互联网上弱监督文本分类的持续学习方法

CL-WSTC: Continual Learning for Weakly Supervised Text Classification on the Internet

Miaomiao Li, Jiaqi Zhu*, Xin Yang*, Yi Yang, Qiang Gao and Hongan Wang

Published in *the ACM Web Conference 2023 (WWW 2023), Research Track, pp. 1489-1499, 2023.*

Contact: Jiaqi Zhu, zhujq@ios.ac.cn, 13683257241

Motivation

Problem: How to build a reasonable framework to integrate WSTC models and continual learning techniques?

- The open and evolving Internet involves constant semantic change of known topics and the appearance of unknown topics.
- Text annotations are hard to access in time for each period.
- WSTC only requires few seed words for each category.

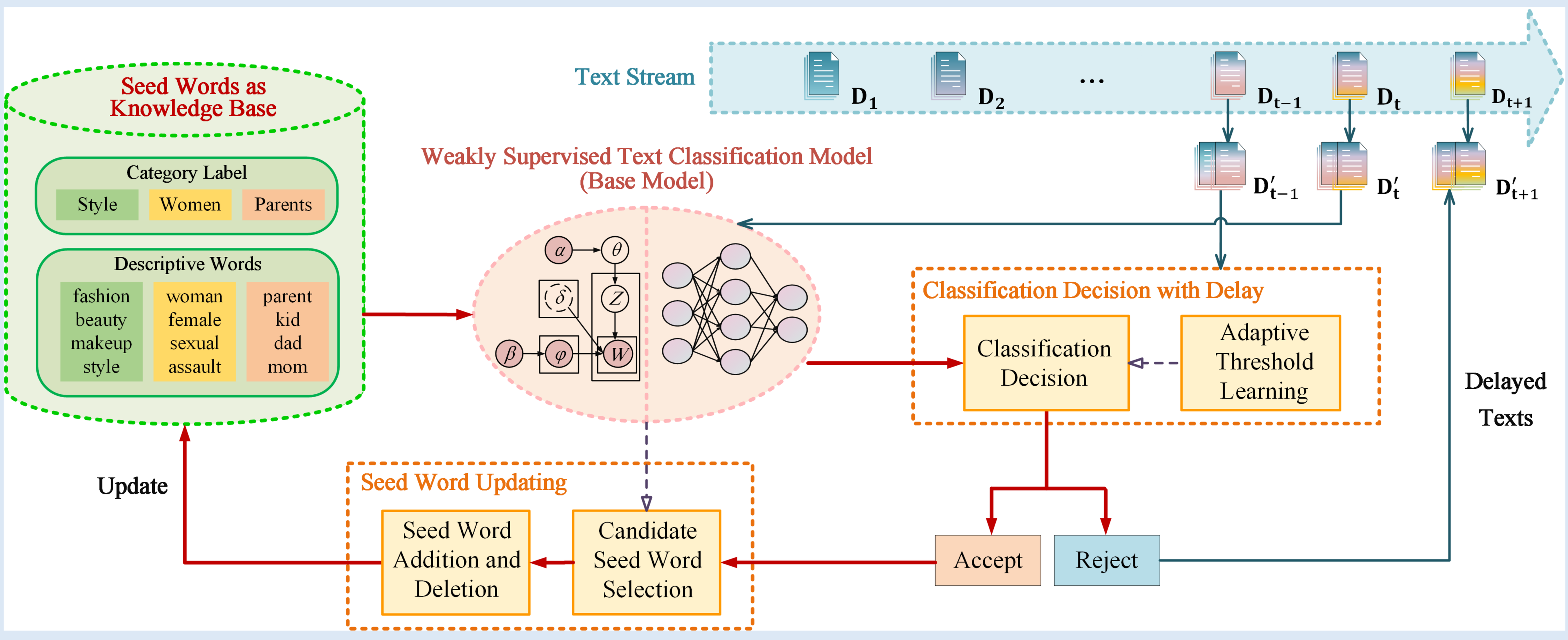
Challenges:

- Delayed Decision: Make a good trade-off between classification accuracy and decision timeliness.
- Knowledge Updating: Balance the maintenance of old knowledge (stability) and the discovery of new knowledge (adaptability).

Contribution

- This is the first attempt to formulate and study the problem of weakly supervised (dataless) text classification in an open and dynamic environment, through building a continual learning paradigm.
- We employ seed words as a new kind of information for knowledge transfer to enhance the robustness of the model and alleviate catastrophic forgetting.
- A novel framework for continual weakly supervised classification on Internet text streams is proposed. The approach can also be seamlessly extended to identify unknown categories by introducing buffer topics to the topic-model based method.
- The approach is applicable to any static base model of the two main kinds, and realizes a fine trade-off between classification accuracy and decision timeliness, with intuitively interpretable seed word transition.

Framework



Approach CL-WSTC

Classification Decision with Delay

Two levels: Period / Category

Supervisory Information in Weakly Supervised Paradigm:
The acceptance/rejection results in the last period

Loss Function: Balance the open space risk and the empirical risk

$$L_b = \frac{1}{N}\sum_{i=1}^{N}\left[\boxed{\delta_i\cdot\delta'_i\cdot(P_t(d_i^{t-1})-\epsilon)} + \boxed{\delta_i\cdot(1-\delta'_i)\cdot(\epsilon-P_t(d_i^{t-1}))}\right]$$

$$L_{acc} = \frac{1}{N}\sum_{i=1}^{N}\left[\boxed{\delta_i\cdot(1-\delta'_i)\cdot(\epsilon-P_t(d_i^{t-1}))} + \boxed{(1\text{-}\delta_i)\cdot\delta'_i\cdot(P_t(d_i^{t-1})-\epsilon))}\right]$$

$$\delta_i = \begin{cases}1, & d_i^{t-1} \text{ is a positive sample,}\\ 0, & d_i^{t-1} \text{ is a negative sample.}\end{cases} \qquad \delta'_i = \begin{cases}1, & P_t(d_i^{t-1}) \geq \epsilon,\\ 0, & P_t(d_i^{t-1}) < \epsilon.\end{cases}$$

The two losses are combined with a loss weight γ to learn an appropriate decision threshold.
To make sure ϵ is between 0 and 1, we use Sigmoid function to establish the mapping between ϵ and a new parameter $\hat{\epsilon}$.

$$\begin{aligned} L &= \gamma\cdot L_b + (1-\gamma)\cdot L_{acc} \\ &= \frac{1}{N}\sum_{i=1}^{N}[\gamma\cdot\delta_i\cdot\delta'_i\cdot(P_t(d_i^{t-1})-\epsilon) \\ &\quad + \delta_i\cdot(1-\delta'_i)\cdot(\epsilon-P_t(d_i^{t-1})) \\ &\quad + (1-\gamma)\cdot(1-\delta_i)\cdot\delta'_i\cdot(P_t(d_i^{t-1})-\epsilon))] \end{aligned}$$

$$\epsilon = \frac{1}{1+e^{-\hat{\epsilon}}} \qquad \hat{\epsilon} = \hat{\epsilon} - \eta\frac{\partial L}{\partial\hat{\epsilon}}$$

$$\frac{\partial L}{\partial\hat{\epsilon}} = (-\gamma\cdot n_1 + n_2 - (1-\gamma)\cdot n_3)\cdot \frac{1}{1+e^{-\hat{\epsilon}}}\cdot(1-\frac{1}{1+e^{-\hat{\epsilon}}})$$

Seed Word Updating

Weakly Supervised Manner: Unsupervised evaluation: An empirically certified measure C_{UMass}
Simplified Reinforcement Learning: Immediate rewards for performance promotion

Algorithm 1 Evaluating and Updating Knowledge Base.

Input: Ordered sets S_t, S_{t-1} of seed words in the last two periods, ordered set W of current candidate words, text sets D_{t-2}, D_{t-1}, D_t.
Output: Updated seed word set S_{t+1}.

1: Initialize $S_{t+1}, S' \leftarrow S_t$;
2: **while** $W \neq \emptyset$ **do** ▷ Seed word addition
3: Extract the first element $w_1 \in W$;
4: **if** $C_{\text{UMass}}(S_{t+1}\cup\{w_1\}, D_t) > C_{\text{UMass}}(S_{t+1}, D_t)$ **then**
5: $S_{t+1} \leftarrow S_{t+1}\cup\{w_1\}$;
6: $W \leftarrow W - \{w_1\}$;

Seed Word Addition: Accepted texts are better to obtain candidate seed word set than all texts. Two aspects of constraints can be imposed: Scope + Number

Seed Word Deletion: Additional condition: The measure is also better if the deletion happens in either of the last two periods.
→ Leverage more evidence to alleviate catastrophic forgetting.

7: **while** $S' \neq \emptyset$ **do** ▷ Seed word deletion
8: Extract the first element $s_1 \in S'$;
9: **if** $C_{\text{UMass}}(S_{t+1} - \{s_1\}, D_t) > C_{\text{UMass}}(S_{t+1}, D_t)$ **then**
10: **if** $C_{\text{UMass}}(S_t - \{s_1\}, D_{t-1}) > C_{\text{UMass}}(S_t, D_{t-1}) \vee$
11: $C_{\text{UMass}}(S_{t-1} - \{s_1\}, D_{t-2}) > C_{\text{UMass}}(S_{t-1}, D_{t-2})$ **then**
12: $S_{t+1} \leftarrow S_{t+1} - \{s_1\}$;
13: $S' \leftarrow S' - \{s_1\}$;
14: **return** S_{t+1};

Experiments

Overall Results

Seed Words	Approach	AGnews F1	AGnews Period	HuffN8 F1	HuffN8 Period
1	SeedBTM+independent	61.5	3	43.9	3.4
	SeedBTM+data-transfer	62.2	3	47.2	3.4
	SeedBTM+together	61.8	5	48.9	5
	SeedBTM+CL-WSTC	72.1	4.2	48.7	4.4
3	SeedBTM+independent	69.3	3	51.4	3.4
	SeedBTM+data-transfer	72.2	3	54.8	3.4
	SeedBTM+together	73.1	5	57.4	5
	SeedBTM+CL-WSTC	73.4	4.3	57.7	4.4
5	SeedBTM+independent	70.9	3	52.3	3.4
	SeedBTM+data-transfer	73.4	3	57.1	3.4
	SeedBTM+together	74.6	5	60.6	5
	SeedBTM+CL-WSTC	74.5	4.3	58.7	4.5
Label Name	LOTClass+independent	38.4	3	29.5	3.4
	LOTClass+data-transfer	49.3	3	44.6	3.4
	LOTClass+together	49.8	5	45.9	5
	LOTClass+CL-WSTC	46.5	4.1	46.9	4.8

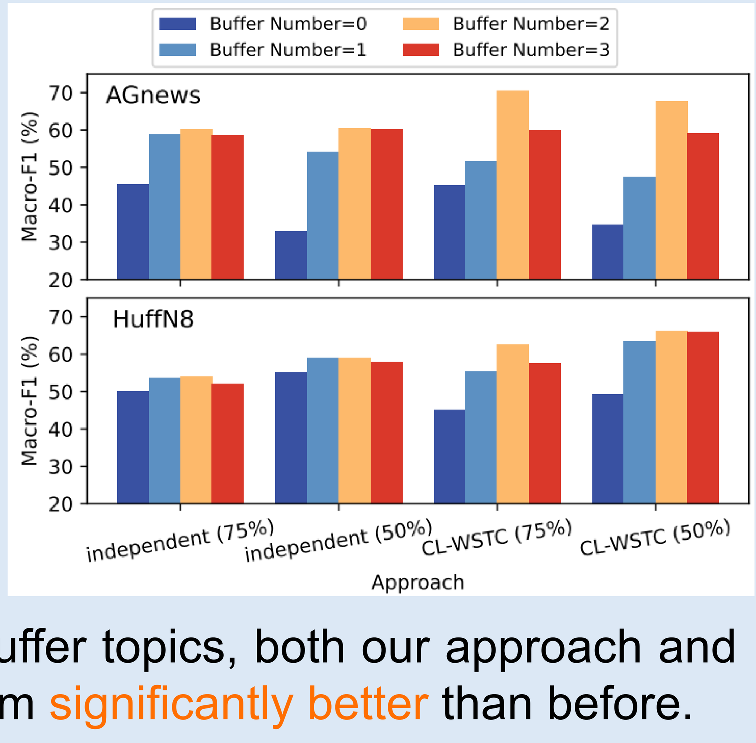
✓ From the perspective of trade-off between accuracy and timeliness, CL-WSTC has its obvious advantages.

Qualitative analysis

✓ With LOTClass on the AGnews dataset, for the category “science”, the seed words “technology”, “software” and “device” have been added, while “chemistry” and “design” have been deleted.

✓ The results demonstrate that CL-WSTC can indeed supplement the up-to-date seed words and remove out-of-date seed words.

Approach Extension



✓ After introducing buffer topics, both our approach and the baseline perform significantly better than before.

Applications

Public Opinion Analysis: For emerging hot events on social media such as Twitter and Weibo, classify the opinions of related posts in a continual manner without labeled data and discover the target content the government or companies concern.

Case Investigation: From chat messages, discover suspicious users and identify the criminal behaviors with self-updating prior knowledge.