Improving Multi-task Stance Detection with Multi-task Interaction Network

Heyan Chai[1], **Siyu Tang**[1], **Jinhao Cui**[1], **Ye Ding**[2], **Binxing Fang**[1,3] and **Qing Liao**[1,3*]
[1] Harbin Institute of Technology, Shenzhen, China
[2] Dongguan University of Technology, China
[3] Peng Cheng Laboratory, Shenzhen, China
{chaiheyan,tangsiyu999}@stu.hit.edu.cn,gcjh777@163.com
dingye@dgut.edu.cn,fangbx@cae.cn,liaoqing@hit.edu.cn

（EMNLP-2022)

code：None

gesis
Leibniz-Institut für Sozialwissenschaften

Reported by Zhaoze Gao

1. Introduction

2. Approach

3. Experiments

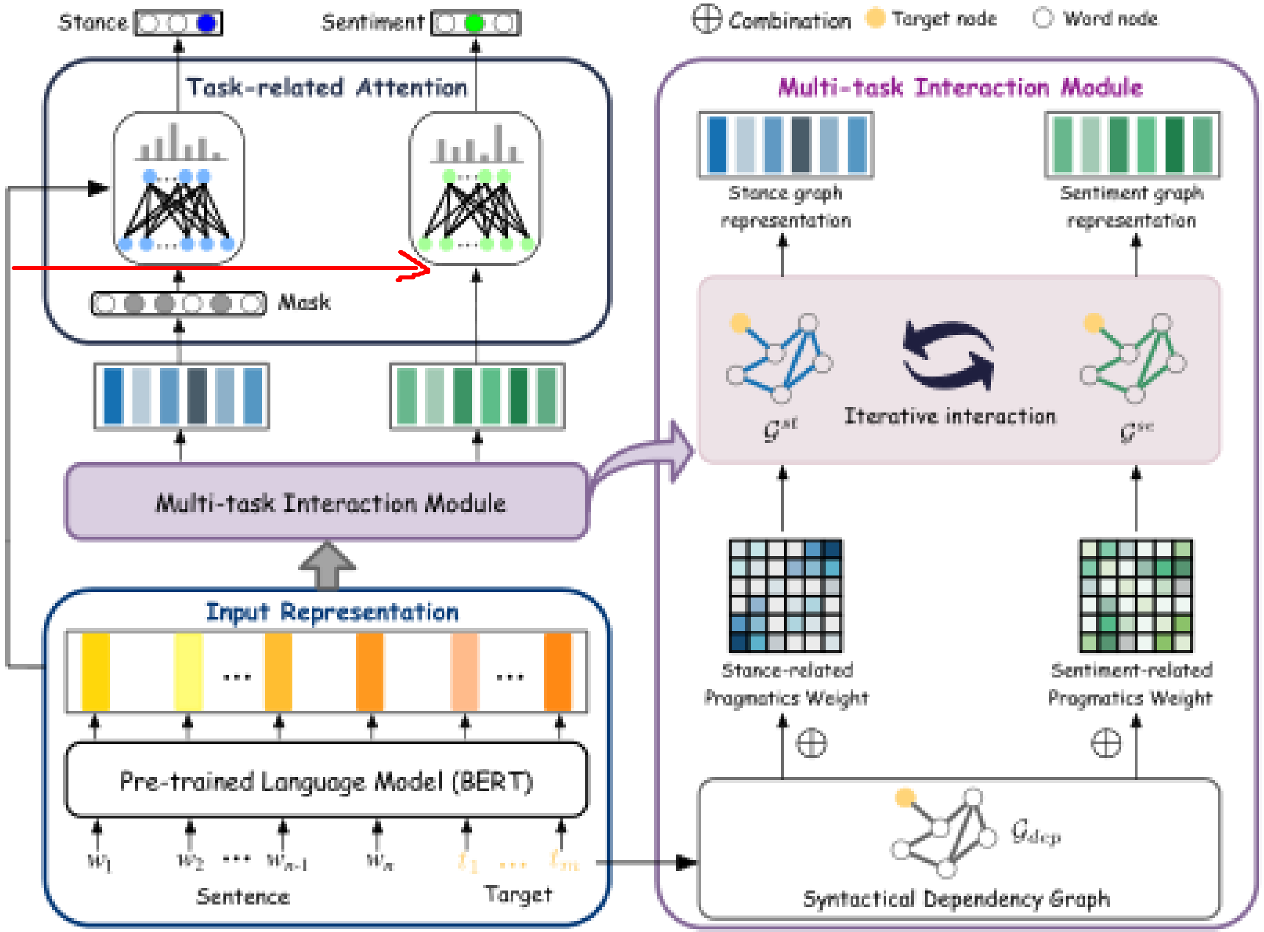
Introduction

	Tweet	Target	Stance	Sentiment
Example 1	Pregnant people have feelings, and the ability to make decisions about their health	Legalization of abortion	Favor	Positive
	They have not the ability and shouldn't make decisions that involve their health	Legalization of abortion	Favor	Negative
Example 2	I have an immune system that works fine, masks harm our immune system	Wearing a face mask	Against	Positive
	I have next to no immune system right now so thanks to all wearing masks	Wearing a face mask	Favor	Positive

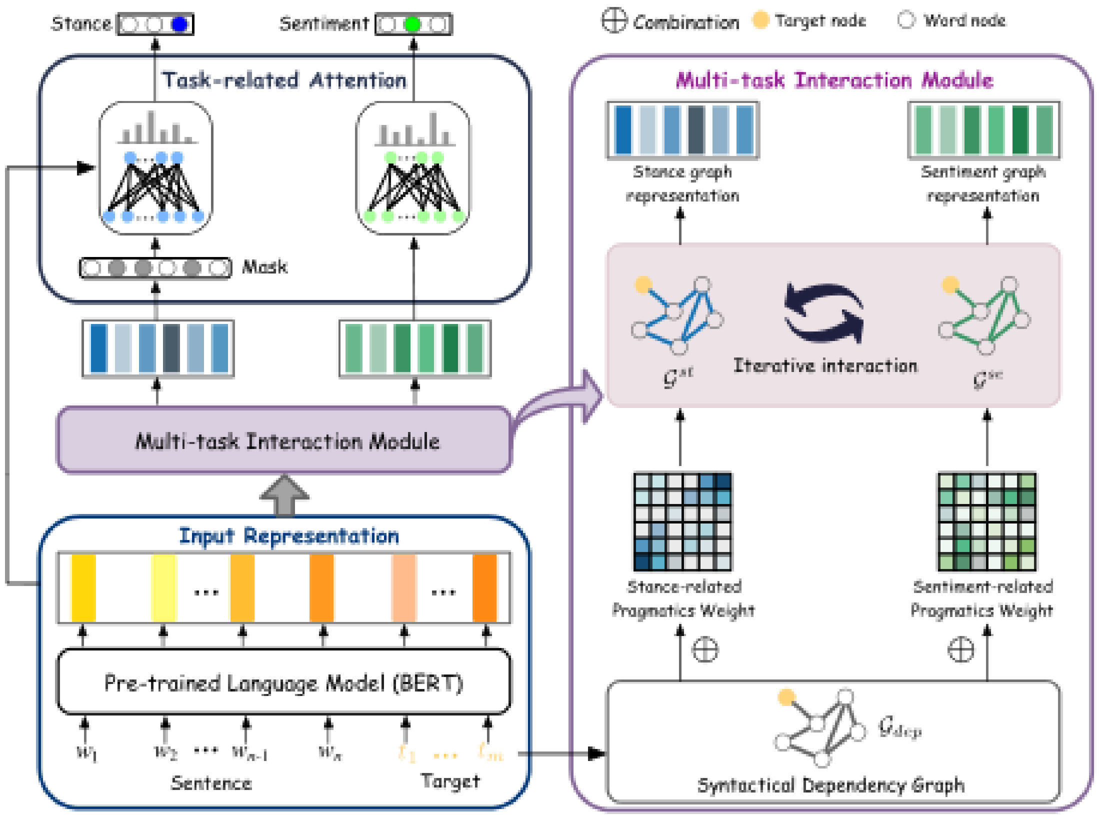
Figure 1: Examples of MTL stance detection.

Above MTL models only focus on extracting contextual task representations and sentence-based features as the shared information across different tasks to identify stances and classify sentiments of sentences, ignoring the word-level task-specific information.

Approach


Stance
Sentiment
Task-related Attention
Mask
Multi-task Interaction Module
Input Representation
Pre-trained Language Model (BERT)
Sentence
Target
Combination
Target node
Word node
Multi-task Interaction Module
Stance graph representation
Sentiment graph representation
Iterative interaction
Stance-related Pragmatics Weight
Sentiment-related Pragmatics Weight
Syntactical Dependency Graph

Approach



$$\mathcal{D}_s = \{(s_i, a_i, y_i^{st}, y_i^{se})\}_{i=1}^{N_s}$$

$$\mathcal{D}_q = \{(s_i, a_i)\}_{i=1}^{N_q}$$

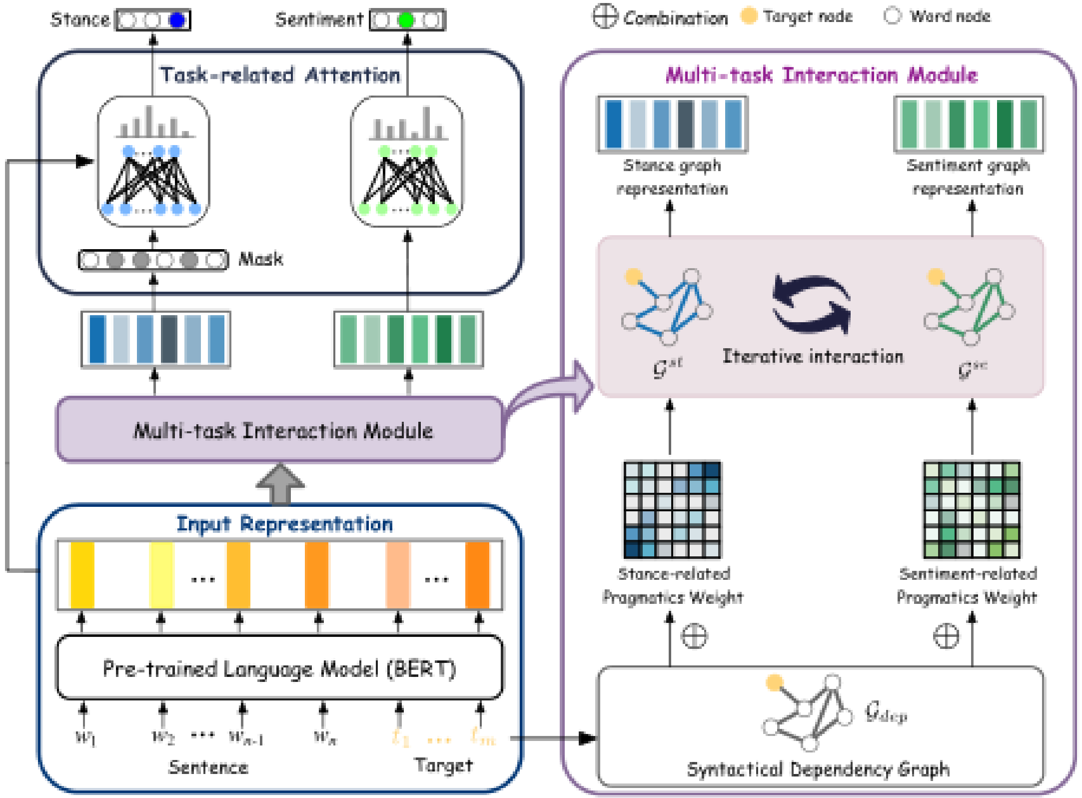
$$\mathcal{F}(s_i, a_i) = \{\hat{y}_i^{se}, \hat{y}_i^{st}\} \approx \{y_i^{se}, y_i^{st}\}.$$

$$s = \{w_i\}_{i=1}^{n} \qquad a = \{w_i\}_{i=1}^{m}$$

$$\boldsymbol{h} = BERT([CLS]s[SEP]a[SEP]) \quad (1)$$

$$\boldsymbol{h} = \{\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_{(n+m)}\} \quad \boldsymbol{h} \in \mathbb{R}^{(n+m)\times d_m}$$

Approach



$$\mathcal{A}_{dep}[i,j] = \begin{cases} 1 & \text{if } \mathcal{T}(w_i, w_j) \text{ or } \mathcal{T}(w_j, w_i), \\ 1 & \text{if } i = j, \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$p(w_i) = \frac{N(w_i)}{N}, \ \varphi(w_i) = \frac{p(w_i) - \mu(p(\cdot))}{\sigma(p(\cdot))} \tag{3}$$

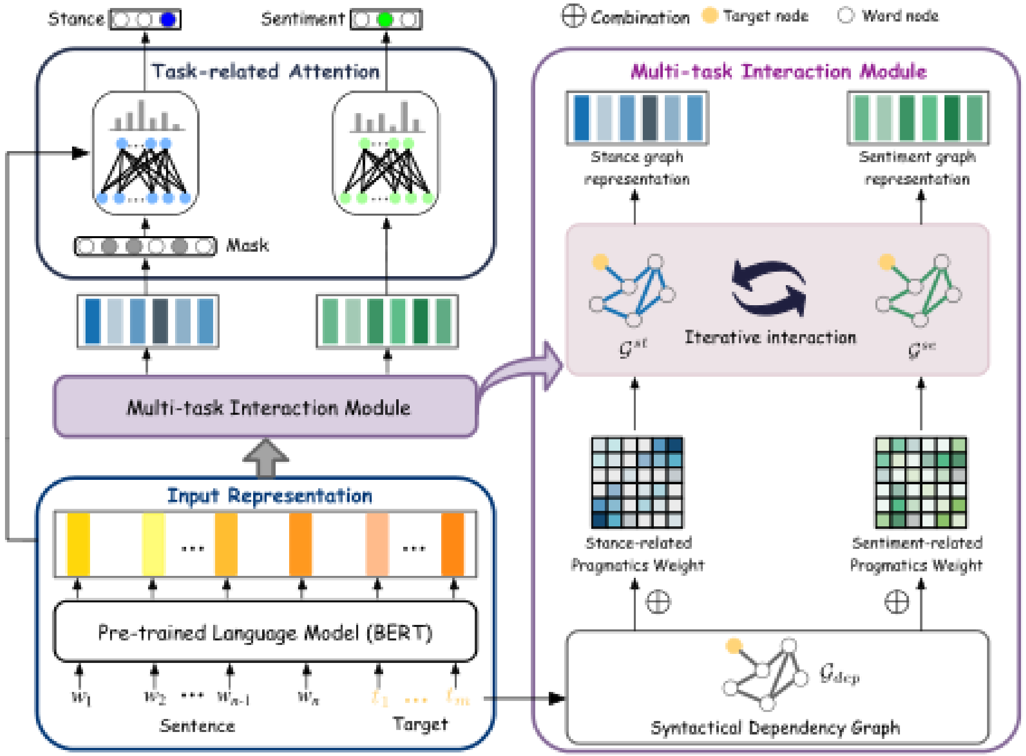
$$\rho^t(w_i) = \left| \frac{N^t(w_i, label_+)}{N^t(label_+)} - \frac{N^t(w_i, label_-)}{N^t(label_-)} \right| \tag{4}$$

$$\phi^t(w_i) = 1 + \frac{\rho^t(w_i) - \mu(\rho^t)}{\sigma(\rho^t)}, \quad w_i \in s \tag{5}$$

$$\mathcal{A}^t[i,j] = \begin{cases} \sum_{k \in \{i,j\}} \phi^t(w_k)\varphi(w_k) & \text{if } \mathcal{A}_{dep}[i,j] = 1, \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$$t \in \{st, se\}$$

Approach



$$e_i = \begin{cases} 1 - \frac{n+1-i}{n+m} & 1 \leq i < n+1 \\ 0 & n+1 \leq i \leq n+m \end{cases} \tag{7}$$

$$\hat{\boldsymbol{h}}_i = e_i \boldsymbol{h}_i.$$

$$\hat{\boldsymbol{h}} = \{\hat{\boldsymbol{h}}_1, \hat{\boldsymbol{h}}_2, \cdots, \hat{\boldsymbol{h}}_{n+m}\}$$

$$\boldsymbol{g}^{st,l} = \mathrm{ReLU}\left(\boldsymbol{E}_{se}^{\frac{1}{2}} \tilde{\mathcal{A}}^{se} \boldsymbol{E}_{se}^{-\frac{1}{2}} \boldsymbol{g}^{se,l-1} \boldsymbol{W}_{st}^{l}\right) \tag{8}$$

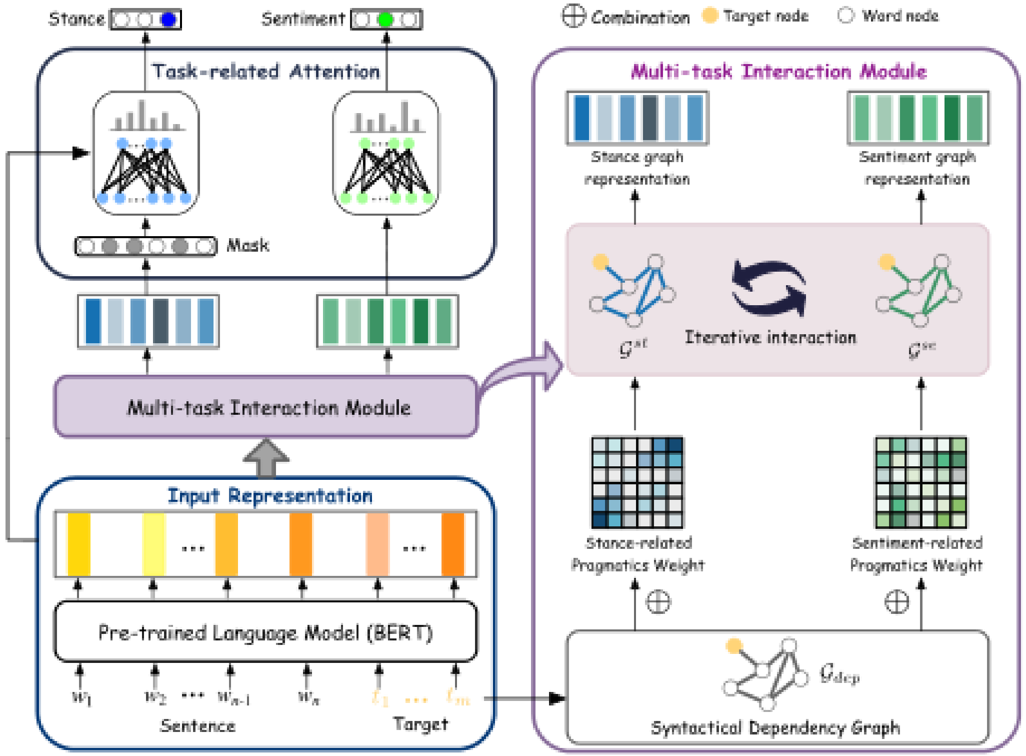
$$\boldsymbol{g}^{se,l} = \mathrm{ReLU}\left(\boldsymbol{E}_{st}^{\frac{1}{2}} \tilde{\mathcal{A}}^{st} \boldsymbol{E}_{st}^{-\frac{1}{2}} \boldsymbol{g}^{st,l} \boldsymbol{W}_{se}^{l}\right) \tag{9}$$

$$\tilde{\mathcal{A}}^t = \mathcal{A}^t + \boldsymbol{I}, \text{ where } \boldsymbol{E}_t = \sum_j (\mathcal{A}_j^t + \boldsymbol{I})$$

$$\boldsymbol{g}^{se,0} = \hat{\boldsymbol{h}} = \{\hat{\boldsymbol{h}}_1, \cdots, \hat{\boldsymbol{h}}_{n+m}\}$$

$$t \in \{st, se\}$$

Approach



$$mask_i = \begin{cases} 0 & 1 \leq i < n+1 \\ 1 & n+1 \leq i \leq n+m \end{cases} \quad (10)$$

$$\hat{\boldsymbol{g}}^{st,L} = mask \times \boldsymbol{g}^{st,L}$$

$$\beta_k = \sum_{i=1}^{m+n} \boldsymbol{h}_k^\top \hat{\boldsymbol{g}}_i^{st,L}, \ \alpha_k = \frac{\exp(\beta_k)}{\sum_{i=1}^{n} \exp(\beta_i)} \quad (11)$$

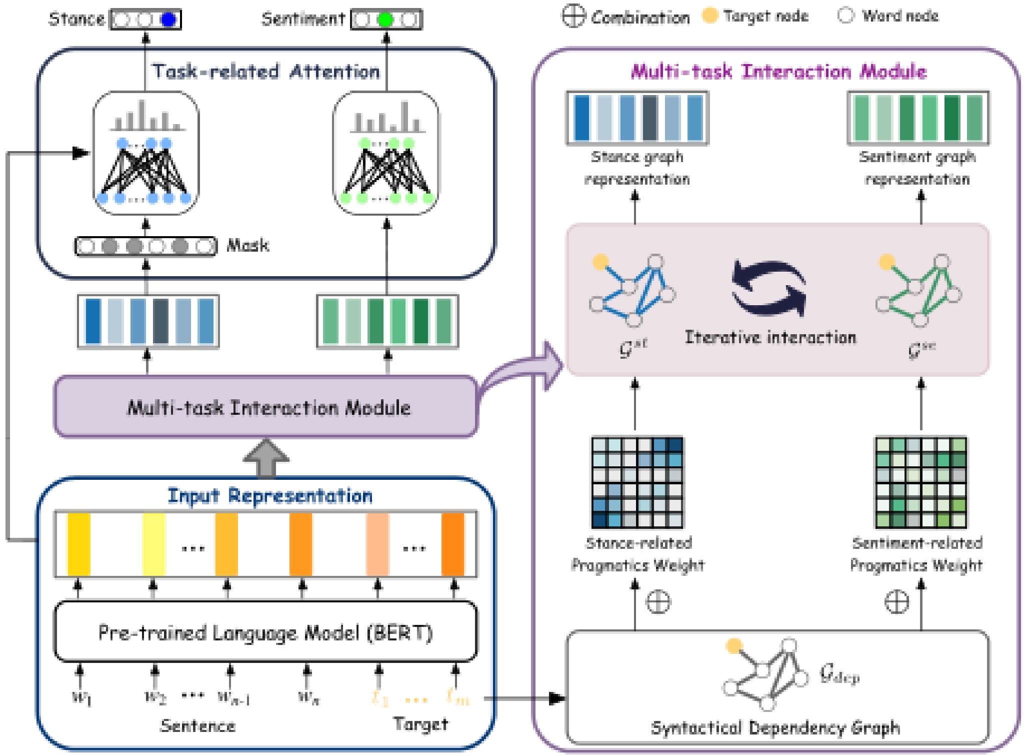
$$\boldsymbol{r}^{st} = \sum_{k=1}^{m+n} \alpha_k \boldsymbol{h}_k$$

$$\beta_k' = \sum_{i=1}^{m+n} \boldsymbol{h}_k^\top \boldsymbol{g}_i^{se,L}, \ \alpha_k' = \frac{\exp(\beta_k')}{\sum_{i=1}^{n} \exp(\beta_i')} \quad (12)$$

$$\boldsymbol{r}^{se} = \sum_{k=1}^{m+n} \alpha_k' \boldsymbol{h}_k$$

$$\hat{\boldsymbol{y}}^t = softmax(\boldsymbol{W}^t \boldsymbol{r}^t + \boldsymbol{b}^t)$$

Approach



$$\mathcal{L} = -\sum_{i=1}^{N_s}(\lambda_1 y_i^{st} \log \hat{y}_i^{st} + \lambda_2 y_i^{se} \log \hat{y}_i^{se}) + \lambda_3 \|\Theta\|_2 \quad (13)$$

$$F_{avg} = \frac{F_{\text{favor}} + F_{\text{against}}}{2} \quad (14)$$

$$F_{\text{favor}} = \frac{2P_{\text{favor}} R_{\text{favor}}}{P_{\text{favor}} + R_{\text{favor}}} \qquad F_{\text{against}} = \frac{2P_{\text{against}} R_{\text{against}}}{P_{\text{against}} + R_{\text{against}}}$$

Experiments

Dataset	Category	Stance Task			Sentiment Task		
		Favor	Against	Neither	Positive	Neutral	Negative
SemEval16	Train	753	1395	766	962	189	1763
	Test	304	715	230	368	90	791
COVID19	Train	1252	750	748	451	360	1939
	Test	306	205	176	108	82	497

Table 1: Statistics for the two datasets.

Experiments

Category	Model	SemEval16 $MacF_{avg}$	SemEval16 F_{avg}	COVID19 $MacF_{avg}$	COVID19 F_{avg}
STL	SVM-ngram (Mohammad et al., 2016)	0.580	0.689	0.584	0.654
	MITRE (Mohammad et al., 2016)	0.560†	0.678†	-	-
	pkudblab (Mohammad et al., 2016)	0.586†	0.673†	-	-
	BERT (Devlin et al., 2019)	0.536	0.646	0.619	0.647
	SCN (Yang et al., 2020a)	0.545	0.613	0.512	0.533
	KEMLM (Kawintiranon and Singh, 2021)	0.556	0.641	0.619	0.644
	TextGCN (Yao et al., 2019)	0.599	0.653	0.557	0.588
	TextING (Zhang et al., 2020)	0.597	0.658	0.589	0.609
	FCS (Sun et al., 2019)	0.602	0.692	-	-
	BERTtoCNN (Li et al., 2021a)	-	0.678	-	-
MTL	AT-JSS (Li and Caragea, 2019)	0.513	0.600	0.566	0.595
	Tchebycheff (Mao et al., 2020)	0.504	0.573	0.581	0.605
	BanditMTL (Mao et al., 2021)	0.540	0.608	0.601	0.619
Ours	MTIN *w/o* SE	0.578	0.645	0.622	0.638
	MTIN-BiLSTM	<u>0.643</u>	<u>0.689</u>	<u>0.647</u>	<u>0.666</u>
	MTIN (Ours)	**0.649**	**0.703**	**0.653**	**0.679**

Table 2: Experimental results of stance detection task on two datasets. Average $MacF_{avg}$ and F_{avg} over 3 runs with random initialization. The best and second-best results are in bold and underlined, respectively. The results with † are retrieved from semantic evaluation (SemEval-2016).

Experiments

Category	Model	SemEval16 ACC_{avg}	SemEval16 $F1\text{-}score_{avg}$	COVID19 ACC_{avg}	COVID19 $F1\text{-}score_{avg}$
STL	BERT (Devlin et al., 2019)	0.776	0.791	0.687	0.734
	ABCDM (Basiri et al., 2021)	0.685	0.651	0.633	0.562
	ISA (Barnes et al., 2021)	0.662	0.632	0.664	0.606
	TextGCN (Yao et al., 2019)	0.775	0.790	0.728	0.695
	TextING (Zhang et al., 2020)	0.774	0.753	0.736	0.564
MTL	AT-JSS (Li and Caragea, 2019)	0.709	0.644	0.669	0.562
	Tchebycheff (Mao et al., 2020)	0.699	0.565	0.604	0.501
	BanditMTL (Mao et al., 2021)	0.654	0.594	0.658	0.580
Ours	MTIN *w/o* ST	0.773	0.798	0.741	0.718
	MTIN-BiLSTM	<u>0.781</u>	**0.802**	<u>0.761</u>	<u>0.746</u>
	MTIN (Ours)	**0.807**	<u>0.796</u>	**0.785**	**0.762**

Table 3: Experimental results of sentiment analysis task on two datasets. Average $F1\text{-}score_{avg}$ and ACC_{avg} over 3 runs with random initialization. The best and second-best results are in bold and underlined, respectively.

Experiments

Model	ST task		SE task	
	$MacF_{avg}$	F_{avg}	ACC_{avg}	$F1\text{-}score_{avg}$
MTIN with *cos sim*	0.573	0.640	0.737	0.725
MTIN with *TF-IDF*	0.526	0.594	0.730	0.727
MTIN with *Fixed*	0.582	0.657	0.771	0.771
MTIN (Ours)	**0.649**	**0.703**	**0.807**	**0.796**

Table 4: Ablation study results of ST and SE task on SemEval16 dataset.

Experiments

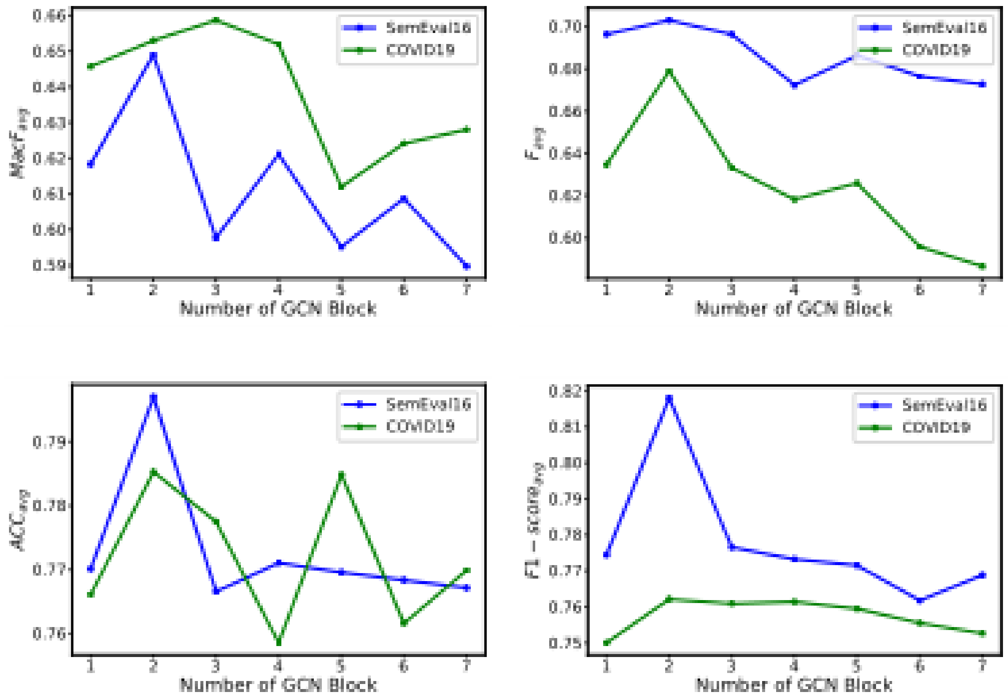
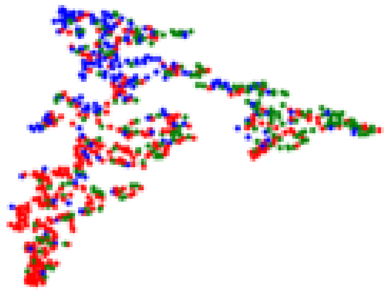
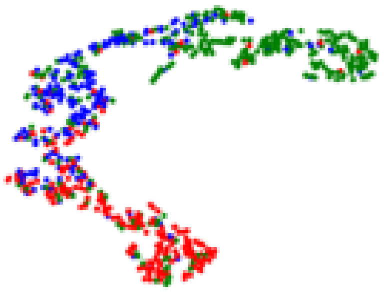


Figure 3: Performance of setting different numbers of interactions between task-related graphs.

Experiments



(a) BERT model



(b) MTIN (**Ours**)

Figure 4: Visualization of intermediate vector representations. Red=*Against*, blue=*Neither*, green=*Favor*.

Experiments

Model	Target	Attention Visualization	Prediction	Label
TextGCN	FM	I believe that every woman should have their own rights .	Against✗	Favor
	WFM	Not wearing a mask because I am in my own greenhouse .	Favor✗	Against
BERT	FM	I believe that every woman should have their own rights .	Favor✓	Favor
	WFM	Not wearing a mask because I am in my own greenhouse .	Favor✗	Against
MTIN	FM	I believe that every woman should have their own rights .	Favor✓	Favor
	WFM	Not wearing a mask because I am in my own greenhouse .	Against✓	Against

Table 5: Case study. Visualization of attention scores from BERT, TextGCN, and MTIN on testing examples of SemEval16 and COVID19 datasets, along with their predictions and corresponding ground truth labels. The target FM (Feminist Movement) and WFM (Wearing a Face Mask) are from SemEval16 and COVID19 datasets respectively. The marker ✓ and ✗ indicate the correct and incorrect predictions, respectively.

Thank you！