Latent Opinions Transfer Network for Target-oriented Opinion Words Extraction

Fei Zhao
Natural Language Processing Laboratory
Nanjing University

May 14, 2020

Outline

Outline

Target-oriented Opinion Words Extraction

- Target-oriented Opinion Words Extraction (TOWE): extract the corresponding opinion words for a given **target** from a sentence.

Target-oriented Opinion Words Extraction

- Target-oriented Opinion Words Extraction (TOWE): extract the corresponding opinion words for a given **target** from a sentence.

Waiters are very friendly and the pasta is out of this world.

Given opinion target: **waiters** Corresponding opinion words: **friendly**
Given opinion target: **pasta** Corresponding opinion words: **out of this world**

Figure 1: An example of TOWE. The words highlighted in red are two given opinion targets. TOWE task aims to extract the spans in blue as opinion words for the given targets.

Formulation

- Given a sentence $\{w_1, w_2, \cdots, w_n\}$ consisting n words.
- Give a target w_t(we notate an opinion target as one word for simplicity).
- The goal is to tag each word w_i in s with a label $y_i \in \{B, I, O\}$ (B: Beginning, I: Inside, O: Others).

1.	Waiters/O are/O very/O friendly/B and/O the/O pasta/O is/O out/O of/O this/O world/O ./O
2.	Waiters/O are/O very/O friendly/O and/O the/O pasta/O is/O out/B of/I this/I world/I ./O

Figure 2: Different labeling results of a sentence when given different opinion targets. The opinion targets are highlighted in red and the opinion words/phrases are in blue.

Outline

Motivation

- The difficulty of annotation causes the datasets of TOWE to be insufficient, which heavily limits the performance of neural models.

Motivation

- The difficulty of annotation causes the datasets of TOWE to be insufficient, which heavily limits the performance of neural models.
- Enormous review sentiment classification data such as Amazon and Yelp are easily accessible online. These reviews contain substantial opinions information and semantic patterns.

Motivation

- The difficulty of annotation causes the datasets of TOWE to be insufficient, which heavily limits the performance of neural models.
- Enormous review sentiment classification data such as Amazon and Yelp are easily accessible online. These reviews contain substantial opinions information and semantic patterns.

⬇

- we propose to transfer latent opinion knowledge from large-scale sentiment classification datasets to the low-resource task TOWE.

Challenges

- The opinions information such as opinion words in sentiment classification datasets are latent and unannotated, we need to find them explicitly before transferring them(Attention Mechanism).

Challenges

- The opinions information such as opinion words in sentiment classification datasets are latent and unannotated, we need to find them explicitly before transferring them(Attention Mechanism).
- Since sentiment classification for reviews does not consider the target information, the latent opinion information obtained is global and independent of the target. Thus, this information cannot be used directly by TOWE(Transformation Method).

Outline

Latent Opinions Transfer Network

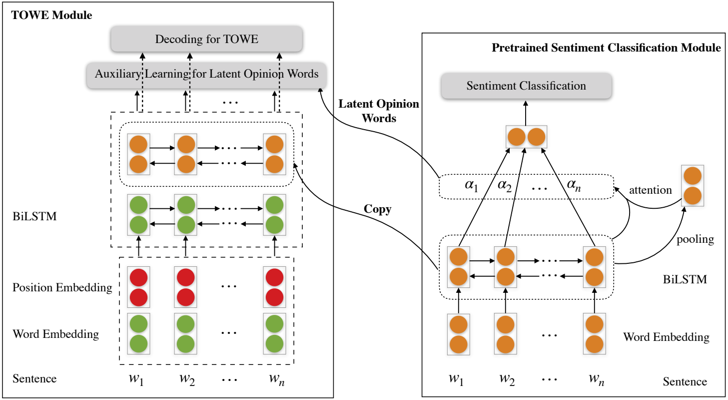

Figure 3: The architecture of Latent Opinions Transfer Network. Different opinion targets in a sentence have different position embeddings.

Base Model: Position Embedding based BiLSTM

The representation $\mathbf{e}_i$ of each word w_i is formed by concatenating the word vector and the corresponding position vector:

$$\mathbf{e}_i = \left[\mathbf{E}_{emb}(w_i); \mathbf{E}_{pos}(l_i)\right], \tag{1}$$

We employ a BiLSTM network to capture the contextual information of each word. The simplified update rule can be written as follows:

$$\mathbf{h}_i^t = \mathrm{BiLSTM}(\mathbf{h}_{i-1}^t, \mathbf{e}_i, \theta_t), \tag{2}$$

In the base model, the context representation $\mathbf{h}^t$ can be used for predicting the opinion words of the given target.

Pretrained Sentiment Classification Model

- We employ a BiLSTM network to capture the contextual information for each word, and outputs a sequence of hidden vectors$\{\mathbf{h}_1^{sc}, \mathbf{h}_2^{sc}, \cdots, \mathbf{h}_m^{sc}\}$.
- The attention mechanism is employed to capture the latent and global opinion words that are significant to sentiment classification.

$$u(\mathbf{h}_i^{sc}, \mathbf{h}_{avg}^{sc}) = \mathbf{h}_i^{sc} \cdot \mathbf{W}_u \cdot \mathbf{h}_{avg}^{sc} + b_u, \tag{3}$$

$$\alpha_i = \frac{\exp(u(\mathbf{h}_i^{sc}, \mathbf{h}_{avg}^{sc}))}{\sum_{j=1}^{m} \exp(u(\mathbf{h}_j^{sc}, \mathbf{h}_{avg}^{sc}))}, \tag{4}$$

Transferring Pretrained Encoder

- From the semantic level view, the encoder of the pretrained sentiment classification module holds substantial implicit opinion information
- We integrate it into the TOWE module by concatenating two hidden states:

$$\mathbf{r}_i = \left[\mathbf{h}_i^t; \mathbf{h}_i^{sc}\right], \tag{5}$$

Transferring Latent Opinion Words

Transformation Method

- we introduce the opinion target information into the attention distribution by a target-relevant distance weight c_i:

$$\alpha_i{}' = c_i \cdot \alpha_i, \tag{6}$$

$$c_i = 1 - \frac{|i - t|}{n}, \tag{7}$$

- To regain the probabilistic attention distribution, the target-dependent attention weight $\alpha_i{}'$ is re-normalized:

$$\beta_i = \frac{\alpha_i{}'}{\sum_{j=1}^{n} \alpha_j{}'}. \tag{8}$$

- we use a heuristic strategy to convert the normalized attention weight β_i into the binary latent opinion words by the threshold $\frac{1}{n}$:

$$y_i^a = \begin{cases} 1 & \text{if } \beta_i \geq \frac{1}{n}, \\ 0 & \text{otherwise }, \end{cases} \tag{9}$$

Transferring Latent Opinion Words

Auxiliary Learning Signal

- We integrate these latent opinions into TOWE module by auxiliary learning signal:

$$\hat{\mathbf{y}}_i^a = softmax(\mathbf{W_a r_i} + \mathbf{b_a}), \tag{10}$$

$$\mathcal{L}_a = -\sum_{i=1}^{n}\sum_{k=0}^{1} \mathbb{I}(y_i^a = k)\log(\hat{y}_{i,k}^a), \tag{11}$$

Decoding and Training

We use a linear layer and a softmax layer to compute prediction probaility $\hat{\mathbf{y}}_i$:

$$\hat{\mathbf{y}}_i = softmax(\mathbf{W_t r_i} + \mathbf{b_t}), \tag{12}$$

The cross-entropy loss of TOWE task can be defined as follows:

$$\mathcal{L}_t = -\sum_{i=1}^{n}\sum_{k=0}^{2}\mathbb{I}(y_i = k)\log(\hat{y}_{i,k}), \tag{13}$$

LOTN also integrates latent opinions through auxiliary learning signal $\mathcal{L}_a$. Thus the final loss is defined as follows:

$$J = \mathcal{L}_t + \lambda\mathcal{L}_a, \tag{14}$$

Outline

Datasets

- We evaluate our model on four benchmark datasets(Fan et al., 2019). The statistics of the datasets are summarized in Table 2.

Datasets		#sentences	#targets
14res	Train	1,627	2,643
	Test	500	864
14lap	Train	1,158	1,634
	Test	343	482
15res	Train	754	1,076
	Test	325	436
16res	Train	1,079	1,512
	Test	329	457

Figure 4: Statistics of TOWE datasets. A sentence may contain multiple opinion targets.

Datasets

- To pretrain the sentiment classification model, we use the two datasets respectively from Amazon Review and Yelp Review. Table 3 shows the statistics of Amazon Review and Yelp Review.

Datasets	#positive	#negative	#total
Yelp Review	266,041	177,218	443,259
Amazon Review	277,228	277,769	554,997

Figure 5: Statistics of the two datasets Amazon Review and Yelp Review.

Overall Performance Comparison

Models	14res			14lap			15res			16res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Distance-rule	58.39	43.59	49.92	50.13	33.86	40.42	54.12	39.96	45.97	61.90	44.57	51.83
Dependency-rule	64.57	52.72	58.04	45.09	31.57	37.14	65.49	48.88	55.98	76.03	56.19	64.62
LSTM	52.64	65.47	58.34	55.71	57.53	56.52	57.27	60.69	58.93	62.46	68.72	65.33
BiLSTM	58.34	61.73	59.95	64.52	61.45	62.71	60.46	63.65	62.00	68.68	70.51	69.57
Pipeline	77.72	62.33	69.18	72.58	56.97	63.83	74.75	60.65	66.97	81.46	67.81	74.01
TC-BiLSTM	67.65	67.67	67.61	62.45	60.14	61.21	66.06	60.16	62.94	73.46	72.88	73.10
IOG	82.38	78.25	80.23	73.43	**68.74**	70.99	72.19	**71.76**	71.91	84.36	79.08	81.60
PE-BiLSTM	80.10	76.51	78.26	72.01	64.20	67.83	70.36	65.73	67.96	82.27	74.95	78.43
LOTN	**84.00**†	**80.52**†	**82.21**†	**77.08**†	67.62	**72.02**†	**76.61**†	70.29	**73.29**†	**86.57**†	**80.89**†	**83.62**†

Figure 6: Main experiment results(%). Best results are in bold (P, R, and F1-score, the larger is the better). The marker † represents that LOTN outperforms other methods significantly ($p < 0.01$) .

Summary

- Compared to its base version PE-BiLSTM, LOTN obtains about 4%∼5% improvements in F1-score.
- LOTN outperforms the previous state-of-the-art method IOG by 1.98% and 2.02% F1-score respectively in the datasets 14res and 16res.

Effects of Tranferring Encoder and Opinion Words

Models	14res			14lap			15res			16res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
PE-BiLSTM	80.10	76.51	78.26	72.01	64.20	67.83	70.36	65.73	67.96	82.27	74.95	78.43
+transferred encoder	84.57	79.54	81.97	77.50	67.47	72.13	75.90	69.00	72.26	86.05	79.81	82.79
+auxiliary learning	84.10	77.20	80.49	75.63	66.42	70.71	76.31	68.67	72.29	86.77	79.46	82.93
LOTN	84.00	80.52	82.21	77.08	67.62	72.02	76.61	70.29	73.29	86.57	80.89	83.62

Figure 7: Experiment results of adding the transferred encoder or auxiliary learning on PE-BiLSTM(%).

Summary

- Compared to the base model PE-BiLSTM, we can find that PE-BiLSTM+transferred encoder and PE-BiLSTM+auxiliary learning both achieve significant and consistent improvements on all datasets.
- The results indicate that the proposed two methods are useful for the final model LOTN and they transfer opinions knowledge from different perspectives.

Effect of the Hyper-parameter λ

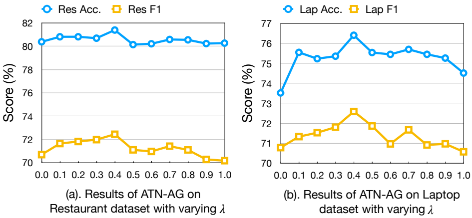

Figure 8: The effect of different hyper-parameter λ.

Summary

- LOTN achieves the relatively stable performance with varying λ on the datasets 14res, 15res and 16res, which indicates the robustness of our method.
- The performance of LOTN has a downward trend with an increase of λ since the bigger λ has a negative effect on the decoding of the model.

Case Study

<table>
<tr><th rowspan="2">Sentence</th><th rowspan="2">Distance-rule</th><th rowspan="2">BiLSTM</th><th rowspan="2">IOG</th><th rowspan="2">PE-BiLSTM</th><th colspan="2">LOTN</th></tr>
<tr><th>Latent Opinion Words</th><th>Target Decoding</th></tr>
<tr><td>The bread is top notch as well.</td><td>top✗</td><td>top notch✓</td><td>top notch✓</td><td>top notch✓</td><td>top notch</td><td>top notch✓</td></tr>
<tr><td>Great food but the service was dreadful!</td><td>Great✓</td><td>dreadful✗</td><td>Great✓</td><td>Great✓</td><td>Great</td><td>Great✓</td></tr>
<tr><td>Great food but the service was dreadful!</td><td>dreadful✓</td><td>dreadful✓</td><td>dreadful✓</td><td>dreadful✓</td><td>dreadful</td><td>dreadful✓</td></tr>
<tr><td>Good for a quick sushi lunch.</td><td>quick✗</td><td>Good, quick✓</td><td>quick✗</td><td>quick✗</td><td>Good</td><td>Good, quick✓</td></tr>
<tr><td>Their twist on pizza is healthy, but full of flavor.</td><td>full✓</td><td>healthy✗</td><td>healthy, full✗</td><td>NULL✗</td><td>full</td><td>full✓</td></tr>
</table>

Figure 9: Examples of the extracted results in different methods. The opinion targets are in red and the corresponding golden opinion words are in blue. The "NULL" represents the prediction is empty.

Summary

- Review sentiment classification model can reduce the errors of TOWE model to a certain extent

Error Analysis

Models	NULL	Under-extracted	Over-extracted	Others	Total
PE-BiLSTM	76	107	49	34	266
LOTN	65	85	62	31	243

Figure 10: Statistics of different error types for PE-BiLSTM and LOTN in the dataset 14res.

Summary

- PE-BiLSTM and LOTN do not extract any opinion words in more than a quarter of error cases.
- Compared to PE-BiLSTM, LOTN makes fewer mistakes in the NULL and under-extracted type. In contrast, PE-BiLSTM makes fewer over-extracted predictions.

Outline

Conclusions

- Insufficiency of labeled data heavily restricts the effectiveness of the neural models for TOWE.
- We propose a novel model to transfer latent opinions knowledge from resource-rich review sentiment classification datasets to improve the low-resource task TOWE.
- Results from numerous experiments indicate that our approach achieves better performance than other state-of-the-art methods. Extensive analysis also demonstrates the effectiveness of our model.

End.

Thanks!

Zhifang Fan, Zhen Wu, Xinyu Dai, Shujian Huang, and Jiajun Chen. 2019. Target-oriented opinion words extraction with target-fused neural sequence labeling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2509–2518.