Zero-shot Text Classification via Reinforced Self-training

Zhiquan Ye, Yuxia Geng, Jiaoyan Chen, Xiaoxiao Xu, Suhang Zheng, Feng Wang, Jingmin Chen, Jun Zhang, Huajun Chen

ACL 2020
Motivation

- Zeroshot method
  - Embedding text and label into joint space
  - Matching text and label representation
- A self-training based method to leverage unlabeled data in zero-shot text classification
- A reinforcement learning framework to learn data selection policy automatically instead of using manually designed heuristics
Supervised Learning vs Zeroshot Learning

Class distribution:
\[ P(y_1|x), P(y_2|x), \ldots, P(y_n|x) \]

Matching Score:
\[ f(x, y_1), f(x, y_2), \ldots, f(x, y_n) \]

(a) Traditional Classifier

(b) Standard ZSL Model

Figure 1: Illustration of the traditional classifier and standard ZSL model.
Model overview

Figure 2: Overview of our reinforced self-training framework for zero-shot text classification.

Figure 3: BERT as the base matching model.
Reinforcement Learning for Self-training

- **Self-training:**
  - Predict label on unlabeled data
  - Select samples with high confidence

- **States**
  - Prediction confidence $p_{x,y^*}$
  - Representation of text $c_{x,y^*}$

- **Action**
  - Select instance or not $P(a|s_t)$

- **Reward**
  - Train the model on selected data, evaluate on dev set
  - Dev set contains labeled and unlabeled data

$$r_k = \frac{(F_k^s - \mu^s)}{\sigma^s} + \lambda \cdot \frac{(F_k^u - \mu^u)}{\sigma^u}$$
Reinforcement Learning for Self-training (2)

- Policy network

\[
z_t = \text{ReLU}(W_1^T c_{x,y^*} + W_2^T p_{x,y^*} + b_1), \quad (5)
\]

\[
P(a|s_t) = \text{softmax}(W_3^T z_t + b_2). \quad (6)
\]

- Optimization

\[
J(\phi) = E_{P_\phi(a|s)}[R(s,a)] ,
\]
Algorithm 1 Reinforced self-training for zero-shot text classification

Require: labeled seen data $\mathcal{D}^s = \{(x_i^s, y_i^s)\}_{i=1}^N$, unlabeled unseen data $\mathcal{D}^u = \{(x_i^u)\}_{i=1}^M$, seen validation set $\mathcal{D}^s_{dev}$.

1: Initialize pseudo-labeled data $\mathcal{D}^p \leftarrow \emptyset$
2: for $i = 1 \rightarrow N_1$ do  //iteration $i$
3:   Train matching model $f$ with instances
4:      from $\mathcal{D}^s$ and $\mathcal{D}^p$.
5:   Make prediction on $\mathcal{D}^u$, get confidence $P$.
6:   Get a subset $\Omega$ from $\mathcal{D}^u$ by ranked confidence $P$.
7:  for $j = 1 \rightarrow N_2$ do  //episode $j$
8:      if early stop criteria is met then
9:         break
10:     end if
11:    Shuffle $\Omega = \{B_1, B_2, ..., B_{N_3}\}$.
12:  end for  //for $k = 1 \rightarrow N_3$ do  //batch $k$
13:   for $k = 1 \rightarrow N_3$ do  //batch $k$
14:      Get a batch $B_k$ from $\Omega$.
15:      Decide action for each instance in $B_k$, get selected instances $B_k^p$.
16:      Train model $f'$ with $B_k^p$.
17:      Evaluate on $\mathcal{D}^s_{dev}$ and $\mathcal{D}^u_{dev}$,
18:         get $F_k^s$, $F_k^u$.
19:   end for  //update policy network
20: Compute rewards $\{r_k\}_{k=1}^{N_3}$ by equation 4.
21:   for $k = 1 \rightarrow N_3$ do
22:      $\phi \leftarrow \phi + \eta \frac{r_k}{|B_k|} \sum_{t=1}^{|B_k|} \nabla_{\phi} log P(a_t|s_t)$
23:   end for
24: $\mathcal{D}^p_i \leftarrow \bigcup_{k=1}^{N_3} B_k^p$
25: $\mathcal{D}^p \leftarrow \mathcal{D}^p \cup \mathcal{D}^p_i$
26: $\mathcal{D}^u \leftarrow \mathcal{D}^u \setminus \mathcal{D}^p_i$
27: $\mathcal{D}^u_{dev} \leftarrow \mathcal{D}^p$.
28: end for
## Results

<table>
<thead>
<tr>
<th></th>
<th>Topic</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>I</td>
<td>II</td>
<td>I</td>
</tr>
<tr>
<td>Word2vec</td>
<td>35.50</td>
<td>35.33</td>
<td>4.77</td>
<td>11.45</td>
<td>40.67</td>
</tr>
<tr>
<td>Label similarity</td>
<td>34.62</td>
<td>36.14</td>
<td>10.63</td>
<td>16.89</td>
<td>54.56</td>
</tr>
<tr>
<td>FC</td>
<td>19.45</td>
<td>22.46</td>
<td>27.36</td>
<td>8.31</td>
<td>24.33</td>
</tr>
<tr>
<td>RNN+FC</td>
<td>9.68</td>
<td>13.41</td>
<td>15.45</td>
<td>3.15</td>
<td>15.58</td>
</tr>
<tr>
<td>BERT</td>
<td>57.07</td>
<td>45.50</td>
<td>16.86</td>
<td>10.21</td>
<td>60.23</td>
</tr>
<tr>
<td>BERT+self-training</td>
<td>72.21</td>
<td>62.90</td>
<td>31.96</td>
<td>19.72</td>
<td>69.00</td>
</tr>
<tr>
<td>BERT+RL</td>
<td>73.41</td>
<td>65.53</td>
<td>36.98</td>
<td>19.38</td>
<td>73.14</td>
</tr>
</tbody>
</table>