5-2023

MSRL-Net: A multi-level semantic relation-enhanced learning network for aspect-based sentiment analysis

Zhenda HU

Zhaoxia WANG
Singapore Management University, zxwang@smu.edu.sg

Yinglin WANG

Ah-hwee TAN
Singapore Management University, ahtan@smu.edu.sg

Citation

HU, Zhenda; WANG, Zhaoxia; WANG, Yinglin; and TAN, Ah-hwee. MSRL-Net: A multi-level semantic relation-enhanced learning network for aspect-based sentiment analysis. (2023). *Expert Systems with Applications*. 217, 1-10.
Available at: https://ink.library.smu.edu.sg/sis_research/7794

MSRL-Net: A multi-level semantic relation-enhanced learning network for aspect-based sentiment analysis

Zhenda Hu [a], Zhaoxia Wang [b], Yinglin Wang [a], Ah-Hwee Tan [b]

a. School of Information Management and Engineering, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai, 200433, China
b. School of Computing and Information Systems, Singapore Management University, 80 Stamford Road, Singapore, 178902, Singapore

Published in Expert Systems with Applications (2023 May), 217, 119492. DOI: 10.1016/j.eswa.2022.119492

Abstract: Aspect-based sentiment analysis (ABSA) aims to analyze the sentiment polarity of a given text towards several specific aspects. For implementing the ABSA, one way is to convert the original problem into a sentence semantic matching task, using pre-trained language models, such as BERT. However, for such a task, the intra- and inter-semantic relations among input sentence pairs are often not considered. Specifically, the semantic information and guidance of relations revealed in the labels, such as positive, negative and neutral, have not been completely exploited. To address this issue, we introduce a self-supervised sentence pair relation classification task and propose a model named Multi-level Semantic Relation-enhanced Learning Network (MSRL-Net) for ABSA. In MSRL-Net, after recasting the original ABSA task as a sentence semantic matching task, word dependency relations and word-sentence relations are utilized to enhance the word-level semantic representation for the sentence semantic matching task, while sentence semantic relations and sentence pairs relations are utilized to enhance the sentence-level semantic representation for sentence pair relation classification. Empirical experiments on SemEval 2014 Task 4, SemEval 2016 Task 5 and SentiHood show that MSRL-Net significantly outperforms other baselines such as BERT in terms of accuracy, Macro-F1 and AUC.

Keywords: Aspect-based sentiment analysis, Semantic relation, Sentence pairs, Word dependency

1. Introduction

The aim of aspect-based sentiment analysis (ABSA) is to identify fine-grained opinion polarity towards a specific aspect (Pontiki et al., 2016), which has gained a lot of attention in the field of natural language processing (NLP) in recent years. Compared to traditional text sentiment analysis, ABSA can extract more fine-grained emotional expression, analyze the specific emotional views of users on various objectives and gain a more granular understanding of products, which contribute to providing more accurate decision support for decision makers.

With the development of AI and NLP techniques, machine learning and deep learning algorithms have been widely applied to ABSA. Moreover, neural network models combined with the attention mechanism have achieved good results compared with traditional algorithms (Fan et al., 2018, Liang et al., 2019, Lin et al., 2019, Liu et al., 2018, Majumder et al., 2018, Tang et al., 2020, Xue and Li, 2018, Zhang et al., 2019). More recently, the pre-trained language models, such as ELMo (Peters et al., 2018), OpenAI GPT (Radford, Narasimhan, Salimans, & Sutskever, 2018), and BERT (Devlin, Chang, Lee, & Toutanova, 2019), have been shown to be effective for various NLP tasks. Especially, BERT has obtained pretty good effects in Question Answering (QA) and Natural Language Inference (NLI). In addition, several variants of BERT, such as RoBERTa (Liu et al., 2019), and ALBERT (Lan et al., 2019), have also been proved effective and performed excellently on specific tasks.

However, the improvements for ABSA task are limited with the direct use of BERT. To better utilize pre-trained language models for ABSA, Sun, Huang, and Qiu (2019) leveraged BERT on the generated sentence pairs for ABSA by constructing an auxiliary sentence, which had been demonstrated to be effective. Sentence semantic matching aims to determine the most appropriate label for a pair of sentences (Zhang et al., 2021). Since the original ABSA task is converted to a sentence semantic matching task, relations of sentence pairs become the prediction targets. Different semantic expressions are derived from differently related sentence pairs (Gururangan et al., 2018). Therefore, relation learning should be paid enough attention to. However, deeper relation learning between sentence pairs has not yet been fully explored. In addition, the utilization of label information, such as the labels of relations between sentences, has been proved to be helpful for text classification tasks (Du et al., 2019, Meng et al., 2020, Pappas

& Henderson, 2019; Wang et al., 2018; Zhang et al., 2021; Zhang, Xiao, Chen, Wang, & Jin, 2018), but there is little work utilizing label information in the ABSA task. Therefore, after converting ABSA to a sentence semantic matching task (Sun et al., 2019), this paper designs a new classification task, sentence pair relation classification, by learning the relations between the generated sentence pairs via contrastive learning among different relations, gaining rich semantic information of labels and guidance of relations. Through pairwise relation learning, the semantic information implied in the relations can be measured and explicit label utilization can be explored for ABSA tasks.

Moreover, it has been shown that word dependency information can enhance the performance (Huang & Carley, 2019; Žunić, Corcoran, & Spasić, 2021). For instance, Tian, Chen, and Song (2021) designed a model to complete ABSA with word dependencies using key–value memory networks. Motivated by Tian et al. (2021), we utilize weighted word dependency information for word semantic representation. Specifically, word dependency relations as well as corresponding dependency types are considered through the memory mechanism and combined according to their contributions. Also, when forming the final representation with all words in the raw sentence, the weights can be measured by the semantic relation between each word in the raw sentence and the whole auxiliary sentence.

The major contributions of this work can be summarized as the following aspects: (1) After recasting the original ABSA task as a sentence semantic matching task, a new self-supervised relation classification task is introduced. We classify the relation between sentence pairs by contrastive learning among different relations in consideration of rich semantic information of labels and guidance of relations. (2) A Multi-level Semantic Relation-enhanced Learning Network (MSRL-Net) model is proposed for ABSA in which the intra- and inter-semantic relations among input sentence pairs are learnt through word-level and sentence-level relation modules. (3) MSRL-Net selectively leverages four types of relations at different levels. Word dependency relation and word to sentence relation are utilized to enrich the word-level semantic representation for sentence semantic matching, while sentence semantic relation and sentence pairs relation are utilized to enrich the sentence-level semantic representation for sentence pair relation classification. (4) The experiments on SemEval 2014 Task 4, SemEval 2016 Task 5 and SentiHood datasets demonstrate the effectiveness of MSRL-Net. The results show that the performance of MSRL-Net is markedly better than that of the baselines.

This paper follows the organization: Section 2 discusses the related work. The problem for ABSA is methodologically formulated in details in Section 3, and the proposed MSRL-Net is presented in Section 4. The proposed MSRL-Net is evaluated through extensive experiments, and analysis of the results is discussed in Section 5. Finally, the conclusion is presented in the final section.

2. Related work

Machine learning and deep learning algorithms have been widely leveraged for ABSA. Due to the wide application of attention mechanism, neural network models such as RNN (Fan et al., 2018; Liang et al., 2019; Tang et al., 2020), CNN (Ren, Feng, Xiao, Cai, & Cheng, 2020; Xue & Li, 2018; Zhang et al., 2019), Memory Network (Lin et al., 2019; Liu et al., 2018; Majumder et al., 2018), combined with the attention mechanism have achieved superior results than traditional algorithms. These methods can focus on specific contents more accurately and pay attention to the more important parts of sentences. For example, Xue and Li (2018) proposed an efficient convolutional neural network (CNN) for several sub-tasks of ABSA with the help of gating mechanisms, in which the aspect and sentiment information were extracted by two convolutional layers separately, and the sentiment flow was effectively controlled according to the given aspect information. Majumder et al. (2018) presented a new framework, named IARM, which leveraged recurrent memory networks with multi-hop attention mechanism. Ren et al. (2020) designed a lightweight and efficient model using gated CNN which integrates stacked gated convolution and attention mechanism.

In addition, motivated by the superiority of attention mechanisms in deep learning models, much recent work has incorporated the attention mechanism into graph neural network models (Huang & Carley, 2019; Li et al., 2021; Wang, Shen, Yang, Quan, & Wang, 2020; Zhang et al., 2019; Zhu, Zhu, Guo, Liang, & Dietze, 2021). For instance, Huang et al. Huang and Carley (2019) proposed a novel target-dependent graph attention network (TD-GAT) for ABSA, in which the dependency relationship among words was explicitly utilized and sentiment features were directly propagated from the syntactic context of an aspect target through the dependency graph. Furthermore, by the way of pruning dependency parse tree, a relational graph attention network (R-GAT) was designed to encode the coordinated aspect-aware dependency tree structure for polarity prediction (Wang et al., 2020). Zhu et al. (2021) proposed a variant of graph convolutional networks with global and local dependency which simultaneously utilized both global structure information and local structure information for ASBA.

Recently, as powerful models of self-attention emerge, pre-trained language models such as BERT (Devlin et al., 2019; Xu, Liu, Shu, & Philip, 2019) and RoBERTa (Liu et al., 2019) are extensively applied to ABSA, which brought significant performance improvements. For example, BERT was leveraged on input sentence pairs for ASBA by the way of constructing an auxiliary sentence (Sun et al., 2019). Wu et al. Wu and Ong (2020) evaluated the effect of adding context to self-attention models boosted the performance. Specifically, two variants of Context-Guided BERT were proposed to develop various ways of distributing attention. Dai, Yan, Sun, Liu, and Qiu (2021) showed that the induced tree generated from fine-tuned RoBERTa achieves much higher levels of performance than a parser-provided tree, and the induced tree was found to be more sensitive to sentiment words following further analysis. Also, a refinement network with semantics perception was proposed for ABSA (Song, Wen, Xiao, & Park, 2021).

Moreover, one of the research highlights of pre-training language model for sentiment analysis is how to make the pre-training language model learn related sentiment knowledge in the pre-training stage, in which domain knowledge is added to the design of self-monitoring mode of pre-training. It is a helpful way to enhance the performance of pre-training language model in downstream tasks. Specifically, the knowledge used as self-monitoring signals includes emotion dictionary, syntactic knowledge, external knowledge base and so on. For example, SentiBERT (Yin, Meng, & Chang, 2020) added emotional component information into the pre-training model to guide it to learn more emotional combinations. In this work, the syntactic components and phrase polarity were leveraged to supplement the pre-training signals. SentiLARE (Ke, Ji, Liu, Zhu, & Huang, 2019) considered how to integrate the word polarity of external dictionaries into the pre-training stage, and SentiWordnet (Esuli & Sebastiani, 2006), a large-scale open sentiment dictionary was used to find the most suitable word meaning and its polarity by using the explanation of the word meaning in the dictionary and by matching the text. In the pre-training stage, the self-monitoring signals of sentence polarity, part of speech of mask token and emotional polarity were added. However, the research above did not leverage the sentence-level semantic features, which may cause information loss.

In addition, labels of training samples may contain much useful information which would enhance the model. In the field of computer vision, the information from labels has been exploited in many different ways. However, there are relatively few research works focusing on explicit label utilization in NLP. Some work has shown the effectiveness of explicit label utilization (Du et al., 2019; Meng et al., 2020; Pappas & Henderson, 2019; Wang et al., 2018; Zhang et al., 2021, 2018). For instance, a multi-task label embedding model was developed for finding out implicit correlations and extracting common features (Zhang et al.,

Table 1
An example of Semeval2014 dataset.

Example	Aspect Category	Sentiment Polarity
To be completely fair, the only redeeming factor was the food, which was above average, but could not make up for all the other deficiencies of Teodora. (Pontiki et al., 2016)	Food	Positive
	Service	None
	Price	None
	Ambience	None
	Anecdotes	Negative

Table 2
An example of SentiHood dataset.

Example	Target	Aspect Category	Sentiment Polarity
LOCATION1 is not that bad, its a vibrant and diverse place, which does have crime, but so does LOCATION2	LOC1	General	None
	LOC1	Price	None
	LOC1	Safety	Negative
	LOC1	Transit-location	None
	LOC2	General	None
	LOC2	Price	None
	LOC2	Safety	Negative
	LOC2	Transit-location	None

2018). To analyze how word representations interact with label embeddings, an explicit interaction model was designed, which achieved excellent performance on text classification (Du et al., 2019). Zhang et al. (2021) demonstrated that relations can assist in discovering implicit features and patterns for semantic understanding and matching, and can enhance the performance for NLI (Natural Language Inference) task and PI (Paraphrase Identification) task. However, there are few research works utilizing label information or multi-level semantic information in the ABSA task.

3. Problem description

ABSA aims to analyze the sentiment polarity of a given text towards several specific aspects (Pontiki et al., 2016). For instance, for the comment: "The steak is good, but I dislike the service", it is a positive comment towards the aspect of food, but a negative comment towards the aspect of service. In this section, we state the problem description.

For ABSA, given a sentence $s = \{w_1, \ldots, w_z\}$ which comprises a series of words and a fixed aspect set $A = \{a_1, \ldots, a_m\}$, the problem is to (1) identify which aspect category is mentioned in the given sentence and (2) simultaneously predict the sentiment polarity $p \in P = \{p_1, \ldots, p_n\}$ for the mentioned aspect category. Table 1 shows an example of SemEval-2014 Task 4. "None" denotes that the corresponding aspect category is not stated in s.

For targeted ABSA (TABSA), with a sentence $s = \{w_1, \ldots, w_z\}$ which comprises a series of words, a fixed aspect set $A = \{a_1, \ldots, a_m\}$ and a set of target entities $T = \{t_1, \ldots, t_k\}$, the problem is to (1) identify which target–aspect pair is mentioned in the given sentence and (2) simultaneously predict the sentiment polarity over the full set of the target–aspect pairs $\{(t, a) : t \in T, a \in A\}$. Table 2 shows an example of SentiHood. "None" denotes that the corresponding target–aspect pair is not mentioned in the given sentence.

3.1. Sentence semantic matching

The original ABSA problem is converted into a sentence semantic matching task, which is described in this subsection. Following (Sun et al., 2019), given an input sentence s, by constructing the auxiliary sentence, the original ABSA task is converted into a binary classification problem which identifies whether the input sentence pair indicates the same meaning. The way of constructing the auxiliary sentence is based on each aspect from the aspect set $A = \{a_1, a_2, \ldots, a_m\}$ and each polarity from the polarity set $P = \{p_1, a_2, \ldots, p_n\}$ to form $m * n$ aspect–polarity pairs. Each aspect–polarity pair would produce one sentence. For example, the aspect–polarity pair: "service-negative" would produce the corresponding auxiliary sentence: "the polarity of the aspect service is negative". We give an example of the conversion in Fig. 1. The generated $m * n$ sentences s_b shown in the blue box in Fig. 1 are used to match with the given sentence s to form the sentence pair $\{s_a, s_b\}$, where $s_a = s$.

Due to the limited improvements with the direct use of BERT for ABSA task but excellent performances in QA and NLI tasks, the input of ABSA task can be converted into a pair of text sentences rather than a single text sentence (Sun et al., 2019). That is why we construct the auxiliary sentence to make the input of the original ABSA task similar to that of the QA or NLI task. Consequently, the original ABSA task is formalized as a binary classification task performing on the generated sentence pairs. The task is defined as:

$$P(y \mid s_a, s_b) = C(s_a, s_b), \tag{1}$$

$$y' = argmax_{y \in Y} P(y \mid s_a, s_b), \tag{2}$$

where the true label $y \in Y = \{1, 0\}$ denotes the semantic relation between s_a and s_b in the input sentence pair. In addition, the probability value of label yes is obtained as the matching score. For each aspect with corresponding generated sequences, the highest scoring sequence class would be regarded as the predicted category.

After converting the original ABSA task into a sentence semantic matching task, relations become the prediction targets, so relation learning should be paid enough attention to. It has been observed that different relations between sentence pairs indicate unique semantic expressions (Gururangan et al., 2018). Following sentence semantic matching, a new self-supervised relation classification sub-task is designed, i.e., sentence pair relation classification, which is consistent with (Zhang et al., 2021).

3.2. Sentence pair relation classification

According to Zhang et al. (2018), labels in some tasks are denoted by independent and pointless one-hot vectors. For instance, positive, negative and neutral in ABSA are encoded as [1, 0, 0], [0, 1, 0] and [0, 0, 1], respectively, leading to the loss of potential label information. Moreover, relationships are useful for revealing implicit features or patterns for semantic matching and understanding (Gururangan et al., 2018). Based on these, Zhang et al. (2021) has also demonstrated that a better understanding of implicit common features of different relationships can be obtained via the relation of relation classification for the sentence semantic matching task. Consequently, for properly and fully utilizing relation information, instead of simply classifying the most appropriate relation of s_a and s_b for each input sentence pair, this research aims to explore more implicit knowledge implied in the relation of sentence pairs via analyzing the pairwise semantic relation

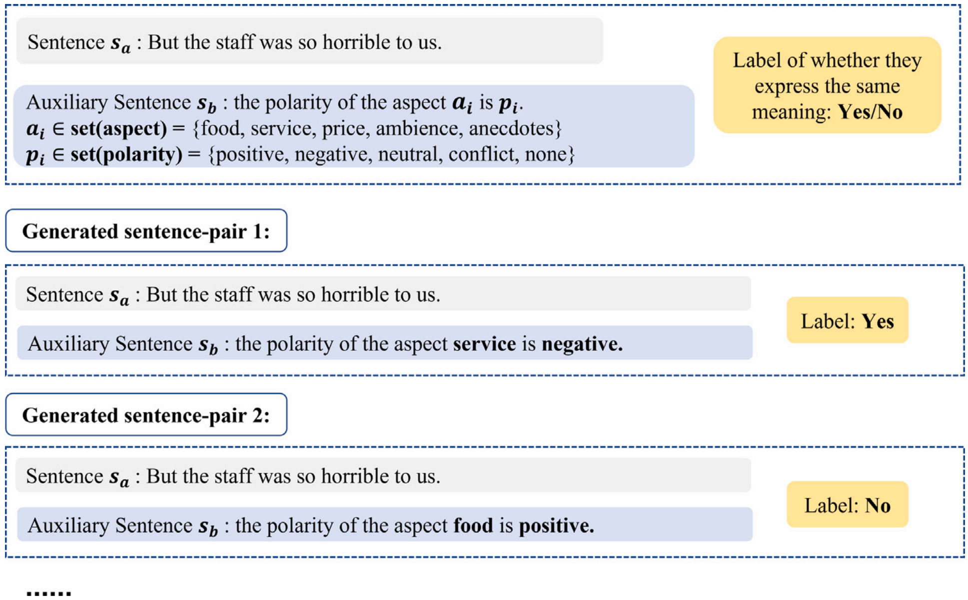

Fig. 1. An example after conversion in Semeval2014.

between the two sentence pairs $s_1 = \{s_a^1, s_b^1\}$ and $s_2 = \{s_a^2, s_b^2\}$. The formula can be expressed as follows:

$$F(s_1 = \{s_a^1, s_b^1\}, s_2 = \{s_a^2, s_b^2\}) = \begin{cases} 1, & if \quad y_1 = y_2, \\ 0, & if \quad y_1 \neq y_2, \end{cases} \tag{3}$$

$$P(z \mid s_1, s_2) = F((s_a^1, s_b^1), (s_a^2, s_b^2)), \tag{4}$$

$$z' = argmax_{z \in Z} P(z \mid s_1, s_2), \tag{5}$$

where s_a^1 and s_a^2 are from different original s_a. y_1 and y_2 represent the labels of two random input sentence pairs s_1 and s_2, respectively. The true label $z \in Z = \{1, 0\}$ indicates the relation between the two sentence pairs s_1 and s_2.

Finally, our problem consists of two sub-tasks, including sentence semantic matching and sentence pair relation classification. Taking Fig. 2 below as an example, sentence-pair 1 and sentence-pair 2 have the same label, while sentence-pair 2 and sentence-pair 3 have different labels. Through the first sub-task (sentence semantic matching), it can be found that a sentence such as "…keep …coming back" often leads to a positive sentiment. However, if the sentence involves relatively complicated sentence structure, for example, "even after a few bad evenings…" in sentence-pair 3, the sentiment may be in conflict (Note: ground truth for sentence-pair 3 is (anecdotes/miscellaneous, conflict)). When two sentence pairs both contain the span "keep …coming back", comparison learning would focus more on the other parts in the sentence to further identify the sentiment involved. Through the second sub-task (sentence pair relation classification), more implicit information can be found that is implied in the relations between the sentence pairs, which becomes an auxiliary subtask based on the first sub-task.

Based on the two classification sub-tasks above, the following critical tips should be answered:

- Because relation of sentences is our predicting objective, how can we take advantage of relation information to enhance the performance in a proper way?
- How can we combine the two sub-tasks effectively for the usage of multi-level relations and performance improvement?

Consequently, we propose a model called Multi-level Semantic Relation-enhanced Learning Network (MSRL-Net) for utilizing relation information to the fullest extent and dealing with the ABSA task.

4. The Multi-level Semantic Relation-enhanced Learning Network (MSRL-Net)

The overall structure of MSRL-Net is presented in Fig. 3, which comprises two main components, namely (1) Sentence Semantic Matching, and (2) Sentence Pair Relation Classification. As illustrated in Fig. 3, the input sentence pairs are fed concurrently into a word-level relation module and a sentence-level relation module, designed for the two sub-tasks, respectively.

4.1. Word-level relation module

The word-level relation module is mainly designed for sentence semantic matching. By using large domain corpus as well as multi-layer transformers, a joint post-training approach Post-training BERT (Xu et al., 2019) was proposed to enhance both the domain and task knowledge based on BERT. Due to the specific domain of the dataset, utilizing the word embeddings from Post-training BERT as word-level semantic representation is an effective way to obtain some domain-aware knowledge.

For the aforementioned input sentence pairs in our task, it is organized to structure a special sequence of $[CLS], s_a, [SEP], s_b, [SEP]$, and subsequently the sequence is fed into a Post-training BERT encoder, expressed as below:

$$h_0, H^a, H^b = Post - training\ BERT(s_a, s_b), \tag{6}$$

where $h_0 \in R^d$ indicates the embedding of the first token $[CLS]$ at the last layer, and $H^a \in R^{l_a \times d}$, $H^b \in R^{l_b \times d}$ represent the words' embedding matrices in s_a and s_b, respectively. After obtaining the initial word-level semantic representation, word dependencies information is then fused into the current representation through the following two types of information by the enhanced Key-Value Memory Networks (KVMN) (Miller et al., 2016).

(1) Word Dependencies Relation

In order to utilize word dependencies relation, dependency parsing is applied to obtain the word dependency embedding information as the external features to enhance the word representation.

Similar to Tian et al. (2021), firstly, all word dependency relations obtained from the parsing results of s_a are collected. As shown in Fig. 4, there is governor-dependent relationship between words. Then key-value memory networks (KVMN) (Miller et al., 2016), a non-graph

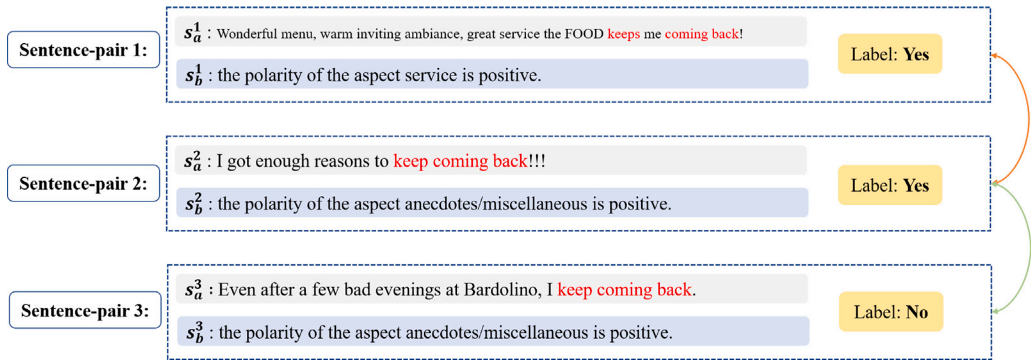

Fig. 2. Several samples from SemEval-2014 Task 4.

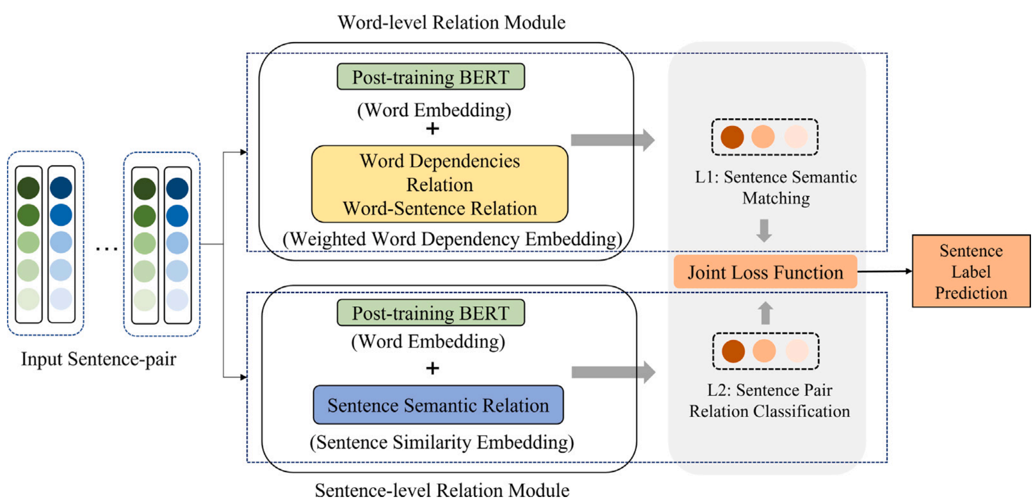

Fig. 3. Architecture of Multi-level Semantic Relation-enhanced Learning Network (MSRL-Net).

structure method, is used to produce the key set, and their corresponding dependency types are mapped to the value set. Consequently, the matrices, K and V can be obtained. Each vector represents a key or a value in the key and value sets, respectively. For each word w_i in s_a, we use o_i to denote its word dependency information outputted by the KVMN model:

$$t_{i,j} = \frac{exp(h_i \cdot e^k_{i,j})}{\sum_{j=1}^{q} exp(h_i \cdot e^k_{i,j})} \tag{7}$$

$$o_i = \sum_{j=1}^{q} t_{i,j} e^v_{i,j}, \tag{8}$$

where h_i represents the hidden vector for w_i from the encoder (i.e., BERT), and $e^k_{i,j}$ and $e^v_{i,j}$ denote the embedding of each element in K and V, respectively. As a result, all the dependency information $e^v_{i,j}$ for w_i can be weighted summed, which can weaken the influence caused by the noise of the dependency parser.

Next, in order to obtain a richer semantic representation of each sentence pair with syntactic dependency information, the dependency relations between words in s_a and the type information of dependency relationship can be added to the initial semantic representation of (s_a, s_b) from Post-training BERT. Upon encoding word dependencies for each w_i in s_a, corresponding o_i are obtained and summed according to the weights. The formula can be expressed as following:

$$h_w = h_0 \oplus \sum_{i=1}^{l} weight_i * o_i, \tag{9}$$

where $h_0 \in R^d$ is the embedding of the first token $[CLS]$, $o_i \in R^{\bar{d}}$ represents the embedding of word dependencies information for each

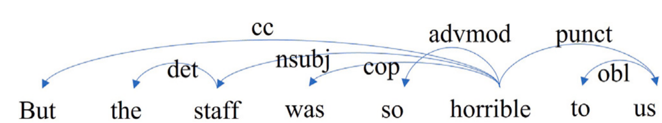

Fig. 4. An example of word dependencies relation.

word w_i in s_a, and $h_w \in R^{d+\bar{d}}$ is the semantic representation of each sentence pair after fusing word dependencies information.

Moreover, different syntactic dependency knowledge is measured by different weights $weight_i \in [0,1]$ to identify the importance in syntactic dependency knowledge, so as to improve the fineness for utilizing syntactic dependency knowledge. The weights are calculated by the word-sentence relation between w_i and s_b, as shown in Eqs. (10)–(12).

(2) Word-Sentence Relation

As mentioned before, the weights are calculated by the word-sentence relation. Specifically, we compute the cosine similarity between each word w_i in s_a and s_b. When adding the word dependencies information to the semantic representations, if different o_i are averaged by the simple way, the word dependencies of each word in s_a would be treated equally, which resulted in the loss of some information between the two sentences. The way of calculation can be expressed as follows:

$$v_b = W_b * H^b + b_b \tag{10}$$

$$l_i = cosine\ similarity(w_i, v_b), \tag{11}$$

$$weight_i = \frac{exp(l_i)}{\sum_{i=1}^{z} exp(l_i)}, \tag{12}$$

where a full connected layer would be applied to obtain the sentence representation ($v_b \in R^d$) of s_b before calculating the cosine similarity $l_i \in [-1, 1]$. As a result, the sentence pair can allow us to utilize the word-sentence relation to enhance the effectiveness of dependency syntax knowledge.

Next, a fully connected layer with softmax activation function is used with a trainable matrix $W \in R^{lab\times(d+\bar{d})}$ and vector b to obtain $P_w \in R^{lab}$ (lab denotes the number of labels), i.e.,

$$P_w = softmax(W * h_w + b). \tag{13}$$

Finally, the cross-entropy loss function is applied to u for predicting the label y' of each sentence pair, i.e.,

$$y' = argmax_{y\in Y} P(y \mid s_a, s_b). \tag{14}$$

Besides the word-level relations utilized for sentence semantic matching, the sentence-level relations are also utilized for sentence pair relation classification, as described in the next sub-section.

4.2. Sentence-level relation module

The sentence-level relation module is mainly designed for sentence pair relation classification. Through exploring the relation of sentence pairs, we can obtain deeper semantic information of sentences and extract the features related to semantic relations to construct better sentence semantic representation. For the sub-task of sentence pair relation classification, two relations including sentence semantic relation and sentence pair relation are utilized in this module.

(1) Sentence Semantic Relation

The sentence semantic relation between s_a and s_b can be obtained from Sentence BERT (SBERT) (Reimers et al., 2019), which was trained on the combination of SNLI (Bowman, Angeli, Potts, & Manning, 2015) and Multi-Genre NLI (Williams, Nangia, & Bowman, 2018) dataset. Since SBERT is quite helpful for the sentence semantic matching task, it is used to obtain semantic textual similarity of two sentences in each sentence pair. In addition, due to the domain of the dataset, Augmented SBERT (in-domain) strategy can be applied to specific datasets. The similarity embedding is defined as:

$$r = [r_1 * r_2, r_1 - r_2], \tag{15}$$

$$r_1 = SBERT_1(s_a, s_b), r_2 = SBERT_2(s_a, s_b), \tag{16}$$

where $r \in R^2$ is the similarity embedding for each sentence pair by concatenating the two semantic similarity scores r_1 and r_2 from two different pre-trained models of Sentence-BERT, respectively.

(2) Relation of Sentence Pairs

Considering the relation of generated sentence pairs by contrastive learning among different relations for the implicit common features, the sentence pair relation classification task is introduced into MSRL-Net. With regard to rich semantic information of labels and guidance of relations, we utilize the relation of sentence pairs to explore the pairwise relation of samples. This process would be described in the following steps.

Firstly, the representations of sentence pair (s_a^1, s_b^1) and sentence pair (s_a^2, s_b^2) can be obtained by Post-training BERT through Eq. (6).

Next, for the pairwise relation between the semantic representations (v_1 for pair (s_a^1, s_b^1), and v_2 for pair (s_a^2, s_b^2)), v_1 and v_2 are transformed by a nonlinear Relu activation function as below:

$$\hat{v}_1 = Relu(v_1), \hat{v}_2 = Relu(v_2). \tag{17}$$

Then, we leverage heuristic matching (Chen et al., 2017) to compute the similarity and difference between $\hat{v}_1 \in R^d$ and $\hat{v}_2 \in R^d$. The difference and element-wise product are concatenated with the original vectors to retain all the information. Next, the constructed similarity embedding $u \in R^{4d}$ and $r \in R^s$ are sent to a linear layer, respectively. These steps can be represented as:

$$u = [\hat{v}_1; \hat{v}_2; (\hat{v}_1 \odot \hat{v}_2); (\hat{v}_1 - \hat{v}_2)], \tag{18}$$

$$\hat{u} = W_u u + b_u, \tag{19}$$

$$\hat{r} = W_r r + b_r. \tag{20}$$

Finally, we concatenate $\hat{u} \in R^{d'}$ and $\hat{r} \in R^{d''}$, and use a MLP for the last classification, i.e.,

$$P(z \mid (s_a^1, s_b^1), (s_a^2, s_b^2)) = MLP(\hat{u}; \hat{r}), \tag{21}$$

$$z' = argmax_{z\in Z} P(z \mid (s_a^1, s_b^1), (s_a^2, s_b^2)), \tag{22}$$

where z' indicates the predicted label of (s_a^1, s_b^1) and (s_a^2, s_b^2).

4.2.1. Joint loss function

As mentioned in Sections 3 and 4, both Sentence Semantic Matching and Sentence Pair Relation Classification are formulated as classification tasks. Consequently, Cross-Entropy is employed as the basic loss function as represented in the following equations:

$$L_{s-label} = -\hat{y}_i log P(y_i \mid s_a, s_b), \tag{23}$$

$$L_{r-label} = -\hat{z}_i log P(z_i \mid ((s_a^1, s_b^1), (s_a^2, s_a^2))_i). \tag{24}$$

Finally, we compute the weighted sum of the two loss functions with a hyper-parameter α as the overall loss function for MSRL-Net. The joint loss function is represented as:

$$L = \alpha L_{s-label} + (1 - \alpha) L_{r-label}. \tag{25}$$

5. Experiments

5.1. Data description

To evaluate the proposed model, three well-known datasets have been used in this research.

(1) SemEval-2014 Task 4

SemEval-2014 Task 4 (Pontiki et al., 2014) is a very common dataset constructed from customer reviews for ABSA. The initial research built the dataset, aiming to detect the categories mentioned and sentiment polarity towards each category. Subtask 3 (Aspect Category Detection) and subtask 4 (Aspect Category Polarity) are jointly evaluated. Specifically, the dataset contains five aspects including food, service, price, ambience, and anecdotes/miscellaneous, as well as four polarities including positive, negative, neutral and conflict. It consists of 3044 training samples and 800 testing samples.

(2) SemEval-2016 Task 5

Besides the above two datasets, SemEval-2016 Task 5 (Pontiki et al., 2016) is also utilized for the experiments, which consists of 2000 training samples and 676 testing samples. Specifically, the dataset contains ten aspect categories including RESTAURANT#GENERAL, RESTAURANT#PRICES, RESTAURANT#MISCELLANEOUS, SERVICE#GENERAL, AMBIENCE#GENERAL, LOCATION#GENERAL, FOOD#QUALITY, FOOD#STYLE _OPTIONS, DRINKS#STYLE_OPTIONS, DRINKS#PRICES, as well as three polarities including positive, negative and neutral.

(3) SentiHood

Another dataset used in this research is SentiHood dataset (Radford et al., 2018), which consists of 3752 training samples and 1879 testing samples. Each sentence would contain several target–aspect pairs t, a as well as the sentiment polarity y. There is a need to detect an aspect a for t and decide y for the mentioned target–aspect pairs, with the given sentence s and the target t. Specifically, the dataset involves four aspects including general, price, safety and transit-location, as well as two polarities including positive and negative.

Table 3
The experimental results on Semeval-2014 task 4.

Model	Aspect Category Detection			Sentiment Polarity		
	Precision	Recall	F1	4-way	3-way	2-way
NRC-Canada (Kiritchenko, Zhu, Cherry, & Mohammad, 2014)	91.04	86.24	88.58	–	–	–
ATAE-LSTM (Wang, Huang, Zhu, & Zhao, 2016)	–	–	–	–	84.00	89.90
BERT-single (Devlin et al., 2019)	92.78	89.07	90.89	83.70	86.90	93.30
BERT-sentence pair (Sun et al., 2019)	93.04	89.56	91.47	85.90	89.90	95.60
BERT-single + KVMN (Tian et al., 2021)	–	–	–	86.10	90.68	94.75
Post-training BERT (Xu et al., 2019)	93.29	90.93	92.09	88.00	91.47	95.56
CG-BERT (Wu & Ong, 2021)	93.12	90.17	91.62	86.90	90.40	94.70
MSRL-Net	**94.44** (±.34)	**92.10** (±.29)	**92.68** (±.31)	**89.42** (±.52)	**92.29** (±.41)	**96.93** (±.36)

5.2. Baselines

Our MSRL-Net model is compared with the following methods. NRC-Canada (Kiritchenko et al., 2014) is a classic lexicon-based method through the use of sentiment lexicons and word clusters. ATAE-LSTM (Wang et al., 2016) is a LSTM based model, which utilizes an extra attention mechanism to perform soft-selection over the context words. SenticLSTM (Ma, Peng, & Cambria, 2018) is also a LSTM based model which introduces external information from SenticNet. Dmu-Entnet (Liu et al., 2018) is a bidirectional EntNet with external "memory chains" with a delayed memory update mechanism to track entities. For BERT based methods, we use the standard uncased BERT-base model with 12 transformer blocks as the baseline model. BERT-single + KVMN (Tian et al., 2021) selectively leverages different dependency results by key–value memory networks. Post-training BERT (Xu et al., 2019) proposes a novel post-training approach on the popular language model BERT. CG-BERT (Wu & Ong, 2021) is Context-Guided BERT that learns to distribute attention under different contexts.

5.3. Training details

We adopt the BERT-base post-trained on the restaurant domain, with 12 Transformer blocks and the hidden layer size is 768. The dropout probability is 0.1. For training, we set the initial learning rate to 2e-5 and set the batch size to 24. Adam optimizer is employed to optimize the training parameters. In addition, we set epoch number to 6 for SemEval-2014 Task 4, and 4 for SentiHood. We run our model 5 times and show the mean values and standard deviations.

For the word dependencies information, Stanford Parser[1] is utilized to construct dependency trees for given sentences. For Sentence-BERT, all-mpnet-base-v2 and multi-qa-mpnet-base-dot-v1[2] are chosen as the pre-trained models to obtain the semantic textual similarity for each sentence-pair. In addition, for the hyper-parameter α in the updated loss function, 0.7, 0.8, 0.9 and 0.99 are adopted. The value of α depends on the performance on the test set. All experiments are implemented by PyTorch, and the model is trained on one Nvidia Tesla-V100 GPU.

5.4. The results

By comparing the proposed model MSRL-Net with the published baselines on three datasets including SemEval-2014 Task 4, SemEval-2016 Task 5 and SentiHood, the results are summarized in Table 3, Table 4 and Table 5, respectively.

(1) SemEval-2014 Task 4: ABSA

Following (Pontiki et al., 2014), for evaluating the sub-task of aspect category detection, Precision, Recall and F1-score are chosen. For the evaluation of the sub-task of aspect category polarity, we use Accuracy with various statistical ways.

From Table 3, we can find that MSRL-Net obtains the best results in both two sub-tasks, (Precision, Recall, F1) = (94.44, 92.10, 92.38) for aspect category detection and (4-way Acc., 3-way Acc., 2-way Acc.) = (89.42, 92.29, 96.93) for sentiment polarity classification. Compared to BERT-sentence pair model, for aspect category detection, MSRL-Net improves by 1.21 percent of F1-score. For polarity classification, the improvement is more significant, by 3.52 percent of the 4-way accuracy. The results show that the proposed MSRL-Net utilizing both word dependencies knowledge and label information obtained from input sentence pairs can significantly improve the performance of ABSA task.

(2) SemEval-2016 Task 5: ABSA

Similar to SemEval-2014 Task 4, for evaluating the sub-task of aspect category detection, Precision, Recall and F1-score are used for SemEval-2016 Task 5 as well. For the evaluation of the sub-task of aspect category polarity, we use accuracy with various statistical measures.

From Table 4, we find that MSRL-Net obtains the best results in both sub-tasks. Specifically, for aspect category detection, Precision, Recall and F1 reach 94.11, 79.69 and 86.30, respectively. For sentiment polarity classification, 3-way Acc. and 2-way Acc. are 90.38 and 94.64, respectively. Compared to Post-training BERT model, for aspect category detection, MSRL-Net delivers an improvement on F1 score by 1.16 percent. For polarity classification, the improvement is more significant, by 2.62 percent in 2-way accuracy. The results demonstrate that the proposed MSRL-Net utilizing both word dependencies knowledge and label information obtained from input sentence pairs can make a positive impact on improving the performance of ABSA task.

(3) SentiHood: TABSA

To be consistent with (Sun et al., 2019), for the evaluation metrics of SentiHood, accuracy, Macro-F1 and macro-average AUC are used to assess the performance. According to Table 5, the proposed model MSRL-Net outperforms other baselines obviously, (Acc., F1, AUC) = (80.8, 87.1, 97.3) for aspect category detection and (Acc., AUC) = (93.8, 97.3) for sentiment polarity classification. Specifically, compared to BERT-sentence pair, the accuracy has improved 2% and approximately 1% for aspect detection and sentiment polarity classification, respectively. Compared to Post-training BERT, there is a 4.1% improvement of F1-score for aspect category detection and a 2.8% improvement of accuracy for sentiment polarity classification. The results demonstrate that the utilization of multi-level information and updated loss function can contribute to learning more implied information and boosting the model's efficiency.

It is worth noting that Post-training BERT does not achieve the performance as good as that on Semeval-2014 task 4. That is because the reviews in SentiHood dataset focus on describing the location, but not describing the restaurant, which may cause the domain deviation.

5.5. Discussion

It is observed from the experiment results above that the proposed MSRL-Net achieves best performance in both datasets compared with the baselines. The results also demonstrate that BERT-base model

[1] https://nlp.stanford.edu/software/lex-parser.shtml
[2] https://www.sbert.net/docs/pretrained_models.html

Table 4
The experimental results on Semeval-2016 task 5.

Model	Aspect Category Detection			Sentiment Polarity	
	Precision	Recall	F1	3-way	2-way
ATAE-LSTM (Wang et al., 2016)	–	–	–	81.53	83.56
BERT-single (Devlin et al., 2019)	92.67	74.97	83.14	81.90	85.23
BERT-sentence pair (Sun et al., 2019)	93.93	75.30	83.59	83.85	87.41
BERT + KVMN (Tian et al., 2021)	–	–	–	84.56	88.40
Post-training BERT (Xu et al., 2019)	93.01	78.98	85.42	88.84	92.02
MSRL-Net	**94.11** (±.26)	**79.69** (±.31)	**86.58** (±.28)	**90.38** (±.39)	**94.64** (±.32)

Table 5
The experimental results on SentiHood.

Model	Aspect			Sentiment	
	Acc.	F1	AUC	Acc.	AUC
SenticLSTM (Ma et al., 2018)	67.4	78.2	–	89.3	–
Dmu-Entnet (Liu et al., 2018)	73.5	78.5	94.4	91.0	94.8
BERT-single (Devlin et al., 2019)	73.7	81.0	96.4	85.5	84.2
BERT-sentence pair (Sun et al., 2019)	78.8	83.9	96.6	92.9	96.8
BERT-single + KVMN (Tian et al., 2021)	–	–	–	90.9	96.5
Post-training BERT (Xu et al., 2019)	79.1	83.0	96.5	91.1	96.1
MSRL-Net	**80.8** (.3)	**87.1** (.3)	**97.3** (.2)	**93.8** (.4)	**97.3** (.2)

Table 6
Cases from Semeval-2014 task 4 (Note: ✗ indicates incorrect prediction).

Case	Aspect	MSRL-Net	BERT
Even after a few bad evenings at Bardolino, I keep coming back.	Anecdotes	Conflict	Negative (✗)
After so many great reviews here, my bf and I went to Esca to celebrate my birthday last night.	Anecdotes	Neutral	Positive (✗)
The service is great (maybe even borderline nagging but at least you get attention), the desserts are excellent and the coffee is so very good...	Service	Conflict	Positive (✗)
The food did take a few extra minutes to come, but the cute waiters' jokes and friendliness made up for it.	Service	Positive (✗)	Positive (✗)

outperforms BERT-free models significantly, which is consistent with previous work. With the design of multi-layer transformers, the BERT based models can learn sentence semantics from huge amount of text. However, existing BERT based models focus on the input sequence, which may underestimate the rich semantic information implied by different relations. Consequently, MSRL-Net is proposed to fully utilize the rich embedding information by incorporating the semantic information of the four multi-level relations. Word dependency relation and word to sentence relation are utilized to enrich the word-level semantic representation for sentence semantic matching, while sentence semantic relation and sentence pairs relation are utilized to enrich the sentence-level semantic representation for sentence pair relation classification. The experiment results demonstrate the powerful ability of MSRL-Net in terms of the evaluation metrics.

The superiority of MSRL-Net mainly comes from two aspects: post-training BERT and multi-level semantic relation utilization. Especially for the latter, experimental results show that MSRL-Net utilizing various types of semantic relations can enhance the performance for ABSA tasks.

In addition, the label information provides guidance for exploring the relation of sentence pairs. The results show the effectiveness of MSRL-Net for relation learning in this research. The utilization of contrastive relation learning in the proposed model can also be applied to other related tasks.

5.6. Case study

From Table 6, the prediction results of several cases from Semeval-2014 task 4 are shown. The aspect category of Case 1 and Case 2 is "anecdotes/miscellaneous", with conflict sentiment and neutral sentiment respectively. The aspect category of Case 3 and Case 4 is "service", with conflict and neutral sentiments, respectively. Compared with BERT, our MSRL-Net model can predict the correct sentiments for Case 1, Case 2 and Case 3. For Case 1 and Case 2, BERT returns negative and positive sentiments as it may focus too much on "a few bad evenings" and "so many great reviews", respectively. MSRL-Net is able to prevent the sentence from not being overly affected by unimportant sentiment words through multi-level semantic relations. The last two cases are both relatively complex sentences, which may contain multiple sentiments. For Case 4, both BERT and MSRL-Net do not predict correct sentiments. This typically involves complicated sentence structures, which requires advanced language understanding capability.

5.7. Ablation study

Whereas the experimental results as described above have proved the superiority of MSRL-Net, it is still unclear how each part of MSRL-Net influences the performance improvement. So, an ablation study is performed to validate and compare the effectiveness of each part in MSRL-Net, including word dependencies information, sentence semantic relation and sentence-pair relation. Note that BERT-sentence pair is selected as the baseline in the ablation study.

The role of word relation is the key of the Word-level Relation Module, which is designed for the Sentence Semantic Matching task. This type of relation collects word dependency relations obtained from the parsing results. MSRL-Net without word relation would weaken the

Table 7
Ablation performance of MSRL-Net for SemEval-2014.

Model	Aspect Category Detection			Aspect Category Polarity		
	Precision	Recall	F1	4-way	3-way	2-way
BERT-sentence pair	93.04	89.56	91.47	85.90	89.90	95.60
MSRL-Net (w/o word relation)	93.12	91.34	92.24	88.00	91.47	95.56
MSRL-Net (w/o sentences relation)	92.99	91.90	92.44	88.29	90.66	95.90
MSRL-Net (w/o sent-pair relation)	93.29	90.93	92.09	86.15	90.75	94.83
MSRL-Net	**94.44**	**92.10**	**92.68**	**89.42**	**92.29**	**96.93**

Table 8
Ablation performance of MSRL-Net for Sentihood.

Model	Aspect			Sentiment	
	Acc.	F1	AUC	Acc.	AUC
BERT-sentence pair	78.8	83.9	96.6	92.9	96.8
MSRL-Net (w/o word relation)	80.5	87.0	97.1	92.8	96.7
MSRL-Net (w/o sentences relation)	**81.1**	86.7	96.4	92.4	96.4
MSRL-Net (w/o sentence-pair relation)	79.1	83.0	96.5	91.1	96.1
MSRL-Net	80.8	**87.1**	**97.3**	**93.8**	**97.3**

ability of MSRL-Net on the Sentence Semantic Matching task, the 2-way Acc. of which deceases 1.37 percent as shown in Table 7.

In addition, the role of sentences relation and sentence-pair relation are the core of the Sentence-level Relation Module, which is designed for the Sentence Pair Relation Classification task. Sentences relation is used to obtain semantic textual similarity of two sentences in each sentence pair. MSRL-Net without sentences relation would result in MSRL-Net not being able to explore the similarity of two sentences in each sentence pair. Sentence-pair relation is mainly for predicting the relations of generated sentence pairs by contrastive learning among different relations. MSRL-Net without sentence-pair relation would result in MSRL-Net not being able to learn different relations between sentence pairs, and weaken the ability of MSRL-Net on the Sentence Pair Relation Classification task, the 2-way Acc. of which deceases 2.1 percent as shown in Table 7.

According to the ablation performance shown in Tables 7 and 8, it can be seen that any part of MSRL-Net would influence the model performance. Sentence-pair relation has the most significant impact, while sentences semantic relation has relatively little effect on the model performance among all the components. The reason would be that sentence semantic relation comes from the external knowledge, which may contain some noise. These observations indicate that the relation utilization of sentence pairs by contrastive learning in consideration of rich semantic information of labels and guidance of relations is very necessary.

6. Conclusion

This paper develops a novel Multi-level Semantic Relation-enhanced Learning Network model, named MSRL-Net, for ABSA. Not only the initial ABSA problem is converted into a sentence-pair classification task, but also multi-level intra- and inter-semantic relations among input sentences and sentence pairs are explored. Specifically, four types of relation information including word dependency relation, word to sentence relation, semantic relation of sentences and relation of sentence pairs are utilized to enhance both the word-level semantic representation and sentence-level semantic representation for ABSA. The proposed model can capture abundant semantic relation information to achieve significant improvements over previous models without requiring any other extra data. The research results demonstrate that the proposed MSRL-Net is effective, which indicates a potential direction of utilizing semantic information.

Going forward, the introduction of external sentiment knowledge, such as domain sentiment lexicons, to ABSA is our on-going work. Leveraging knowledge graph to find structured commonsense knowledge, such as relations between sentiment words and non-sentiment words, to further improve the performance of the proposed MSRL-Net will be explored. In addition, exploring human-interpretable learning approach for ABSA is also our future work.

CRediT authorship contribution statement

Zhenda Hu: Conceptualization, Methodology, Software, Formal analysis, Validation, Writing – original draft, Writing – review & editing. **Zhaoxia Wang:** Methodology, Supervision, Investigation, Writing – review & editing. **Yinglin Wang:** Supervision, Investigation, Writing – review & editing. **Ah-Hwee Tan:** Supervision, Investigation, Writing – review & editing.

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.

Data availability

Data will be made available on request.

Acknowledgments

We would like to thank the National Natural Science Foundation of China (Grant No. 61375053) for part of the financial support of this research. This research was supported by A*STAR, Singapore under its Advanced Manufacturing and Engineering (AME) Programmatic Grant (Award No.: A19E2b0098) and the Jubilee Technology Fellowship awarded to Prof. Ah-Hwee Tan by Singapore Management University.

References

Bowman, S., Angeli, G., Potts, C., & Manning, C. D. (2015). A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 632–642).

Chen, Q., Zhu, X., Ling, Z.-H., Wei, S., Jiang, H., & Inkpen, D. (2017). Enhanced LSTM for natural language inference. In *Proceedings of the 55th Annual meeting of the association for computational linguistics (volume 1: long papers)* (pp. 1657–1668).

Dai, J., Yan, H., Sun, T., Liu, P., & Qiu, X. (2021). Does syntax matter? A strong baseline for aspect-based sentiment analysis with roberta. In *Proceedings of the 2021 Conference of the North American chapter of the association for computational linguistics: human language technologies* (pp. 1816–1829).

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT (1)*.

Du, C., Chen, Z., Feng, F., Zhu, L., Gan, T., & Nie, L. (2019). Explicit interaction model towards text classification. In *Proceedings of the AAAI conference on artificial intelligence, vol. 33 no. 01* (pp. 6359–6366).

Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the fifth international conference on language resources and evaluation (LREC'06)*.

Fan, F., Feng, Y., & Zhao, D. (2018). Multi-grained attention network for aspect-level sentiment classification. In *Proceedings of the 2018 Conference on empirical methods in natural language processing* (pp. 3433–3442).

Gururangan, S., Swayamdipta, S., Levy, O., Schwartz, R., Bowman, S. R., & Smith, N. A. (2018). Annotation artifacts in natural language inference data. In *NAACL-HLT (2)*.

Huang, B., & Carley, K. M. (2019). Syntax-aware aspect level sentiment classification with graph attention networks. In *Proceedings of the 2019 Conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 5469–5477).

Ke, P., Ji, H., Liu, S., Zhu, X., & Huang, M. (2019). SentiLARE: Sentiment-aware language representation learning with linguistic knowledge. arXiv preprint arXiv:1911.02493.

Kiritchenko, S., Zhu, X., Cherry, C., & Mohammad, S. (2014). NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th International workshop on semantic evaluation (SemEval 2014)* (pp. 437–442).

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.

Li, R., Chen, H., Feng, F., Ma, Z., Wang, X., & Hovy, E. (2021). Dual graph convolutional networks for aspect-based sentiment analysis. In *Proceedings of the 59th Annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: long papers)* (pp. 6319–6329).

Liang, B., Du, J., Xu, R., Li, B., & Huang, H. (2019). Context-aware embedding for targeted aspect-based sentiment analysis. In *Proceedings of the 57th Annual meeting of the association for computational linguistics* (pp. 4678–4683).

Lin, P., Yang, M., & Lai, J. (2019). Deep mask memory network with semantic dependency and context moment for aspect level sentiment classification.. In *IJCAI* (pp. 5088–5094).

Liu, F., Cohn, T., & Baldwin, T. (2018). Recurrent entity networks with delayed memory update for targeted aspect-based sentiment analysis. In *Proceedings of the 2018 Conference of the north american chapter of the association for computational linguistics: human language technologies, volume 2 (short papers)* (pp. 278–283).

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., et al. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

Ma, Y., Peng, H., & Cambria, E. (2018). Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In *Thirty-Second AAAI Conference on artificial intelligence*.

Majumder, N., Poria, S., Gelbukh, A., Akhtar, M. S., Cambria, E., & Ekbal, A. (2018). IARM: Inter-aspect relation modeling with memory networks in aspect-based sentiment analysis. In *Proceedings of the 2018 Conference on empirical methods in natural language processing* (pp. 3402–3411).

Meng, Y., Zhang, Y., Huang, J., Xiong, C., Ji, H., Zhang, C., et al. (2020). Text classification using label names only: A language model self-training approach. arXiv preprint arXiv:2010.07245.

Miller, A., Fisch, A., Dodge, J., Karimi, A.-H., Bordes, A., & Weston, J. (2016). Key-value memory networks for directly reading documents. In *Proceedings of the 2016 Conference on empirical methods in natural language processing* (pp. 1400–1409).

Pappas, N., & Henderson, J. (2019). Gile: A generalized input-label embedding for text classification. *Transactions of the Association for Computational Linguistics*, *7*, 139–155. http://dx.doi.org/10.1162/tacl_a_00259.

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., et al. (2018). Deep contextualized word representations. In *Proceedings of NAACL-HLT* (pp. 2227–2237).

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., et al. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In *International workshop on semantic evaluation* (pp. 19–30).

Pontiki, M., Papageorgiou, H., Galanis, D., Androutsopoulos, I., Pavlopoulos, J., & Manandhar, S. (2014). SemEval-2014 task 4: Aspect based sentiment analysis. *SemEval 2014*, 27.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding with unsupervised learning*: *Technical report*, OpenAI.

Reimers, N., Gurevych, I., Reimers, N., Gurevych, I., Thakur, N., Reimers, N., et al. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. In *Proceedings of the 2019 Conference on empirical methods in natural language processing*. association for computational linguistics.

Ren, F., Feng, L., Xiao, D., Cai, M., & Cheng, S. (2020). DNet: a lightweight and efficient model for aspect based sentiment analysis. *Expert Systems with Applications*, *151*, Article 113393. http://dx.doi.org/10.1016/j.eswa.2020.113393, URL https://www.sciencedirect.com/science/article/pii/S0957417420302177.

Song, W., Wen, Z., Xiao, Z., & Park, S. C. (2021). Semantics perception and refinement network for aspect-based sentiment analysis. *Knowledge-Based Systems*, *214*, Article 106755. http://dx.doi.org/10.1016/j.knosys.2021.106755, URL https://www.sciencedirect.com/science/article/pii/S0950705121000186.

Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In *Proceedings of NAACL-HLT* (pp. 380–385).

Tang, H., Ji, D., Li, C., & Zhou, Q. (2020). Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification. In *Proceedings of the 58th Annual meeting of the association for computational linguistics* (pp. 6578–6588).

Tian, Y., Chen, G., & Song, Y. (2021). Enhancing aspect-level sentiment analysis with word dependencies. In *Proceedings of the 16th Conference of the european chapter of the association for computational linguistics: main volume* (pp. 3726–3739).

Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2016). Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2016 Conference on empirical methods in natural language processing* (pp. 606–615).

Wang, G., Li, C., Wang, W., Zhang, Y., Shen, D., Zhang, X., et al. (2018). Joint embedding of words and labels for text classification. In *Proceedings of the 56th Annual meeting of the association for computational linguistics (volume 1: long papers)* (pp. 2321–2331).

Wang, K., Shen, W., Yang, Y., Quan, X., & Wang, R. (2020). Relational graph attention network for aspect-based sentiment analysis. In *Proceedings of the 58th Annual meeting of the association for computational linguistics* (pp. 3229–3238).

Williams, A., Nangia, N., & Bowman, S. (2018). A broad-coverage challenge corpus for sentence understanding through inference. In *Proceedings of the 2018 Conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long papers)* (pp. 1112–1122).

Wu, Z., & Ong, D. C. (2020). Context-guided bert for targeted aspect-based sentiment analysis. *Association for the Advancement of Artificial Intelligence*, 1–9. http://dx.doi.org/10.1609/aaai.v35i16.17659, URL https://ojs.aaai.org/index.php/AAAI/article/view/17659.

Wu, Z., & Ong, D. C. (2021). Context-guided BERT for targeted aspect-based sentiment analysis. In *Proceedings of the AAAI conference on artificial intelligence, vol. 35 no. 6* (pp. 14094–14102).

Xu, H., Liu, B., Shu, L., & Philip, S. Y. (2019). BERT post-training for review reading comprehension and aspect-based sentiment analysis. In *Proceedings of the 2019 Conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 2324–2335).

Xue, W., & Li, T. (2018). Aspect based sentiment analysis with gated convolutional networks. In *Proceedings of the 56th Annual meeting of the association for computational linguistics (volume 1: long papers)* (pp. 2514–2523).

Yin, D., Meng, T., & Chang, K.-W. (2020). SentiBERT: A transferable transformer-based architecture for compositional sentiment semantics. In *Proceedings of the 58th Annual meeting of the association for computational linguistics* (pp. 3695–3706).

Zhang, C., Li, Q., & Song, D. (2019). Aspect-based sentiment classification with aspect-specific graph convolutional networks. In *Proceedings of the 2019 Conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 4568–4578).

Zhang, K., Wu, L., Lv, G., Wang, M., Chen, E., & Ruan, S. (2021). Making the relation matters: Relation of relation learning network for sentence semantic matching. In *Proceedings of the AAAI conference on artificial intelligence, vol. 35 no. 16* (pp. 14411–14419).

Zhang, H., Xiao, L., Chen, W., Wang, Y., & Jin, Y. (2018). Multi-task label embedding for text classification. In *Proceedings of the 2018 Conference on empirical methods in natural language processing* (pp. 4545–4553).

Zhu, X., Zhu, L., Guo, J., Liang, S., & Dietze, S. (2021). GL-GCN: Global and local dependency guided graph convolutional networks for aspect-based sentiment classification. *Expert Systems with Applications*, *186*, Article 115712. http://dx.doi.org/10.1016/j.eswa.2021.115712, URL https://www.sciencedirect.com/science/article/pii/S0957417421010940.

Žunić, A., Corcoran, P., & Spasić, I. (2021). Aspect-based sentiment analysis with graph convolution over syntactic dependencies. *Artificial Intelligence in Medicine*, *119*, Article 102138. http://dx.doi.org/10.1016/j.artmed.2021.102138, URL https://www.sciencedirect.com/science/article/pii/S0933365721001317.