Multi-Granularity Fusion Text Semantic Matching Based on WoBERT

Hongchun Yu, Wei Pan*, Xing Fan and Hanqi Li
School of Computer Science, China West Normal University, Nanchong, China
{yhc,lihq}@stu.cwnu.edu.cn, panwei@cwnu.edu.cn, faith_8888@163.com

Abstract

Text semantic matching is crucial in natural language processing, applied in information retrieval, question answering, and recommendation systems. Traditional Chinese text matching methods struggle with semantic nuances in short text. Recent advancements in multi-granularity representation learning have led to increased interest in improving text semantic matching models. We propose a novel multi-granularity fusion model that harnesses WoBERT, a pre-trained language model, to enhance the accuracy of text semantic information capture. Initially, we process text using WoBERT to acquire semantic representations, effectively capturing individual text semantic nuances. Next, we employ a soft attention alignment mechanism, enabling multi-granularity fusions among characters, words, and sentences, thus further improving matching performance. Our approach was evaluated through experiments on common Chinese short text matching datasets, BQ and LCQMC. Results reveal a noticeable improvement in performance compared to traditional methods, particularly in terms of accuracy.

Keywords: text analytics, Chinese, text semantic matching, multi-granularity, WoBERT

1. Introduction

With the rise of social media and search engines, the importance of semantic matching in short texts is gaining increasing prominence. However, traditional text matching methods face challenges due to the limited information and complexity of semantic expression in short texts. They commonly depend on statistical text matching models, such as the Bag of Words model (BOW) (Harris, 1954) or Term Frequency-Inverse Document Frequency (TF-IDF) (Sparck Jones, 1972), which are evaluated by comparing word frequencies.Although these methods perform well on certain long texts, they face limitations on short texts. For example, they have difficulty capturing the semantic information of short texts due to the lack of contextual information.

In recent years, with the rapid evolution of deep learning, there has been significant progress in short text semantic matching methods rooted in deep learning. These methods benefit from the remarkable representational capabilities of deep models, excelling not only in modeling semantic relationships among texts but also offering greater flexibility and adaptability across diverse text matching applications. Word2vec (Mikolov et al., 2013) is an early deep learning text matching method, which relies on Skip-gram or Continuous Bag of Words (CBOW) to build neural word embedding. Besides Word2vec models, Convolutional Neural Networks (CNN) (Kim, 2014) and Recurrent Neural Networks (RNN) (Elman, 1990) were also extensively used in text processing. CNN employs convolution kernels to extract features from different parts of the text, enhancing short text matching by capturing local context. However, fixed-size and structured convolution kernels are not ideal for variable-length text. On the other hand, RNN's cyclic structures are adaptable to varying input data lengths and time steps. Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997), Gate Recurrent Unit (GRU) (Cho et al., 2014), and enhanced RNN structures were developed to mitigate challenges associated with vanishing and exploding gradients, which are common issues in traditional RNN. While LSTM and GRU are well-suited for handling long-term dependencies, they may still encounter memory constraints when dealing with exceptionally lengthy sequences, particularly without appropriate measures in place. To address these limitations, researchers have introduced progressively innovative and efficient deep learning models. Among these, pretrained models have excelled in text matching tasks. These models undergo pretraining on extensive language datasets to acquire text representations and are subsequently fine-tuned for specific tasks. This approach has infused fresh energy and vitality into the field of natural language processing.

Despite the notable achievements of current methods in text matching tasks, they still encounter the following challenges: (1) The precision of short text semantic matching remains insufficient, and (2) the disregard for multi-granularity semantic information. To elaborate, short texts, due to their brevity, often lack the necessary context. As a result, word choice and order play a crucial role in

*corresponding author

semantics. Furthermore, Chinese short texts encompass various semantic levels, including characters, words, and sentences. Multi-granularity semantic information contributes to a more comprehensive and accurate text understanding. However, current text matching methods primarily emphasize the representation of a single granularity, lacking a holistic consideration of multi-granularity semantic relationships in short texts. In light of the aforementioned challenges, this paper proposes a multi-granularity fusion text semantic matching model based on WoBERT. Combining the rich semantic information of the pre-training model and the comprehensive consideration of multi-level semantics by the multi-granularity mechanism, the WoBERT model was fine-tuned to align the information of character, word and sentence granularity to get a more comprehensive semantic representation. The key contributions of this paper are as follows:

- We introduce a novel multi-granularity fusion method for short text semantic matching based on WoBERT, aimed at enhancing the precision of short text similarity calculation.
- The utilization of a multi-granularity fusion mechanism facilitates the integration of semantic information across various levels, including characters, words, and sentences, leading to a substantial improvement in text semantic matching performance.
- A comprehensive experimental assessment on several widely employed text matching benchmark datasets demonstrates a substantial performance enhancement in our approach compared to traditional methods.

This paper is structured as follows: Section 2 provides a review of relevant research on pretrained models and multi-granularity text matching methods. Section 3 describes the proposed method. Section 4 presents the experimental results and analysis. Finally, in Section 5, we conclude our work.

2. Related Work

2.1. Text semantic matching based on pre-training model

Pretrained language models represent a pivotal technique in the realm of natural language processing (NLP), encompassing renowned models such as BERT (Kenton and Toutanova, 2019), RoBERTa (Liu et al., 2019), ALBERT (Lan et al., 2019), WoBERT (Su, 2020), and more. These models undergo pretraining on extensive text corpora, endowing them with the ability to grasp the grammatical, semantic, and contextual nuances of natural language. Typically, these models utilize the encoder component of the Transformer (Vaswani et al., 2017) architecture as their foundation. They hold a pivotal role in addressing text semantic matching tasks, equipping the field of natural language processing with robust tools and methodologies for a multitude of applications. Wu et al. (2021) used the integrated method of BERT and the ascending tree model to explore the matching relationship between Chinese medical Q&A data. Zhang et al. (2022) used BERT model and multi-feature convolution to extract semantic features to achieve text semantic matching tasks. Zou et al. (2022) used RoBERTa as the backbone model to carry out text semantic matching in a divide-and-conquer approach by decomposing keywords and intentions. Reusch et al. (2021) conducted similarity analysis of mathematical answer retrieval based on ALBERT, and explored the ability of ALBERT model in text modeling. The WoBERT model primarily focuses on word-level information. In the work of Dong (2023), OpenHowNet is harnessed as prior knowledge to guide semantic fusions, empowering the model to access multi-level semantic information. Wen et al. (2021) utilized WoBERT and RoBERTa pretrained language models, in conjunction with intradomain training methods, to extract medical knowledge from unlabeled medical texts. These innovations significantly bolster the performance and effectiveness of short text matching models when processing Chinese text.

However, most of these methods are based on single-granularity text semantic matching, only from one level of semantic analysis, can not accurately measure the semantic relationship between texts. To address these issues, researchers began exploring methods for multi-granularity fusions.

2.2. Multi-granularity fusion text semantic matching

The multi-granularity fusion text semantic matching method offers an effective solution to the challenges encountered in short text matching tasks. It aims to measure semantic similarity and matching degrees between texts in a more profound and comprehensive manner by considering different levels or granularity of semantic information. This approach combines various levels of feature representation to offer more comprehensive and precise semantic information. Character-level representations excel at capturing nuanced features and morphological intricacies within words, while word-level representations emphasize semantic relevance and contextual information. Additionally, within the realm of word-level analysis, words may

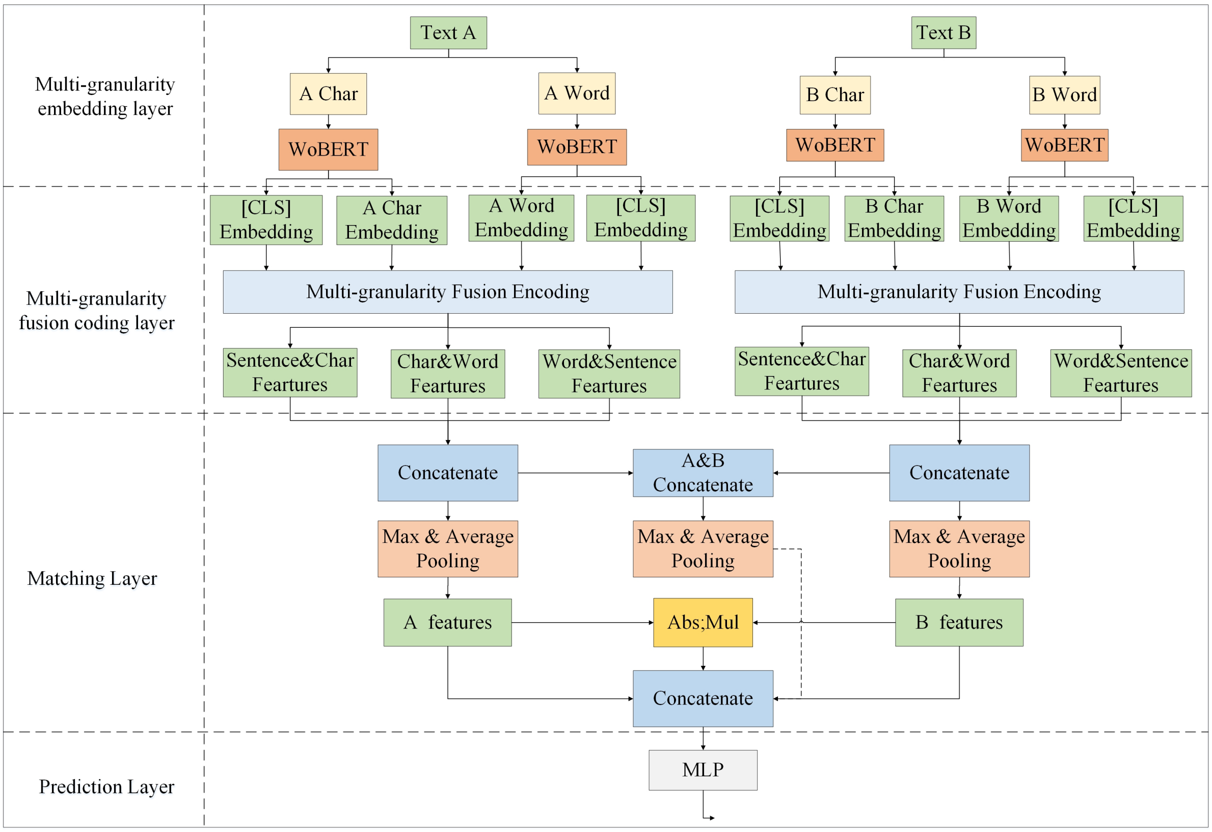

Figure 1: Model structure of WSTM

exhibit multiple meanings. Sentence-level representations aid the model in acquiring a more comprehensive grasp of the overall context, thus enhancing its capacity to disambiguate and select the most suitable meaning. Various researchers have introduced a range of methods to address diverse text matching issues. Li et al. (2019) and Chang et al. (2023) proposed to realize semantic similarity analysis through the fusion of words and phrases. Zhang et al. (2015) and Huang et al. (2017), which decompose text into smaller units to better capture the internal structure and meaning of words. In addition, there are many models that use character and word granularity for deep fusion, such as the work by Yu et al. (2021), which effectively leverages interactive information to address the challenging problem of modeling highly general text pairs across various tasks and languages. Wang and Yang (2022) simultaneously consider deep semantic similarity and shallow semantic similarity of input sentences, while also taking into account granularity at the lexical and character levels to thoroughly explore the similarity information between sentences. Zhang et al. (2020a) employ a soft alignment attention mechanism to enhance the local information of sentences at different levels, allowing them to capture the feature information of sentences and the correlation between sentences at various levels from multiple angles.While these methods leverage multi-granularity information to comprehensively understand text semantics from various perspectives, they all rely on the Word2Vec model for text embedding representation. Consequently, they may struggle to handle vocabularies not present in the training data, as well as adapt to semantic changes between different tasks or contexts. This limitation can lead to suboptimal performance when dealing with rare words or domain-specific terms.

3. Model

We have developed a multi-granularity fusion text semantic matching model called WSTM, which is based on WoBERT. The overall model structure is illustrated in Figure 1 and comprises four main components: a multi-granularity embedding layer, a multi-granularity fusion coding layer, a matching layer, and a prediction layer.To begin with, we utilize the WoBERT model to encode the input text and obtain a multi-granularity embedded representation. Subsequently, through the multi-granularity fusion coding layer, embedded representations of varying granularities interact with each other using a soft attention alignment mechanism to enrich the semantic information of the text.Following this, the matching layer is employed to compute the similarity between the extracted features. Finally, the prediction layer utilizes multiple layers of perceptrons to make the ultimate output predictions.

3.1. Multi-granularity embedding layer

Suppose there are two sentences A and B, of length m and n respectively, represented by$\{A_1, A_2, \quad A_3 \ldots A_N\}$, $\{B_1, B_2, \quad B_3 \ldots B_N\}$. where A_i and B_j represents the position of each character or word in the sentence. Firstly, the text was processed into a form suitable for WoBERT model to receive, as shown in Table 1. Subsequently, following data preprocessing, these texts are represented at different granularities as A^c, A^w, B^c, B^w . The WoBERT model is then employed to convert the text sequences of characters and words into embedded representations.

Sentence A1	逾期还款会影响征信吗 Will late repayment affect credit
A1-Character	逾/期/还/款/会/影/响/征/信/吗
A1-Word	逾期/还款/会/影响信/吗
Sentence B1	我可以贷款吗 Can I get a loan
B1- Character	我/可/以/贷/款/吗
B1-Word	我/可以/贷款/吗

Table 1: Examples of sentence segmentation with different granularity

3.2. Multi-granularity fusion coding layer

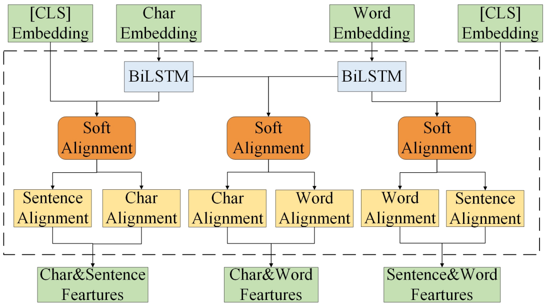

Figure 2: Multi-granularity fusion coding layer structure

We utilize a technique that combines and encodes semantic information at the character, word, and sentence levels to establish connections across multiple semantic layers. The specific structural representation is depicted in Figure 2. Initially, we utilize a Bidirectional Long Short-Term Memory (BiLSTM) to encode the character-level and word-level vectors of sentences A and B. BiLSTM has the capability to process text sequences bidirectionally, taking into account both forward and backward contextual information simultaneously. This ensures a comprehensive understanding of the contextual relevance of each word within a text sequence, thereby enhancing our grasp of word semantics and context. Moreover, this approach assists in optimizing the alignment of context, allowing the model to identify matching segments within two texts. We utilize variables A_m^c and A_l^w to represent the output states of the encoding module for character embeddings and word embeddings of A after passing through BiLSTM. Similarly, for B, we use B_n^c and B_o^w to represent them, as exemplified in equations 1-4:

$$A_m^c = BiLSTM\left(A^c, m\right), m \in (1, 2, \ldots, N) \quad (1)$$

$$A_l^w = BiLSTM\left(A^w, l\right), l \in (1, 2, \ldots, N) \quad (2)$$

$$B_n^c = BiLSTM\left(B^c, n\right), n \in (1, 2, \ldots, N) \quad (3)$$

$$B_o^w = BiLSTM\left(B^w, o\right), o \in (1, 2, \ldots, N) \quad (4)$$

To capture the significant relationships between various levels within the input sequence, we employ a soft attention alignment mechanism (Chen et al., 2017) to compute a weighted correlation between different sequences or vectors. This allows us to more effectively consider these associations when transferring information or merging representations. At the first, we calculate attention scores between the character-level and word-level, aiming to identify connections between specific character-level components and their corresponding word-level counterparts. This is demonstrated in equations 5 and 6:

$$e_{A_{ml}} = {A_m^c}^T \cdot A_l^w \quad (5)$$

$$e_{B_{no}} = {B_n^c}^T \cdot B_o^w \quad (6)$$

Next, employ the dot product method to calculate the attention scores, denoted as $e_{A_{ml}}$ and $e_{B_{no}}$. Subsequently, we utilize the softmax function in sentence A to perform normalization and weighted summation, establishing fusion relationships between characters and words, as shown in Equation 7 and Equation 8:

$$M_A^{cw} = \sum_{l=1}^{N} softmax\left(e_{A_{ml}}\right) \cdot A_l^w \quad (7)$$

$$M_A^{wc} = \sum_{m=1}^{N} softmax\left(e_{A_{ml}}\right) \cdot A_m^c \quad (8)$$

In these equations, M_A^{cw} represents the weighted sum of the m-th character in sentence A based on character granularity and the corresponding content of all words in sentence A based on word granularity. Similarly, M_A^{wc} represents the weighted sum of the l-th word in sentence A based on word granularity and the relevant content of all characters in sentence A based on character granularity. The operation for obtaining the fusion results between character and word granularity in sentence B follows the same procedure, as demonstrated in equations 9-10:

$$M_B^{cw} = \sum_{o=1}^{N} softmax\left(e_{B_{no}}\right) \cdot B_o^w \tag{9}$$

$$M_B^{wc} = \sum_{n=1}^{N} softmax\left(e_{B_{no}}\right) \cdot B_n^c \tag{10}$$

Through this process, we generate a multi-granular representation of sentences A and B based on character and word fusions. Furthermore, the [CLS] tag in WoBERT, like in BERT, is situated at the beginning of the input sequence and serves to represent the semantic information of the entire sentence or text. We opt to extract the [CLS] encodings of character and word sequences, denoted as $A_{[CLS]}^c$, $A_{[CLS]}^w$, $B_{[CLS]}^c$, and $B_{[CLS]}^w$, to represent sentence-level information. These representations are then aligned with characters and words, facilitating multi-granularity fusions between characters and sentences, as well as words and sentences. This is exemplified in equations 11-22:

$$e_{A_{sc}} = {A_{[CLS]}^c}^T \cdot A_m^c \tag{11}$$

$$e_{A_{sw}} = {A_{[CLS]}^w}^T \cdot A_l^w \tag{12}$$

$$e_{B_{sc}} = {B_{[CLS]}^c}^T \cdot B_n^c \tag{13}$$

$$e_{B_{sw}} = {B_{[CLS]}^w}^T \cdot B_o^w \tag{14}$$

$$M_A^{cs} = softmax\left(e_{A_{sc}}\right) \cdot A_{[CLS]}^c, \tag{15}$$

$$M_A^{sc} = \sum_{m=1}^{N} softmax\left(e_{A_{cs}}\right) \cdot A_m^c \tag{16}$$

$$M_A^{ws} = softmax\left(e_{A_{sw}}\right) \cdot A_{[CLS]}^w \tag{17}$$

$$M_A^{sw} = \sum_{l=1}^{N} softmax\left(e_{A_{sw}}\right) \cdot A_l^w \tag{18}$$

$$M_B^{cs} = softmax\left(e_{B_{sc}}\right) \cdot B_{[CLS]}^c \tag{19}$$

$$M_B^{sc} = \sum_{n=1}^{N} softmax\left(e_{B_{cs}}\right) \cdot B_n^c \tag{20}$$

$$M_B^{ws} = softmax\left(e_{B_{sw}}\right) \cdot B_{[CLS]}^w \tag{21}$$

$$M_B^{sw} = \sum_{o=1}^{N} softmax\left(e_{B_{sw}}\right) \cdot B_o^w \tag{22}$$

In the given context, M_A^{cs} and M_A^{sc} represent the outcomes of soft attention alignment between sentences and characters in text A, while M_A^{ws} and M_A^{sw} signify the results of soft attention alignment between sentences and words in text A. Similarly, M_B^{cs} and M_B^{sc} correspond to the soft attention alignment outcomes between sentence-level and character-level fusions in text B, and M_B^{ws} and M_B^{sw} denote the results of soft attention alignment between sentence-level and word-level fusions in text B. Through these operations, we obtain fusion characteristic information at three levels of granularity: character, word, and sentence.

3.3. Matching Layer

The multi-granularity fusion results of A and B are concatenated separately to derive their respective feature representations for A and B:

$$C_A = \left[M_A^{cw}; M_A^{wc}; M_A^{cs}; M_A^{sc}; M_A^{ws}; M_A^{sw}\right] \tag{23}$$

$$C_B = \left[M_B^{cw}; M_B^{wc}; M_B^{cs}; M_B^{sc}; M_B^{ws}; M_B^{sw}\right] \tag{24}$$

Next, we connect the features of A and B to get the correlation representation between sentences:

$$C_{AB} = \left[C_A; C_B\right] \tag{25}$$

Ultimately, the features described above are processed through maximum pooling and average pooling operations to yield their final semantic representations, as depicted in formulas 26-28.

$$C_{A_{pool}} = Max\left(C_A\right) + Mean\left(C_A\right) \tag{26}$$

$$C_{B_{pool}} = Max\left(C_B\right) + Mean\left(C_B\right) \tag{27}$$

$$C_{AB_{pool}} = Max\left(C_{AB}\right) + Mean\left(C_{AB}\right) \tag{28}$$

After acquiring the ultimate semantic representations of text A and B, we establish a matching operation to capture the their correspondence, as demonstrated in equations 29 and 30:

$$abs = \left|C_{A_{pool}} - C_{B_{pool}}\right| \tag{29}$$

$$mul = C_{A_{pool}} \odot C_{B_{pool}} \tag{30}$$

Finally, we concatenate the matching results to represent them as F, as shown in equation 31：

$$F = \left[C_{A_{pool}}; C_{B_{pool}}; C_{AB_{pool}};\ abs;\ mul\right] \tag{31}$$

3.4. Prediction Layer

In the prediction layer, a multi-layer Perceptron (MLP) classifier is employed to predict the final match. This MLP comprises three dense sublayers, with the first two using the ReLU activation function (Nair and Hinton, 2010), and the final dense layer utilizing the sigmoid activation function. Furthermore, our model employs the improved binary cross-entropy (MBCE) (Su, 2017) as the loss function, as demonstrated in formulas 32-34:

Dataset	Train	Dev	Test
BQ	100,000	10,000	10,000
LCQMC	238,766	8,802	12,500

Table 2: Distribution of BQ and LCQMC datasets

$$\theta\left(x\right)=\begin{cases}1, & x>0\\ \frac{1}{2}, & x=0\\ 0, & x<0\end{cases} \tag{32}$$

$$\begin{aligned}\lambda\left(y_{true},y_{pred}\right)&=1-\theta\left(y_{true}-m\right)\theta\left(y_{pred}-m\right)\\ &-\theta\left(1-m-y_{true}\right)\theta\left(1-m-y_{pred}\right)\end{aligned} \tag{33}$$

$$\begin{aligned}L=-\sum_{y}\lambda(y_{\text{true}},y_{\text{pred}})\,\big(y_{\text{true}}\log y_{\text{pred}}\\ +\,(1-y_{\text{true}})\log\left(1-y_{\text{pred}}\right)\big)\end{aligned} \tag{34}$$

Here, $\theta(x)$ represents the step function, y_{true} is the true label, and y_{pred} is the predicted label generated by our model. The parameter 'm' regulates the distance threshold between sample pairs, determining whether the pairs are similar. In most cases, two sample pairs are considered similar if their similarity score is less than 'm', otherwise, they are classified as dissimilar.

4. Experiment

4.1. Datasets

Our experiment involved training and evaluating on two publicly available Chinese datasets: BQ (Chen et al., 2018) and LCQMC (Liu et al., 2018). The LCQMC dataset is a substantial dataset designed for the Chinese intent-matching problem, comprising 260,068 problem pairs, each categorized as either a positive (match) or negative (mismatch) relationship. This dataset is commonly employed to train and assess natural language processing models, especially text matching models, for assessing the alignment between two problems.

The BQ dataset, on the other hand, is a Chinese corpus utilized for recognizing semantic equivalence in sentences. It encompasses 120,000 question pairs derived from online banking customer service logs, primarily used to support sentence semantic equivalence recognition in natural language processing tasks. The division of BQ and LCQMC datasets is detailed in Table 2, while Figure 3 and Figure 4 display the density distributions of text lengths for BQ and LCQMC, respectively.

4.2. Parameter Settings

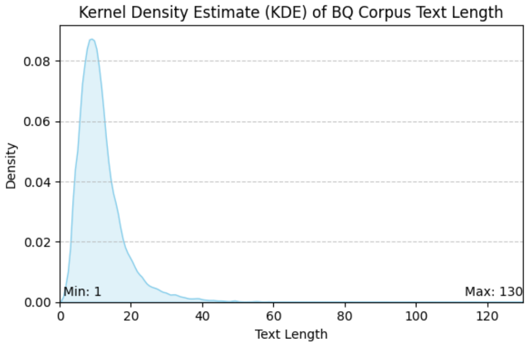

Figure 3: Text length probability distribution in BQ dataset.

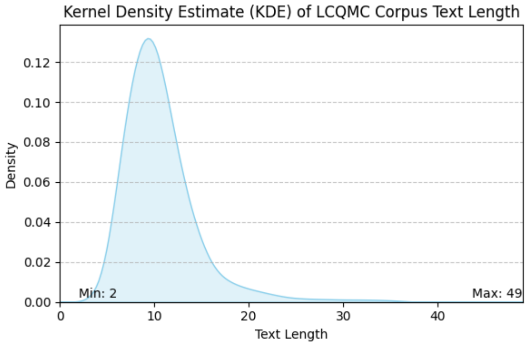

Figure 4: Text length density distribution in LCQMC dataset

Our experimental setup utilized an NVIDIA GeForce RTX 3090 (24GB) graphics card. The framework for building the model was based on PyTorch. Table 3 provides an overview of the parameter settings for our experiment, with 'threshold' denoting the threshold value for the loss function.

4.3. Experimental results and analysis

4.3.1. Comparative Experiment

We conducted a comparative analysis of the two datasets using multiple models, and the experimental results are presented in Tables 4 and 5. Table 4 clearly demonstrates that models built upon multi-granularity fusion, such as ESIM (Huang et al., 2019), MGF (Zhang et al., 2020b), LET (Lyu et al., 2021), MSEM (Huang et al., 2019) and GMN (Chen et al., 2020), outperformed those rely-

	BQ	LCQMC
max_length	30	30
Batch_size	64	128
dropout	0.5	0.5
threshold	0.5	0.74

Table 3: Parameter Settings

	BQ		LCQMC	
Model	ACC	F1	ACC	F1
Text-CNN	68.52	69.17	72.80	75.70
BiLSTM	73.51	72.68	73.50	77.5
BIMPM	81.85	81.73	83.30	84.90
ESIM	81.93	81.87	82.58	84.49
MGF	82.86	81.21	85.83	86.72
LET	83.22	83.03	84.81	86.08
MSEM	83.47	83.62	84.33	85.68
GMN	84.21	84.11	84.62	86.00
WSTM(ours)	**85.84**	**85.65**	**89.05**	**89.06**

Table 4: Experimental results in BQ and LCQMC datasets

ing on single-granularity methods like Text-CNN, BiLSTM (OscarTäckström and Das, 2017), and BiMPM (Wang et al., 2017). This underscores the feasibility of enhancing matching performance through the fusion of multi-granularity information.

Additionally, for matching models founded on multi-granularity fusion, our WSTM method showed considerable improvements. On the BQ dataset, our method achieved an increase of 1.63% $\sim$ 3.91% in accuracy and 1.54% $\sim$ 4.44% in F1 score when compared to other models. On the LCQMC dataset, the improvements were even more pronounced, with an increase of 3.22% $\sim$ 6.47% in accuracy and 2.34% $\sim$ 4.57% in F1 score. These improvements are attributed to our utilization of the pre-trained WoBERT model as opposed to the conventional Word2Vec model, which substantially enhanced our model's performance.

To assess the effectiveness of multi-granularity, we conducted a performance comparison involving several BERT-based pre-training models, as presented in Table 5. These models have traditionally been designed for text semantic matching tasks based on individual granularities. In this context, BERT (Kenton and Toutanova, 2019), BERT-WWM (Cui et al., 2021), and ERNIE (Sun et al., 2019) operate at the character level, SBERT (Reimers and Gurevych, 2019) operates at the sentence level, and WoBERT operates at the word level. Our findings reveal that our model outperforms other single-grained pre-training models, underscoring the significance of incorporating multi-grained text information for enhancing semantic matching tasks.

Given that our model employs an MLP in the prediction layer, the choice of threshold in the sigmoid activation function within its final layer is a crucial parameter. This threshold effectively determines whether the model's predicted output is considered positive or negative. Figures 5 and 6 illustrate the impact of varying thresholds on our model's performance on the two datasets. In the case of the BQ dataset, the model achieves its optimal overall performance when the threshold is set at 0.5. On the other hand, for the LCQMC dataset, the model performs best when the threshold is set to 0.74. These discrepancies can be attributed to differences in the characteristics and distribution of the two datasets. Factors such as the varying proportions of positive and negative samples and the degree of separation between these categories contribute to the selection of the model's optimal threshold.

	BQ		LCQMC	
Model	ACC	F1	ACC	F1
BERT	84.61	84.36	85.73	86.86
BERT-wwm	84.89	84.29	86.80	87.78
ERNIE	84.67	84.20	87.04	88.06
SBERT	-	-	87.28	-
WoBERT	-	-	88.47	88.21
WSTM(ours)	**85.84**	**85.65**	**89.05**	**89.06**

Table 5: Experimental results of pre-trained models in BQ and LCQMC datasets

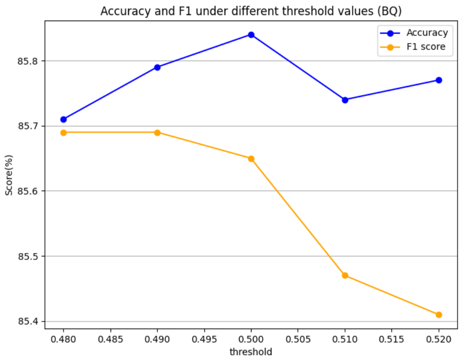

Figure 5: Effects of different thresholds on BQ dataset

4.3.2. Ablation experiment

Figures 7 and 8 illustrate the impact of removing specific elements from our WSTM model on the BQ and LCQMC datasets. In these figures, 'WSTM-char' represents the removal of character granularity, 'WSTM-word' signifies the elimination of word granularity, 'WSTM-sentence' denotes the absence of sentence granularity, and 'WSTM-wobert' corresponds to the removal of the WoBERT model in favor of the traditional Word2Vec model. It is evident that removing the WoBERT model had the most significant impact on performance. This resulted in a 2.24% decrease in

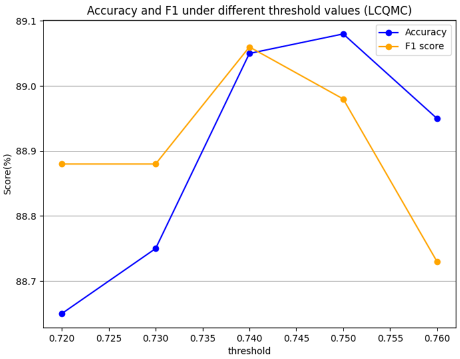

Figure 6: Effects of different thresholds on LCQMC dataset

accuracy in the BQ dataset, along with a 1.62% reduction in F1 score. In the case of the LCQMC dataset, the removal of WoBERT led to a substantial 3.64% decrease in accuracy and a 2.52% drop in F1 score. When character granularity, word granularity, and sentence granularity were removed from the BQ dataset, the accuracy of the model decreased by 1.35%, 1.43%, and 1.33%, and the F1 score dropped by 1.27%, 1.37%, and 1.56%, respectively, compared to our original model. In LCQMC, the accuracy decreased by 0.96%, 1.23%, and 0.91%, respectively, and the F1 scores dropped by 0.59%, 0.71%, and 0.70%, respectively. These results emphasize the feasibility and importance of applying the WoBERT model to a multi-granularity fusion text semantic matching task.

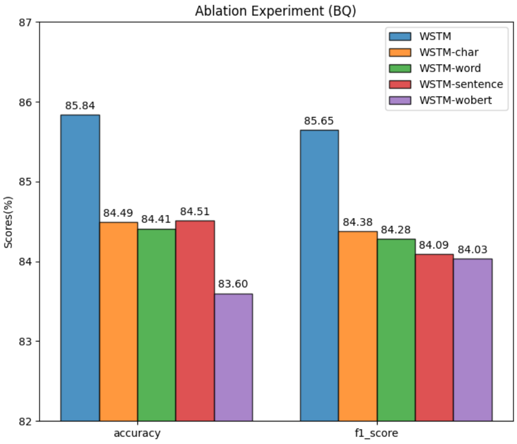

Figure 7: Ablation results of BQ dataset

5. Conclusion

In this study, we delve into a text semantic matching method for multi-granularity fusion leveraging the WoBERT pre-trained model. Our findings highlight several key observations. Firstly, by incorporating a multi-granularity fusion mechanism that integrates information at the character, word, and sentence levels while conducting semantic fusion across multiple levels, we achieved improved performance in text matching tasks compared to traditional single-granularity matching methods, to some extent. Secondly, replacing the Word2Vec model with the large-scale pre-trained language model, WoBERT, enhanced the model's ability to comprehend semantic relationships within text. In conclusion, our research demonstrates that our model produces satisfactory results across two entirely unrelated datasets, indicating the universality and extensibility of this approach.

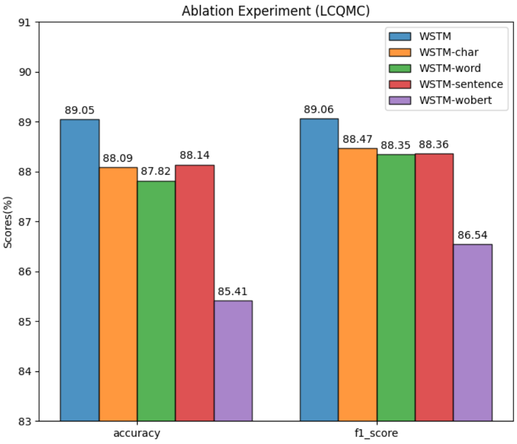

Figure 8: Ablation results of LCQMC dataset

Of course, this method also has some limitations and shortcomings. To begin with, the pre-trained model WoBERT was directly used to obtain the feature embedding, and the effect of the model was greatly affected by the pre-trained model. Moreover, the way of feature fusion is also relatively simple, and the pertinence of different levels of features is not enough. In addition, there are shortcomings in the polysemous use of the term and in the processing of domain-specific data. Therefore, in the following work, we will study from the above aspects, optimize the model, and further improve the effect.

6. Acknowledgments

This work was partially supported by the scientific foundation of China West Normal University (No.17E064,No.17YC159). The authors also would like to thank the anonymous reviewers for their help.

7. References

Guanghui Chang, Weihan Wang, and Shiyang Hu. 2023. Matchacnn: A multi-granularity deep matching model. *Neural Processing Letters*, 55(4):4419–4438.

Jing Chen, Qingcai Chen, Xin Liu, Haijun Yang, Daohe Lu, and Buzhou Tang. 2018. The bq corpus: A large-scale domain-specific chinese corpus for sentence semantic equivalence identification. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pages 4946–4951.

Lu Chen, Yanbin Zhao, Boer Lyu, Lesheng Jin, Zhi Chen, Su Zhu, and Kai Yu. 2020. Neural graph matching networks for chinese short text matching. In *Proceedings of the 58th annual meeting of the Association for Computational Linguistics*, pages 6152–6158.

Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Si Wei, Hui Jiang, and Diana Inkpen. 2017. Enhanced lstm for natural language inference. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1657–1668.

Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, page 1724. Association for Computational Linguistics.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-training with whole word masking for chinese bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3504–3514.

Jiaming Dong. 2023. Chinese short text matching model based on wobert word embedding representation and priori knowledge. In *2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, pages 298–302. IEEE.

Jeffrey L Elman. 1990. Finding structure in time. *Cognitive science*, 14(2):179–211.

Zellig S Harris. 1954. Distributional structure. *Word*, 10(2-3):146–162.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Jiangping Huang, Shuxin Yao, Chen Lyu, and Donghong Ji. 2017. Multi-granularity neural sentence model for measuring short text similarity. In *Database Systems for Advanced Applications: 22nd International Conference, DASFAA 2017, Suzhou, China, March 27-30, 2017, Proceedings, Part I 22*, pages 439–455. Springer.

Qiang Huang, Jianhui Bu, Weijian Xie, Shengwen Yang, Weijia Wu, and Liping Liu. 2019. Multitask sentence encoding model for semantic retrieval in question answering systems. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2019. Albert: A lite bert for self-supervised learning of language representations. In *International Conference on Learning Representations*.

Xu Li, Chunlong Yao, Qinyang Zhang, and Guoqi Zhang. 2019. Semantic similarity modeling based on multi-granularity interaction matching. *International Journal of Innovative Computing, Information and Control*, 15(5):1685–1700.

Xin Liu, Qingcai Chen, Chong Deng, Huajun Zeng, Jing Chen, Dongfang Li, and Buzhou Tang. 2018. Lcqmc: A large-scale chinese question matching corpus. In *Proceedings of the 27th international conference on computational linguistics*, pages 1952–1962.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach.

Boer Lyu, Lu Chen, Su Zhu, and Kai Yu. 2021. Let: Linguistic knowledge enhanced graph transformer for chinese short text matching. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13498–13506.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

Vinod Nair and Geoffrey E Hinton. 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814.

GauravSinghTomar ThyagoDuque OscarTäckström and Jakob Uszkoreit Dipanjan Das. 2017. Neural paraphrase identification of questions with noisy pretraining. *EMNLP 2017*, page 142.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992.

Anja Reusch, Maik Thiele, and Wolfgang Lehner. 2021. An albert-based similarity measure for mathematical answer retrieval. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1593–1597.

Karen Sparck Jones. 1972. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1):11–21.

J. Su. 2017. Text emotion classification (iv): Better loss function. https://spaces.ac.cn/archives/4293. Accessed 30 March 2017.

Jianlin Su. 2020. Wobert: Word-based chinese bert model-zhuiyiai. *Technical report*.

Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xin Tian, Danxiang Zhu, Hao Tian, and Hua Wu. 2019. Ernie: Enhanced representation through knowledge integration. *arXiv preprint arXiv:1904.09223*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Xin Wang and Huimin Yang. 2022. Mgmsn: Multi-granularity matching model based on siamese neural network. *Frontiers in Bioengineering and Biotechnology*, 10:839586.

Zhiguo Wang, Wael Hamza, and Radu Florian. 2017. Bilateral multi-perspective matching for natural language sentences. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 4144–4150.

Chaojie Wen, Tao Chen, Xudong Jia, and Jiang Zhu. 2021. Medical named entity recognition from un-labelled medical records based on pretrained language models and domain dictionary. *Data Intelligence*, 3(3):402–417.

Ziming Wu, Jun Liang, Zhongan Zhang, and Jianbo Lei. 2021. Exploration of text matching methods in chinese disease q&a systems: A method using ensemble based on bert and boosted tree models. *Journal of biomedical informatics*, 115:103683.

Chuanming Yu, Haodong Xue, Yifan Jiang, Lu An, and Gang Li. 2021. A simple and efficient text matching model based on deep interaction. *Information Processing & Management*, 58(6):102738.

Xiang Zhang, Junbo Zhao, and Yann LeCun. 2015. Character-level convolutional networks for text classification. *Advances in neural information processing systems*, 28.

Xu Zhang, Yifeng Li, Wenpeng Lu, Ping Jian, and Guoqiang Zhang. 2020a. Intra-correlation encoding for chinese sentence intention matching. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5193–5204.

Xu Zhang, Wenpeng Lu, Guoqiang Zhang, Fangfang Li, and Shoujin Wang. 2020b. Chinese sentence semantic matching based on multi-granularity fusion model. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 246–257. Springer.

Zhe Zhang, Yiyang Zhang, Xiang Li, Yurong Qian, and Tao Zhang. 2022. Bmcsa: multi-feature spatial convolution semantic matching model based on bert. *Journal of Intelligent & Fuzzy Systems*, 43(4):4083–4093.

Yicheng Zou, Hongwei Liu, Tao Gui, Junzhe Wang, Qi Zhang, Meng Tang, Haixiang Li, and Daniell Wang. 2022. Divide and conquer: Text semantic matching with disentangled keywords and intents. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 3622–3632.