Semantics-Preserved Data Augmentation for Aspect-Based Sentiment Analysis

Ting-Wei Hsu ¹, Chung-Chi Chen ¹, Hen-Hsen Huang ²,³, Hsin-Hsi Chen ¹,³
¹ Department of Computer Science and Information Engineering, National Taiwan University, Taiwan
² Institute of Information Science, Academia Sinica, Taiwan
³ MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan
{twhsu,cjchen}@nlg.csie.ntu.edu.tw, hhhuang@iis.sinica.edu.tw, hhchen@ntu.edu.tw
### Motivation

| Original review: | But the staff was so horrible to us. |
|------------------|------------------------------------|
| BT               | But the staff *were* so awful for us. |
| EDA              | But *so* staff was *the ugly to uranium*. |
| C-BERT           | But the *situation* was *being good to me*. |
| AS-SPM & AE      | But the staff was *always* horrible to me. |
| Senti-SPM & AE   | But the *situation* was so horrible to *me*. |
Semantics-Preserved Data Augmentation

**Insert Auxiliary Sentence**

Input review: the staff was so kind

The polarity of the service is positive [SEP] the staff was so kind

**Token Replacement**

Replaced token: always

Unimportant token: [MASK]

$ED(E', E_i) < \text{Avg. } ED$
Experimental Settings

• Tasks
  • Aspect category sentiment classification (ACSC)
  • Aspect term sentiment classification (ATSC)
  • Aspect term extraction (ATE)
  • Multilingual Scenarios
  • Multi-Aspect Multi-Sentiment Scenarios
  • Stock Price/Risk Movement Prediction

• Baselines
  • Back Translation (BT)
  • Easy Data Augmentation (EDA)
  • C-BERT (Random Masking Strategy)
# Aspect-Based Sentiment Analysis (English)

| Model                  | ACSC | ATSC | ATE      |
|------------------------|------|------|----------|
|                        | Rest14 | Lap14 | Rest15 | Rest16 | Rest14 | Lap14 | Rest15 | Rest16 |
| Bert<sub>base</sub>    | 82.98±0.78 | 79.48±0.64 | 75.32±1.08 | 81.62±1.07 | 86.58±0.56 | 86.44±0.49 | 78.49±1.38 | 66.10±4.61 | 72.42±2.38 |
| + BT                   | 82.45±0.62 | 79.98±0.51 | 75.76±1.19 | 82.61±0.60 | 86.22±0.58 | 86.57±0.48 | 80.66±2.27 | 70.34±1.65 | 74.23±0.64 |
| + EDA                  | 82.82±0.15 | 79.82±0.58 | 76.11±0.58 | 81.77±1.43 | 85.65±0.53 | - | - | - | - |
| + C-BERT               | 83.45±1.14 | 79.67±0.80 | 76.45±0.90 | 80.37±2.56 | 85.57±1.69 | 86.73±0.15 | 81.00±1.68 | 69.21±1.14 | 75.19±0.57 |
| + AS-SPM & AE          | 83.14±0.98 | 80.55±0.42 | 76.33±1.19 | 83.91±0.98 | 87.85±0.38 | 87.18±0.63 | 82.86±1.50 | 70.68±1.15 | 75.62±0.64 |
| + Senti-SPM & AE       | 84.07±0.36 | 80.50±0.80 | 77.21±0.61 | 84.28±0.64 | 87.61±0.40 | - | - | - | - |
| + AS-SPM & Seq2Seq     | **84.17±0.94** | **81.19±0.65** | **77.93±0.43** | **84.46±0.22** | **87.55±0.45** | 87.04±0.54 | 81.51±1.07 | 69.27±0.87 | 75.24±0.58 |
| + Senti-SPM & Seq2Seq  | 83.39±1.03 | **81.50±0.47** | 77.55±1.31 | 83.74±1.25 | 87.81±0.54 | - | - | - | - |
Influence of Augmentation Size
# Multilingual Scenario

| Model                      | Language  |  |  |  |  |  |  |
|---------------------------|-----------|---|---|---|---|---|---|
|                           | AR  | CH  | DU  | FR  | RU  | ES  | TU  |
| **Bert_{base}**           | 88.48\,0.78 | 93.79\,0.87 | 85.37\,3.36 | 85.98\,2.47 | 90.78\,1.86 | 81.66\,1.17 | 66.81\,1.73 |
| + BT                      | 88.24\,0.87 | 94.58\,1.40 | **88.20\,1.66** | 87.66\,3.93 | 93.90\,1.50 | 84.00\,1.70 | 72.89\,3.92 |
| + C-BERT                  | 87.88\,2.24 | 94.20\,1.58 | 84.74\,2.82 | 88.41\,1.69 | 92.96\,1.46 | 80.16\,2.72 | 71.59\,6.34 |
| + AS-SPM & AE             | 89.20\,1.01 | 95.48\,1.17 | 86.32\,1.30 | 88.41\,1.56 | **94.21\,0.42** | **84.66\,1.51** | 72.31\,2.13 |
| + Senti-SPM & AE          | 87.41\,0.88 | 94.31\,0.32 | 86.16\,1.54 | 88.97\,1.02 | 92.81\,1.28 | 83.00\,2.32 | 71.73\,4.06 |
| + AS-SPM & Seq2Seq        | **90.28\,1.34** | 95.31\,0.74 | 86.63\,1.25 | **89.15\,1.41** | 93.75\,1.23 | 82.83\,1.51 | 71.44\,2.74 |
| + Senti-SPM & Seq2Seq     | 88.48\,1.76 | **95.58\,0.71** | 87.26\,1.95 | 88.78\,2.19 | 91.40\,0.95 | 83.66\,2.80 | **73.62\,3.01** |
## Sentiment Analysis & Multi-Aspect Multi-Sentiment

| Model                  | MR   | SST-2  |
|------------------------|------|--------|
| Bert<sub>base</sub>    | 85.64<sub>0.77</sub> | 90.39<sub>0.81</sub> |
| + BT                   | 85.90<sub>0.37</sub> | 90.82<sub>0.54</sub> |
| + EDA                  | 85.54<sub>0.41</sub> | 90.53<sub>0.88</sub> |
| + C-BERT               | 85.02<sub>1.38</sub> | 90.16<sub>0.46</sub> |
| + Senti-SPM & AE       | 85.75<sub>0.50</sub> | 90.91<sub>0.34</sub> |
| + Senti-SPM & Seq2Seq  | 86.52<sub>0.59</sub> | 91.51<sub>0.42</sub> |

| Model                  | ATSC | ACSC   |
|------------------------|------|--------|
| Bert<sub>base</sub>    | 82.23<sub>0.41</sub> | 73.45<sub>1.38</sub> |
| + BT                   | 82.73<sub>0.37</sub> | 73.60<sub>1.02</sub> |
| + EDA                  | 82.78<sub>0.20</sub> | 74.53<sub>1.40</sub> |
| + C-BERT               | 82.34<sub>0.48</sub> | 74.22<sub>0.92</sub> |
| + AS-SPM & AE          | 82.09<sub>0.41</sub> | 75.29<sub>0.93</sub> |
| + Senti-SPM & Seq2Seq  | 82.33<sub>0.72</sub> | 73.89<sub>1.08</sub> |
| + AS-SPM & AE          | 83.00<sub>0.80</sub> | 76.15<sub>0.79</sub> |
| + Senti-SPM & Seq2Seq  | 83.17<sub>0.51</sub> | 75.27<sub>0.69</sub> |
## Stock Price/Risk Movement Prediction

| Model               | Aspect          | Auxiliary Sentence           | ACC.   |
|---------------------|-----------------|------------------------------|--------|
|                     | Return          |                              |        |
| **Bert\textsubscript{base}** | 50.57 \(\pm\) 2.99 | [MASK]                       | 50.57 \(\pm\) 2.68   |
| + BT                | 51.24 \(\pm\) 0.95 | Risk [MASK]                  | 51.51 \(\pm\) 2.56   |
| + EDA               | 51.84 \(\pm\) 1.83 | Risk will [MASK]             | 52.12 \(\pm\) 2.82   |
| + C-BERT            | 52.13 \(\pm\) 1.25 | Market risk will [MASK]      | **55.32 \(\pm\) 2.96** |
| + Return-SPM & AE   | 52.87 \(\pm\) 0.78 |                              |        |
| + Return-SPM & Seq2Seq | **54.04 \(\pm\) 0.77** |                              |        |
| + Risk-SPM & AE     | 51.98 \(\pm\) 0.68 |                              |        |
| + Risk-SPM & Seq2Seq| 52.02 \(\pm\) 0.54 |                              |        |
Conclusion

• We present a controllable augmentation for ABSA, which is controllable to generate reasonable reviews without converting aspect-level polarity.

• We propose SPM to measure the impact of the related words on deciding specific aspect and sentiment, and adopt two replacement strategies to ABSA tasks.

• The exploration in the financial application scenario also supports the usefulness of the proposed method

Code & Datasets: https://github.com/Quant-NLP/SPDAug-ABSA
Related Works and Events

• Related Works
  • From Opinion Mining to Financial Argument Mining. (Springer Nature – Open Access)
    • http://springer.nlpfin.com/

• Related Events
  • Call for Paper: Financial Technology on the Web @ ACM TWEB
    • https://acmfinweb.nlpfin.com/
  • FinNum-3 Shared Task @ NTCIR-2022 – Investor's and Manager’s Fine-grained Claim Detection

• EMNLP-2021 Tutorial: Financial Opinion Mining

• The Workshop of Financial Technology and Natural Language Processing (FinNLP @ IJCAI)
• The Workshop on Financial Technology on the Web (FinWeb @ WWW)

Join Mail List for Latest News ➔ https://forms.gle/RB9Qq9ok6z5exu1G6
Thank you for your attention!

Feel free to contact us if you have any questions.

Chung-Chi Chen: cjchen@nlg.csie.ntu.edu.tw

http://cjchen.nlpfin.com/

Join Mail List for Latest News

https://forms.gle/RB9Qq9ok6z5exu1G6