Target Really Matters: Target-aware Contrastive Learning and Consistency Regularization for Few-shot Stance Detection

Rui Liu[1,2], Zheng Lin[1,2*], Huishan Ji[1,2], Jiangnan Li[1,2], Peng Fu[1] and Weiping Wang[1]
[1]Institute of Information Engineering, Chinese Academy of Sciences
[2]School of Cyber Security, University of Chinese Academy of Sciences
{liurui1995,linzheng,jihuishan,lijiangnan}@iie.ac.cn
{fupeng,wangweiping}@iie.ac.cn

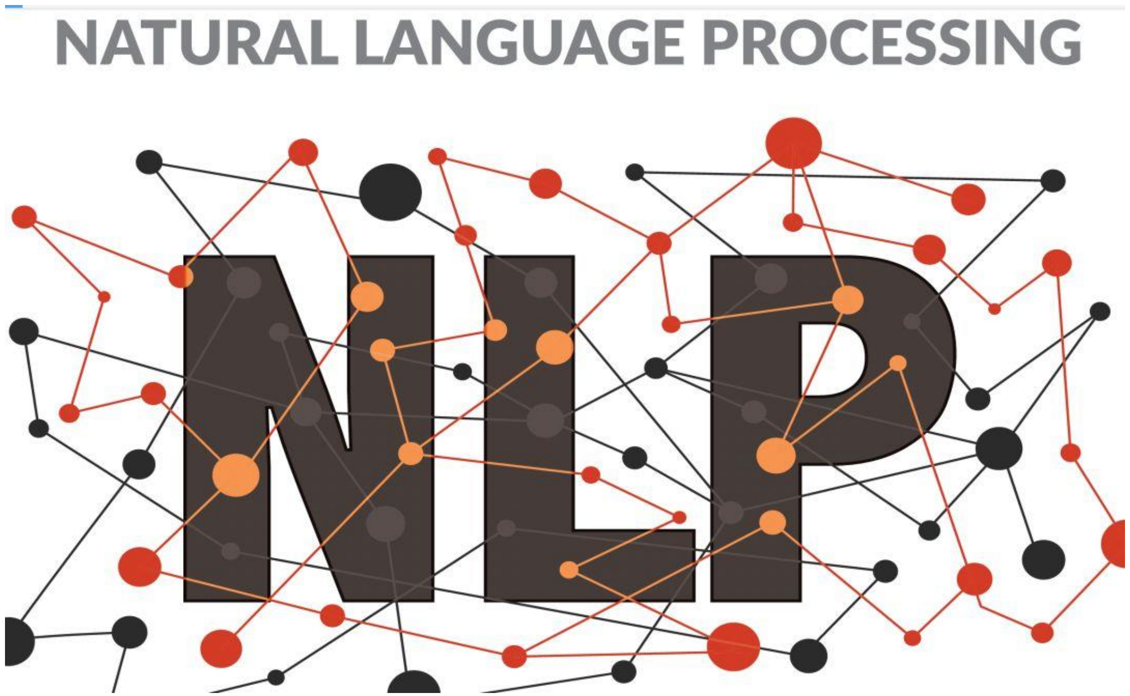
code: https://github.com/monolith-v1/STCC

2023. 4. 27 • ChongQing

2022_COLING

Reported by Junhao Cao

gesis Leibniz-Institut für Sozialwissenschaften



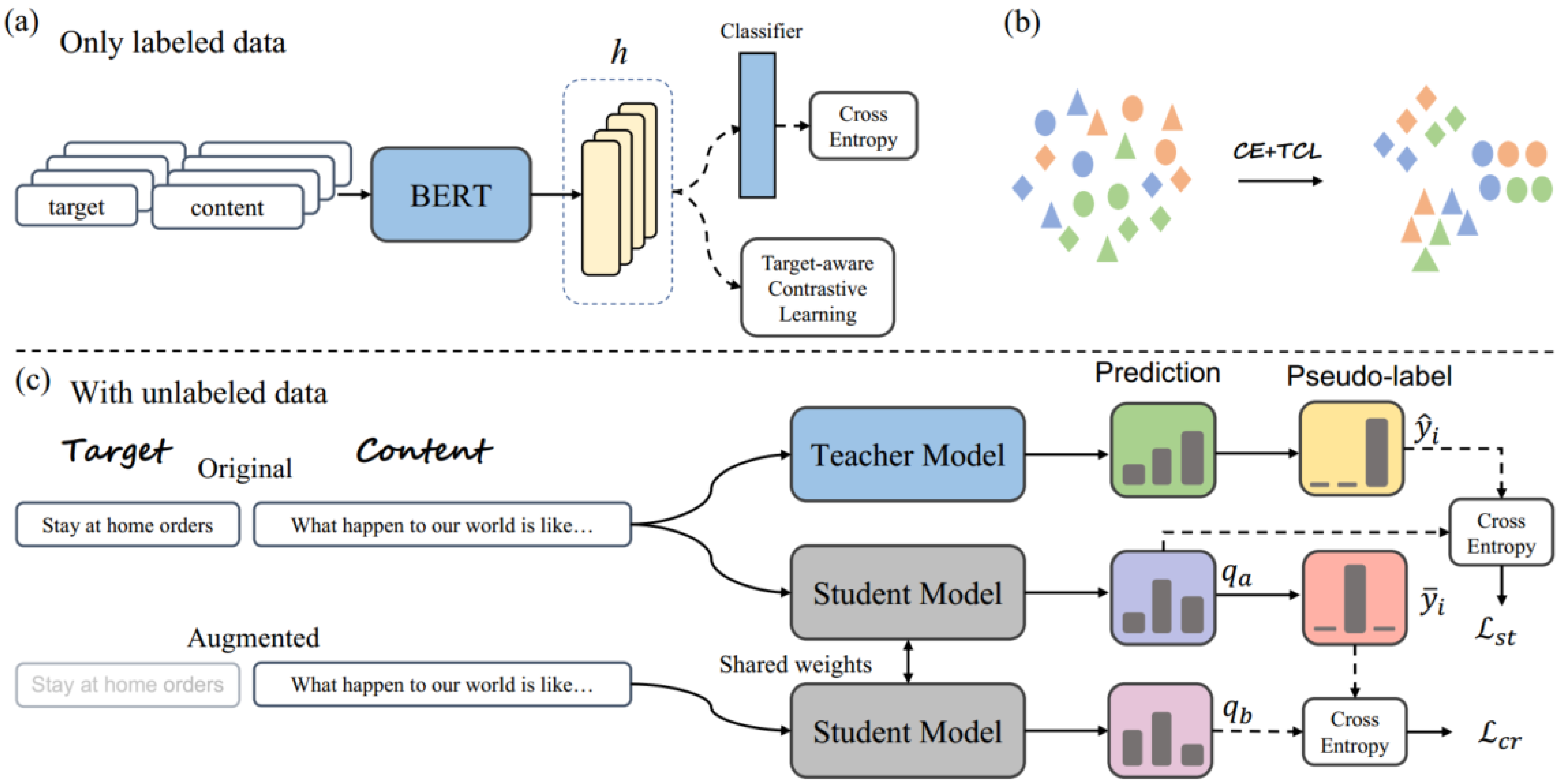
1. Introduction
2. Method
3. Experiments

gesis
Leibniz-Institut
für Sozialwissenschaften

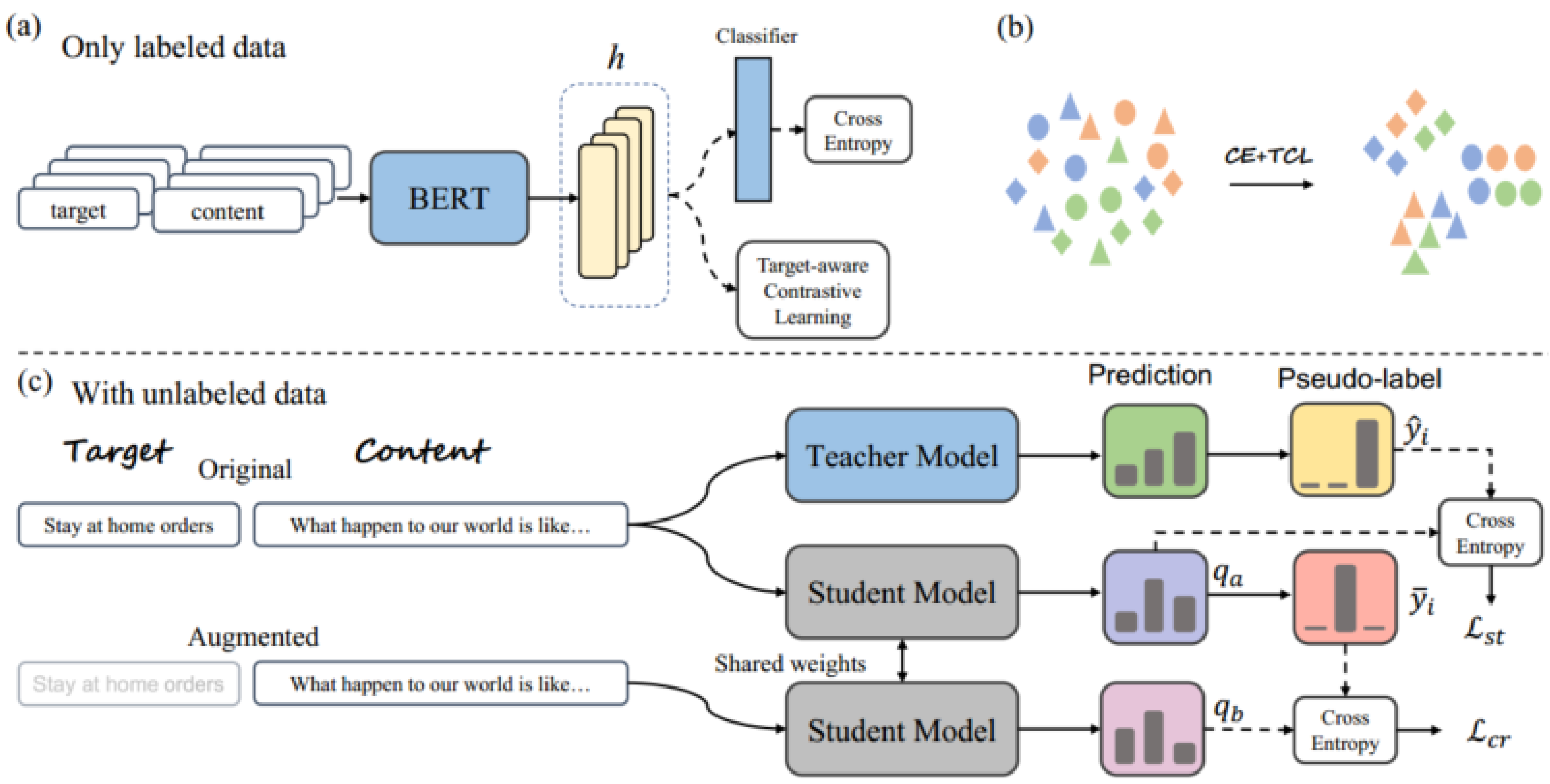
Introduction

it is extremely time-consuming and budget unfriendly to collect sufficient high-quality labeled data for every new target under fully supervised learning, whereas unlabeled data can be collected easier.

Method


(a)
Only labeled data
target
content
BERT
h
Classifier
Cross Entropy
Target-aware Contrastive Learning
(b)
CE+TCL
(c)
With unlabeled data
Target
Original
Content
Stay at home orders
What happen to our world is like…
Augmented
Stay at home orders
What happen to our world is like…
Teacher Model
Student Model
Shared weights
Student Model
Prediction
Pseudo-label
$\hat{y}_i$
q_a
$\bar{y}_i$
q_b
Cross Entropy
$\mathcal{L}_{st}$
Cross Entropy
$\mathcal{L}_{cr}$

Method



BERT Model

$$\mathcal{X} = \{(x_i, t_i, y_i)\}_{i=1}^{N_l}$$

$$\mathcal{U} = \{x_i, t_i\}_{j=1}^{N_u}$$

[CLS] $\boldsymbol{t}_i$ [SEP] $\boldsymbol{x}_i$ [SEP]

$$p(\hat{y}_i|x_i, t_i) = softmax(W_h h_i^{[CLS]})$$

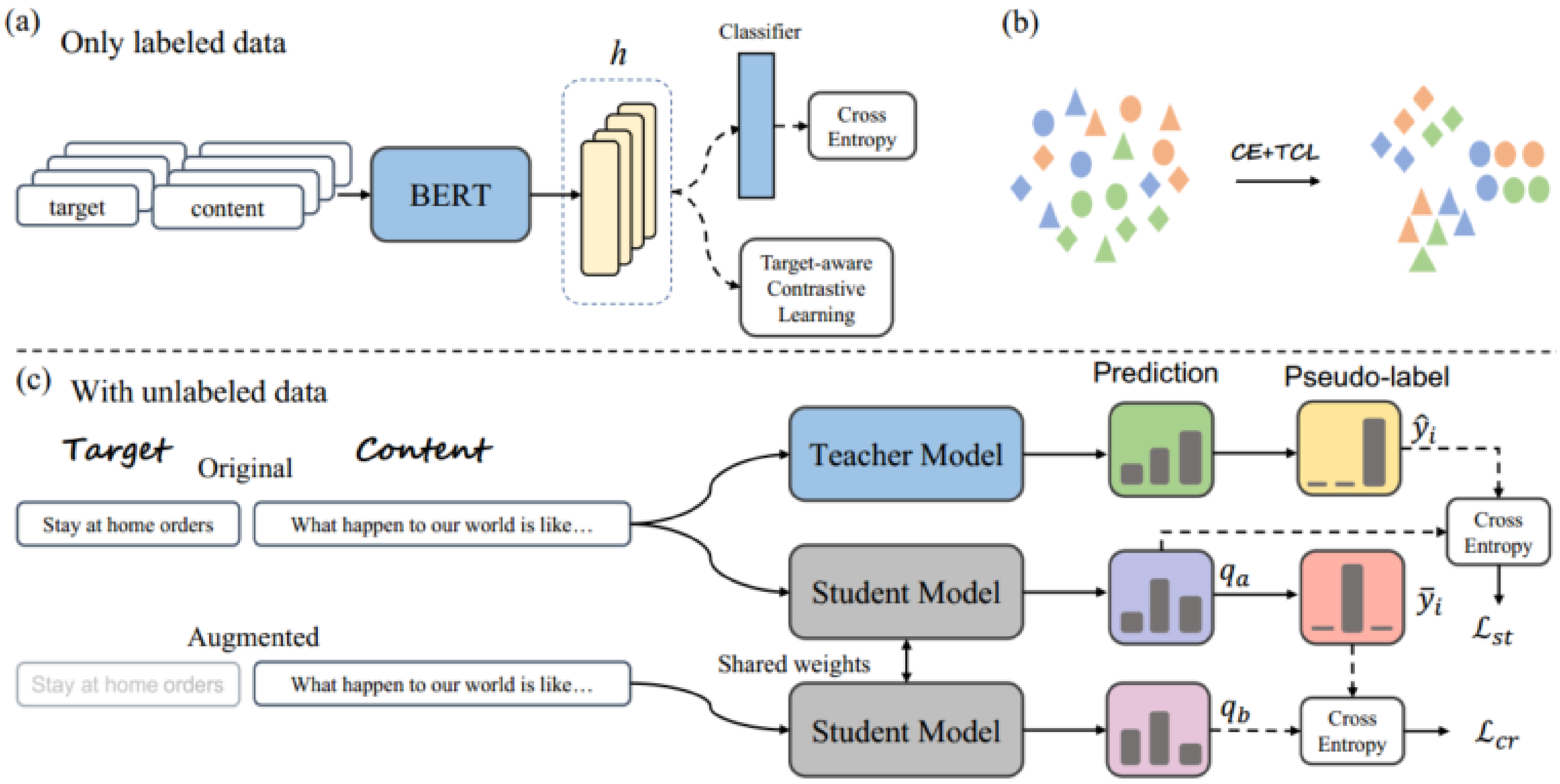
$$\mathcal{L}_{ce} = -\frac{1}{|\mathcal{X}|}\sum_{i}^{|\mathcal{X}|} CE(p(\hat{y}_i|x_i, t_i), y_i), \quad (1)$$

$$\mathcal{L}_{sup}^{i} = -\sum_{i=1}^{B} \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \log \frac{e^{sim(\boldsymbol{h}_i, \boldsymbol{h}_j/\tau)}}{\sum_{k \in C(i)} e^{sim(\boldsymbol{h}_i \cdot \boldsymbol{h}_k/\tau)}}, \quad (2)$$

$$\mathcal{L}_{tcl} = -\sum_{i=1}^{B} \mathcal{L}_{tcl}^{i} \quad (3)$$

$$\mathcal{L}_{tcl}^{i} = \frac{1}{|\mathcal{T}_i|} \sum_{j \in \mathcal{T}_i} \log \frac{e^{sim(\boldsymbol{h}_i, \boldsymbol{h}_j/\tau)}}{\sum_{k \in C(i)} e^{sim(\boldsymbol{h}_i \cdot \boldsymbol{h}_k/\tau)}}, \quad (4)$$

Method



Target-Aware Consistency Regularization in Semi-supervised Learning

$$q_a = p_s(\bar{y}_i | x_i, t_i)$$

$$\bar{y}_i = argmax(q_a)$$

$$q_b = p_s(\tilde{y}_i | x_i)$$

$$\mathcal{L}_{cr} = -\frac{1}{|\mathcal{U}|}\sum_{i}^{|\mathcal{U}|} CE(q_b, \bar{y}_i), \quad (5)$$

$$\mathcal{L}_{st} = -\frac{1}{|\mathcal{U}|}\sum_{i}^{|\mathcal{U}|} CE(q_a, \hat{y}_i), \quad (6)$$

Experiment

Model	SemEval2016 5	SemEval2016 10	SemEval2016 20	COVID19 5	COVID19 10	COVID19 20
without unlabeled data						
CrossNet (Xu et al., 2018)	29.82	33.85	35.37	31.20	34.93	44.32
BERT (Devlin et al., 2019)	41.12	44.45	49.72	32.45	36.85	50.83
ProtoNets(Snell et al., 2017)	41.50	44.13	48.72	33.90	40.50	48.44
SCL (Gunel et al., 2021)	**48.02**	49.40	52.22	37.40	42.23	52.83
PT-HCL (Liang et al., 2022)	34.72	39.56	45.22	-	-	-
Prompt (Schick and Schütze, 2021)	37.88	41.80	43.74	34.96	37.46	47.52
TCL (Ours)	47.32	**51.41**	**53.47**	**40.27**	**46.80**	**53.52**
with unlabeled data						
Prompt (Schick and Schütze, 2021)	37.93	42.41	43.80	34.96	49.42	47.11
UDA (Xie et al., 2020a)	46.86	46.77	50.87	40.27	47.02	53.52
ST (Glandt et al., 2021)	48.35	51.12	55.01	42.08	47.66	55.67
UPS (Rizve et al., 2021)	43.45	48.11	52.73	41.45	44.37	53.87
CL (Cascante-Bonilla et al., 2021)	48.92	51.34	55.42	40.96	50.22	56.86
STCC (Ours)	**52.84**	**55.00**	**57.11**	**44.38**	**52.26**	**58.06**
BERT w/ full data		68.34			73.12	

Table 1: Summary of test results for few-shot stance detection using the shot size of 5, 10, 20 for training. The best results are in bold.

Experiment

Model	5-shot	10-shot	20-shot
STCC	52.84	55.00	57.11
-TCR	51.32	53.53	56.48
-TCL	49.91	52.32	56.18
-ST	43.45	47.70	52.65
-ST&TCR	47.32	51.41	53.47
-ST&TCL	45.19	46.29	48.96
-TCR&TCL	48.92	51.34	55.42
-TCR&TCL + SCL	49.32	51.40	55.92
-ST&TCR&TCL	41.12	44.45	49.72

Table 2: Ablation results on SemEval-2016. "ST" means self-training procedure, and "TCR" means target-aware consistency regularization.

Experiment

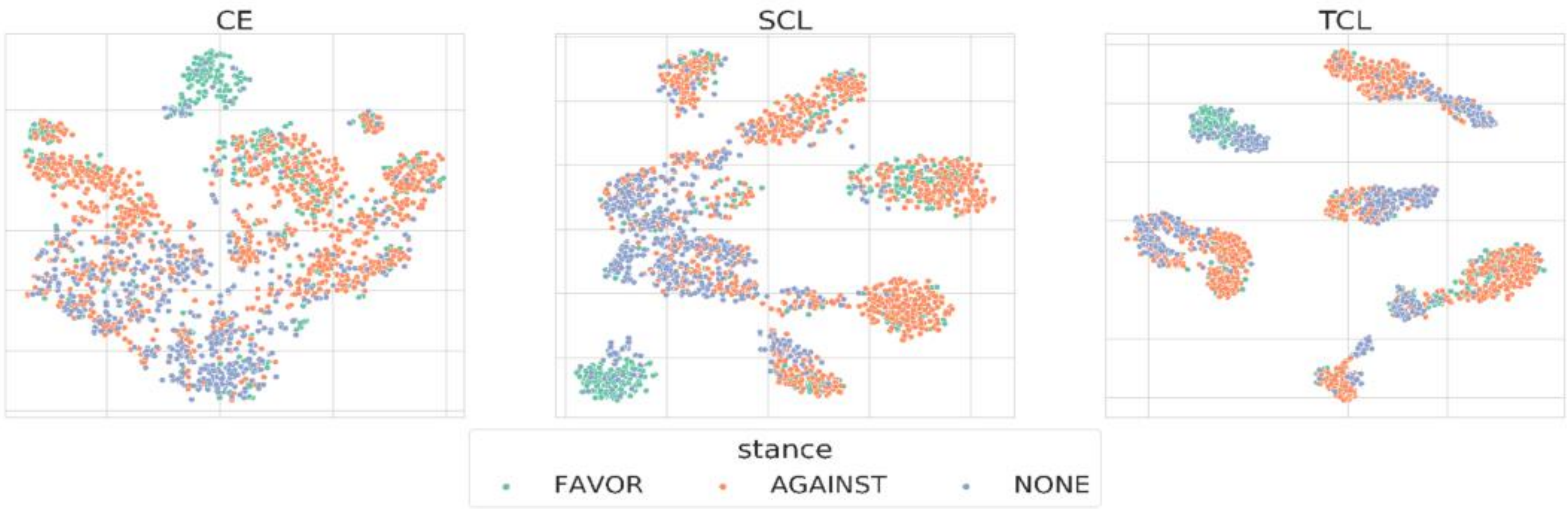


Figure 2: 2D t-SNE plots of the learned [CLS] representations on the unlabeled data of the SemEval2016 from the BERT models only trained by 20-shot labeled samples for every target, which are fine-tuned on different objectives CE (left), SCL (middle), and TCL (right).

Experiment

Model	5-shot	10-shot	20-shot
NT	52.84	55.00	57.11
BT	52.13	52.21	55.96
SR	52.26	54.28	55.23
RD	50.92	52.35	55.99

Table 3: The performance of different data augmentation in semi-supervised learning on SemEval2016. “NT”: masking the target, “BT”: back translation, “SR”: synonym replacement, “RD”: random deletion.

Experiment

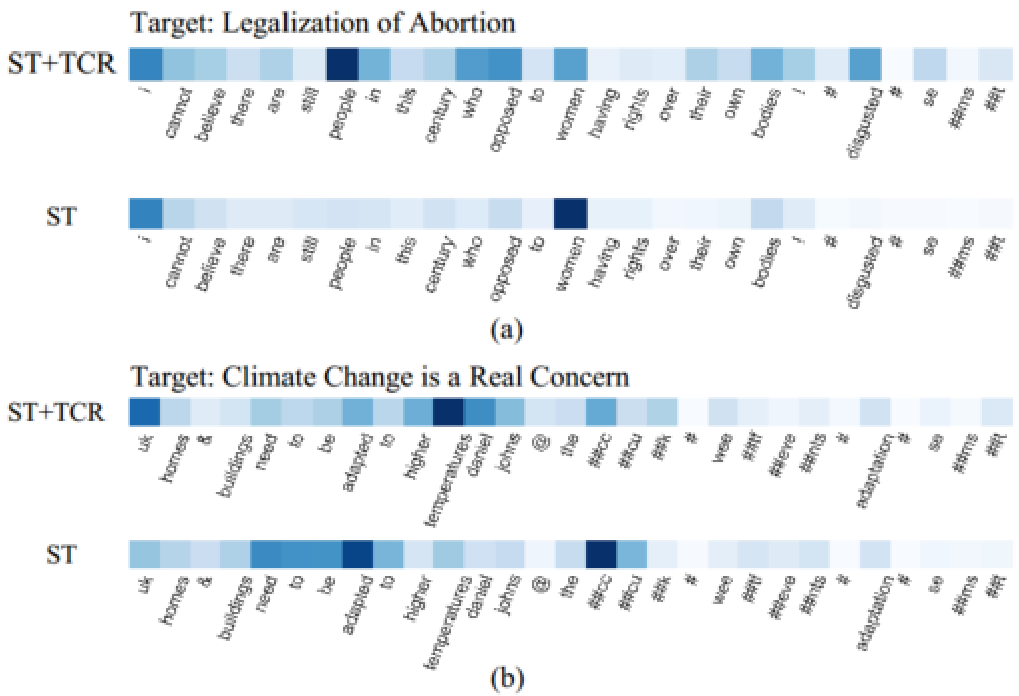


Figure 3: The heatmap of the attention weights of [CLS] towards each subword in the content for "ST+TCR" and "ST" under semi-supervised setting. The attention weights are averaged from the multi-heads in the top layer. The darker the color, the greater the weight.

Thank you!

Leibniz-Institut
für Sozialwissenschaften