Chongqing University of Technology

ATAI
Advanced Technique of Artificial Intelligence

Borrowing Human Senses: Comment-Aware Self-Training for Social Media Multimodal Classification

Chunpu Xu, Jing Li*
Department of Computing, The Hong Kong Polytechnic University, China
chun-pu.xu@connect.polyu.hk
jing-amelia.li@polyu.edu.hk

(EMNLP-2022)

code: https://github.com/cpaaax/Multimodal_CAST

gesis Leibniz-Institut für Sozialwissenschaften

Forschungszentrum · Research Center L3S

Reported by Zhaoze Gao

Introduction

***Text*:**
was going to take mollie for a walk in the park today . pennsylvania weather cooperated as per usual

Retrieved comments:
1. damn ! that 's a lot of snow to still be around for this time of year , no
2. snow in your area we 'd love to see pictures of it !
3. heavier snow now shifting to your east

…

Figure 1: A sample tweet with its image on the left. On the right, the tweet text is shown on the top, followed by the comments retrieved from similar tweets. The word "snow" (in blue) in comments helpfully hint the implicitly shared semantics between image and text.

Inspired by that, we propose "borrowing" the senses from human readers and modeling user comments to learn the hinting features therein to bridge the image-text gap.

Approach

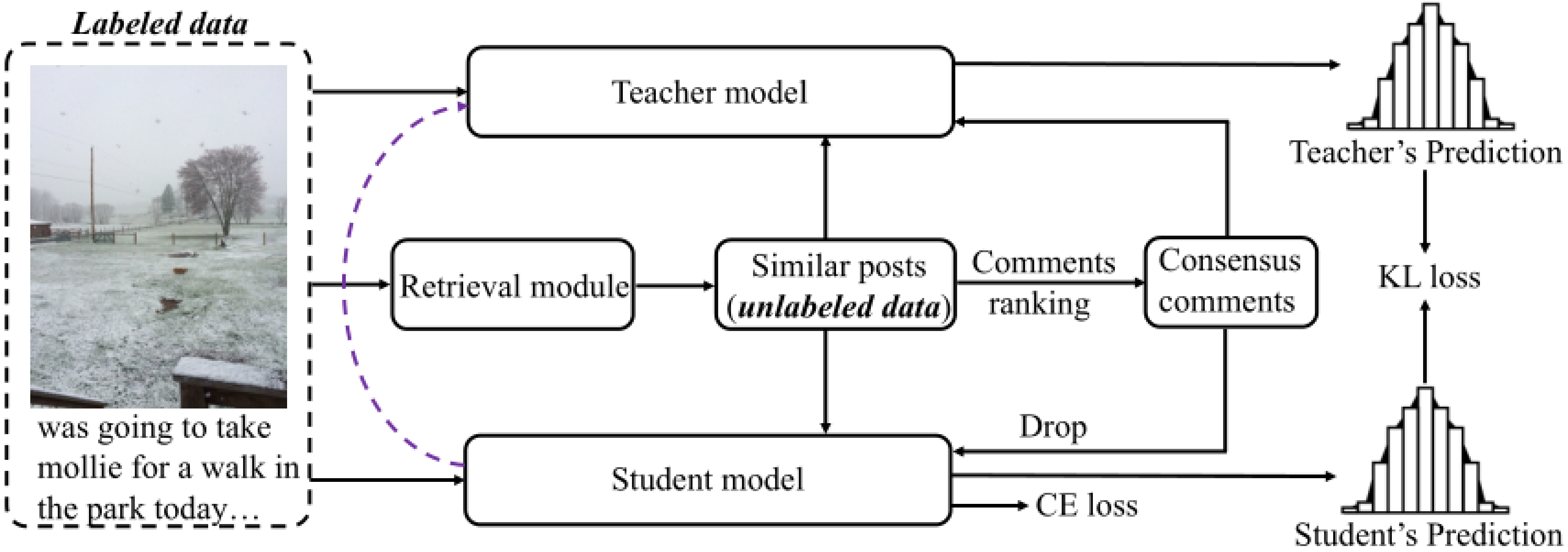
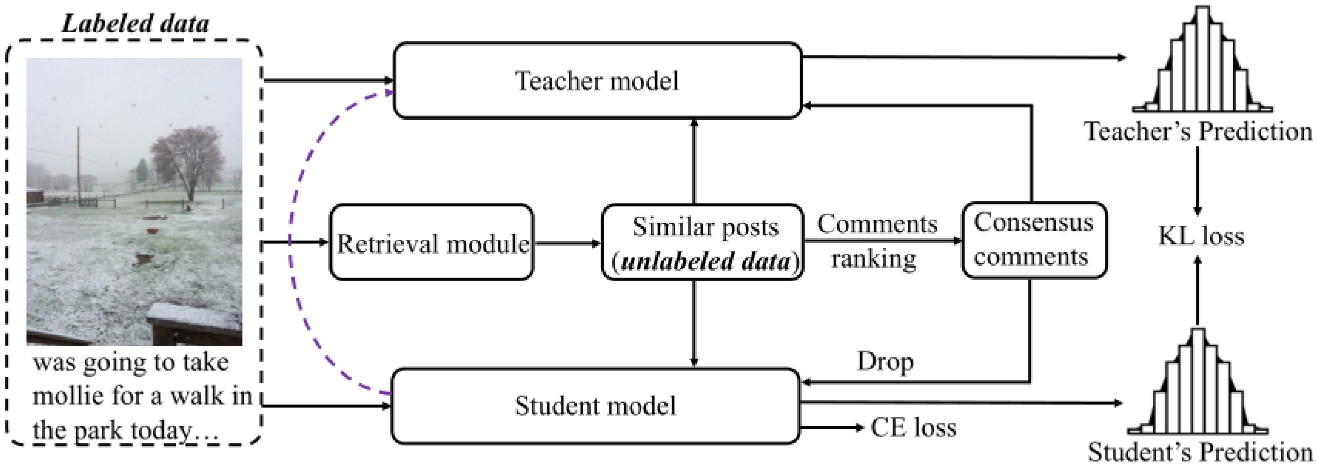


Figure 2: The workflow of comment-aware self-training. Given a post (image-text pair), we first query similar posts and their comments in a retrieval module. Then the retrieved data is employed in teacher-student training as unlabeled data, where student model is trained with CE (cross-entropy) and Kullback–Leibler (KL) divergence loss.

Approach



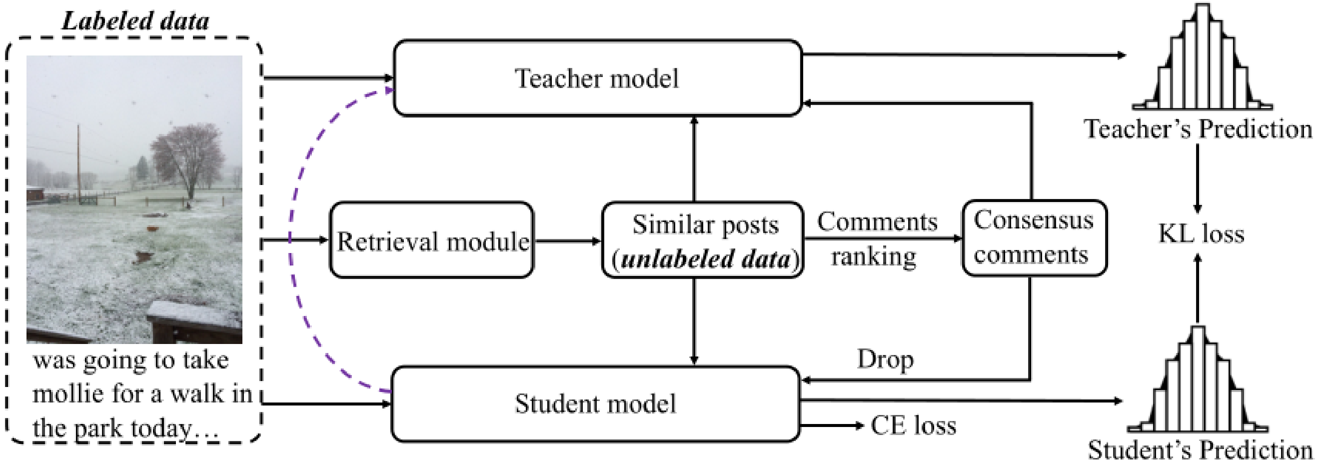
$$s_i = \alpha s_i^I + (1-\alpha) s_i^T \tag{1}$$

$$\alpha = \frac{T_{mean}}{I_{mean} + T_{mean}} \tag{2}$$

$$I_{mean} = \frac{1}{MK} \sum_{m=1}^{M} \sum_{k=1}^{K} p_{m,k} \tag{3}$$

$$T_{mean} = \frac{1}{MK} \sum_{m=1}^{M} \sum_{k=1}^{K} q_{m,k} \tag{4}$$

Approach



$$q_i = \frac{1}{|P|} \sum_{p' \in P} Sim\left(p_i, p'\right) \qquad (5)$$

Early Fusion Scheme

$$u = \sum_{n=1}^{N} \beta_n h_n^c \qquad (6)$$

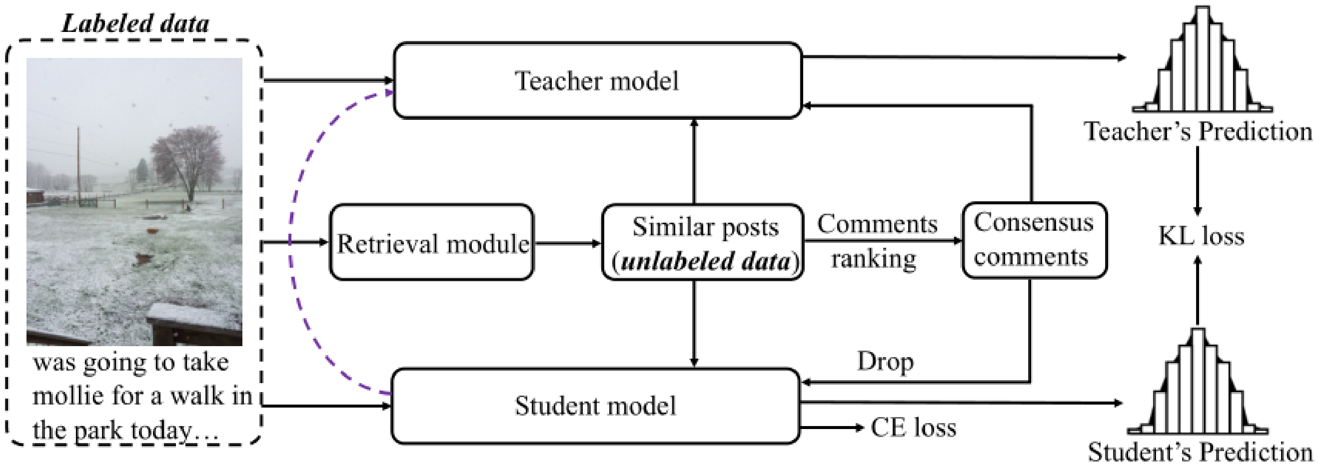
$$\beta_i = \frac{\exp(z_n)}{\sum_{n=1}^{N} \exp(z_n)}; \quad z_n = \sigma(h^f, h_n^c) \qquad (7)$$

Late Fusion Scheme

h^v, h^t, and h^f

$\{h_1^c, \ldots, h_N^c\}$

Approach



$L = \{x_i, c_i, y_i\}_{i=1}^{l}$, where x_i is an image-text pair, c_i indicates the retrieved N comments, and y_i a label specified by the task.

$$U = \{x_i', c_i\}_{i=1}^{Kl}$$

$$\mathcal{L} = \frac{1}{|L|}\sum_{i \in L} y_i \log y_i + \frac{1}{|U|}\sum_{i \in U} \mathrm{KL}(t_i || s_i) \qquad (8)$$

Experiments

Year	Num	Text len	Com len	Com num
2014	3,178,845	12.88	8.81	3.27
2015	6,373,198	13.24	8.19	3.32
2016	6,230,437	13.69	8.98	3.13
2017	4,583,203	11.37	8.55	3.37
2018	3,370,186	10.95	9.01	3.41
2019	4,048,959	10.46	8.35	3.06
Total	27,784,828	12.31	8.62	3.25

Table 1: Statistics of the wild dataset for retrieval. **Text len** and **Com num** indicate the average length (token number) in text and average comment number per tweet. **Com len** is the average length per comment.

Dataset	#Train	#Val	#Test	#All
MVSA	3,611	451	451	4,511
ITR	3,575	447	449	4,471
MSD	19,816	2,410	2,409	24,635
MHP	3,998	500	502	5,000

Table 2: Statistics of the evaluation datasets.

Experiments

Methods	Acc	F1
MultiSentiNet	69.84	69.63
CoMN	70.51	70.01
MMMU-BA	68.72	68.35
Self-MM	72.37	71.96
CoMN-BERT	71.33	70.66
CoMN-BERT (full)	**73.71**	**72.83**

Table 3: Comparison results on the MVSA dataset.

Methods	Pre	Rec	F1
MMSD	76.57	84.15	80.18
D&R Net	77.97	83.42	80.60
Res-BERT	78.87	84.46	81.57
Att-BERT	80.87	85.08	82.92
CMGCN	83.63	84.69	84.16
MMSD-BERT	83.57	84.52	84.04
MMSD-BERT (full)	**85.50**	**85.92**	**85.70**

Table 5: Comparison results on the MSD dataset.

Methods	Pre	Rec	F1
LSTM	42.33	48.55	38.77
CNN	37.11	47.22	35.99
LSTM-CNN	48.21	50.78	44.58
BERT	44.65	48.78	40.39
BERT-CNN	50.31	50.60	49.72
BERT-CNN (full)	**53.69**	**54.42**	**53.38**

Table 4: Comparison results on the ITR dataset.

Methods	Pre	Rec	F1
Xception	56.0	54.5	54.4
LSTM	70.7	73.7	71.9
RoBERTa	75.9	76.5	75.4
MMBT	76.3	78.5	77.1
MMBT (full)	**79.15**	**79.88**	**78.76**

Table 6: Comparison results on the MHP dataset.[6]

Experiments

Model	MVSA		ITR			MSD			MHP		
	Acc	F1	Pre	Rec	F1	Pre	Rec	F1	Pre	Rec	F1
Base Classifier	71.33	70.66	50.31	50.60	49.72	83.57	84.52	84.04	76.30	78.50	77.10
Base+Com	72.34	71.57	52.08	52.67	51.64	84.76	85.19	84.98	77.31	78.29	77.67
Base+ST	73.33	71.63	51.26	51.89	50.64	84.32	85.45	84.88	77.72	78.49	77.85
Base+Com+ST	73.11	72.29	53.06	53.45	52.32	85.42	85.24	85.33	78.45	78.20	78.29
Full Model	73.71	72.83	53.69	54.42	53.38	85.50	85.92	85.70	79.15	79.88	78.76

Table 7: Ablation results on the four datasets. Our Full Model outperform all the ablations measured by all metrics.

Experiments

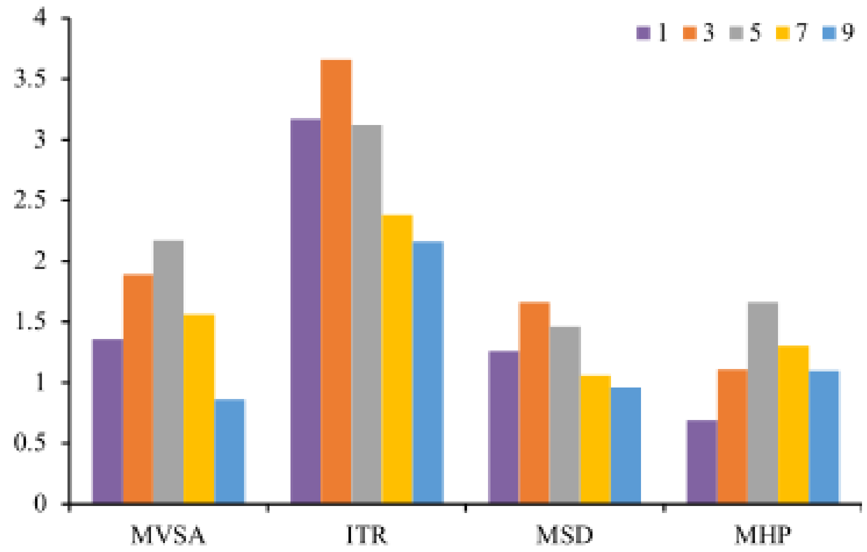


Figure 3: Performance gain observed from self-training given varying number of unlabeled retrieved posts (and their associated comments). X-axis: within each dataset, the bars from left to right indicate self-training with varying number of posts (K); y-axis: the difference in F1 between our Full Model and Base Classifier.

Experiments

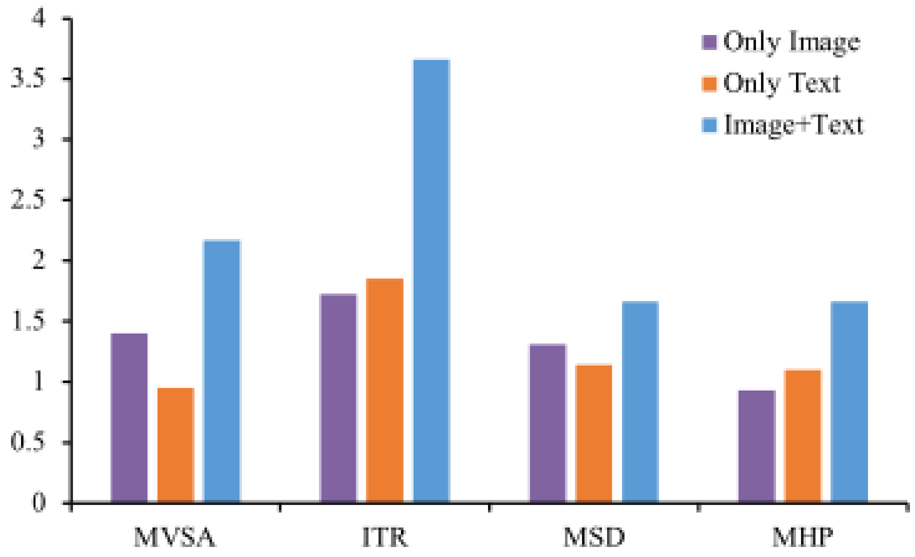


Figure 4: F1 gain compared to the Base Classifier (y-axis) over varying datasets. For each dataset, the bars from left to right indicate the retrieval with image only, text only, and both image and text (Image+Text).

Experiments

Text: snowy owl came too close to traffic camera and he is being ticketed for not reading the signs

Comments:

1. owl photo bomb.
2. a traffic camera in montreal caught this amazing pic of an owl in flight ! Amazing.
3. fbn owl ! rush is a band.
4. this picture is just a year old independent environment snowy owl picture snow weather.
5. harfang des neiges that 's the name of that very beautifull owl.

Label: positive (MSD task)

Figure 5: Visualization of attention heatmaps over the retrieved comments for the MSD benchmark. Deeper colors indicate higher attention weights.

Experiments

Figure 6: Examples of major error types from comment and post retrieval. The top indicates the general comments and the bottom semantically unrelated posts.

Thank you！