long-distance text correlation information. This shows that the combination of a modified DSSM and deep learning network is effective in long text matching in the QA matching task.

Author Contributions: Conceptualization, methodology, resources, S.C.; software, validation, formal analysis, investigation, data curation, T.X.; writing—original draft preparation, T.X.; writing—review and editing, S.C.; visualization, T.X.; supervision, T.X.; project administration, S.C.; funding acquisition, S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data were mainly crawled from the well-known psychological counseling platforms in China, including “One Psychology” and “Consulting China”.

Acknowledgments: This work was supported by the GDUPS Laboratory of Language Engineering and Computing Bidding Item (grant No. LEC2018ZBKT002 and No. LEC2020ZBKT002), National Natural Science Foundation of China (grant No. 61772146), the Science and Technology Planning Project of Guangzhou (grant No. 202002030239).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Ling, P. *Research on Construction of Automatic Question Answering System Based on Neural Network Semantic Matching*; Nanjing University of Posts and Telecommunications: Nanjing, China, 2020. (In Chinese)
- Huanjong, Z.; Yujian, L.; Yixin, Z. A New Method for Text Similarity Computing. *Comput. Sci.* **2002**, *29*, 92–93. (In Chinese)
- Hou-Feng, X.G.W. The Development of Topic Models in Natural Language Processing. *Chin. J. Comput.* **2011**, *8*, 9. (In Chinese)
- Wang, Z.Z.; He, M.; Du, Y.P. Text Similarity Computing Based on Topic Model LDA. *Comput. Sci.* **2013**, *40*, 229–232.
- Qi, L.; Zhang, Y.; Liu, T. SCIR-QA at SemEval-2017 Task 3: CNN Model Based on Similar and Dissimilar Information between Keywords for Question Similarity. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), Vancouver, BC, Canada, 3–4 August 2017; pp. 305–309.
- Liu, T.C.; Wu, Y.H.; Lee, H.Y. Query-based Attention CNN for Text Similarity Map. *arXiv* **2017**, arXiv:1709.05036.
- Zhang, G.; Fan, X.; Jin, C.; Wu, M. Open-Domain Document-Based Automatic QA Models Based on CNN and Attention Mechanism. In Proceedings of the 2019 IEEE International Conference on Big Knowledge (ICBK), Beijing, China, 10–11 November 2019; pp. 326–332.
- Hou, Y.; Xiang, Y.; Tang, B.; Chen, Q.; Wang, X.; Zhu, F. Identifying High Quality Document–Summary Pairs through Text Matching. *Information* **2017**, *8*, 64.
- Xiong, C.Z.; Su, M. IARNN-Based Semantic-Containing D-Level Embedding Bi-LSTM for Question-and-Answer Matching. *Comput. Intell. Neurosci. March* **2019**, *9*, 6074840. [[CrossRef](#)] [[PubMed](#)]
- Bing, B.; Weston, J.; Grangier, D.; Collobert, B.; Sadamasa, K.; Qi, Y.; Chapelle, O.; Weinberger, K. Supervised semantic indexing. In *Proceedings of the ACM Conference on Information & Knowledge Management*; ACM: Hong Kong, China, 2009.
- Wei, W.; Lu, Z.; Hang, L. Learning Bilinear Model for Matching Queries and Documents. *J. Mach. Learn. Res.* **2013**, *14*, 2519–2548.
- Gao, J.; Toutanova, K.; Yih, W.T. Clickthrough-Based Latent Semantic Models for Web Search. In *SIGIR 2011, Proceeding of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, Beijing, China, 25–29 July 2011*; ACM: New York, NY, USA, 2011.
- Qiu, X.; Huang, X. Convolutional neural tensor network architecture for community-based question answering. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.
- Yang, R.; Zhang, J.; Gao, X.; Ji, F.; Chen, H. Simple and effective text matching with richer alignment features. *arXiv* **2019**, arXiv:1908.00300.
- Peng, S.; Cui, H.; Xie, N.; Li, S.; Zhang, J.; Li, X. Enhanced-RCNN: An Efficient Method for Learning Sentence Similarity. In Proceedings of the Web Conference 2020, Taipei, Taiwan, 20–24 April 2020; pp. 2500–2506.
- Liu, B.; Niu, D.; Wei, H.; Lin, J.; He, Y.; Lai, K.; Xu, Y. Matching Article Pairs with Graphical Decomposition and Convolutions. *arXiv* **2018**, arXiv:1802.07459.
- Minwei, F.; Bing, X.; Michael, R. Glass Applying Deep Learning to Answer Selection: A Study and an Open Task. *arXiv* **2015**, arXiv:1508.01585.
- Chen, Q.; Zhu, X.; Ling, Z.; Wei, S.; Jiang, H.; Inkpen, D. Enhanced LSTM for Natural Language Inference. *arXiv* **2016**, arXiv:1609.06038.
- Wang, Z.; Hamza, W.; Florian, R. Bilateral multi-perspective matching for natural language sentences. In Proceedings of the IJCAI, Melbourne, Australia, 15–24 August 2017.

20. Huang, P.S.; He, X.; Gao, J.; Deng, L.; Acero, A.; Heck, L. Learning deep structured semantic models for web search using clickthrough data. In *Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management*; ACM: New York, NY, USA, 2013.
21. Shen, Y.; He, X.; Gao, J.; Deng, L.; Mesnil, G. A latent semantic model with convolutional-pooling structure for information retrieval. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, Shanghai, China*; ACM: New York, NY, USA, 2014.
22. Palangi, H.; Deng, L.; Shen, Y.; Gao, J.; He, X.; Chen, J.; Song, X.; Ward, R. Semantic modelling with long-short-term memory for information retrieval. *arXiv* **2014**, arXiv:1412.6629 2014.
23. Jiwei, L.; Minh-Thang, L.; Dan, J. A Hierarchical Neural Autoencoder for Paragraphs and Documents. *arXiv* **2015**, arXiv:1506.01057.
24. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [[CrossRef](#)]
25. Cao, Y.; Li, T.; Jia, Z.; Yin, C. BiGRU: New Method of Chinese Text Sentiment Analysis. *J. Front. Comput. Sci. Technol.* **2019**, *13*, 973–981. [[CrossRef](#)]
26. Wang, N.; Li, S.L.; Liu, T.L.; Zhao, W. Attention-Based BiGRU on Tendency Analysis of Judgment Results. *Comput. Syst. Appl.* **2019**, *28*, 191–195. (In Chinese)