Chongqing University of Technology

JointCL: A Joint Contrastive Learning Framework for Zero-Shot Stance Detection

Bin Liang[1,2]*, Qinglin Zhu[1]*, Xiang Li[1]*, Min Yang[3], Lin Gui[4], Yulan He[4,5], Ruifeng Xu[1,6]†

[1]School of Computer Science and Technology,
Harbin Institute of Technology, Shenzhen, China
[2] Joint Lab of HITSZ and China Merchants Securities, Shenzhen, China
[3] SIAT, Chinese Academy of Sciences, Shenzhen, China
[4] Department of Computer Science, University of Warwick, U.K
[5] The Alan Turing Institute, UK, [6] Peng Cheng Laboratory, Shenzhen, China
{bin.liang, zhuqinglin, xiangli}@stu.hit.edu.cn
min.yang@siat.ac.cn, {lin.gui, Yulan.He}@warwick.ac.uk
xuruifeng@hit.edu.cn

https://github.com/HITSZ-HLT/JointCL

—— ACL2022

Reported by Yuyang Lai

1.Introduction

2.Method

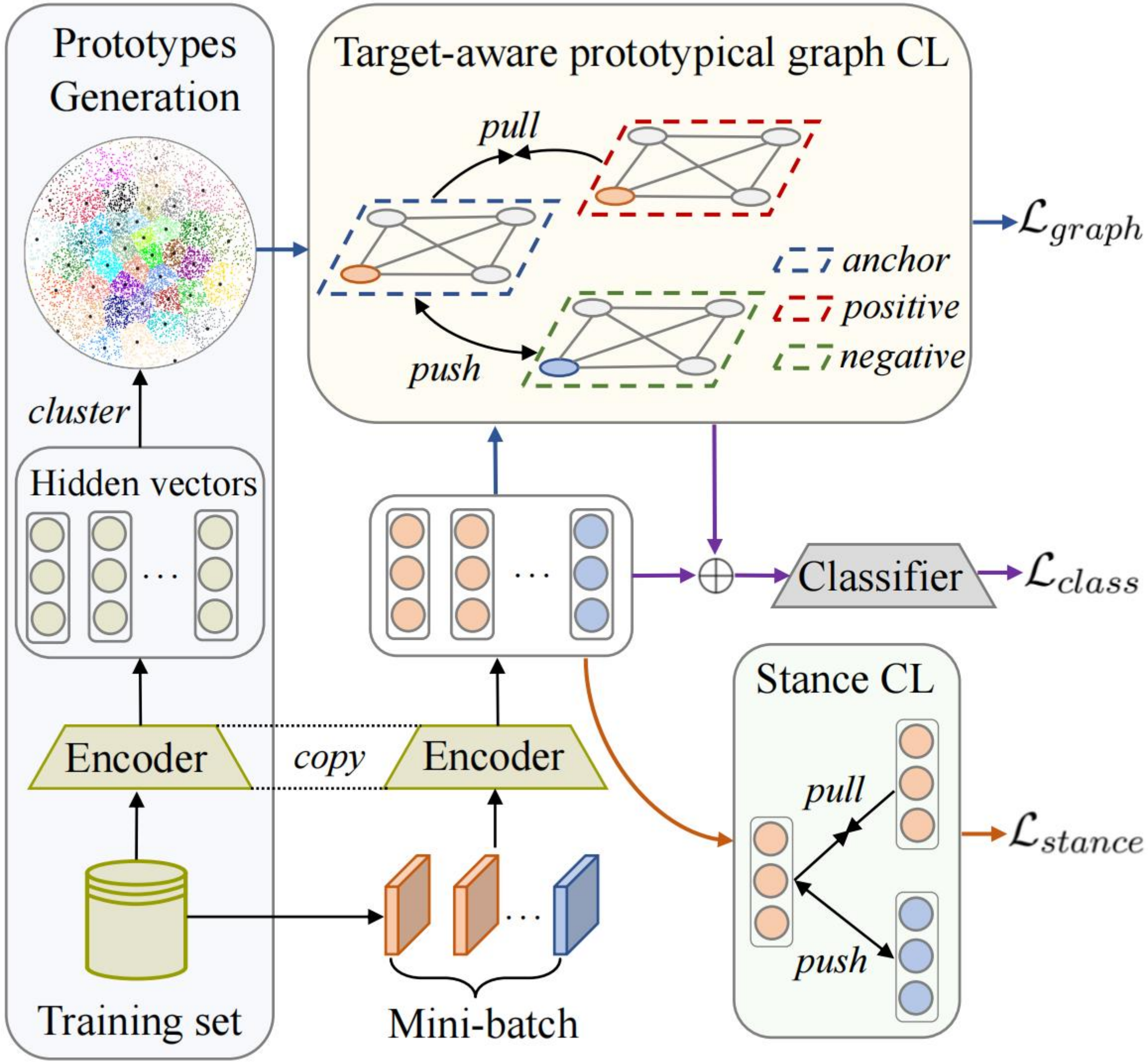
3.Experiments

Introduction

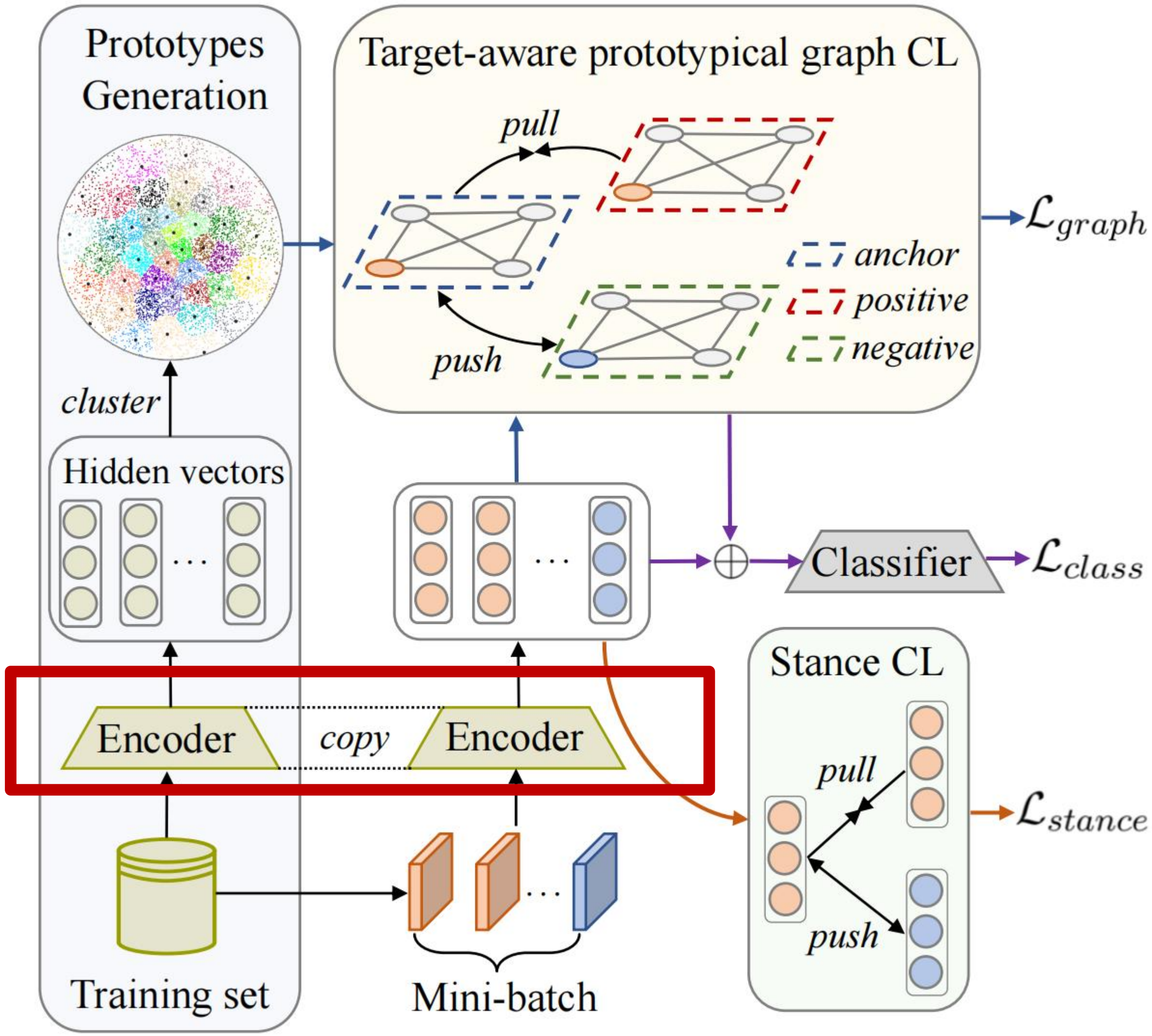
context-aware ——Stance Contrastive Learning

target-aware ——Target-Aware Prototypical Graph Contrastive Learning

Method

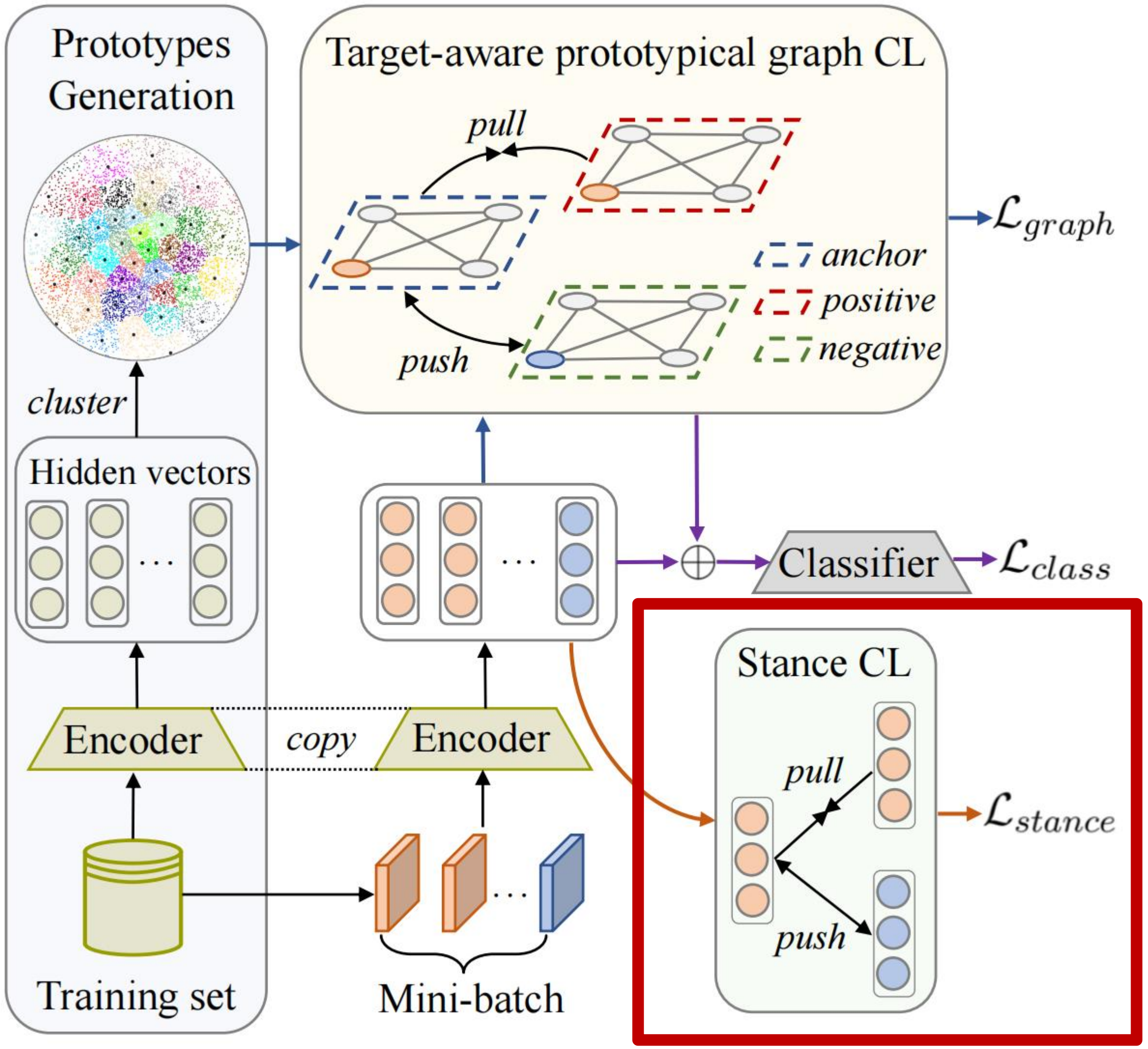

Prototypes Generation
Target-aware prototypical graph CL
pull
push
anchor
positive
negative
$\mathcal{L}_{graph}$
cluster
Hidden vectors
Classifier
$\mathcal{L}_{class}$
Stance CL
pull
push
$\mathcal{L}_{stance}$
Encoder
copy
Encoder
Training set
Mini-batch

Method



$$\boldsymbol{h} = \mathrm{BERT}([CLS]r[SEP]t[SEP])_{[CLS]} \quad (1)$$

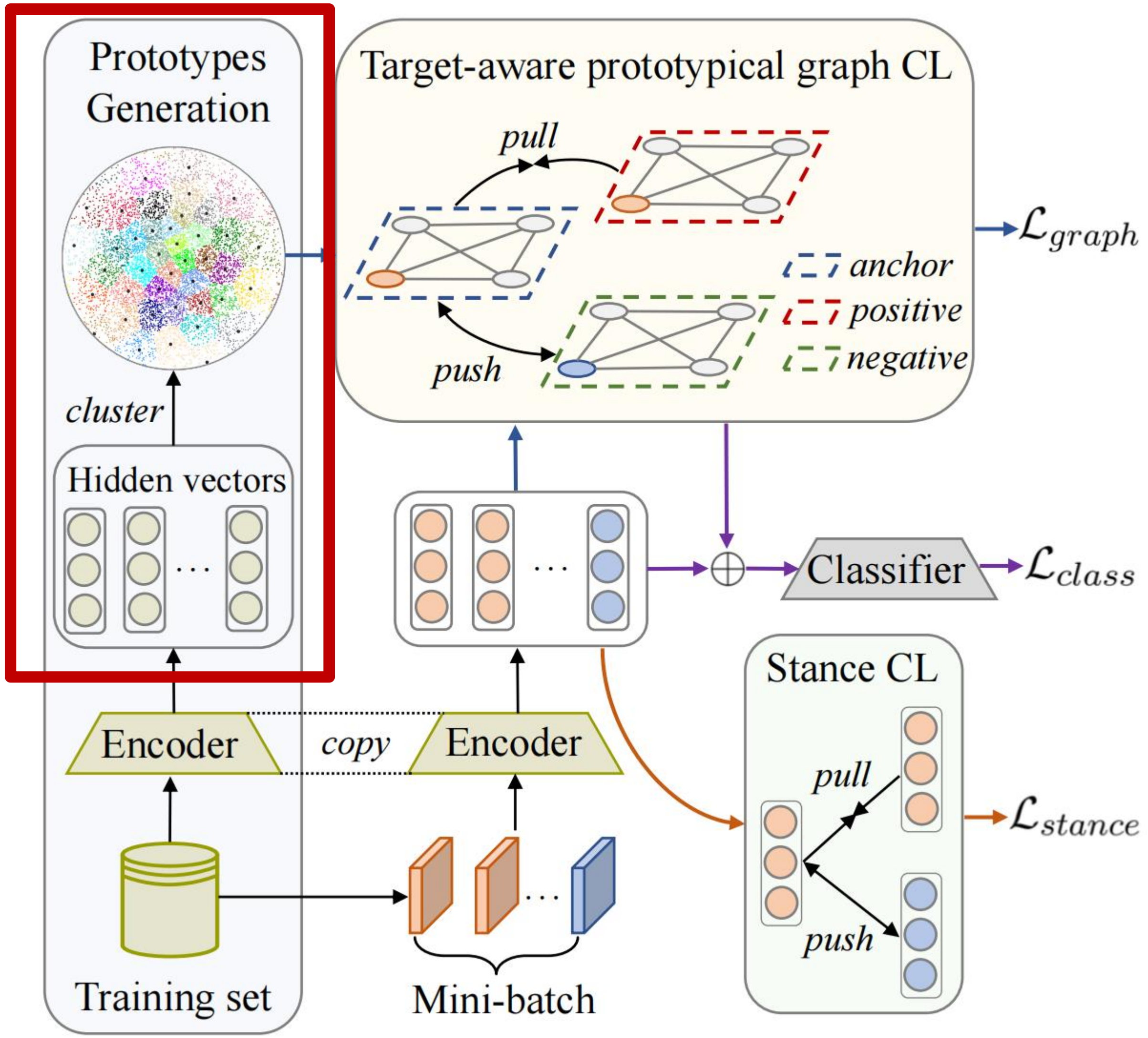
Method



$$\mathcal{L}_{stance} = \frac{-1}{N_b} \sum_{\boldsymbol{h}_i \in \mathcal{B}} \ell^s(\boldsymbol{h}_i) \tag{2}$$

$$\ell^s(\boldsymbol{h}_i) = \log \frac{\sum_{j \in \mathcal{B} \setminus i} \mathbb{1}_{[y^i = y^j]} \exp(f(\boldsymbol{h}_i, \boldsymbol{h}_j)/\tau_s)}{\sum_{j \in \mathcal{B} \setminus i} \exp(f(\boldsymbol{h}_i, \boldsymbol{h}_j)/\tau_s)} \tag{3}$$

Method



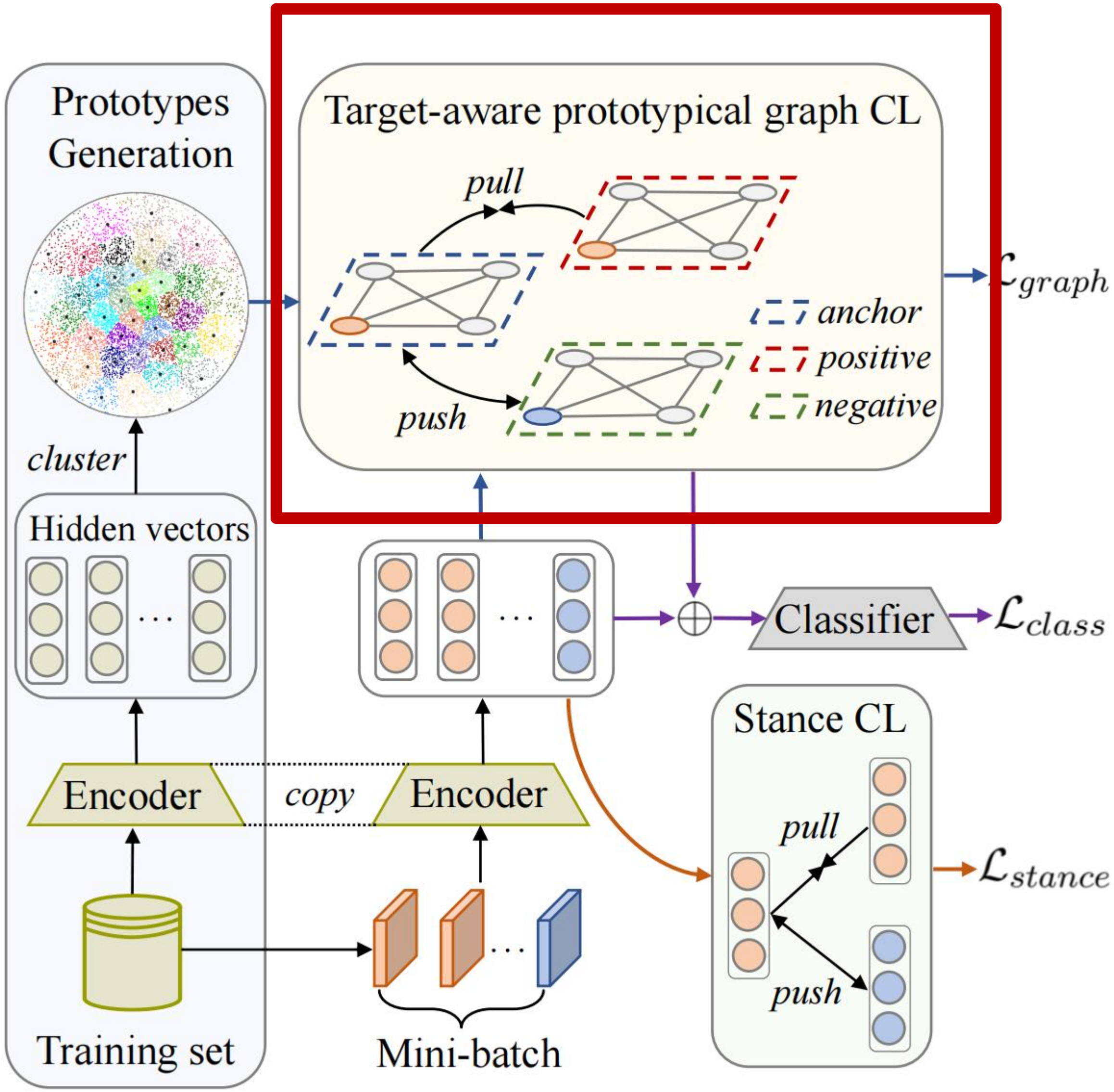
$$\mathcal{H} = \{\boldsymbol{h}_i\}_{i=1}^{N_s}$$

$$\mathcal{C} = \{\boldsymbol{c}_i\}_{i=1}^{k}$$

$$\boldsymbol{X} = [\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_k, \boldsymbol{h}_i],$$

$$\mathcal{G} \in \mathbb{R}^{(k+1)\times(k+1)}$$

Method



$$\boldsymbol{\alpha}_i = a(\text{GAT}(\boldsymbol{X}; \mathcal{G})) \tag{4}$$

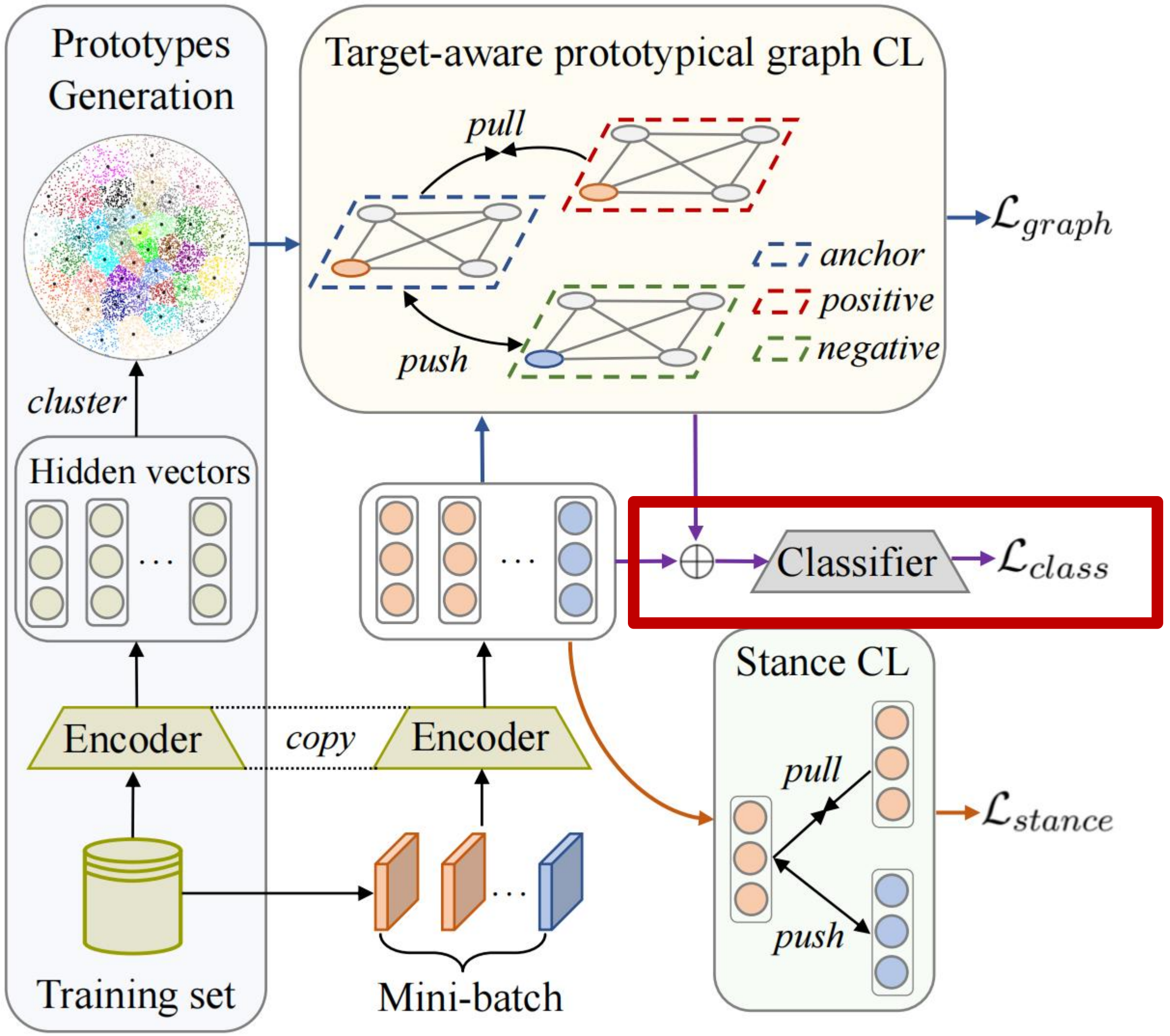
$$\boldsymbol{z}_i = f(\text{GAT}(\boldsymbol{X}; \mathcal{G})) \tag{5}$$

$$\mathcal{L}_{graph} = \frac{-1}{N_b} \sum_{\boldsymbol{\alpha}_i \in \mathcal{B}} \ell^g(\boldsymbol{\alpha}_i) \tag{6}$$

$$\ell^g(\boldsymbol{\alpha}_i) = \log \frac{\sum_{j \in \mathcal{B} \backslash i} \Phi(i,j) \exp(f(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j)/\tau_g)}{\sum_{j \in \mathcal{B} \backslash i} \exp(f(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j)/\tau_g)} \tag{7}$$

$$\Phi(i,j) = \begin{cases} 1 & \text{if } y^i = y^j \text{ and } p^i = p^j \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Method

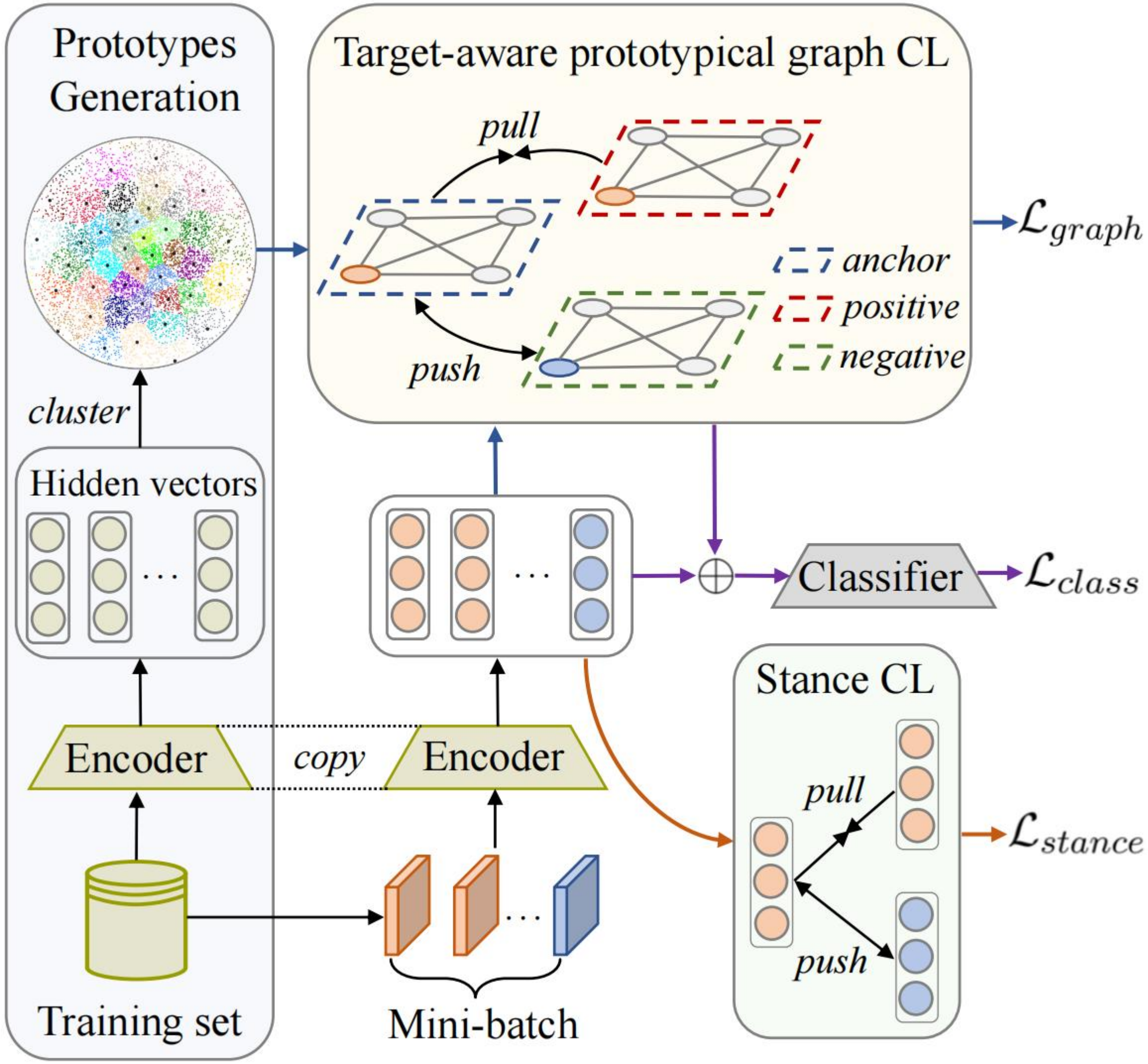


$$v_i = h_i \oplus z_i \tag{9}$$

$$\hat{y}_i = \mathrm{softmax}(W v_i + b) \tag{10}$$

$$\mathcal{L}_{class} = -\sum_{i=1}^{N_b}\sum_{j=1}^{d_y} y_i^j \log \hat{y}_i^j \tag{11}$$

Method



$$\mathcal{L} = \gamma_c \mathcal{L}_{class} + \gamma_s \mathcal{L}_{stance} + \gamma_g \mathcal{L}_{graph} + \lambda ||\Theta||^2 \quad (12)$$

Experiments

	Train	Dev	Test
# Examples	13477	2062	3006
# Unique Comments	1845	682	786
# Zero-shot Topics	4003	383	600
# Few-shot Topics	638	114	159

Table 1: Statistics of VAST dataset.

Dataset	Target	Favor	Against	Neutral	Unrelated
SEM16	DT	148	299	260	-
	HC	163	565	256	-
	FM	268	511	170	-
	LA	167	544	222	-
	A	124	464	145	-
	CC	335	26	203	-
WT-WT	CA	2469	518	5520	3115
	CE	773	253	947	554
	AC	970	1969	3098	5007
	AH	1038	1106	2804	2949

Table 2: Statistics of SEM16 and WT-WT datasets.

Experiments

Model	VAST (%)				SEM16 (%)						WT-WT (%)			
	Pro	Con	Neu	All	DT	HC	FM	LA	A	CC	CA	CE	AC	AH
BiCond	44.6$^{\natural}$	47.4$^{\natural}$	34.9†	42.8$^{\natural}$	30.5‡	32.7‡	40.6‡	34.4‡	31.0‡	15.0‡	56.5$^{\sharp}$	52.5$^{\sharp}$	64.9$^{\sharp}$	63.0$^{\sharp}$
CrossNet	46.2$^{\natural}$	43.4$^{\natural}$	40.4†	43.4$^{\natural}$	35.6	38.3	41.7	38.5	39.7	22.8	59.1$^{\sharp}$	54.5$^{\sharp}$	65.1$^{\sharp}$	62.3$^{\sharp}$
SiamNet	47.5	43.3	39.6	43.5	36.9	37.5	44.3	41.6	41.2	25.6	58.3$^{\sharp}$	54.4$^{\sharp}$	68.7$^{\sharp}$	67.7$^{\sharp}$
SEKT	50.4†	44.2†	30.8†	41.8†	-	-	-	-	-	-	-	-	-	-
TPDG	53.7	49.6	52.3	51.9	47.3	50.9	53.6	46.5	48.7	32.3	66.8$^{\flat}$	65.6$^{\flat}$	74.2$^{\flat}$	73.1$^{\flat}$
TOAD	42.6	36.7	43.8	41.0	49.5‡	51.2‡	**54.1**‡	46.2‡	46.1‡	30.9‡	55.3	57.7	58.6	61.7
BERT	54.6$^{\natural}$	58.4$^{\natural}$	85.3†	66.1$^{\natural}$	40.1‡	49.6‡	41.9‡	44.8‡	**55.2**‡	37.3‡	56.0$^{\flat}$	60.5$^{\flat}$	67.1$^{\flat}$	67.3$^{\flat}$
TGA Net	55.4$^{\natural}$	58.5$^{\natural}$	85.8†	66.6$^{\natural}$	40.7	49.3	46.6	45.2	52.7	36.6	65.7	63.5	69.9	68.7
BERT-GCN	58.3†	60.6†	86.9†	68.6†	42.3	50.0	44.3	44.2	53.6	35.5	67.8	64.1	70.7	69.2
CKE-Net	61.2†	61.2†	88.0†	70.2†	-	-	-	-	-	-	-	-	-	-
JointCL (ours)	**64.9***	**63.2***	**88.9***	**72.3***	**50.5***	**54.8***	53.8	**49.5***	54.5	**39.7***	**72.4***	**70.2***	**76.0***	**75.2***

Table 3: Experimental results on three ZSSD datasets. The results with ♮ are retrieved from (Allaway and McKeown, 2020), † from (Liu et al., 2021), ‡ from (Allaway et al., 2021), ♯ from (Conforti et al., 2020), and ♭ from (Liang et al., 2021a). Best scores are in bold. Results with ⋆ denote the significance tests of our JointCL over the baseline models at $p-$value < 0.05.

Experiments

Model	VAST (%)				SEM16 (%)						WT-WT (%)			
	Pro	Con	Neu	All	DT	HC	FM	LA	A	CC	CA	CE	AC	AH
JointCL (ours)	**64.9**	**63.2**	**88.9**	**72.3**	**50.5**	**54.8**	**53.8**	**49.5**	**54.5**	**39.7**	**72.4**	**70.2**	**76.0**	**75.2**
w/o $\mathcal{L}_{stance}$	61.6	60.7	87.2	69.8	46.2	51.4	51.2	45.3	52.5	36.3	69.4	67.8	72.1	71.4
w/o $\mathcal{L}_{graph}$	62.5	62.1	87.8	70.7	48.8	52.7	51.5	48.2	53.2	38.1	70.5	68.3	74.7	73.6
w/o *graph*	60.8	62.3	87.7	70.3	46.5	50.3	49.7	45.6	52.3	37.4	69.8	68.7	73.2	71.7
w/o *cluster*	59.6	62.2	86.8	69.5	47.4	53.1	52.3	48.6	53.7	38.8	70.9	69.2	74.9	72.6
w/o *edge*	63.3	62.5	88.4	71.4	49.2	53.4	53.1	48.9	53.5	39.2	71.2	69.5	75.2	74.2

Table 4: Experimental results of ablation study.

$$\mathcal{L}_{graph} = \frac{-1}{N_b} \sum_{\boldsymbol{\alpha}_i \in \mathcal{B}} \ell^g(\boldsymbol{\alpha}_i) \qquad (6)$$

$$\ell^g(\boldsymbol{\alpha}_i) = \log \frac{\sum_{j \in \mathcal{B} \setminus i} \Phi(i,j) \exp(f(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j)/\tau_g)}{\sum_{j \in \mathcal{B} \setminus i} \exp(f(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j)/\tau_g)} \qquad (7)$$

w/o graph:

$$\mathcal{L}_{graph} = \frac{-1}{N_b} \sum_{\boldsymbol{h}_i \in \mathcal{B}} \ell^g(\boldsymbol{h}_i) \qquad (13)$$

$$\ell^g(\boldsymbol{h}_i) = \log \frac{\sum_{j \in \mathcal{B} \setminus i} \mathbb{1}_{[t^i = t^j]} \exp(f(\boldsymbol{h}_i, \boldsymbol{h}_j)/\tau)}{\sum_{j \in \mathcal{B} \setminus i} \exp(f(\boldsymbol{h}_i, \boldsymbol{h}_j)/\tau)} \qquad (14)$$

w/o edge:

$$\mathcal{L}_{graph} = \frac{-1}{N_b} \sum_{\boldsymbol{z}_i \in \mathcal{B}} \ell^g(\boldsymbol{z}_i) \qquad (15)$$

$$\ell^g(\boldsymbol{z}_i) = \log \frac{\sum_{j \in \mathcal{B} \setminus i} \mathbb{1}_{[p^i = p^j]} \exp(f(\boldsymbol{z}_i, \boldsymbol{z}_j)/\tau)}{\sum_{j \in \mathcal{B} \setminus i} \exp(f(\boldsymbol{z}_i, \boldsymbol{z}_j)/\tau)} \qquad (16)$$

Experiments

Model	Pro	Con	Neu	All
BiCond	45.4	46.3	25.9	39.2
Cross-Net	50.8	50.5	41.0	47.4
SEKT	51.0	47.9	21.5	47.4
BERT	54.4	59.7	79.6	64.6
TGA Net	58.9	59.5	80.5	66.3
BERT-GCN	62.8	63.4	83.0	69.7
CKE-Net	**64.4**	62.2	83.5	70.1
JointCL (ours)	63.2	**66.7**	**84.6**	**71.5**

Table 5: Experimental results of few-shot condition. Results of baselines are retrieved from (Liu et al., 2021).

Model	HC→DT	DT→HC	FM→LA	LA→FM
BiCond	29.7	35.8	45.0	41.6
CrossNet	43.1	36.2	45.4	43.3
BERT	43.6	36.5	47.9	33.9
SEKT	47.7	42.0	53.6	51.3
TPDG	50.4	52.9	58.3	54.1
JointCL (ours)	**52.8**	**54.3**	**58.8**	**54.5**

Table 6: Experimental results of cross-target condition. “HC→DT” denotes training on HC and testing on DT, etc. Results of baselines are retrieved from (Liang et al., 2021a).

Experiments

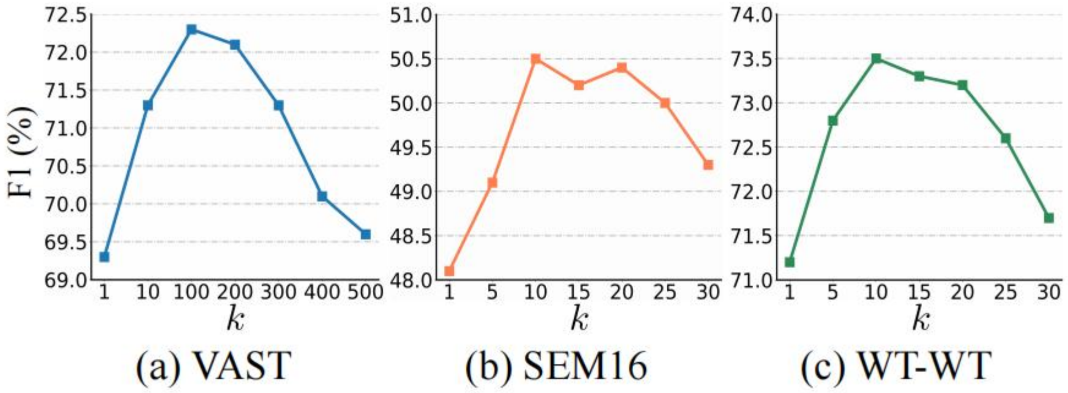


Figure 2: Experimental results of different values of k.

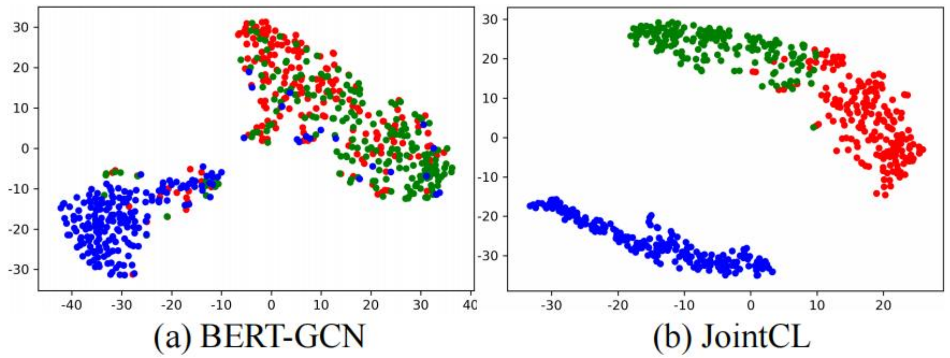


Figure 3: Visualization of intermediate embeddings. Red dots denote Pro examples, green dots denote Con examples, and blue dots denote $Neutral$ examples.

Thank you!