IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 11, 2021, accepted April 9, 2021, date of publication April 21, 2021, date of current version April 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3074747

Leveraging Grammatical Roles for Measuring Semantic Similarity Between Texts

MD. ATABUZZAMAN [1], MD SHAJALAL [1], M. ELIUS AHMED [1], MASUD IBN AFJAL [1], (Graduate Student Member, IEEE), AND MASAKI AONO [2], (Member, IEEE)

[1]Department of Computer Science and Engineering, Hajee Mohammad Danesh Science and Technology University, Dinajpur 5200, Bangladesh
[2]Department of Computer Science and Engineering, Toyohashi University of Technology, Toyohashi 441-8580, Japan

Corresponding author: Md. Shajalal (shajalal@hstu.ac.bd)

This work was supported in part by the Aono Laboratory funding at Toyohashi University of Technology, Japan.

ABSTRACT Semantic similarity between texts can be defined based on their meaning. Assessing the textual similarity is a prerequisite in almost all applications in the field of language processing and information retrieval. However, the diversity in the sentence structure makes it formidable to estimate the similarity. Some sentences pairs are lexicographically similar but semantically dissimilar. That is why the trivial lexical overlapping is not enough for measuring the similarity. To attain the semanticity of sentences, the context of the words and the structure of the sentence should be considered. In this paper, we propose a new method for capturing the semantic similarity between sentences based on their grammatical roles through word semantics. First, the sentences are divided grammatically into different parts where each part is considered as a grammatical role. Then multiple new measures are introduced to estimate the role-based similarity exploiting word semantics considering the sentence structure. The proposed similarity measures focus on inter-role and intra-role similarity between the sentence-pair. The word-level semantic information is extracted from a pre-trained word-embedding model. The performance of the proposed method was verified by conducting a wide range of experiments on the SemEval STS dataset. The experimental results indicated the effectiveness of the proposed method in terms of different standard evaluation metrics and outperformed some known related works.

INDEX TERMS Semantic similarity, sentence structure, word-embedding, word semantics.

I. INTRODUCTION

Semantic Textual Similarity (STS) between sentences is indispensable and beneficial for Information Retrieval (IR) and Natural Language Processing (NLP) tasks. It is being used in many fields of NLP such as text summarization [1], [2], machine translation [3], paraphrase detection [4], [5], question-answering [6], dialog and conversational systems, sentiment analysis, and clinical information extraction [7]. There are some other applications such as relevance feedback, text classification [8], word sense disambiguation [9], subtopic mining, web search [10] and so on [11]–[13]. STS can be defined as a process that takes two sentences as input and returns a similarity score in the range [0,1] based on their meaning. When the sentence-pair is completely semantically similar, the score will be 1. On the other hand, the score will be 0 when the sentence-pair is dissimilar. The basis of similarity scoring for each sentence-pair is summarized in Table 1.

Traditionally, the similarity between texts is measured using the string matching technique. String matching is not good enough to capture the semantic similarity between sentence-pair. Because string matching cannot deal with the semanticity (meaning) of the text. However, some sentences are lexically similar but not semantically such as *"He appreciates your teacher"* and *"He is your teacher"*. Short sentences might carry very few contextual information of the words. Moreover, idiomatic phrases are often used in sentences that might change the meaning of the sentences in a different direction. For example, lets consider two sentences *Sentence_A* : *"The new car is not very expensive."* and *Sentence_B* : *"The new car costs **an arm and a leg**."* where *Sentence_B* contains an idiomatic phrase that leads the meaning of *Sentence_A* towards a different direction.

However, generally, sentence structures are very diverse. The semantic similarity measures without considering the

The associate editor coordinating the review of this manuscript and approving it for publication was Arianna Dulizia.

TABLE 1. The basis of similarity score for sentence-pair.

Basis	Score
Two sentences are completely equivalent, as they mean the same thing	100%
Two sentences are mostly equivalent, but some unimportant details differ	80%
Two sentences are roughly equivalent, but some important information differs/missing	60%
Two sentences are not equivalent, but share some details	40%
Two sentences are not equivalent, but are on the same topic	20%
Two sentences are completely dissimilar	0%

sentence structure may not be able to capture the original similarity beyond a trivial level. Because different parts of the sentence have different roles that reflect the meaning of the sentence. Therefore, the similarity between two parts from a particular sentence-pair that plays the same role should be measured. This paper investigates the grammatical role-based similarity measures to compute the semantic similarity between two sentences. Let's consider a sentence-pair, *"He is our English teacher"* and *"This is the man who teaches us English"*. The structure of these two sentences is different. But both sentences are semantically similar. These diversities make measuring semantic similarity a formidable task.

To tackle the sentences' structures diversity challenge, this paper proposes a new method that estimates the similarity considering grammatical structure through the semantic information of words. The hypothesis behind our method is that if two sentences are built up with similar grammatical roles and the parts from two sentences playing the same role are semantically similar then those two sentences might have the same meaning. To measure the similarity between intra-roles and inter-roles similarity, multiple new similarity functions are introduced with the help of word-level semantic information extracted exploiting word-embedding and WordNet. However, there are a considerable amount of sentences, especially on the web and social sites that are not written following the exact grammatical rules. Considering this, we also propose different new measures that are capable to compute the similarity using word-level semantic information without splitting the sentences into grammatical roles. The performance of our proposed measures has been validated by a wide range of experiments on the SemEval STS dataset. The experimental results demonstrated that the proposed method achieved effectiveness in measuring semantic textual similarity and outperformed some known related works. The results also concluded that the combination of the proposed similarity measures can boost up the performance of measuring similarity. The contributions of this research are listed below:

1) We proposed a novel *sentence structure-based* method to estimate the semantic similarity between texts. In this regard, we introduce a *novel algorithm* and *three new similarity measures* considering the grammatical structure of sentences.
2) Our proposed method achieved new state-of-the-art result and outperformed some related works. Moreover, our proposed measures can also be applied to estimate the similarity between texts in some other applications.

In the remainder of this paper, we discuss some related works on semantic textual similarity in section II. Then we present the proposed method in section III. The experiments and evaluation of the proposed method are presented and discussed in section IV. Finally, section V concluded our proposed method with some future directions.

II. RELATED WORKS

This section presents some prominent research works on semantic textual similarity. SemEval has organized different tasks for measuring the semantic textual similarity of monolingual and cross-lingual texts in recent years [14]–[19]. One of the best performing methods in SemEval STS2017, ECNU [20] leveraged kernel-based traditional features used in natural language processing tasks and feed them into neural networks for building a universal model for multilingual and cross-lingual sentences. Zhuang and Chang [21] proposed a method using an attention-based recurrent neural network to predict the degree of equivalence between texts. Duma and Menzel [22] suggested a knowledge-free approach that made use of paragraph vector to attain semantic similarity between sentence-pair. They applied three widely used classical features with different dimensions for estimating the similarity. Lee *et al.* [23] exploited the paraphrase and event-embedding with regression model to compute the textual similarity.

Word-embedding is widely being used for estimating semantic textual similarity. Kenter and Rijke [24] used word-embedding for measuring the semantic similarity of short texts. They extracted average high dimensional vector for each text and estimated the similarity employing cosine similarity. Shajalal and Aono [25] measured the semantic similarity by using word-embedding and WordNet. For word-embedding based measures, they considered the average vector and parts-of-speech (POS) tags of the words. They also combined word-embedding, WordNet, TF, etc. and ranked different measures according to the importance. Shajalal and Aono [26] also applied word-embedding for measuring the similarity of Bengali texts. They utilized the cosine similarity between the average vectors as the textual similarity. Tian and Lan [27] employed the sentence embedding and word-embedding to measure the semantic textual similarity. Some other works also used word-embedding for semantic similarity measurement such as [28], [29]. The mentioned related works did not consider the grammatical structure of the sentences to estimate the semantic similarity, but the grammatical structure holds different properties of the sentences that might be useful to compute the similarity.

Recently, Gazpio et al. [30] proposed an attention model using the word n-grams to measure the similarity between texts. Wang [31] demonstrated that a particular linguistic measure might have a different effect on different corpus due to the great difference in sentence structure and vocabulary. Adouane et al. [32] introduced an LSTM-based neural model to detect the binary similarity label. Brychcín [33] introduced a new transformation technique to project monolingual semantic spaces using a bilingual dictionary. Lenz et al. [34] proposed new supervised and unsupervised measures in the context of a graph-based similarity for argument graphs. BERT encoder is applied in a fully unsupervised cross-lingual semantic similarity measures for identifying parallel data [35]. Tien et al. [36] modeled the sentence leveraging multiple pre-trained word-embeddings and multi-level comparison. Hay et al. [37] automatically selected the complementary vector representation of the word. An unsupervised semantic textual similarity method is suggested by Hassan et al. [38]. They first used a synset-oriented word aligner that relied on a huge multilingual semantic network. They also proposed three unsupervised STS approaches, including string kernel-based (SK), alignment-based (AL), and weighted alignment-based (WAL) similarity. The sentence meta-embedding based technique is also employed to model the sentence similarity [39]. Nowadays, semantic consistency and cross-model attention is being used for image STS [40].

Various WordNet-based similarity measurements have been proposed by researchers. Mihalcea et al. [41] measured the semantic textual similarity considering the term-level similarity of corresponding texts. They employed different types of knowledge-based and corpus-based measures. Ferreira et al. [42] measured the semantic similarity considering the lexical, syntactical, and semantic features. WordNet is also being used for computing textual similarity from earlier in [43]–[46]. Some works considered the POS tags of each word in measuring the semantic similarity [25], [41]. They computed the similarity by employing the similarity of words having identical POS tags. They did not consider the contextual information of the words.

But in this paper, we used a word-embedding model to consider the contextual information of a word that reflects the meaning. This paper also employs role-based measures that might contain more semantic information of a sentence than word-level similarity. Many researchers also used WordNet to measure the similarity of similar kind of phrases [47]–[49]. Li et al. [47] proposed a method using a shallow parsing to divide each sentences into noun phrases, verb phrases and preposition phrases. Then they assessed the sentence similarity computing the phrase-level similarity. Oliva et al. [48] proposed a syntax-based measure that estimates the similarity between subject to subject, verb to verb, and object to object. The similarity between two sentences is then calculated as a sum of the similarities between the heads of these three phrases. Lee [49] introduced a two-phase algorithm where they computed similarity after categorizing the words into the noun and verb sets. The proposed method of this paper is different from them for the following reasons. Firstly, the proposed method used contextual information of the words using word-embedding to hold the meaning of the words, but they did not consider the contextual information to capture the meaning of the words. Secondly, the proposed inter-role and intra-role based similarity measures are employed to capture similarity considering grammatical roles but they only consider the intra-roles to measure the STS. Moreover, their method cannot be applicable to the grammatically unstructured sentences, but the proposed method is applicable to grammatically unstructured texts as some proposed measures do not consider the grammatical roles to estimate the STS.

III. PROPOSED METHOD

A sentence-pair might be semantically similar but they may be different in grammatical structure. Due to huge diversity in sentences structure, the task of measuring semantic similarity is much more challenging. We hypothesize that if two sentences have the same types of grammatical roles and they are semantically similar, then the similarity between those corresponding sentences is higher. The high-level overview of our proposed grammatical role-based measures is illustrated in Fig. 1. Given a pair of sentences, first, their grammatical structure is studied by splitting into different grammatical parts. Each part represents a grammar role of the sentence. Then, each role is preprocessed to filter unnecessary items and the lemmatizer is applied to find the root word. After that, the proposed inter-role and intra-role similarity measures are applied to find out the semantic similarity. In both cases, the word-level semantic information is exploited in the proposed similarity measures. Finally, all the variants of the proposed measures are combined to compute the semantic similarity between the sentence-pair.

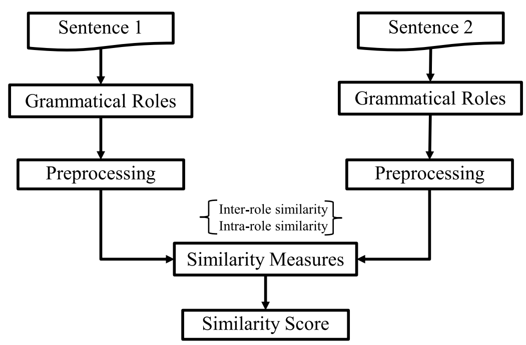

FIGURE 1. Overview diagram of our proposed role-based method.

A. SPLITTING SENTENCE INTO GRAMMATICAL ROLES

Let S_1 and S_2 be two sentences. Considering the grammatical structure of the sentences, the roles of the sentences are extracted. A sentence consists of different grammatical parts including subject, noun phrase (NP), verb phrase (VP), and object. A sentence can have more than one NP and VP. Generally, NP plays the role of the subject and object of

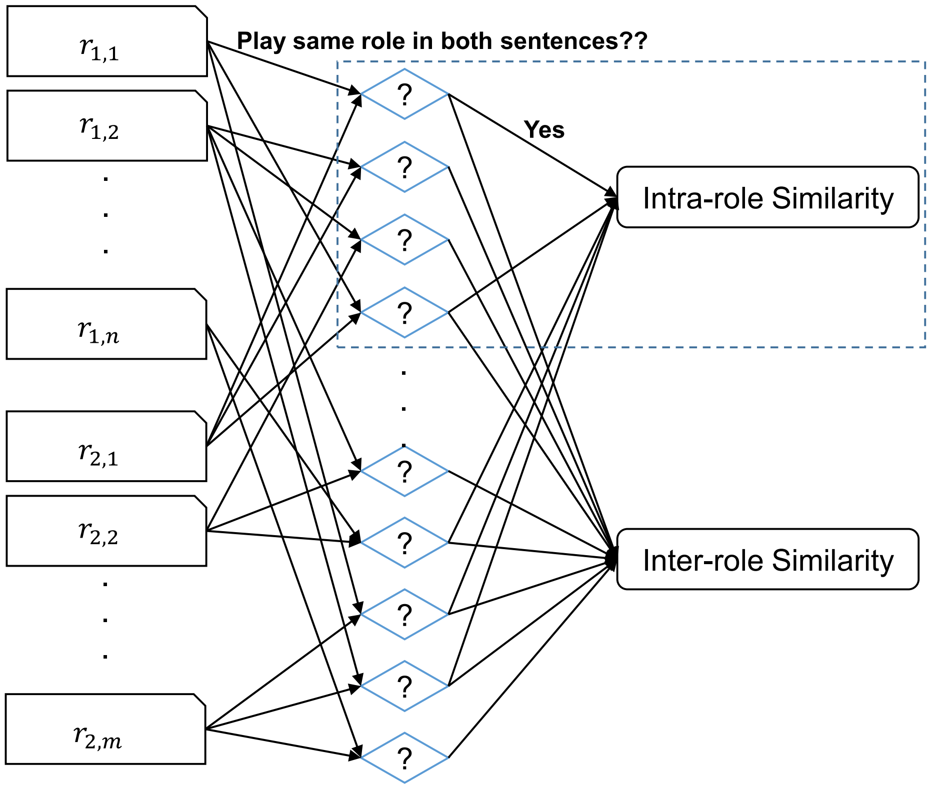

FIGURE 2. Strategy for inter-role and intra-role similarity.

a sentence. Consider a sentence for clarification, "*A poor fisherman is catching fishes*". Here, "*A poor fisherman*" is subject and "*fishes*" is object of this sentence. After splitting the sentence, the phrases are NPs: [NP_1: "A poor fisherman", NP_2:"fishes"] and VP: ["is catching"]. The sentences are split into NPs and VPs for similarity estimation. As noted earlier, this paper investigates the influence of grammatical roles to calculate the sentence similarity. According to our hypothesis, if the corresponding sentences share the same types of grammatical roles and the semantic similarity among the same roles is higher, then those two sentences might have higher similarity. Therefore, the sentences are split into different grammatical parts and preprocessed each part. In the preprocessing step, the punctuation marks and stopwords are removed. Here, stopwords mean the words which have very little impact on the meaning of a sentence. Then, the words are lemmatized using WordNet Lemmatizer to convert them into their base form.

Let $R1 = \{r_{1,1}, r_{1,2}, r_{1,3}, \ldots, r_{1,n}\}$ and $R2 = \{r_{2,1}, r_{2,2}, r_{2,3}, \ldots, r_{2,m}\}$ be the sets of grammatical roles extracted from the two corresponding sentences S_1 and S_2. The proposed inter-role and intra-role similarity measures are applied according to a specific strategy depicted in Fig. 2. When two grammatical parts from two different sentences play the same role then intra-role similarity measures are applied. Suppose two sentences contain NPs and VPs. Then our intra-role similarity is applied to similar phrases such as NP-NP and VP-VP. The sentence-pair might have multiple roles that are both similar and dissimilar. Then the inter-roles similarity measures are employed. For the inter-role similarity measure, each role of a sentence is compared with all the roles of the other sentence. Because NP of a sentence can play the role as NP or VP in another sentence. Therefore, NP-NP, NP-VP, and VP-VP similarity scores are estimated using the proposed inter-role based similarity measures.

B. SIMILARITY MEASURES

This paper proposed several similarity measures to compute the similarity semantically. Considering the grammatical structure of the sentences, multiple measures are introduced based on the particular roles in the sentence-pair.

1) ROLE-BASED MAXIMUM SIMILARITY (RM_Sim)

As we noted earlier, one particular role of a sentence might play a different role in another sentence. Let's consider an example: S_1: "Walking is good for people's health." and S_2: "People are walking for good health." in this sentence-pair "walking" plays the role of NP and VP in S_1 and S_2, respectively. For this kind of scenarios, a system might need approaches that can assess the inter-role and intra-role similarity. In this measure, an inter-role based similarity estimation algorithm is introduced. The maximum role-based similarity algorithm is summarized in Algo. 1.

Considering the sentences structure, the grammatical roles are extracted and stored in lists. The list which has maximum number of roles is considered as main list for

Algorithm 1: Maximum Role-Based Similarity **MRole_Sim**(R_1,R_2)

Input: Sets of roles R_1 and R_2 for two sentences S_1 and S_2
Output: Semantic Similarity Score
Initialisation :
1: $n_1 \leftarrow |R_1|$; // Number of roles in Sentence 1
2: $n_2 \leftarrow |R_2|$; // Number of roles in Sentence 2
3: $R_{min}, R_{max} \leftarrow max(n_1, n_2)$; // finding maximum roles
4: $l_{r_{max\_sim}} \leftarrow []$; // empty list
LOOP Process
5: **for** each $r_i \in R_{max}$ **do**
6: $l_{r_{sim}} \leftarrow []$; // empty list
LOOP Process
7: **for** each $r_i \in R_{max}$ **do**
8: $\vec{\mathbf{s}}_{v1} \leftarrow resource(get\_phrase(r_i))$; // semantic vector for role
9: $\vec{\mathbf{s}}_{v2} \leftarrow resource(get\_phrase(r_j))$; // semantic vector for role
10: $r_{sim} \leftarrow compute\_sim(\vec{\mathbf{s}}_{v1}, \vec{\mathbf{s}}_{v2})$
11: $l_{r_{sim}} \leftarrow Append(l_{r_{sim}}, r_{sim})$
12: **end for**
13: $r_{max\_sim} \leftarrow max(l_{r_{sim}})$
14: $l_{r_{max\_sim}} \leftarrow Append(l_{r_{max\_sim}}, r_{max\_sim})$
15: **end for**
16: $s\_score \leftarrow 0$
17: $tc \leftarrow 0$
LOOP Process
18: **for** each $sim \in l_{r_{max\_sim}}$ **do**
19: **if** $sim \geqslant max(l_{r_{max\_sim}})/2$ **then**
20: $s\_score \leftarrow s\_score + sim$
21: $tc \leftarrow tc + 1$
22: **end if**
23: **end for**
24: $similarity\_score \leftarrow \frac{s\_score}{tc}$
25: **return** $similarity\_score$

similarity measurement. The inter-role similarity is then estimated by exploiting the semantic information from the pre-trained word-embedding model. If any role contains more than one word, all the corresponding vectors of the words are summed up. The similarity between the two roles is then computed by applying cosine distance. The similarities of one role of a sentence with all roles of other sentence are computed. In every iteration, the maximum similarity of one role of a sentence to all other roles in another sentence is stored (statement 14). Statements [5-15] summarize the above processes. There might some roles that are not similar anymore between sentence-pair and as a result their similarity score can be smaller. These smaller scores may influence the final similarity score. Therefore, only those scores having higher than half of their maximum are considered. Finally, the average of the selected scores is used as the role-based maximum similarity (statement 18 to 24).

2) MAXIMUM WORD-LEVEL ROLE-BASED SIMILARITY (MWLR_Sim)

This is an intra-role similarity measure. The similarity between roles having the same action in both sentences is estimated. In other words, when two phrases from two corresponding sentences are similar in type (both are noun phrase or verb phrase), they will be considered as roles having the same action. The intuition behind this measure is that if the words from two roles share the similar semantic information, they might be similar semantically. Therefore, the word level semantic contextual information is exploited for measuring the similarity. To capture the contextual information of words belong to each role, the word-level semantic information is extracted from word-embedding model. If there are more than one roles having same type in a sentence, they are combined into one role. The contribution of NP and VP in the meaning of a sentence might be different. That is why a weight α is employed to emphasize the particular intra-role similarity.

$$sim_{maxR}(S_1, S_2) = \alpha \; sim_{maxR_{NP}}(R_{NP_{S_1}}, R_{NP_{S_2}}) + (1-\alpha) \; sim_{maxR_{VP}}(R_{VP_{S_1}}, R_{VP_{S_2}}) \quad (1)$$

where α indicates the weighting threshold ranges in [0,1]. The threshold gives priority to one grammatical role and also demoting the other. The value of α is selected empirically by conducting experiments. $sim_{max_{R_{NP}}}(R_{NP_{S_1}}, R_{NP_{S_2}})$ indicates the intra-role similarity for NP is defined as the following:

$$sim_{max_{R_{NP}}}(R_{NP_{S_1}}, R_{NP_{S_2}}) = \frac{1}{2}\left[\sum_{w\in R_{NP_{S_1}}} maxSim(w, R_{NP_{S_2}}) + \sum_{w\in R_{NP_{S_2}}} maxSim(w, R_{NP_{S_1}})\right]$$

The definition of $sim_{max_{R_{VP}}}(R_{VP_{S_1}}, R_{VP_{S_2}})$ is analogous as above equation. The function $maxSim(w, R_{NP_{S_2}})$ returns the maximum similarity of a word of S_1 from role $R_{NP_{S_1}}$ with respect to all the words of same type of role $R_{NP_{S_2}}$ from S_2 (analogous for $maxSim(w, R_{NP_{S_1}})$).

3) AVERAGE VECTOR-BASED GRAMMATICAL ROLES SIMILARITY (AVR_Sim)

This measure attempts to capture the similarity of grammatical roles using feature vector. To do this, the common grammatical roles of sentences S_1 and S_2 are utilized. The average vector of words employing pre-trained word-embedding model is estimated for same kind of roles separately. Then the similarity of the same kinds of roles of the sentences is estimated as the following equation.

$$sim_{avgR}(S_1, S_2) = \alpha\ Cosine(\vec{v}_{R_{NP_{S_1}}} \cdot \vec{v}_{R_{NP_{S_2}}}) + (1-\alpha)\ Cosine(\vec{v}_{R_{VP_{S_1}}} \cdot \vec{v}_{R_{VP_{S_2}}}) \quad (2)$$

where $\vec{v}_R$ denotes the average vector of a particular role of a sentence. The dot product returns the cosine similarity of two roles.

4) PATH BASED ROLE-SIMILARITY (PR_Sim)

Path similarity in WordNet is calculated based on shortest path between two words' in the ontology tree. The smaller the path length between nodes (words) in the ontology tree, the larger the similarity. If multiple roles belong to the same type of a sentence, they are combined into one role. The path similarity is exploited to find out the role-based similarity as follows:

$$sim_{pathR}(S_1, S_2) = \alpha\ sim_{R_{NP}}(R_{NP_{S_1}}, R_{NP_{S_2}}) + (1-\alpha)\ sim_{R_{VP}}(R_{VP_{S_1}}, R_{VP_{S_2}}) \quad (3)$$

where,

$$sim_{R_{NP}}(R_{NP_{S_1}}, R_{NP_{S_2}}) = \frac{1}{|w|}\sum_{w\epsilon R_{NP_{S_1}}} max(path\_sim_{v\epsilon R_{NP_{S_2}}}(w, v))$$

where $max(path\_sim_{v\varepsilon R_{NP_{S_2}}}(w, v))$ returns maximum score denoting how similar two word senses are based on the shortest path that connects the senses in the (hypernym / hypnoym) taxonomy of two words from same kind of grammatical roles in both sentences. The definition of $sim_{R_{VP}}(R_{VP_{S_1}}, R_{VP_{S_2}})$ is analogous as the above Eq. 3.

However, there are some sentences on the web that do not have specific grammatical structures. That is why it is quite difficult to identify the grammatical roles accurately of the unstructured sentences. To overcome this drawback, we propose sentence similarity estimation approaches that do not consider the grammatical structure of the sentence. The semantics of a sentence depend on the words. The word-level similarity between two sentences has an impact on the similarity measurement. When two words have the same context, their meaning might be similar. Some proposed measures can be applied without considering the grammatical roles. The maximum word-level similarity as like as maximum role-level similarity is applied to estimate the sentences similarity. The similarity of the sentences using their average feature vectors is calculated as like as our average vector-based role-level similarity. In that case, the weighting threshold α and the role-specific similarity are not applied. But the word-to-word similarity is utilized. The efficiency of similar kinds of measures of without considering role-level similarity has been validated elsewhere in [25], [41].

IV. EXPERIMENTS AND EVALUATION

Multiple experiments are carried out on the STS-2017 dataset to evaluate the performance of our proposed method in terms of Pearson's correlation coefficient (r) and Spearman's rank correlation coefficient (ρ). The next subsections present the dataset, evaluation metrics, experimental setup, experimental results, and performance comparison and discussion.

A. DATASET

The experiments are carried out on a benchmark dataset to evaluate the performance of the proposed method. The dataset is collected from the SemEval 2017 Task 1. SemEval 2017 task 1 dataset was collected from Stanford Natural Language Inference (SNLI) corpus. There are 250 sentence pairs in the dataset. In the dataset, some sentences are too long and some sentences are very short. There is a long sentence along with a short sentence in some sentence-pairs. Each sentence of this dataset consists of an average of 8.7 words. This dataset has an average similarity score of 2.3 for the sentence-pair [19]. The organizers provided a similarity score for each sentences pairs, which was estimated by the human assessors' judgment. The human assessors have given the similarity score considering semantic labels depicted in Table 2 in the range of 0 to 5. The details of similarity labels are depicted in Table 2 and explained elsewhere in [19], [25]. Word2vec[1] is used as a pre-trained word-embedding model. Word2vec is able to capture contextual information of words and this contextual information is essential to measure the STS of a text. Because the contextual information reflects the meaning of the word. The effectiveness of word2vec as word-embedding model to capture the contextual information is shown by some works [24], [25], [50]. The used word2vec was pre-trained on Google News Corpus [50]. The word2vec has 300 dimensional vector and was trained on 3,000,000 English vocabularies. Sentences are split into grammatical roles using the python spaCy library. Using the spaCy Chunking process,

[1] Word2Vec: https://code.google.com/p/word2vec/

TABLE 2. Labels of similarity score.

Label	Score
On different topics	0
Not similar but share few common details	1
Not similar but share some common details	2
Roughly similar	3
Similar	4
Completely similar	5

the grammatical roles mainly NPs and VPs of the sentences are extracted. NLTK tokenizer is applied to tokenize the sentences into words. To remove the valueless words (stopwords) from the word list, NLTK stopwords are used.

B. EVALUATION METRICS

The performance of the proposed method has been tested in terms of different evaluation metrics including *Pearson Correlation Coefficient*[2] (r) and *Spearman Rank Correlation Coefficient*[3] (ρ). Pearson correlation coefficient is the official metric to test the performance of the methods in SemEval STS2017 [19]. Given that $X = \{x_1, x_2, x_3 \ldots x_n\}$ and $Y = \{y_1, y_2, y_3 \ldots y_n\}$ be the two sets of scores for n sentence-pairs estimated by the system and gold-standard, respectively. Each x_i (or y_i) in X (or in Y) indicates the semantic similarity of i-th sentence-pair. The higher the value of correlation coefficients (both Pearson's (r) and Spearman's (ρ)) between X and Y is, the better the system is.

C. EXPERIMENTAL SETUP

The experiments are carried out in multiple experimental settings to validate the performance of the proposed method. The experiments are divided into two types based on the similarity measures. First, the measures considering the grammatical structure are applied. Then the introduced measures also applied without splitting the sentence into grammatical roles. Finally, all the variants of the proposed measures are combined to obtain the similarity score. The experimental setup is summarized below.

1) SENTENCE STRUCTURE-BASED SIMILARITY

The proposed inter-role and intra-role based similarity measures are applied to calculate the similarity.

- **RM_sim:** The proposed algorithm 1 that estimates the role-based maximum similarity is applied in this setting.
- **MWLR_Sim:** In this setting, the maximum word-level role-based similarity measure is employed (Eq. 1).
- **AVR_Sim:** The average feature vector based grammatical roles similarity (Eq. 2) is applied in this experimental setting.
- **PR_sim:** Utilizing the WordNet path similarity, this setting utilized the path-based role-similarity (Eq. 3).

[2]https://en.wikipedia.org/wiki/Pearson_correlation_coefficient
[3]https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient

2) SIMILARITY WITHOUT GRAMMATICAL STRUCTURE

The proposed measures are applied here without splitting the sentence into grammatical roles.

- **MWL_Sim:** The maximum word-level similarity (Eq. 1) except considering the grammatical roles is applied in this setting.
- **AV_Sim:** Similarly, the average feature vector-based similarity is applied in this setting ignoring the grammatical structure (Eq. 2).

3) COMBINATION

The combination of different settings are applied in this part.

- **Com_GS:** This setting applied the arithmetic average of the above four grammatical structure-based similarity measures *MWLR_Sim*, *RM_sim*, *AVR_Sim* and *PR_sim*.
- **Com_WGS:** This is the arithmetic average of measures ignoring the grammatical structure *MWL_Sim* and *AV_Sim*.
- **Com_ALL:** In this setting, the arithmetic average of *Com_WGS* and *Com_GS* is employed.

4) BASELINE

The performance of different experimental settings are compared with two baselines. The first baseline *Cos_Sim* is the cosine of binary vectors of the sentences, where each dimension represent whether an individual word appears in that sentence or not. In other words, it is the lexical matching based on terms overlap [19]. The second baseline *WE_GN* employed a word-embedding (word2vec) based approach to measure STS and the word2vec was pre-trained on Google News Corpus [25]. They also used parts-of-speech (POS) tags to estimate word-level similarity and average feature vectors corresponding to each sentence. The first one is used as a baseline by SemEval 2017 task 1 organizers [19] and the second one is a word-embedding-based method introduced by [25].

5) PARAMETER SETTING

The value of the weighting threshold, α in the proposed measures (Eq. 1, 2, 3) is selected empirically. Experiments are conducted using the different values of α ranging from 0 to 1 by an interval of 0.05. The best experimental results are selected when the value of α is 0.75. This selected empirical value of the threshold weighting parameter α also concluded that the importance of a noun phrase in estimating sentence similarity is more than the importance of the verb phrase.

D. EXPERIMENTAL RESULTS

The performance of all the proposed approaches is presented in Table 3. This table indicates that the proposed grammatical role-based measures can capture the semantics similarity of the sentence-pair. Among all proposed grammatical role-based similarity measures, *MWLR_Sim* (Eq. 1) achieved better performance. It considered the maximum word-level similarity of all intra-roles of the sentence-pair.

TABLE 3. Performance of our proposed measures in terms of Pearson's $r * 100$ and Spearman's $\rho * 100$ on STS-2017 dataset. Best result is in bold.

Techniques	Measures	Pearson's $r * 100$	Spearman's $\rho * 100$
Baseline	Cos_Sim [19]	53.7	54.3
	WE_GN [25]	65.09	66.13
Considering Role	$MWLR\_Sim$	75.15	76.89
	RM_sim	73.28	74.34
	AVR_Sim	71.06	73.45
	PR_sim	72.44	73.40
Without Role	MWL_Sim	79.18	80.02
	AV_Sim	77.37	79.82
Combination	Com_GS	77.93	79.06
	Com_WGS	79.70	81.33
	Com_ALL	**80.42**	**81.51**

Role-based maximum similarity *RM_sim* calculated by the proposed algorithm (Algo. 1) did a promising performance. For *RM_sim*, every type of grammatical role of the sentence pairs is considered. This measure considered each roles' contribution to the semantic similarity of the sentences. Hence, it is capable to extract the semantic information in role-level similarity. When the sentences are written according to the grammatical structure, *RM_sim* can capture better similarity between texts. Similarly, the other two role-based similarities *AVR_Sim* and *PR_Sim* which utilized the average feature vector extracted from word-embedding (Eq. 2) and WordNet path similarity (Eq. 3), respectively showed promising performance. Arithmetic average of all proposed role-based measures, *Com_WGS* performed better than any individual role-based measures.

Table 3 also demonstrates that, the proposed similarity measures without considering the grammatical structure achieved better results in terms of both evaluation metrics than the earlier ones. Considering the contextual information of words, the average vectors of the sentence-pair stay closed in the vector space if they are semantically similar. That is why *AV_Sim* is able to hold the sentence meaning of short sentences more accurately. For *MWL_Sim*, both sentences are considered as the main sentence to calculate the maximum word-to-word level similarity. For this reason, every word's influence is received on the semantic similarity of the sentence-pairs. That is why *MWL_Sim* shows satisfactory and better performance. *Com_WGS* which is the combination of proposed two measures without considering the grammatical roles also achieved the satisfactory performance. Some sentences are too short and they may not be properly structured with grammatical roles. For these kinds of sentence-pairs, the proposed measures without considering the grammatical roles performed better. However, the combination of all the measures *Com_ALL* showed the best performance, because it considered all types of sentences and the impact of grammatical roles and words.

To visualize the strength and weakness of the proposed method, some example sentence-pairs are presented in Table 4 with their estimated similarity score as well as the gold-standard (GS) similarity. The table reflects that the proposed method successfully returned the nearest similarity score as compared to the GS similarity score. The table also indicates that the grammatical structures varied widely across the sentences. When the sentence-pairs consists of many stop words, the proposed method shows weakness to estimate the similarity score. When the short sentence-pair consists of different words that are very closed in semantic space, the proposed method failed to measure the similarity score; such as for example 14. But the shortness of sentences-pairs is tackled when the sentences consist of similar words: such as for example 15. Hence, the similarity measures ignoring the grammatical structure might lose some semantic information. But extracting the grammatical roles by splitting the sentences and estimating the similarity through the proposed intra-role and inter-roles measures can boost up the model at predicting the semantic similarity.

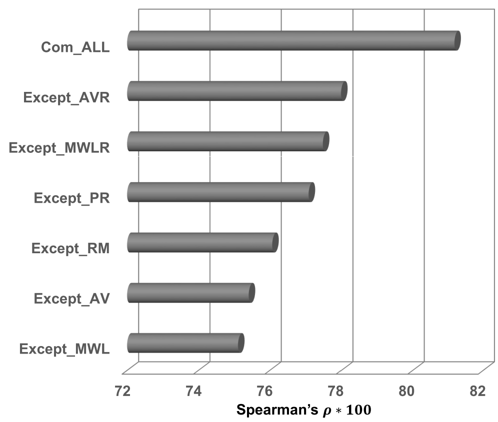

FIGURE 3. The performance our method except particular feature in terms of Spearman's $\rho * 100$.

E. CONTRIBUTION OF INDIVIDUAL PROPOSED MEASURES

To visualize the individual contribution of the proposed similarity measures considering the structure of grammatical roles, we conduct experiments combining all measures except the proposed ones. First, we drop our proposed *RM_Sim* from the combination and denote this setting as *Except_RM*. Similarly, we drop each role-based similarity measure from the combination and conduct the experiments. The performance based on the experimental results is illustrated in Fig. 3. The figure reflects that the contribution and performance of the

TABLE 4. Comparative performance of our proposed method with gold-standard (GS) score on example sentence-pairs.

SN	Sentence 1	Sentence 2	Our Score	GS Score
1	Some cyclists stop near a sign.	Two men stop to talk near a sign outside.	3.25	3.4
2	These cooks in the white are busy in the kitchen making dinner for their customers.	The women are preparing dinner in their kitchen.	3.1	3.0
3	A woman is bungee jumping.	A girl is bungee jumping.	4.35	4.2
4	A person filming the outdoors.	There is a couple outdoors.	1.5	1.6
5	A person sitting in an office takes a picture.	A woman is sitting in an office.	3.3	3.2
6	A woman paints a picture of a large building which can be seen in the background.	A person paints a picture of a large building which can be seen in the background.	4.15	4.2
7	A boy is at school taking a test.	The boy is taking a test at school.	5.0	5.0
8	Zombies parading around eating brains.	The zombies are eating flesh together.	3.3	3.4
9	A crowd of men wearing paper numbers on their shirts run in a race.	Several men running down a grass field wearing numbers on the front of their shirts.	3.55	3.6
10	The people are running a marathon.	People are running a marathon.	5.0	5.0
11	The woman is kneeling next to a cat.	A girl is standing next to a man.	2.85	0.2
12	There is a young girl.	There is a young boy with the woman.	4.20	1.0
13	A man is throwing a penny into a fountain.	A little boy is throwing a man in water.	3.15	0.6
14	The gate is blue.	The gate is yellow.	4.35	1.6
15	Man on steps.	A man sits on steps.	4.1	4.0

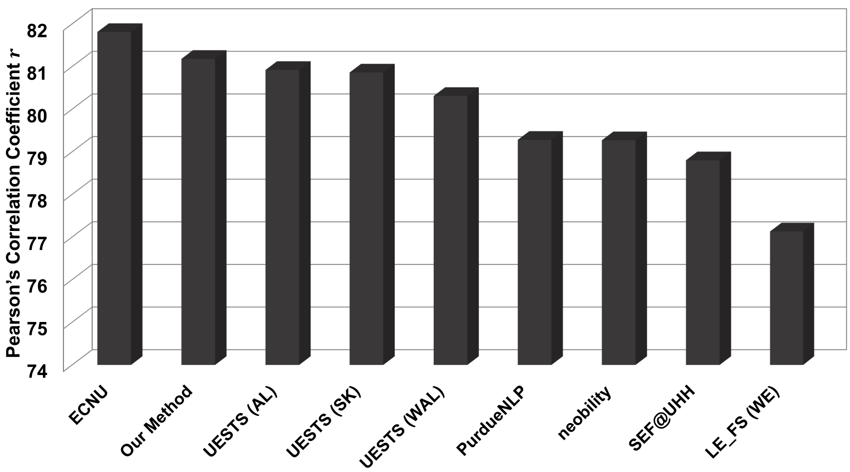

FIGURE 4. The performance comparason (in term of Pearson's correlation coefficient *r*) of the proposed method with some known related works [21]–[23], [25], [38].

proposed measures are varied widely. We can see that after dropping the proposed measure *MWL_Sim*, the performance of *Except_MWL* is lower than any other setting. Hence, the proposed *MWL_Sim* measure contributed more as compared to any other measures. Similarly, the proposed measure *RM_Sim* that considered the grammatical roles of sentences contributed effectively in estimating the similarity. More precisely, the performance of the *Except_RM* decreased almost 5% after dropping role-based maximum similarity *RM_Sim* (Algo. 1).

F. PERFORMANCE COMPARISON AND DISCUSSION

The performance of the proposed method is compared with some known related methods that conducted experiments on the same dataset and used the same evaluation metrics. The comparison is presented in Fig. 4. This figure depicts that the grammatical role-based similarity method outperformed some other known methods except ECNU [20]. ECNU applied cross-lingual textual similarity measures that employed a deep learning framework to utilize a wide range of features. This can be a plausible reason for the superiority of that method. However, the grammatical role-based method still performed almost equally though it relied only on some classical resources. Neobility [21] proposed an attention-based recurrent neural network model for learning the semantic similarity. But our proposed method outperformed neobility. The proposed method also does better performance than a word alignment-based unsupervised method (UESTS) with different variants including string kernel-based (SK), alignment-based (AL), and weighted alignment-based (WAL) similarity [38]. In terms of the pearson's correlation coefficient, the proposed method achieved better accuracy than a regression-based model (PurdueNLP) [23]. The computational cost of PurdueNLP is also higher than the proposed method. This paper used average feature vector based method [25] as one of the baselines (*WE_GN* of Table 3). The proposed method outperformed the baseline because it employed role-level similarity with the word-level semantics that might capture better semantic information than the word-level similarity. The combination of grammatical role-based and without grammatical role-based measures can capture better similarity semantically than SEF@UHH [22] which applied different classical similarity measures. The performance comparison and the consistency in achieving better performance conclude the superiority of the proposed method. However, the proposed method considered the grammatical roles of each sentence and leveraged inter-role and intra-role similarity in the new proposed measures. Therefore, the proposed method achieved new state-of-the-art results in measuring the semantic textual similarity.

V. CONCLUSION AND FUTURE DIRECTIONS

This paper proposed a new method to estimate the degree of equivalence in terms of meaning between a pair of sentences considering the sentence structure. The sentences are divided into different grammatical parts that indicate the particular roles in the sentence. The role-wise semantic similarity score is investigated to estimate the overall sentence similarity. In this regard, the intra-role similarity measures are proposed that applied in the grammatical parts from corresponding sentences which share the common grammatical roles namely noun phrase and verb phrase. The inter-role similarity measures are also introduced with the help of word's semantics extracted from different resources including word-embedding and WordNet. The proposed similarity measures can also be applied without splitting the sentence grammatically. In total, this paper introduce four new similarity measures considering the sentence structure.

The experimental results indicated the effectiveness of the proposed method and outperformed some known related works. The proposed similarity measures individually performed effectively. But combining all the introduced measures, the best performance is obtained in terms of two different official evaluation metrics. The most plausible reason behind this is, it considered both grammatical roles level similarity and word-level measures for assessing the similarity. The experiments on the weighting threshold α indicated that the contribution of noun-phrase based similarity is more than the importance of the verb-phrase at estimating sentence similarity. However, the contribution of the intra-role and inter-role similarity measures is validated in this study, and the difference are significant in percentage. The proposed method shows some limitations when sentence pairs contain many stopwords. Moreover, the findings from the comparative performance analysis indicated that this proposed method can capture better semantic similarity compared to some prominent research works. The proposed role-based similarity measures and algorithm helped to improve the effectiveness in measuring semantic textual similarity.

In the future, the proposed method will be applied to predict semantic textual similarity of cross-lingual texts. The multi-lingual word-embedding can be an interesting resource to be applied in the same purpose. The proposed method will also be applied in some other applications of NLP and IR including query suggestion generation, query disambiguation, document summarization, paraphrase, and plagiarism detection.

REFERENCES

[1] C.-Y. Lin and E. Hovy, "Automatic evaluation of summaries using N-gram co-occurrence statistics," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol. (NAACL)*, 2003, pp. 150–157.

[2] M. Mohd, R. Jan, and M. Shah, "Text document summarization using word embedding," *Expert Syst. Appl.*, vol. 143, Apr. 2020, Art. no. 112958.

[3] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: A method for automatic evaluation of machine translation," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002, pp. 311–318.

[4] S. Fernando and M. Stevenson, "A semantic similarity approach to paraphrase detection," in *Proc. 11th Annu. Res. Colloq. UK Special Interest Group Comput. Linguistics*, 2008, pp. 45–52.

[5] B. Dolan, C. Quirk, and C. Brockett, "Unsupervised construction of large paraphrase corpora: Exploiting massively parallel news sources," in *Proc. 20th Int. Conf. Comput. Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2004, p. 350.

[6] D. Lin and P. Pantel, "Discovery of inference rules for question-answering," *Natural Lang. Eng.*, vol. 7, no. 4, pp. 343–360, Dec. 2001.

[7] Y. Wang, N. Afzal, S. Fu, L. Wang, F. Shen, M. Rastegar-Mojarad, and H. Liu, "MedSTS: A resource for clinical semantic textual similarity," *Lang. Resour. Eval.*, vol. 54, no. 1, pp. 57–72, Mar. 2020.

[8] J. Rocchio, "Relevance feedback in information retrieval," in *The Smart Retrieval System-Experiments in Automatic Document Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1971, pp. 313–323.

[9] H. Schütze, "Automatic word sense discrimination," *Comput. Linguistics*, vol. 24, no. 1, pp. 97–123, 1998.

[10] M. Shajalal and M. Aono, "Coverage-based query subtopic diversification leveraging semantic relevance," *Knowl. Inf. Syst.*, vol. 62, no. 7, pp. 2873–2891, Jul. 2020.

[11] H. Li and J. Xu, "Semantic matching in search," *Found. Trends Inf. Retr.*, vol. 7, no. 5, pp. 343–469, 2014.

[12] D. Metzler, S. Dumais, and C. Meek, "Similarity measures for short segments of text," in *Proc. Eur. Conf. Inf. Retr.* Berlin, Germany: Springer, 2007, pp. 16–27.

[13] M. Shajalal, M. Z. Ullah, A. N. Chy, and M. Aono, "Query subtopic diversification based on cluster ranking and semantic features," in *Proc. Int. Conf. Adv. Inform., Concepts, Theory Appl. (ICAICTA)*, Aug. 2016, pp. 1–6.

[14] E. Agirre, D. Cer, M. Diab, and A. Gonzalez-Agirre, "SemEval-2012 task 6: A pilot on semantic textual similarity," in *Proc. 1st Joint Conf. Lexical Comput. Semantics, Main Conf. Shared Task, 6th Int. Workshop Semantic Eval. (SemEval)*, vol. 1. Montréal, QC, Canada: Association for Computational Linguistics, Jun. 2012, pp. 385–393.

[15] E. Agirre, D. Cer, M. Diab, A. Gonzalez-Agirre, and W. Guo, ''SEM 2013 shared task: Semantic textual similarity,'' in *Proc. 2nd Joint Conf. Lexical Comput. Semantics (SEM), Main Conf. Shared Task, Semantic Textual Similarity*, vol. 1. Atlanta, GA, USA: Association for Computational Linguistics, Jun. 2013, pp. 32–43.

[16] E. Agirre, C. Banea, C. Cardie, D. Cer, M. Diab, A. Gonzalez-Agirre, W. Guo, R. Mihalcea, G. Rigau, and J. Wiebe, ''SemEval-2014 task 10: Multilingual semantic textual similarity,'' in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*. Dublin, Ireland: Association for Computational Linguistics, Aug. 2014, pp. 81–91.

[17] E. Agirre, C. Banea, C. Cardie, D. Cer, M. Diab, A. Gonzalez-Agirre, W. Guo, I. Lopez-Gazpio, M. Maritxalar, R. Mihalcea, G. Rigau, L. Uria, and J. Wiebe, ''SemEval-2015 task 2: Semantic textual similarity, English, Spanish and Pilot on interpretability,'' in *Proc. 9th Int. Workshop Semantic Eval. (SemEval)*, 2015, pp. 252–263.

[18] E. Agirre, C. Banea, D. Cer, M. Diab, A. G. Agirre, R. Mihalcea, G. R. Claramunt, and J. Wiebe, ''SemEval-2016 task 1: Semantic textual similarity, monolingual and cross-lingual evaluation,'' in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, San Diego, CA, USA. Stroudsburg, PA, USA: Association for Computational Linguistics, Jun. 2016, pp. 497–511.

[19] D. Cer, M. Diab, E. Agirre, I. Lopez-Gazpio, and L. Specia, ''SemEval-2017 task 1: Semantic textual similarity–multilingual and cross-lingual focused evaluation,'' 2017, *arXiv:1708.00055*. [Online]. Available: http://arxiv.org/abs/1708.00055

[20] J. Tian, Z. Zhou, M. Lan, and Y. Wu, ''ECNU at SemEval-2017 task 1: Leverage kernel-based traditional NLP features and neural networks to build a universal model for multilingual and cross-lingual semantic textual similarity,'' in *Proc. 11th Int. Workshop Semantic Eval. (SemEval)*, 2017, pp. 191–197.

[21] W. Zhuang and E. Chang, ''Neobility at SemEval-2017 task 1: An attention-based sentence similarity model,'' 2017, *arXiv:1703.05465*. [Online]. Available: http://arxiv.org/abs/1703.05465

[22] M.-S. Duma and W. Menzel, ''Sef@ UHH at SemEval-2017 task 1: Unsupervised knowledge-free semantic textual similarity via paragraph vector,'' in *Proc. 11th Int. Workshop Semantic Eval. (SemEval)*, 2017, pp. 170–174.

[23] I.-T. Lee, M. Goindani, C. Li, D. Jin, K. M. Johnson, X. Zhang, M. L. Pacheco, and D. Goldwasser, ''PurdueNLP at SemEval-2017 task 1: Predicting semantic textual similarity with paraphrase and event embeddings,'' in *Proc. 11th Int. Workshop Semantic Eval. (SemEval)*, 2017, pp. 198–202.

[24] T. Kenter and M. de Rijke, ''Short text similarity with word embeddings,'' in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2015, pp. 1411–1420.

[25] M. Shajalal and M. Aono, ''Semantic textual similarity between sentences using bilingual word semantics,'' *Prog. Artif. Intell.*, vol. 8, no. 2, pp. 263–272, Jun. 2019.

[26] M. Shajalal and M. Aono, ''Semantic textual similarity in bengali text,'' in *Proc. Int. Conf. Bangla Speech Lang. Process. (ICBSLP)*, Sep. 2018, pp. 1–5.

[27] J. Tian and M. Lan, ''ECNU at SemEval-2016 task 1: Leveraging word embedding from macro and micro views to boost performance for semantic textual similarity,'' in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 621–627.

[28] L. Han, A. L. Kashyap, T. Finin, J. Mayfield, and J. Weese, ''UMBC_EBIQUITY-CORE: Semantic textual similarity systems,'' in *Proc. 2nd Joint Conf. Lexical Comput. Semantics (SEM), Main Conf. Shared Task, Semantic Textual Similarity*, vol. 1, 2013, pp. 44–52.

[29] J. Barrow and D. Peskov, ''UMDeep at SemEval-2017 task 1: End-to-end shared weight LSTM model for semantic textual similarity,'' in *Proc. 11th Int. Workshop Semantic Eval. (SemEval)*, 2017, pp. 180–184.

[30] I. Lopez-Gazpio, M. Maritxalar, M. Lapata, and E. Agirre, ''Word n-gram attention models for sentence similarity and inference,'' *Expert Syst. Appl.*, vol. 132, pp. 1–11, Oct. 2019.

[31] C. Wang, I. Castellón, and E. Comelles, ''Linguistic analysis of datasets for semantic textual similarity,'' *Digit. Scholarship Humanities*, vol. 35, no. 2, pp. 471–484, Jun. 2020.

[32] W. Adouane, J.-P. Bernardy, and S. Dobnik, ''Neural models for detecting binary semantic textual similarity for Algerian and MSA,'' in *Proc. 4th Arabic Natural Lang. Process. Workshop*, 2019, pp. 78–87.

[33] T. Brychcín, ''Linear transformations for cross-lingual semantic textual similarity,'' *Knowl.-Based Syst.*, vol. 187, Jan. 2020, Art. no. 104819.

[34] M. Lenz, S. Ollinger, P. Sahitaj, and R. Bergmann, ''Semantic textual similarity measures for case-based retrieval of argument graphs,'' in *Proc. Int. Conf. Case-Based Reasoning*. Cham, Switzerland: Springer, 2019, pp. 219–234.

[35] C.-K. Lo and M. Simard, ''Fully unsupervised crosslingual semantic textual similarity metric based on BERT for identifying parallel data,'' in *Proc. 23rd Conf. Comput. Natural Lang. Learn. (CoNLL)*, 2019, pp. 206–215.

[36] N. H. Tien, N. M. Le, Y. Tomohiro, and I. Tatsuya, ''Sentence modeling via multiple word embeddings and multi-level comparison for semantic textual similarity,'' *Inf. Process. Manage.*, vol. 56, no. 6, Nov. 2019, Art. no. 102090.

[37] J. Hay, T. Van de Cruys, P. Muller, B.-L. Doan, F. Popineau, and O. Ait-Elhara, ''Automatically selecting complementary vector representations for semantic textual similarity,'' in *Advances in Knowledge Discovery and Management*. Cham, Switzerland: Springer, 2019, pp. 45–60.

[38] B. Hassan, S. E. Abdelrahman, R. Bahgat, and I. Farag, ''UESTS: An unsupervised ensemble semantic textual similarity method,'' *IEEE Access*, vol. 7, pp. 85462–85482, 2019.

[39] N. Poerner, U. Waltinger, and H. Schütze, ''Sentence meta-embeddings for unsupervised semantic textual similarity,'' 2019, *arXiv:1911.03700*. [Online]. Available: http://arxiv.org/abs/1911.03700

[40] X. Xu, T. Wang, Y. Yang, L. Zuo, F. Shen, and H. T. Shen, ''Cross-modal attention with semantic consistence for image–text matching,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 12, pp. 5412–5425, Dec. 2020.

[41] R. Mihalcea, C. Corley, and C. Strapparava, ''Corpus-based and knowledge-based measures of text semantic similarity,'' in *Proc. AAAI*, vol. 6, 2006, pp. 775–780.

[42] R. Ferreira, R. D. Lins, F. Freitas, B. Avila, S. J. Simske, and M. Riss, ''A new sentence similarity method based on a three-layer sentence representation,'' in *Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. (WI), Intell. Agent Technol. (IAT)*, vol. 1, Aug. 2014, pp. 110–117.

[43] C. Leacock and M. Chodorow, ''Combining local context and WordNet similarity for word sense identification,'' *WordNet, Electron. Lexical Database*, vol. 49, no. 2, pp. 265–283, 1998.

[44] Y. Li, Z. A. Bandar, and D. McLean, ''An approach for measuring semantic similarity between words using multiple information sources,'' *IEEE Trans. Knowl. Data Eng.*, vol. 15, no. 4, pp. 871–882, Jul. 2003.

[45] Z. Wu and M. Palmer, ''Verbs semantics and lexical selection,'' in *Proc. 32nd Annu. Meeting Assoc. Comput. Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics, 1994, pp. 133–138.

[46] P. Resnik, ''Using information content to evaluate semantic similarity in a taxonomy,'' 1995, *arXiv:cmp-lg/9511007*. [Online]. Available: https://arxiv.org/abs/cmp-lg/9511007

[47] L. Li, Y. Zhou, B. Yuan, J. Wang, and X. Hu, ''Sentence similarity measurement based on shallow parsing,'' in *Proc. 6th Int. Conf. Fuzzy Syst. Knowl. Discovery*, vol. 7, 2009, pp. 487–491.

[48] J. Oliva, J. I. Serrano, M. D. del Castillo, and Á. Iglesias, ''SyMSS: A syntax-based measure for short-text semantic similarity,'' *Data Knowl. Eng.*, vol. 70, no. 4, pp. 390–405, Apr. 2011.

[49] M. C. Lee, ''A novel sentence similarity measure for semantic-based expert systems,'' *Expert Syst. Appl.*, vol. 38, no. 5, pp. 6392–6399, May 2011.

[50] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, ''Distributed representations of words and phrases and their compositionality,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.

MD. ATABUZZAMAN received the B.Sc. (Eng.) degree in computer science and engineering from the Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh, in 2021. He is currently doing research on semantic textual similarity with the IRLP Research Group. His research interests include natural language processing (NLP) and information retrieval (IR).

MD SHAJALAL received the B.Sc. (Eng.) degree from the Department of Computer Science and Engineering, University of Chittagong, Chittagong, Bangladesh, in 2015, and the master's degree from the Department of Computer Science and Engineering, Toyohashi University of Technology (TUT), Toyohashi, Japan, in 2017. He served as a Faculty Member for Bangladesh Agricultural University and the Bangladesh Army University of Science and Technology. He is currently a Lecturer with the Department of Computer Science and Engineering, Hajee Mohammad Danesh Science and Technology University, Bangladesh. His research interests include Web search, information retrieval, sentiment analysis, and natural language processing.

M. ELIUS AHMED received the B.Sc. (Eng.) degree in computer science and engineering from the Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh, in 2021. He is currently doing research on semantic textual similarity with the IRLP Research Group. His research interests include natural language processing (NLP) and information retrieval (IR).

MASUD IBN AFJAL (Graduate Student Member, IEEE) received the B.Sc. degree in computer science and engineering from the Hajee Mohammad Danesh Science and Technology University (HSTU), Bangladesh, and the M.Sc. degree from the Rajshahi University of Engineering and Technology, (RUET) Bangladesh. He is currently an Assistant Professor with the Department of Computer Science and Engineering, HSTU. His research interests include image processing, information security, big data analysis, and text mining.

MASAKI AONO (Member, IEEE) received the B.S. and M.S. degrees from the Department of Information Science, The University of Tokyo, Tokyo, Japan, and the Ph.D. degree from the Department of Computer Science, Rensselaer Polytechnic Institute, NY, USA. From 1984 to 2003, he was with the IBM Tokyo Research Laboratory. He is currently a Professor with the Department of Computer Science and Engineering, Graduate School of Engineering, Toyohashi University of Technology. His research interests include text and data mining for massive streaming data and information retrieval for multimedia, including 2D images, videos, and 3D shape models. He is a member of ACM and the IEEE Computer Society. He has been a Japanese Delegate of the ISO/IEC JTC1 SC24 Standard Committee, since 1996.