Knowledge-Based Systems 232 (2021) 107454

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Frame-based Multi-level Semantics Representation for text matching

Shaoru Guo [a,*], Yong Guan [a], Ru Li [a,b,*], Xiaoli Li [c], Hongye Tan [a,b]

[a] School of Computer & Information Technology, Shanxi University, China
[b] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, China
[c] Institute for Infocomm Research, A*Star, Singapore

ARTICLE INFO

Article history:
Received 13 October 2020
Received in revised form 26 August 2021
Accepted 27 August 2021
Available online 1 September 2021

Keywords:
Text matching
Frame semantics
Multi-level semantic representation

ABSTRACT

Text matching is a fundamental and critical problem in natural language understanding (NLU), where *multi-level semantics matching* is the most challenging task. Human beings can always leverage their semantic knowledge, while neural computer systems first learn sentence semantic representations and then perform text matching based on learned representation. However, without sufficient semantic information, computer systems will not perform very well. To bridge the gap, we propose a novel *Frame-based Multi-level Semantics Representation* (FMSR) model, which utilizes frame knowledge to extract multi-level semantic information within sentences explicitly for the text matching task. Specifically, different from existing methods that only rely on the sophisticated architectures, FMSR model, which leverages both frame and frame elements in FrameNet, is designed to integrate multi-level semantic information with attention mechanisms to learn better sentence representations. Our extensive experimental results show that FMSR model performs better than the state-of-the-art technologies on two text matching tasks.

1. Introduction

Text matching aims to identify whether two sentences are semantically equivalent or not, which is a core research area in Natural Language Understanding (NLU). In this task, a model takes two sequences as input and predicts a category indicating their relationship, e.g. *duplicate* or *not duplicate*.

Text matching is one of the most critical tasks in many application domains, including, but not limited to, question answering, information retrieval, news recommendation and dialog systems. Specifically, text matching is an important subtask of information retrieval, which aims to predict the probability of the document being relevant to users' query. In news recommendation, system generates personalized news recommendations by calculating the distance between the real-time news and historical news that user has read recently. In addition, text matching helps users identify best answers that match given questions for question answering and dialog systems.

One of the intrinsic challenges for text matching is *multi-level semantics learning* in three different levels, i.e., at sentence level, clause level and phrase/word level. Considering the sentences listed below:

S1: ***What is the reason*** for rising unemployment in India?
S2: ***How can we solve*** unemployment in India?
S3: How much would it cost to ***hire a high school math tutor*** for 2 hours?
S4: How much does it cost to ***record a professional video*** for a 2 h presentation?
S5: How do you draw a ***cat*** step-by-step?
S6: How do you draw an ***angel*** step-by-step?

For S1 and S2, their overall semantic scenario are different at sentence level. For S3 and S4, their overall semantic scenario are same, but local semantics scenario at clause level are different. For S5 and S6, both overall and local semantic scenario are same, although their semantic scenario differ in word level, i.e. *cat* and *angel*. We can see that different levels of semantic inference are needed to address the challenging text matching task. Human beings also need multiple rounds of analysis to truly understand the overall semantics, local semantics, as well as the relationship between two sentences through comparison.

Recently, deep neural network based methods, e.g. pre-trained language model [1–3], have achieved promising results for text

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

* Corresponding authors. School of Computer & Information Technology, Shanxi University, China.
E-mail addresses: guoshaoru0928@163.com (S. Guo), guanyong0130@163.com (Y. Guan), liru@sxu.edu.cn (R. Li), xlli@i2r.a-star.edu.sg (X. Li), tanhongye@sxu.edu.cn (H. Tan).

https://doi.org/10.1016/j.knosys.2021.107454
0950-7051/

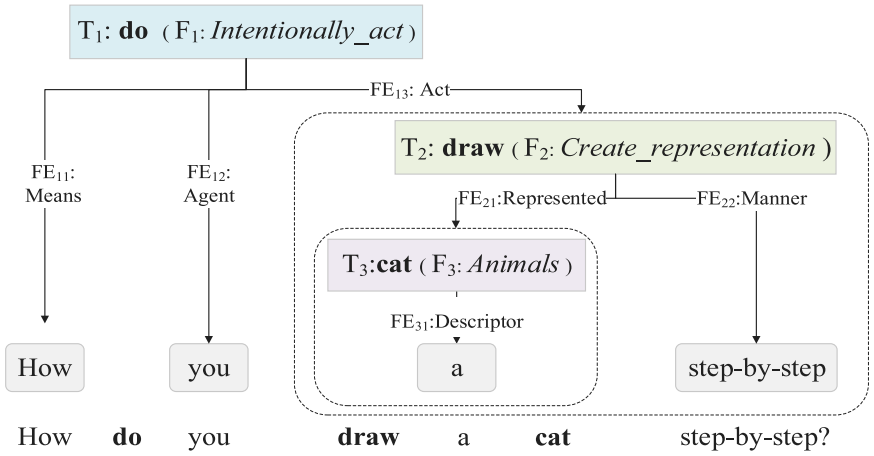

Fig. 1. An example showing frame-based multi-level semantics representation.

matching task. They first utilize a sequence encoder to learn representations of two input sentences, and subsequently calculate their similarity. However, they do not model the interaction between input sentences during the encoding procedure. While many neural network models were proposed to match sentences from multi-level of granularity [4,5], their architectures completely rely on large number of model parameters and huge training data, which need massive computation and do not really understand the natural language text [6]. As such, a new research explores to incorporate semantic knowledge of PropBank for text understanding [7], which directly connected multiple predicate–argument structures to obtain the joint representation. However, PropBank only focuses on verbs and their arguments. Thus, in this paper, we utilize *frame* semantics to model multiple granular semantic information, including both nouns and more meaningful semantic roles.

FrameNet [8,9], as a knowledge base, provides multi-level schematic scenario representation that could be potentially leveraged to better understand sentences. In particular, *Frame* is defined as a composition of *Lexical Units* (LUs) and a set of *Frame Elements* (FEs). FEs are the basic semantic units of a Frame and defined specifically to each frame. Given a sentence, if its certain word evokes a frame by matching a LU, then it is called *Target*. It is worth mentioning that many sentences could have more than one target words that evoke **multiple level frames** [10]. Fig. 1 provides an example of intuitive Frame-based multi-level semantic structures, where target word **do** in the sentence *How do you draw a cat step-by-step* evokes a frame ***Intentionally_act***, and the clause *draw a cat step-by-step* belongs to FE *Act*. In this clause, target word **draw** evokes a frame ***Create_representation***, and *a cat* maps to FE *Represented*. The noun word **cat** can act as target word and evoke a frame ***Animals***. Note Zhang et al. [11] proposed Frame-GBDT method to utilize FrameNet for text matching. But it is less effective due to three reasons: (1) ignores semantically meaningful Frame Elements (FEs), (2) its frame embeddings are initialized randomly, (3) does not focus on critical multi-level semantic modeling.

To address the above problems, we propose a novel model Frame-based Multi-level Semantics Representation (FMSR) model, which leverages rich frame semantic knowledge, including frame and FEs, to extract multi-level semantics from sentences explicitly. We first employ an automatic semantic role labeling system to process sentences. Then, we design attention mechanism to model the multi-level frame semantics structures. Finally, the labels of sentences are generated according to the hidden states of both pre-trained model and explicit semantic structure information. Extensive experiments were conducted on both Quora Question Pair [12] and Stanford Natural Language Inference [13], widely used benchmark datasets for text matching, showing promising results. The key contributions of this work are summarized as follows:

1. To our best knowledge, our work is the first attempt to leverage frame knowledge to extract multi-level semantics from sentences explicitly for text matching task.
2. We propose a novel FMSR model, which encodes frames and frame elements, as well as integrates multi-level frame semantic information for richer sentence representation. In particular, we design novel *co-attention* to model inter-sentence semantic interaction, as well as *self-attention* to model intra-sentence semantic interaction.

2. Related work

Existing work on text matching task can be roughly categorized into three classes, namely, Feature Engineering, Neural Network, Pre-Trained Language Model.

2.1. Feature engineering

Early works usually focus on integrating shallow features into the text matching task [14–16]. The common features used are sentence length, bag of words (BOW), longest contiguous matching subsequence, term frequency and inverse document frequency (TF IDF), unigrams and bigrams. Later, more sophisticated models are used for better measuring text similarity, such as knowledge graph, topic model, dependence parser [17–20]. These features focus on n-gram overlapping, word reordering and syntactic alignment phenomena. In addition, this category of approaches can work well on a specific task or dataset, but it is usually time consuming to construct good features and it is difficult to generalize well in other relevant tasks.

2.2. Neural network

With the renaissance of the neural network, neural-based frameworks have been proposed for the task of text matching. Some methods use neural networks to represent each text as vectors and the vector distance is regarded as the matching score, such as DSSM, Convolutional DSSM and LSTM-DSSM [21, 22]. Researchers also try to utilize word interaction matrix to better capture the interaction information between text, such as DRMM [23] and MatchPyramid [24]. Some frameworks are based on a *siamese* network [25], which consists of two sub-networks.

While the sub-networks in this framework share parameters, there is no interaction between the two sentences. To cope with the limitations, one way is to use *compare-aggregate* framework, which first compares vector representations of smaller units such as words from these sequences and then aggregates these comparison results to make the final decision [26]. Wang et al. [5] proposed BiMPM, which utilizes an advanced bilateral multi-perspective matching operation. The other way to enhance the model is by using the *attention-based* framework, which models the interdependence between sentences in a sentence pair. CSRAN [27] performs multi-level attention refinement with dense connections among multiple levels and MwAN [28] utilizes multiple heterogeneous attention functions to compute the alignment degree. RE2 [29] is a strong neural architecture with multiple alignment processes for text matching. For multi-label text categorization task, works try to capture both the global and the local textual semantics and to model high-order label correlations, for example, CNN–RNN for multi-label text categorization [30]. Capsule networks [31] and Sentic LSTM [32] are also utilized to improve the performance of text classification tasks. HGAT [33] adapts graph attention networks for short text, which based on a dual-level attention mechanism. In addition, Generative models for better sentence generation can be used to help train better classifiers [34]. These methods rely on sophisticated architectures, which need massive computation and might not really understand the natural language texts [6].

2.3. Pre-trained language model

Pre-Trained Language Model are the most popular choice for text matching nowadays. Contextualized word vectors from a language model trained on a large text corpus, such as ELMo [1], GPT [35], BERT [2], XLnet [36], ERNIE 2.0 [3], have been shown to be effective for textual similarity task. Among the context-sensitive language models, Bert has taken the NLP world by storm. So many optimized versions have been proposed. For instance, SemBert [7] incorporates PropBank [37] semantic roles to train an improved language representation model. SpanBert [38] extends BERT by masking contiguous random spans, rather than random tokens. RoBERTa ensemble [39] uses a novel dataset for pre-training to improve performance on downstream tasks. StructBERT [40] aims to make the most of the sequential order of words and sentences. All these methods first encode the sentences into vectors and then compute the distance between the vectors, which mainly rely on uncontrollable model parameters and neglect the level of different semantics granularity.

There are mainly two limitations for modeling "multiple granular semantic information": (1) How to discover the implicated multiple granular semantic information with respect to the central meaning of the sentence. (2) How to model the multiple granular semantic information for reply relationship modeling. To overcome the two limitations, we propose FMSR model, which integrates the information from the pre-trained model and semantic knowledge to predict the label of sentences. In particular, we make use of FrameNet to extract multiple semantic structure in a sentence. In addition, pre-trained language model is used to build representation for structural semantic information, and attention mechanism is employed to model the mutual semantic interactions.

The work most related to our work is the Frame-GBDT approach [11]. The differences between our FMSR model and Frame-GBDT are as follows: (1) Our FMSR model aims to extract frame semantic structures of sentences and further utilize attention mechanism to aggregate multi-level frame semantic structure representations of sentences. In particular, co-attention is built to model the mutual semantic interactions of sentence. In addition, self-attention is used to generate a better sentence representation by emphasizing most important frames within a sentence. We should like to point out that the existing work, i.e. Frame-GBDT, does not use FrameNet in such innovative manner. (2) Frame-GBDT only focuses on frame information, while our FMSR model involves not only frame information, but also frame elements information to model sentence. Especially, frame elements are basic semantic units of FrameNet, and provide additional information to the semantic structure of a sentence, which can help the model to extract essential constituents and find semantic related constituents to compare. (3) The frame embeddings in Frame-GBDT are initialized randomly, while we employ BERT to encode the definition of frame from FrameNet to get frame vector.

3. The proposed frame-based multi-level semantics representation model (FMSR)

In this section, we present our FMSR, considering both context and frame semantics information. The basic idea of our FMSR model is to enrich the sentence representation with multiple frame semantic structure information. Particularly, we first make use of an automatic semantic role labeling system to distill multi-level frame semantics of sentences. Then a pre-trained language model is used to build representation for input raw texts and frame semantic information. Attention mechanism is employed to model the multi-level frame semantics interactions. Finally, we integrate the text representation and multi-level frame semantics representation to predict the labels of sentences.

3.1. Problem formulation and model overview

Formally, a general text matching task can be defined as: $(P, Q) \rightarrow y$, where $P = (p_1, \ldots, p_M)$ is a sentence with a length M, $Q = (q_1, \ldots, q_N)$ is the second sentence with a length N, y is the corresponding label vector, i.e. text matching task is formalized as a classification problem. Particularly, for a *paraphrase identification* task, P and Q are two sentences, $y \in \{0, 1\}$ shows that P and Q are duplicate or not. For a *natural language inference* task, P is a premise, Q is a hypothesis, and $y \in \{0, 1, 2\}$ indicates that P and Q belong to one of the three categories: entailment, contradiction, neutral.

Fig. 2 shows the overall framework of our proposed FMSR model, consisting of three key components:

(1) The **Context Encoder** computes deep and context-aware representations for the source context.

(2) The **Frame Semantic Representation** encodes the Frame and Frame Elements with definitions and takes full advantage their representations to model the multi-level semantic structures of sentences.

(3) The **Answer Predictor** integrates frame semantics and source context representations to classify given sentences P and Q. Here we use natural language inference task as an example, i.e. $y \in \{0, 1, 2\}$. Next, we will introduce these three key components.

3.2. Context encoder

Context encoder computes deep and context-aware representations for the two sentences P and Q. Given the input context C = [CLS] P [SEP] Q [SEP], BERT is employed to encode context information as:

$$c^r = BERT(C) \tag{1}$$

where c^r is the contextual representation using the BERT encoder. For sentence pair S5 and S6 in our running example, we will simply take the concatenation of S5 and S6 as input. Specifically, the input sequence can be denoted as: [CLS] How do you draw a cat step-by-step? [SEP] How do you draw an angel step-by-step? [SEP].

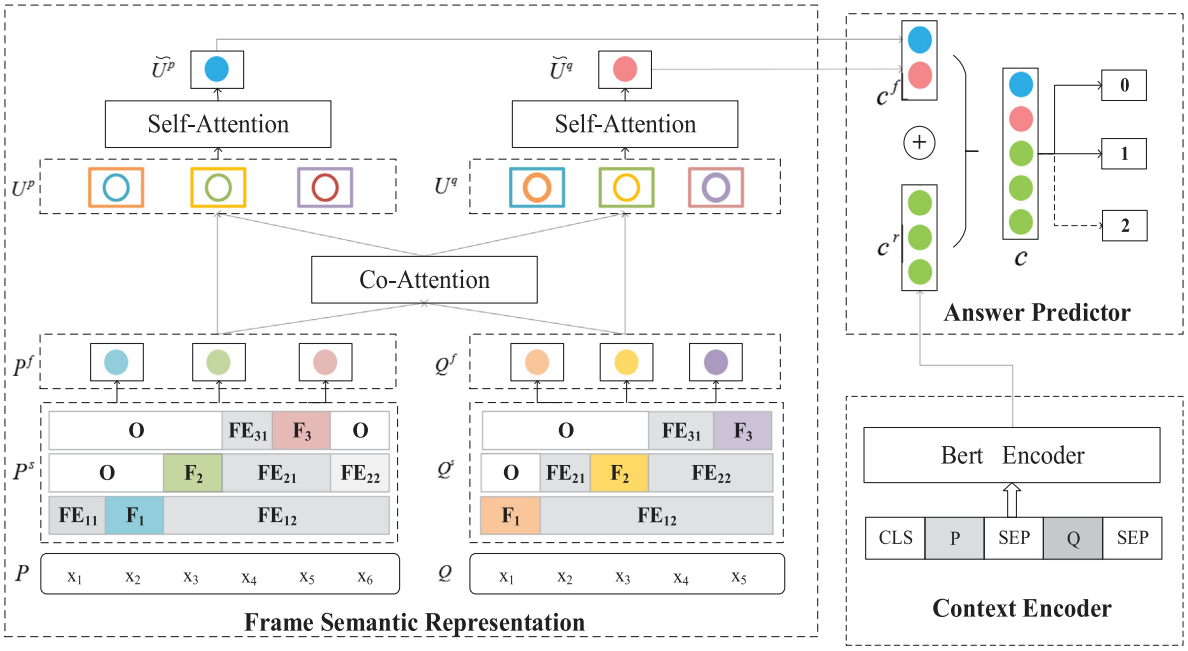

Fig. 2. The architecture of FMSR network.

3.3. Frame semantic representation

Given a sentence $P = (p_1, p_2, \ldots, p_m, \ldots)$, where p_m represents the mth word in P. Let T_i be the ith frame-evoking target of P, and T_i evokes F_i Frame. FE_{ij} denotes the jth Frame Element of F_i. Thus, the frame semantic structure of P can be formulated as $P^s = \{P_1^s, \ldots, P_i^s, \ldots\}$, where $P_i^s = [F_i, FE_{ij}]$ represents the ith frame semantic structure of P. In Fig. 1, T_1 is "*do*" and P_1^s = [Intentionally_act, {Means, Agent, Act}], i.e. F_1 = Intentionally_act, FE_{11} = Means, FE_{12} = Agent, FE_{13} = Act.

Frame semantic representation provides an effective method to model the multi-level frame semantic structures, which is the key component of our model. Particularly, it consists of three major modules: Frame-Semantic Role Labeler, Frame Semantic Structure Encoder, and Context Structure Encoder.

3.3.1. Frame-semantic role labeler

We employ SEMAFOR [41] to automatically process sentences with multiple semantic annotations [42].

Fig. 1 provides an example sentence with three T, namely *do*, *draw* and *cat*. Each T and its evoked semantic frames enclosed in the block. For each frame, its FE are shown with the arrows. For example, T **do** evokes the ***Intentionally_act*** frame, and has the *Agent*, *Act*, *Means* FEs fulfilled by *you*, *draw an angel step-by-step* and *How*. Correspondingly, we can get its multi-level frame structure $P^s = \{P_1^s, P_2^s, P_3^s\}$, where $P_i^s = [F_i, FE_{ij}]$ $(i = 1, 2, 3)$ represents the ith frame structure of P.

More specifically, the sentence P in Fig. 1 has three Frame semantic structures:

1. P_1^s = [*Intentionally_act*, {*Means*, *Agent*, *Act*}]
2. P_2^s = [*Create_representation*, {*Represented*, *Manner*}]
3. P_3^s = [*Animals*, {*Descriptor*}]

3.3.2. Frame semantic structure encoder

Frame semantic structure encoder is used to learn the representation of frame semantic structure P^s. We take sentence P as an example to illustrate how to encode its frame semantic representation, and clearly, the semantic representation of sentence Q is formed analogously.

We search the definition of frame F_d and the definitions of frame elements FE_d from FrameNet, and use BERT to encode F_d and FE_d to get their vectors f, fe respectively.

$$f = BERT(F_d) \tag{2}$$

$$fe = BERT(FE_d) \tag{3}$$

For example, the ***Intentionally_act*** frame in Fig. 1 describes a common situation in which *A Creator produces a physical object which is to serve as a Representation of an actual or imagined entity or event, the Represented*. And FE *Represented* describes *the entity–which may be a thing, an action or a state–that is represented by the Representation*. We use BERT to encode "Intentionally_act" and "Represented" to get their vectors $f(Intentionally\_act)$, $fe(Represented)$ respectively.

For the ith frame structure of P, $P_i^s = [F_i, FE_{ij}]$, we feed the vector of frame f_i, and FE fe_{ij} into BiLSTM layer to obtain the frame semantic structure representation P_i^f.

$$\overrightarrow{P_i^f} = \overrightarrow{BiLSTM}(f_i, \{fe_{i1}, \ldots, fe_{ij}, \ldots\}) \tag{4}$$

$$\overleftarrow{P_i^f} = \overleftarrow{BiLSTM}(f_i, \{fe_{i1}, \ldots, fe_{ij}, \ldots\}) \tag{5}$$

$$P_i^f = \overrightarrow{P_i^f} + \overleftarrow{P_i^f} \tag{6}$$

Finally, the frame semantic structure representation of P is formulated as: $P^f = \{P_1^f, \ldots, P_i^f, \ldots\}$.

For instance, the frame semantic structure representation of "*S5: How do you draw a cat step-by-step?*" is formulated as: $P^f = \{P^f_{Intentionally\_act}, P^f_{Create\_representation}, P^f_{Animals}\}$.

3.3.3. Context structure encoder

A novel context structure encoder is designed to aggregate multi-level frame semantic structure representations (P^f and Q^f) into an single vector c^f. In particular, to fully model the relationship between the sentence pair P and Q, **co-attention** is built to derive their pairwise representations, by modeling their mutual semantic interactions (i.e. inter-sentence semantic interactions). We take $P^f \in \mathbb{R}^{l*|P|}$ and $Q^f \in \mathbb{R}^{l*|Q|}$ to denote the frame

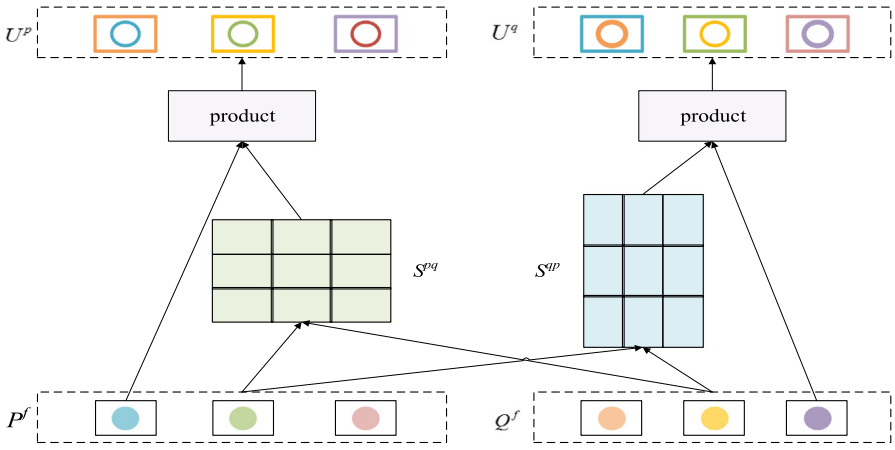

Fig. 3. The proposed co-attention module for FMSR.

semantic structure representation of P and Q respectively, where $|P|$ and $|Q|$ are the lengths of the two sequences, and l is the dimensionality of the frame semantic structure representation. As shown in Fig. 3, co-attention representation between P^f and Q^f can be calculated as:

$$\varphi(P^f, Q^f) = (W^g P^f + b^g \otimes e_{|P|})^T Q^f \tag{7}$$

$$\varphi(Q^f, P^f) = (W^g Q^f + b^g \otimes e_{|Q|})^T P^f \tag{8}$$

$$S^{pq} = softmax(\varphi(P^f, Q^f)) \tag{9}$$

$$S^{qp} = softmax(\varphi(Q^f, P^f)) \tag{10}$$

$$U^p = Q^f (S^{pq})^T \tag{11}$$

$$U^q = P^f (S^{qp})^T \tag{12}$$

where φ is a trainable scalar function that encodes the similarity between its two input matrixes, $W^g \in \mathbb{R}^{l*l}$ and $b^g \in \mathbb{R}^l$ are parameters to be learned. The outer product $(\cdot \otimes e_X)$ produces a matrix by repeating the vector for X times. $S^{pq} \in \mathbb{R}^{|P|*|Q|}$ and $S^{qp} \in \mathbb{R}^{|Q|*|P|}$ are the weight matrices between P^f and Q^f. $U^p \in \mathbb{R}^{l*|P|}$ and $U^q \in \mathbb{R}^{l*|Q|}$ represent co-attention representations for sentence P and Q respectively.

Considering $S5$ and $S6$ in our running example, we will simply use the attention mechanism to match each frame semantic structure representation state of $S6$ to $S5$, and reveal how the frame structure of $S6$ can be aligned to each frame structure of $S5$. In particular, "Intentionally_act" structure of $S5$ corresponding to "Intentionally_act" structure of $S6$ and "Animals" structure of $S5$ corresponding to "angel" structure of $S6$.

Note each sentence typically contains multiple frames. Thus, we design **self-attention** to model the interactions between these frames (i.e. intra-sentence semantic interactions) and generate a better representation by emphasizing most important frames within a sentence. In this paper, we adapt source2token self-attention mechanism [43]. Next, we show how to model the intra-sentence semantic interactions for sentence P, which explores the dependency between $u_i^p \in \mathbb{R}^l$ and the entire sequence $U^p \in \mathbb{R}^{l*|P|}$, and compresses the sequence U^p into a vector $\widetilde{U}^p \in \mathbb{R}^l$, as shown in Fig. 4.

$$f(u_i^p) = W^T \sigma(W_1 u_i^p + b) \tag{13}$$

For each frame structure u_i^p, a softmax function is applied to $f(u_i^p)$, which produces a specific distribution over all dependent frame structure:

$$t_i = softmax(f(u_i^p)) \tag{14}$$

The output of the attention mechanism is a weighted sum of the embedding for all tokens in U^p, i.e.,

$$\widetilde{U}^p = \sum_{i=1}^{|P|} t_i \odot u_i^p \tag{15}$$

where $\odot$ is element-wise multiplication. Similarly we can utilize source2token self-attention mechanism to obtain intra-sentence semantic interactions $\widetilde{U}^q \in \mathbb{R}^l$ of sentence Q.

$$f(u_j^q) = W^T \sigma(W_2 u_j^q + b) \tag{16}$$

$$t_j = softmax(f(u_j^q)) \tag{17}$$

$$\widetilde{U}^q = \sum_{j=1}^{|Q|} t_j \odot u_j^q \tag{18}$$

where $u_j^q \in \mathbb{R}^l$ is the jth frame structure of $U^q \in \mathbb{R}^{l*|Q|}$. t_j is an indicator of which frame structure is important to Q. That is, large t_j means that u_j^q contributes important information to $\widetilde{U}^q$.

In our running example, we use self-attention to compress $U^P = \{U^P_{Intentionally\_act}, U^P_{Create\_representation}, U^P_{Animals}\}$ into a vector representation $\widetilde{U}^q$ from the dependency between each frame structure u_i^p and the entire frame structure U^p.

Finally, $\widetilde{U}^p$ and $\widetilde{U}^q$ are concatenated to obtain context structure representation c^f.

$$c^f = [\widetilde{U}^p; \widetilde{U}^q] \tag{19}$$

where the [;] operator denotes vector concatenation across the rows.

3.4. Answer predictor

Answer predictor module is application-specific (e.g. could be for paraphrase identification or natural language inference task), which is used to make a class prediction. This layer takes the vector representation of the source context representation c^r and frame semantic structure representation c^f as input:

$$c = [c^r; c^f] \tag{20}$$

We apply a linear layer $f(\cdot)$ and a softmax layer on c and predict its class label l.

$$l = softmax(f(c)) \tag{21}$$

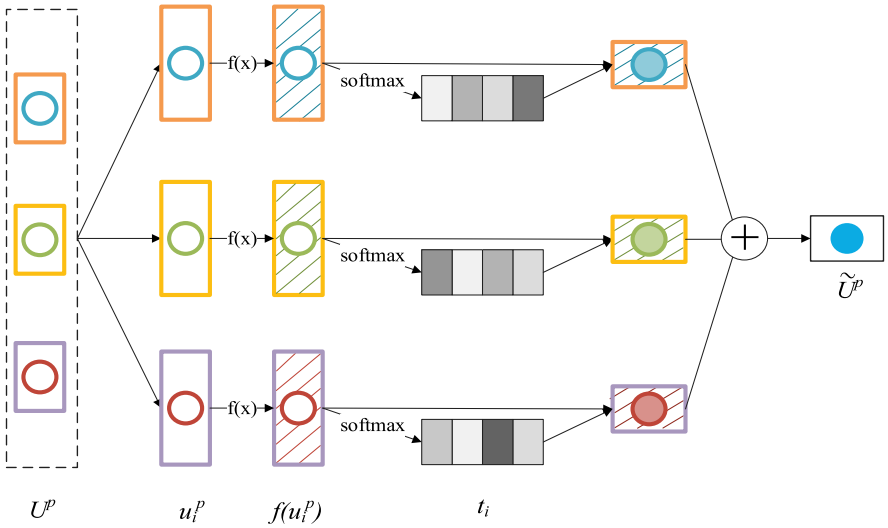

Fig. 4. The proposed self-attention module for FMSR.

Table 1
Statistics of the dataset.

Datasets		QQP	SNLI
Data size (pairs)	Train	323,432	550,152
	Dev	40,429	10,000
	Test	40,429	10,000
Avg sentence len	P (Question1/Premise)	12.3	14.1
	Q (Question2/Hypothesis)	12.5	8.3

For classification tasks, we minimize the training loss, defined by the Cross-Entropy:

$$L(\theta) = -\frac{1}{D}\sum_{d}^{D}\sum_{h}^{H} y_{d,h} log(l_{d,h}) \tag{22}$$

where θ is the set of all parameters in the model, D is the total number of examples in the dataset, $l_{d,h}$ is the predicted probability of class h for example d, and $y_{d,h}$ is the class indicator.

4. Experiments

This section introduces our datasets, experiment results, implementation details and detailed analysis.

4.1. Benchmark datasets

We now introduce two benchmark datasets that have been used in the experiment. The statistics of the two datasets are listed in Table 1.

Quora Question Pair (QQP)[1] [12] is used to identify whether the given question pair is duplicate or not. It consists of over 400,000 question pairs, and each question pair is annotated with a binary value indicating whether the two questions paraphrase or not. We use the same dataset partition, as mentioned in Zhang et al. [11], and split train/dev/test set with a proportion of 8:1:1.

Stanford Natural Language Inference (SNLI)[2] [13] is used to reason the semantic relationship between a premise sentence and a hypothesis sentence. It consists over 570,000 premise–hypothesis pairs, with labels entailment,contradiction and neutral.

Accuracy, widely used for evaluating classification performance, is deployed as our evaluation metric.

[1] https://data.quora.com/First-Quora-Dataset-Release-Question-Pairs.

[2] https://nlp.stanford.edu/projects/snli/.

4.2. Existing models

We compare our model with a number of baseline models. Now we briefly introduce several representative models.

Siamese-CNN/Siamese-LSTM [5] encode two sentences into sentence vectors with a CNN/LSTM encoder, and make a decision based on the cosine similarity between the two sentence vectors.

AI-BLSTM + Frame-GBDT [11] combines frame embedding and word embedding at the input of neural networks, and calculates matching degree of two representations. Frame embeddings are initialized randomly with a uniform distribution between $[-1, 1]$.

BiMPM [5] is a bilateral multi-perspective matching model, which first encodes two sentences with a BiLSTM encoder and then matches the two encoded sentences in two directions.

RE2 [29] highlights three key features, namely previous aligned features (**R**esidual vectors), original point-wise features (**E**mbedding vectors), and contextual features (**E**ncoded vectors) for inter-sequence alignment.

MT-DNN [44] is a Multi-Task Deep Neural Network (MT-DNN) for learning representations across multiple natural language understanding tasks. They used the pre-trained $BERT_{LARGE}$ to initialize its shared layers, refined the model via multi-tasks, and fine-tuned the model for each task using task-specific data.

SJRC [45] presents a semantic learning framework for jointly considering semantic role labeling (SRL) task and text comprehension task. This work makes attempt to let semantic role labeling (SRL) enhance text comprehension and inference through specifying verbal predicates and their corresponding semantic roles.

SemBERT [7] passes words in the input sequence to PropBank [37] semantic role labeler to fetch multiple predicate derived structures of explicit semantics. And then the word representations and semantic embedding are concatenated to form the joint representation for downstream tasks.

4.3. Implementation details

For Source Context Encoder, we use BERT-large (d = 1024) [2] as encoder, setting most of the hyperparameters as described in the original paper. For Frame Semantic Representation, we set the hidden dimensionality of the Bi-LSTM [11] to 300 to obtain the frame semantic structure.

Adam [46] has been selected as our optimizer with a batch size of 64 and the initial learning rate is set as 5e-5. Our model is trained on a single GPU, Nvidia TITAN Xp with 12G memory.

Table 2
Experimental results on Quora Question Pair. The four rows of third block are taken from [29]. The six rows of fourth block are obtained from: https://gluebenchmark.com/leaderboard.

Model	ACC. (%)
Human baseline	80.4
Siamese-CNN [5]	79.60
Siamese-LSTM [5]	82.58
AI-BLSTM+Frame-GBDT [11]	88.53
BiMPM [5]	88.2
MwAN [28]	89.1
CSRAN [27]	89.2
RE2 [29]	89.4
ERNIE 2.0 [3]	90.6
XLNet ensemble [36]	90.3
BERT-Large [2]	89.3
SemBERT [7]	89.8
SpanBert [38]	89.5
RoBert ensemble [39]	90.2
FMSR	91.59

Table 3
Experimental results on SNLI. Results in the first block are taken from [29]. The second block are obtained from [7].

Model	ACC. (%)
BiMPM [5]	86.9
MwAN [28]	88.3
CSRAN [27]	88.7
RE2 [29]	88.9
BERT-Large [7]	91.1
SemBERT [7]	91.6
MT-DNN [44]	91.6
SJRC [45]	91.3
FMSR	91.65

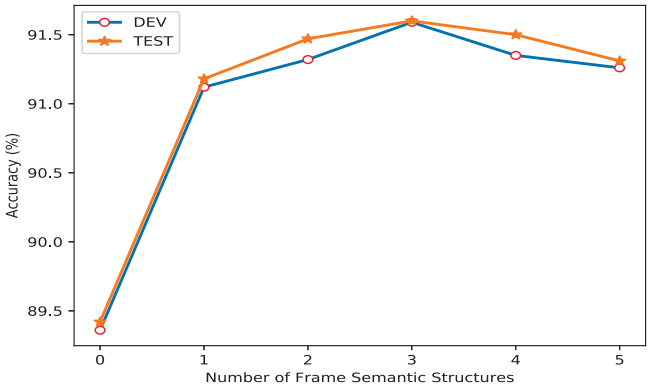

Fig. 5. Results on QQP with different number of frame structures.

4.4. Main results

Both datasets and its associated tasks are quite challenging, especially considering that new development in this area and the latest performance improvement has already become very marginal.

4.4.1. Experiments on Quora Question Pair (QQP)

In this subsection, we investigate paraphrase identification task using QQP dataset. Experimental results on QQP dataset are listed in Table 2, which shows the performance comparison among human baseline, 13 state-of-the-art models and our proposed FMSR model. We have the following four observations: First, Our proposed FMSR achieves the best accuracy, i.e. 91.59%, comparing with other 13 state-of-the-art models. Second, our FMSR model is 2.29% better than Only Bert (BERT-Large), which reveals that integrating frame semantic knowledge can clearly bring advantages and benefits. We would like to clarify that we compare our FMSR to Bert-Large for the following two reasons: (1) The backbone of our FMSR method is Bert-Large. (2) Bert-Large is widely used in many areas and achieves superior performance. Third, compared to SemBERT, our FMSR model improves the performance by 1.79% in accuracy, which indicates that our proposed architecture can capture semantic information more accurately. Note SemBERT simply concatenates predicate–argument structure of PropBank, while our FMSR mainly focuses on model the interaction of multi-frame semantic structures. So we compare our FMSR to SemBERT to verify the effectiveness of our method. Finally, our FMSR model outperforms AI-BLSTM+Frame-GBDT 3.06%, which suggests our method can take full advantage of frame semantics, including valuable frame elements, frame representation and multi-level semantic modeling. The reason why we compare FMSR method to Frame-GBDT is that both of them use frame semantic knowledge to improve text matching performance. In addition, the results of the baselines are produced by our implementation or retrieved from original papers, and we report the better one among them.

4.4.2. Experiments on stanford natural language inference (SNLI)

In this subsection, we study the natural language inference task using SNLI dataset. The detailed experimental results of 8 state-of-the-art models and our proposed FMSR model are listed in Table 3. We can see that the performance of our model is on par with two best models, namely SemBERT and MT-DNN. Therefore, we can conclude that our FMSR model is generic and also very effective for natural language inference task.

4.4.3. Number of frame semantic structures

To investigate the influence of the number of frame semantic structure $P^s = \{P_1^s, \ldots, P_i^s, \ldots\}$, we change the number of structures, i, among {0,1,2,3,4,5}, and keep the other options unchanged. Fig. 5 shows the results. We can see that $i = 3$ gives the best performance on both DEV and TEST of QQP. Too high dimension would cause severe over-fitting results, while too low dimension would cause under-fitting results. Note $i = 0$ corresponds to the performance of our baseline model (Only-BERT). Even if we only utilize one frame ($i = 1$), our model performance better than the baseline, indicating frame semantic structure is really effective for text matching.

We employ SEMAFOR [41] to automatically process sentences. Many sentences could have more than one target words that evoke multiple frames. The number of sentences belong to different number of frame semantic structures in both QQP and SNLI are shown in Fig. 6. These statistics suggest that largest percentage of sentences have 3 frame semantic structures, which verifies the rationality to choose i = 3 as the number of frame semantic structures.

4.4.4. Ablation study

To evaluate the contributions of key components/factors in our FMSR model, a series of ablation studies are performed on the QQP dev and test set.

Attention Mechanism. To evaluate the effectiveness of attention mechanism, we build three different models to integrate the multi-level frame semantic structures:

(1) -*CoAttention*, which directly uses P^f and Q^f to represent U^p and U^q;

(2) -*SelfAttention*, which performs a simple sum function on U^p and U^q to get $\widetilde{U}^p$ and $\widetilde{U}^q$;

(3) -*Attention*, which performs a simple sum function on P^f and Q^f to get $\widetilde{U}^p$ and $\widetilde{U}^q$.

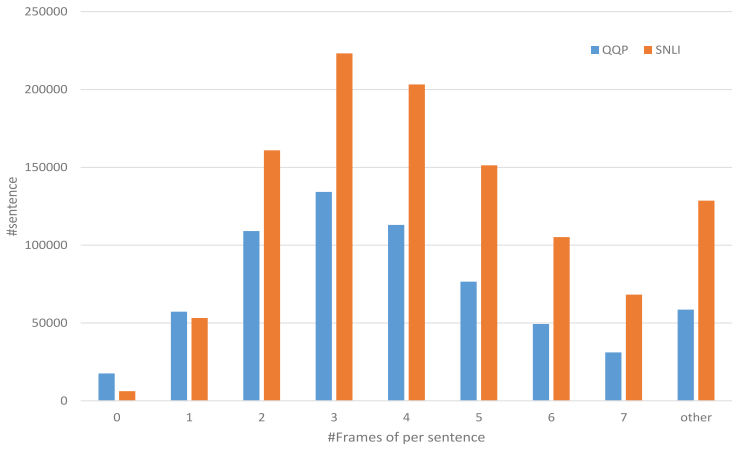

Fig. 6. Statistics of frames of per sentence on QQP.

Table 4
Comparison with different attention mechanisms on QQP.

Model	Dev	Test
Full model (MFSR)	91.60	91.59
-CoAttention	91.20	90.81
-SelfAttention	91.30	91.17
-Attention	90.62	90.03

Table 5
Comparison with different frame semantic information on QQP.

Model	Dev	Test
Full model (MFSR)	91.60	91.59
-Frame	91.07	90.72
-Frame elements	91.22	91.05
-Definition embedding	90.18	89.98

Table 4 shows the performance on the QQP. We observe both CoAttention and SelfAttention contribute to the overall performance. Comparing the three models with the Full Model, we can observe that no Attention hurts the performance most. Therefore, integrating multi-level frame semantic information with attention mechanism is important for achieving better performance.

Frame Semantic Information. Frame semantics are core part of our architecture. We conduct experiments to study how different parts of frame affects the performance of our method.

(1) -*Frame*, which only use frame elements, and frame are replaced with target word;

(2) -*Frame elements*, which only use frame, and frame elements are replaced with corresponding word in the sentence;

(3) -*Definition Embedding*, frame and frame element embeddings are initialized randomly with a uniform distribution between $[-1, 1]$.

The results shown in Table 5. The ablation results show that without richer semantic features, the performance degrade significantly. We observe both frame and frame elements contribute to the overall performance. The performance of random initialization embeddings degrades, indicating it is better to train the frame and frame element vectors instead of direct random initialization, which has been used in Frame-GBDT model.

4.4.5. Model complexity and statistical significance

Model Complexity. It is commonly observed that the performance of neural networks is highly dependent on the model complexity, which is measured by the model size (parameters) and computational consumption (FLOPs) [47]. Table 6 shows the complexity of the MFSR method as compared to the baselines on QQP.

Table 6
Parameters, FLOPs and accuracy for different models on QQP.

Model	Parameters(M)	FLOPs(G)	Accuracy (%)
Frame-GBDT(Our)	43.58	0.024	88.29
BERT-base(Our)	109.48	10.19	88.70
FMSR(BERT-base)	164.59	11.27	90.10
BERT-large(Our)	335.14	36.24	89.15
FMSR(BERT-large)	392.6	37.34	91.59

Table 7
Experiment results on QQP. FMSR achieves a statistically significant improvements compared to the baselines with $p < 0.01$ under t-test. Standard deviation is reported in the parentheses.

Model	ACC. (%)
Frame-GBDT	88.29 ± (0.19)
FMSR	91.59 ± (0.44)
p-value	0.0001
BERT-Large	89.15 ± (0.13)
FMSR	91.59 ± (0.44)
p-value	0.0003

From Table 6, we can observe that: (1) Our FMSR (BERT-large) outperforms all the other baselines and with the complexity close to BERT-Large, which is efficient and acceptable. In addition, our FMSR (BERT-base) outperforms the strong BERT-Large baseline with FLOPs being 3× smaller and parameter size being 2× smaller than BERT-Large, indicating the efficiency of our model. (2) Our proposed model, leveraging a more powerful backbone model with more parameters to model complicated relationships and build more accurate mapping functions, significantly improved its accuracy. For example, FMSR (BERT-base) achieves 90.10% accuracy on QQP and the network contains about 164.6M parameters. FMSR (BERT-large) contains 392.6M parameters and significantly improves the accuracy to 91.59%.

Statistical Significance. Furthermore, to verify statistical significance on accuracy difference between our framework and compared baseline frameworks (BERT-Large, Frame-GBDT), we perform statistical significance test using the t-test over 5 runs. Particularly, we compare our FMSR model with BERT-Large and Frame-GBDT, because BERT-Large achieves comparable results and Frame-GBDT is most related to our FMSR model. Table 7 reports the classification performance on three methods. We can see that FMSR achieves a statistically significant improvement over both the BERT-Large and Frame-GBDT baselines ($p < 0.01$).

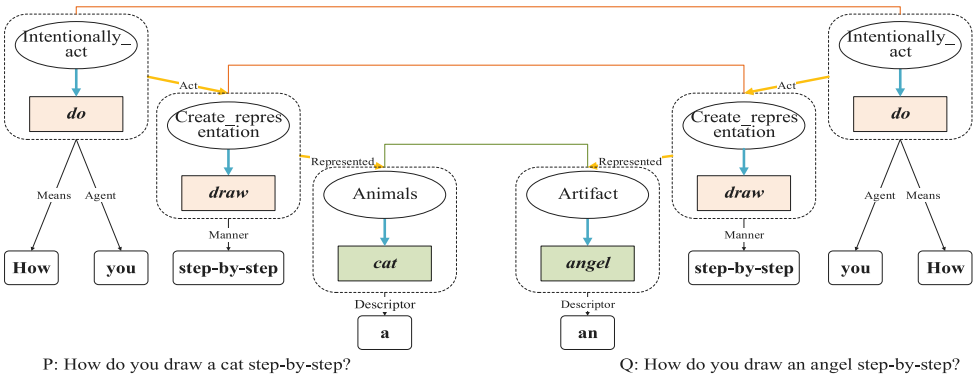

Fig. 7. An example of paraphrase identification.

4.4.6. Case study

In order to show the effectiveness of our model clearly, an example from QQP are shown in Fig. 7. (1) Frame-Semantic Role Labeler. We first process sentences with multiple frame semantic annotations. P and Q both have three frame semantic structures, i.e., $P^s = \{P^s_{Intentionally\_act}, P^s_{Create\_representation}, P^s_{Animals}\}$, $Q^s = \{Q^s_{Intentionally\_act}, Q^s_{Create\_representation}, Q^s_{Artifact}\}$. (2) Frame Semantic Structure Encoder. We then make use of bi-directional LSTMs to pre-process the frame semantic structures, $P^f = \{P^f_{Intentionally\_act}, P^f_{Create\_representation}, P^f_{Animals}\}$ and $Q^f = \{Q^f_{Intentionally\_act}, Q^f_{Create\_representation}, Q^f_{Artifact}\}$. (3) Context Structure Encoder. We further take full advantage of the attention mechanism to aggregate multi-level frame semantic structure representations of P and Q. Co-attention is built to learn the correlation between P^f and Q^f. Equipped with co-attention learning, we can obtain $U^P = \{U^P_{Intentionally\_act}, U^P_{Create\_representation}, U^P_{Animals}\}$ and $U^Q = \{U^Q_{Intentionally\_act}, U^Q_{Create\_representation}, U^Q_{Artifact}\}$, where $U^P_{Animals}$ and $U^Q_{Artifact}$ provide complementary information to each other. Self-attention is utilized to generate a better representation by emphasizing most important frames within a sentence, which helps the model compress U^P into $\widetilde{U}^q$ and U^Q into $\widetilde{U}^q$ and pays more attention on $U^P_{Animals}$ and $U^Q_{Artifact}$ respectively. (4) Answer Predictor. The target words *Do* and *Draw* in the given sentences belong to the *Intentionally_act* and *Create_representation* frame, while *Cat* and *Angel* evoke two very different frames *Animals* and *Artifact* respectively. Finally, utilizing the multi-level semantic information in FrameNet facilitates us to identify the relationship between the sentences.

Limitations. Although MFSR method performs well on some benchmark datasets, it still has some limitations.

(1) MFSR only focuses on modeling the frame and frame elements while ignoring Frame-to-Frame relations, and it is worth mentioning that FrameNet arranges relevant frames into a network by defining Frame-to-Frame relations.

(2) MFSR ignores the world knowledge of text, which can improve prediction performance.

(3) Our model is designed to evaluate text matching systems. And it is necessary to adapt our model to other language understanding systems.

5. Conclusion and future work

In this paper, we have introduced an innovative FMSR model, which, to our best knowledge, is the first work to take full advantages of frame, frame elements to model the multi-level semantic structure of sentences. We show, via extensive experiments, that our FMSR model achieves very competitive performance comparing with state-of-the-art methods (including those which have been enhanced by the latest BERT model) proposed for two text matching tasks, i.e. paraphrase identification and natural language inference. Ablation studies validate the effectiveness of co-attention and self-attention mechanisms, and frame semantic structure.

There are three interesting future research directions: (1) We will explore more advanced methods to improve the text matching performance, for example, capsule networks [31] and deep belief networks [48], which will accurately represent the frame relation and structure information. (2) It is desirable to further design methods that can leverage world knowledge to include content that is not covered by the text, inspired by related hybrid networks [32]. (3) Nevertheless, given the promising results that have been achieved in this paper, we hope our model can encourage researchers to leverage multi-level frame semantic knowledge to improve machine reading comprehension tasks.

CRediT authorship contribution statement

Shaoru Guo: Conceptualization, Methodology, Software, Writing – original draft. **Yong Guan:** Software, Investigation, Data curation. **Ru Li:** Supervision. **Xiaoli Li:** Writing – review & editing. **Hongye Tan:** Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was sponsored by the National Natural Science Foundation of China (No. 61936012, No. 61772324).

References

[1] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 2018, pp. 2227–2237.

[2] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2018, arXiv preprint arXiv:1810.04805.

[3] Y. Sun, S. Wang, Y. Li, S. Feng, H. Tian, H. Wu, H. Wang, Ernie 2.0: A continual pre-training framework for language understanding, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 8968–8975.

[4] W. Yin, H. Schütze, B. Xiang, B. Zhou, Abcnn: Attention-based convolutional neural network for modeling sentence pairs, Trans. Assoc. Comput. Linguist. 4 (2016) 259–272.
[5] Z. Wang, W. Hamza, R. Florian, Bilateral multi-perspective matching for natural language sentences, in: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, 2017, pp. 4144–4150, http://dx.doi.org/10.24963/ijcai.2017/579.
[6] P.K. Mudrakarta, A. Taly, M. Sundararajan, K. Dhamdhere, Did the model understand the question? 2018, CoRR abs/1805.05492. arXiv:1805.05492. URL http://arxiv.org/abs/1805.05492.
[7] Z. Zhang, Y. Wu, H. Zhao, Z. Li, S. Zhang, X. Zhou, X. Zhou, Semantics-aware BERT for language understanding, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 9628–9635.
[8] C.J. Fillmore, Frame semantics and the nature of language, Ann. New York Acad. Sci. 280 (1) (1976) 20–32, http://dx.doi.org/10.1111/j.1749-6632.1976.tb25467.x, arXiv:https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-6632.1976.tb25467.x. URL https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-6632.1976.tb25467.x.
[9] C.F. Baker, C.J. Fillmore, J.B. Lowe, The Berkeley FrameNet project, in: Proceedings of the 17th International Conference on Computational Linguistics, in: COLING '98, Association for Computational Linguistics, Stroudsburg, PA, USA, 1998, pp. 86–90, http://dx.doi.org/10.3115/980451.980860.
[10] S. Guo, R. Li, H. Tan, X. Li, Y. Guan, H. Zhao, Y. Zhang, A frame-based sentence representation for machine reading comprehension, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 891–896, http://dx.doi.org/10.18653/v1/2020.acl-main.83, URL https://www.aclweb.org/anthology/2020.acl-main.83.
[11] X. Zhang, X. Sun, H. Wang, Duplicate question identification by integrating FrameNet with neural networks, in: AAAI, 2018, pp. 6061–6068, URL https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16308.
[12] Z. Chen, H. Zhang, X. Zhang, L. Zhao, Quora question pairs, 2018.
[13] S.R. Bowman, G. Angeli, C. Potts, C.D. Manning, A large annotated corpus for learning natural language inference, 2015, CoRR abs/1508.05326. arXiv:1508.05326. URL http://arxiv.org/abs/1508.05326.
[14] S. Wan, M. Dras, R. Dale, C. Paris, Using dependency-based features to take the'para-farce'out of paraphrase, in: Proceedings of the Australasian Language Technology Workshop 2006, 2006, pp. 131–138.
[15] Z. Wang, A. Ittycheriah, Faq-based question answering via word alignment, 2015, arXiv preprint arXiv:1507.02628.
[16] B. Liu, X. Li, W.S. Lee, P.S. Yu, Text classification by labeling words, in: AAAI, 4, 2004, pp. 425–430.
[17] G. Zhou, Y. Liu, F. Liu, D. Zeng, J. Zhao, Improving question retrieval in community question answering using world knowledge, in: IJCAI, 2013, pp. 2239–2245, URL http://www.aaai.org/ocs/index.php/IJCAI/IJCAI13/paper/view/6581.
[18] Z. Ji, F. Xu, B. Wang, B. He, Question-answer topic model for question retrieval in community question answering, in: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, in: CIKM '12, Association for Computing Machinery, New York, NY, USA, 2012, pp. 2471–2474, http://dx.doi.org/10.1145/2396761.2398669.
[19] S.D. Gollapalli, X.-L. Li, P. Yang, Incorporating expert knowledge into keyphrase extraction, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 31, 2017.
[20] Y. Zhang, Y. Chang, X. Liu, S.D. Gollapalli, X. Li, C. Xiao, Mike: keyphrase extraction by integrating multidimensional information, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, Association for Computing Machinery, New York, NY, USA, 2017, pp. 1349–1358.
[21] W. Hu, A. Dang, Y. Tan, A survey of state-of-the-art short text matching algorithms, in: Y. Tan, Y. Shi (Eds.), Data Mining and Big Data. DMBD 2019. Communications in Computer and Information Science, Vol. 1071, Springer, Singapore, 2019, pp. 211–219, http://dx.doi.org/10.1007/978-981-32-9563-6_22.
[22] P.-S. Huang, X. He, J. Gao, L. Deng, A. Acero, L. Heck, Learning deep structured semantic models for web search using clickthrough data, in: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, in: CIKM13, Association for Computing Machinery, New York, NY, USA, 2013, pp. 2333–2338, http://dx.doi.org/10.1145/2505515.2505665.
[23] J. Guo, Y. Fan, Q. Ai, W.B. Croft, A deep relevance matching model for ad-hoc retrieval, in: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, in: CIKM '16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 55—64, http://dx.doi.org/10.1145/2983323.2983769.
[24] L. Pang, Y. Lan, J. Guo, J. Xu, S. Wan, X. Cheng, Text matching as image recognition, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 30, 2016.
[25] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, R. Shah, Signature verification using a" siamese" time delay neural network, Adv. Neural Inf. Process. Syst. 6 (1993) 737–744.
[26] S. Wang, J. Jiang, A compare-aggregate model for matching text sequences, 2016, arXiv preprint arXiv:1611.01747.
[27] Y. Tay, A.T. Luu, S.C. Hui, Co-stack residual affinity networks with multi-level attention refinement for matching text sequences, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 4492–4502, http://dx.doi.org/10.18653/v1/D18-1479, URL https://www.aclweb.org/anthology/D18-1479.
[28] C. Tan, F. Wei, W. Wang, W. Lv, M. Zhou, Multiway attention networks for modeling sentence pairs, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18, International Joint Conferences on Artificial Intelligence Organization, 2018, pp. 4411–4417, http://dx.doi.org/10.24963/ijcai.2018/613.
[29] R. Yang, J. Zhang, X. Gao, F. Ji, H. Chen, Simple and effective text matching with richer alignment features, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 4699–4709.
[30] G. Chen, D. Ye, Z. Xing, J. Chen, E. Cambria, Ensemble application of convolutional and recurrent neural networks for multi-label text categorization, in: Proceedings of the International Joint Conference on Neural Networks 2017-May, 2017, pp. 2377–2383.
[31] W. Zhao, H. Peng, S. Eger, E. Cambria, M. Yang, Towards scalable and reliable capsule networks for challenging NLP applications, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 2019, pp. 1549–1559, http://dx.doi.org/10.18653/v1/P19-1150.
[32] Y. Ma, H. Peng, T. Khan, E. Cambria, A. Hussain, Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis, Cogn. Comput. 10 (4) (2018) 639–650, http://dx.doi.org/10.1007/s12559-018-9549-x.
[33] H. Linmei, T. Yang, C. Shi, H. Ji, X. Li, Heterogeneous graph attention networks for semi-supervised short text classification, 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), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 4821–4830, http://dx.doi.org/10.18653/v1/D19-1488.
[34] Y. Li, Q. Pan, S. Wang, T. Yang, E. Cambria, A generative model for category text generation, Inform. Sci. 450 (2018) 301–315, http://dx.doi.org/10.1016/j.ins.2018.03.050.
[35] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, Improving language understanding by generative pre-training.
[36] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R.R. Salakhutdinov, Q.V. Le, Xlnet: Generalized autoregressive pretraining for language understanding, in: Advances in Neural Information Processing Systems, 2019, pp. 5754–5764.
[37] M. Palmer, D. Gildea, P. Kingsbury, The proposition bank: An annotated corpus of semantic roles, Comput. Linguist. 31 (1) (2005) 71–106.
[38] M. Joshi, D. Chen, Y. Liu, D.S. Weld, L. Zettlemoyer, O. Levy, Spanbert: Improving pre-training by representing and predicting spans, Trans. Assoc. Comput. Linguist. 8 (2020) 64–77.
[39] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, RoBERTa: A robustly optimized BERT pretraining approach, 2019, CoRR abs/1907.11692. arXiv:1907.11692. URL http://arxiv.org/abs/1907.11692.
[40] W. Wang, B. Bi, M. Yan, C. Wu, J. Xia, Z. Bao, L. Peng, L. Si, StructBERT: Incorporating language structures into pre-training for deep language understanding, in: International Conference on Learning Representations, 2020, URL https://openreview.net/forum?id=BJgQ4lSFPH.
[41] D. Das, D. Chen, A.F.T. Martins, N. Schneider, N.A. Smith, Frame-semantic parsing, Comput. Linguist. 40 (1) (2014) 9–56, http://dx.doi.org/10.1162/COLI_a_00163, arXiv:http://doi.org/10.1162/COLI_a_00163.
[42] M. Kshirsagar, S. Thomson, N. Schneider, J. Carbonell, N.A. Smith, C. Dyer, Frame-semantic role labeling with heterogeneous annotations, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Association for Computational Linguistics, Beijing, China, 2015, pp. 218–224, http://dx.doi.org/10.3115/v1/P15-2036, URL https://www.aclweb.org/anthology/P15-2036.
[43] T. Shen, T. Zhou, G. Long, J. Jiang, S. Pan, C. Zhang, DiSAN: Directional self-attention network for RNN/CNN-free language understanding, 2017, CoRR abs/1709.04696. arXiv:1709.04696. URL http://arxiv.org/abs/1709.04696.
[44] X. Liu, P. He, W. Chen, J. Gao, Multi-task deep neural networks for natural language understanding, 2019, CoRR abs/1901.11504. arXiv:1901.11504. URL http://arxiv.org/abs/1901.11504.

[45] Z. Zhang, Y. Wu, Z. Li, H. Zhao, Explicit contextual semantics for text comprehension, 2018, CoRR abs/1809.02794. arXiv:1809.02794. URL http://arxiv.org/abs/1809.02794.
[46] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: 3rd International Conference on Learning Representations, ICLR 2015, 2015.
[47] B. Wu, Efficient deep neural networks, 2019, arXiv preprint arXiv:1908.08926.
[48] I. Chaturvedi, Y.-S. Ong, I.W. Tsang, R.E. Welsch, E. Cambria, Learning word dependencies in text by means of a deep recurrent belief network, Knowl.-Based Syst. 108 (2016) 144–154, http://dx.doi.org/10.1016/j.knosys.2016.07.019.