Effective Token Graph Modeling using a Novel Labeling Strategy for Structured Sentiment Analysis

Wenxuan Shi, Fei Li, Jingye Li, Hao Fei, Donghong Ji[†]
Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan, China
{shiwenxuan,lifei_csnlp,theodorelee,hao.fei,dhji}@whu.edu.cn

2022.06.8 • ChongQing

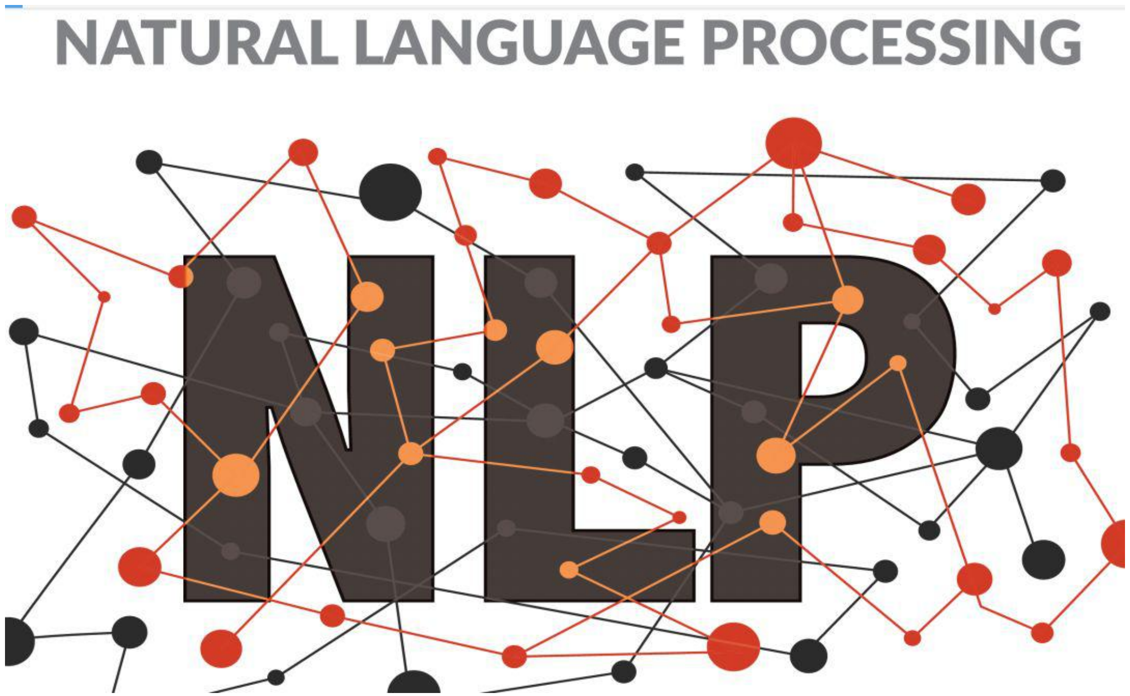
2022_ACL

Leibniz-Institut für Sozialwissenschaften

Reported by Yidan Liu

Code:https://github.com/Xgswlg/TGLS

Artificial
Intelligence



1.Introduction

2.Method

3.Experiments

gesis
Leibniz-Institut
für Sozialwissenschaften

Introduction

- The label proportions for span prediction and span relation prediction are imbalanced.
- The span lengths of sentiment tuple components may be very large in this task, which will further exacerbates the imbalance problem.
- Two nodes in a dependency graph cannot have multiple arcs, therefore some overlapped sentiment tuples cannot be recognized.

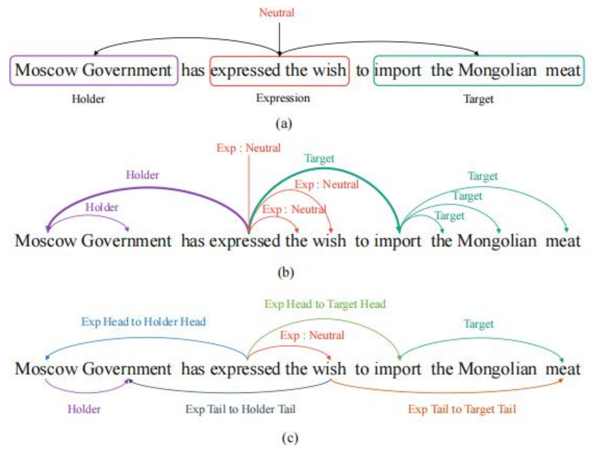


Figure 1: **(a)** An example of structured sentiment analysis. **(b)** The head-first parsing graph proposed by Barnes et al. (2021), where the arcs related to holder(target)-expression linking relations are bold. **(c)** Our proposed *essential label set*, which has more balanced label distribution for holder, target or expression span prediction and their linking relation prediction.

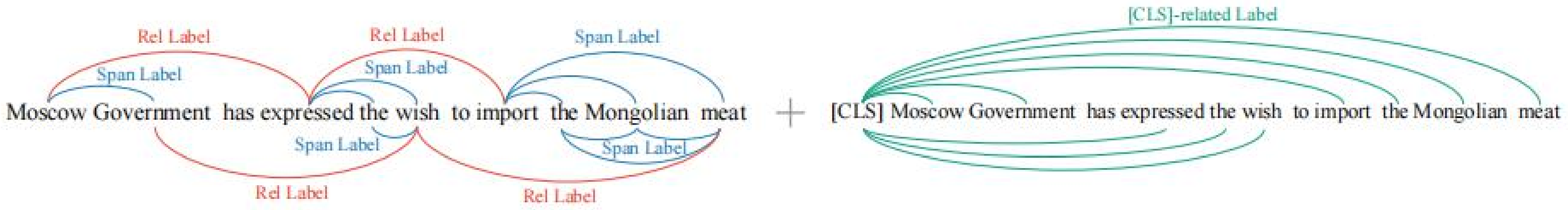


Figure 2: The whole label set contains the labels for span prediction and span relation prediction, as well as the [CLS]-related labels that connect a sentinel [CLS] token with the holder, target and expression tokens.

Method

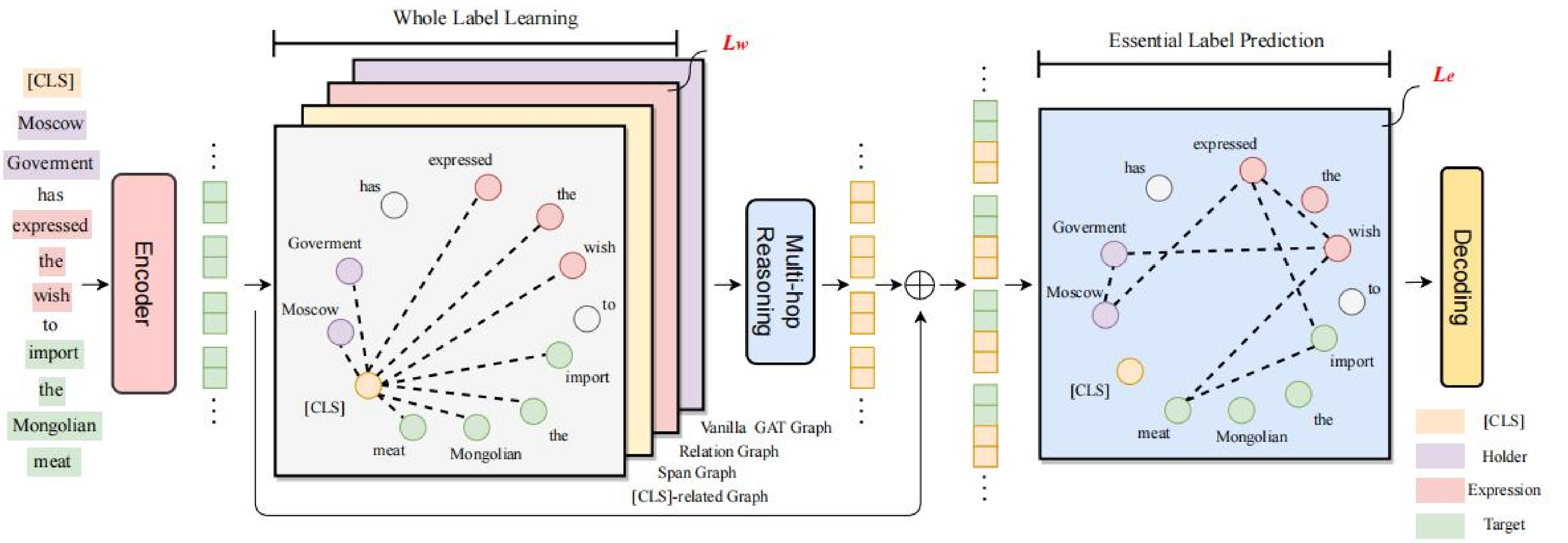
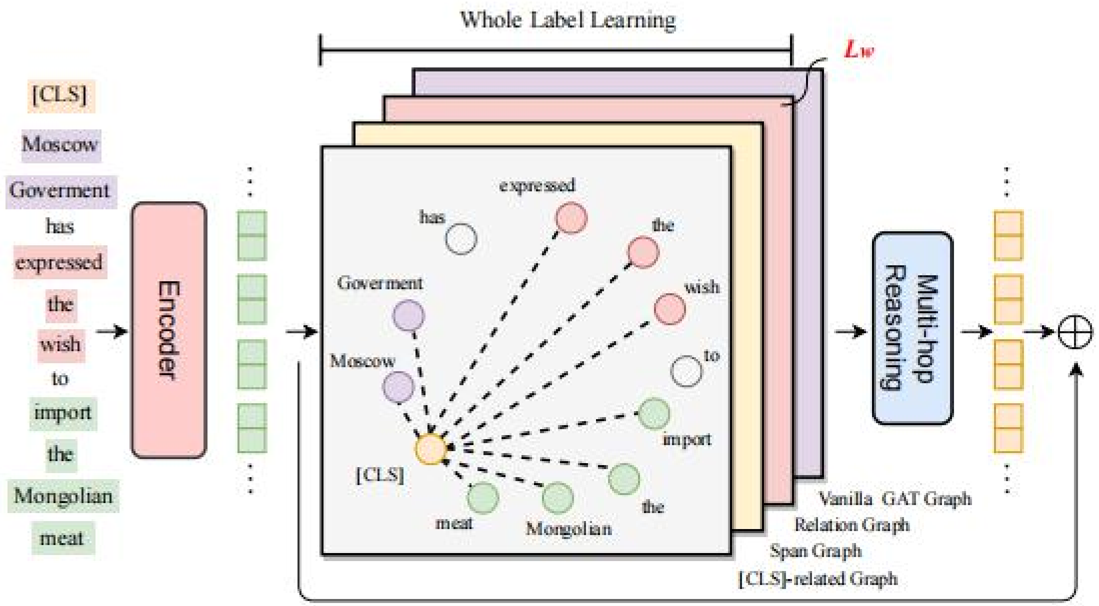


Figure 3: Overall architecture of the our framework. From left to right, the first is an encoder to yield contextualized word representations from input sentences, and the next is a graph layer where we produce attention scoring matrices by whole label prediction. Then we build a multi-hop reasoning layer and refine token representations. Finally, a prediction layer is leveraged for reasoning the relations in essential labels and based on which we decode all components of an opinion tuple.

Method



Encoder Layer

$$\boldsymbol{w}_i = \mathbf{e}_i^{word} \oplus \mathbf{e}_i^{pos} \oplus \mathbf{e}_i^{lemma} \oplus \mathbf{e}_i^{char} \quad (1)$$

$$\boldsymbol{h}_i = \mathrm{BiLSTM}\left(\boldsymbol{w}_i\right) \quad (2)$$

$$\boldsymbol{G} = \left(\mathbf{V}, S_o^{\mathcal{G}}, S_s^{\mathcal{G}}, S_r^{\mathcal{G}}, S_c^{\mathcal{G}}\right) \quad (3)$$

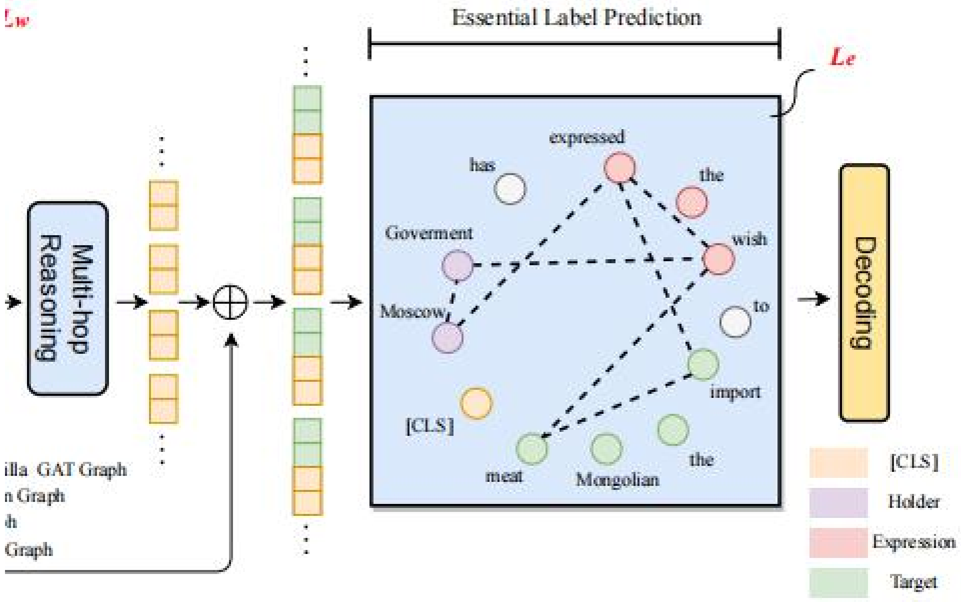
Attention Scoring

$$q_{v,i}, k_{v,j} = MLP_v^q\left(\boldsymbol{h}_i\right), MLP_v^k\left(\boldsymbol{h}_j\right) \quad (4)$$

$$S_{v,ij}^{\mathcal{G}} = \left(q_{v,i}\right)^\top \boldsymbol{R}_{j-i} k_{v,j} \quad (5)$$

$$S^{\mathcal{G}} = \left\{S_{v,ij}^{\mathcal{G}} \mid v \in \{o, s, r, c\}, 1 \leq i, j \leq n\right\} \quad (6)$$

Method



$$\mathcal{L}_w = \sum_i \sum_{j>i} \log \left(e^{TH^{\mathcal{G}}_{ij}} + \sum_{r \in \Omega^{neg}_{ij}} e^{S^{\mathcal{G}}_{r,ij}} \right) + \sum_i \sum_{j>i} \log \left(e^{-TH^{\mathcal{G}}_{ij}} + \sum_{r \in \Omega^{pos}_{ij}} e^{-S^{\mathcal{G}}_{r,ij}} \right) \quad (12)$$

$$\mathcal{L}_{all} = \mathcal{L}_e + \alpha \mathcal{L}_w \quad (13)$$

Multi-hop Reasoning

$$A_v = Softmax\left(S^{\mathcal{G}}_v\right), v \in \{o, s, r, c\} \quad (7)$$

$$\boldsymbol{u}_i^{l+1} = \sigma \left(\frac{1}{N} \sum_v \sum_{j \in \mathcal{N}^v_i} A_{v,ij} \boldsymbol{W}^v_l \boldsymbol{u}^l_j \right) \quad (8)$$

$$\boldsymbol{c}_i = \boldsymbol{h}_i \oplus \boldsymbol{u}_i$$

$$\mathcal{S}^{\mathcal{P}} = \{S^{\mathcal{P}}_r | r \in \mathcal{R}_e\} \quad (9)$$

Adaptive Thresholding

$$\begin{aligned} TH^{\mathcal{P}}_{ij} &= \left(q^{TH}_i\right)^\top \boldsymbol{R}_{j-i} \boldsymbol{k}^{TH}_j \\ TH^{\mathcal{P}} &= \{TH^{\mathcal{P}}_{ij} | 1 \le i, j \le n\} \end{aligned} \quad (10)$$

$$q^{TH}_i = \boldsymbol{W}_q \boldsymbol{h}_i + \boldsymbol{b}_q, k^{TH}_j = \boldsymbol{W}_k \boldsymbol{h}_j + \boldsymbol{b}_k,$$

$$\Omega_{ij} = \{r | S^{\mathcal{P}}_{r,ij} > TH^{\mathcal{P}}_{ij}, r \in \mathcal{R}_e\} \quad (11)$$

Experiments

Dataset	Model	Span				Targeted	Sent. Graph	
		Holder F1	Target F1	Exp. F1	Overall F1	F1	NSF1	SF1
$\text{NoReC}_{\text{Fine}}$	RACL-BERT	-	47.2	56.3	-	30.3	-	-
	Head-first	51.1	50.1	54.4	53.1*	30.5	37.0	29.5
	Head-final	60.4	**54.8**	55.5	55.7*	31.9	39.2	31.2
	TGLS	**60.9**	53.2	**61.0**	**58.1**	**38.1**	**46.4**	**37.6**
$\text{MultiB}_{\text{EU}}$	RACL-BERT	-	59.9	72.6	-	56.8	-	-
	Head-first	60.4	64.0	73.9	69.6*	57.8	58.0	54.7
	Head-final	60.5	64.0	72.1	68.2*	56.9	58.0	54.7
	TGLS	**62.8**	**65.6**	**75.2**	**71.0**	**60.9**	**61.1**	**58.9**
$\text{MultiB}_{\text{CA}}$	RACL-BERT	-	67.5	70.3	-	52.4	-	-
	Head-first	43.0	72.5	71.1	70.5*	55.0	62.0	56.8
	Head-final	37.1	71.2	67.1	70.2*	53.9	59.7	53.7
	TGLS	**47.4**	**73.8**	**71.8**	**71.6**	**60.6**	**64.2**	**59.8**
MPQA	RACL-BERT	-	20.0	31.2	-	17.8	-	-
	Head-first	43.8	51.0	**48.1**	**47.7***	**33.5**	24.5	17.4
	Head-final	**46.3**	49.5	46.0	47.2*	18.6	26.1	18.8
	TGLS	44.1	**51.7**	47.8	47.0	23.3	**28.2**	**21.6**
DS_{Unis}	RACL-BERT	-	44.6	38.2	-	27.3	-	-
	Head-first	28.0	39.9	40.3	40.1*	26.7	31.0	25.0
	Head-final	37.4	42.1	**45.5**	43.0*	29.6	34.3	26.5
	TGLS	**43.7**	**49.0**	42.6	**45.7**	**31.6**	**36.1**	**31.1**

Table 2: Main experimental results of our TGLS model and comparison with previous works. The score marked as bold means the best performance among all the methods. The baseline results with "*" are from our reimplementation, the others are from (Barnes et al., 2021).

Experiments

	Span Overall F1	Targeted F1	SF1
Ours(TGLS)	**58.1**	**38.1**	**37.6**
w/o [CLS]-related graph	57.6	36.9	36.1
w/o span graph	57.2	38.1	37.4
w/o relation graph	57.7	38.0	36.1
w/o vanilla GAT graph	57.8	37.6	36.5
w/o RoPE	57.7	36.4	36.8
w/o adaptive thresholding	56.0	36.3	35.2

Table 3: Experimental results of ablation studies.

	NoReC$_{Fine}$	**MultiB$_{EU}$**	**MultiB$_{CA}$**	**MPQA**	**DS$_{Unis}$**
Head-final	52.3	63.9	67.3	45.0	41.5
TGLS model					
+parsing labels	54.2	65.4	67.5	44.7	43.2
+our labels	**57.8**	**68.7**	**70.1**	**46.1**	**45.7**

Table 4: Experimental results of the relation extraction F1 score, where *parsing labels* denote the dependency-parsing-based labels in head-final setting, *our labels* denote the whole and essential labels.

Experiments

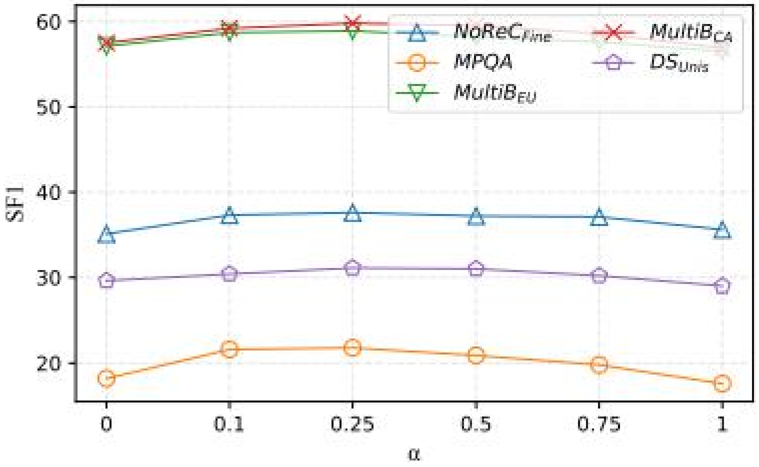


Figure 4: Experimental results (SF1 score) using different α to control the impact of the whole label prediction.

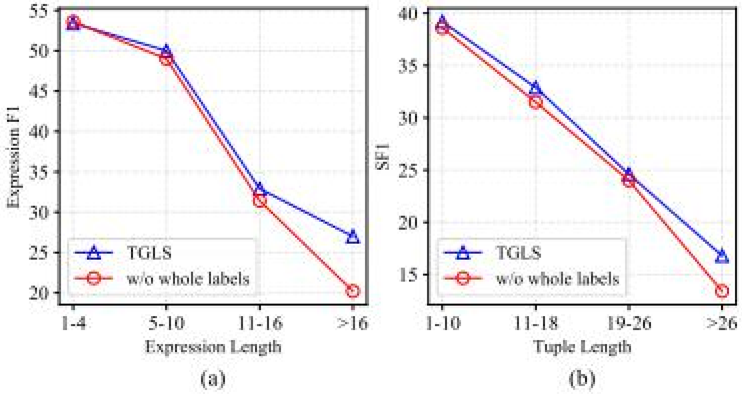


Figure 5: Analysis on the effect of the whole label set for long span identification. **(a)** Expression F1 scores regarding to different expression lengths. **(b)** SF1 scores regarding to different tuple lengths.

Thank you!