# Exploiting Position Bias for Robust Aspect Sentiment Classification

Fang Ma, Chen Zhang, Dawei Song
Beijing Institute of Technology Beijing
{mfang, czhang, dwsong}@bit.edu.cn

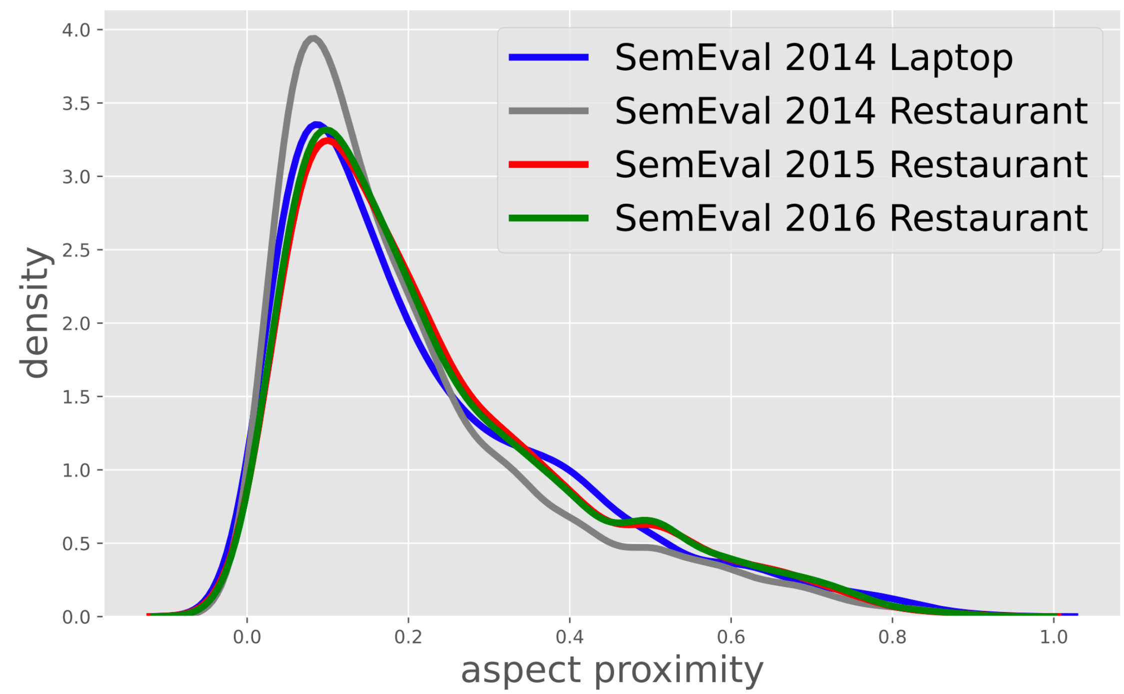
# Background

- Fine-grained opinion mining - aspect sentiment classification (ASC)
  - E.g., *Great food but the service was bad*.
- Pervious ASC Models have achieved remarkable in-domain (I.D.) performance on the ASC task
  - By modeling complex interactions between aspects and contexts.
- State-of-the-art ASC models have been shown to suffer from the lack of robustness. Particularly in two scenarios:
  - out-of-domain (O.O.D.) scenario.
  - adversarial (Adv.) scenario.

| **Scenario** | **Example** | **Pred./Lb.** |
|---|---|---|
| I.D. | Great food but the service was bad ! | neg./neg. |
| O.O.D. | The battery has never worked well . | pos./neg. |
| Adv. | Awful food but the service was great ! | neg./pos. |

# Motivation

- Prior work observes that highlighting words close to a target aspect would boost I.D. performance (termed as ***position bias***)
  - E.g., *Great food but the service was bad*.
  - *Great* is close to *food* and far away from *service*.
- We hypothesize that position bias is also crucial for robust ASC models in O.O.D. and Adv. settings
  - The hypothesis is statistically evidenced by existing benchmarking datasets.



# Method

- Notations
  - Word indices $S = \{w_0, w_1, \ldots, w_\gamma, w_{\gamma+1}, \ldots, w_{n-1}\}$
  - Word representations $V = \{e_0, e_1, \ldots, e_\gamma, e_{\gamma+1}, \ldots, e_{n-1}\}$
  - $\gamma$ denotes the start of the aspect, and the length of the aspect is $m$
- Position-biased weight
  - $$p_i = \begin{cases} 1 - \frac{\gamma - i}{n-m} & 0 \le i < \gamma \\ \frac{1}{n-m} & \gamma \le i < \gamma + m \\ 1 - \frac{i-\gamma-m+1}{n-m} & \gamma + m \le i < n \end{cases}$$
  - Biased word representations $E = \{p_i e_i\}$
- Position-biased dropout
  - $$z_i \sim \mathrm{Bernoulli}(p_i)$$
  - Biased word representations $E = \{z_i e_i\}$

# Experiment

- Datasets
  - SemEval Laptop
  - SemEval Restaurant
  - ARTS Laptop
  - ARTS Restaurant
- Target Models
  - LSTM
  - LSTM-Attn
  - IAN
  - MemNet
  - AOA
  - ROBERTa

| Dataset | | # pos. | # neu. | # neg. |
|---|---|---|---|---|
| SemEval-Lap | train | 930 | 433 | 800 |
| | test | 341 | 169 | 128 |
| | dev | 57 | 27 | 66 |
| SemEval-Rest | train | 2,094 | 579 | 779 |
| | test | 728 | 196 | 196 |
| | dev | 70 | 54 | 26 |
| ARTS-Lap | test | 883 | 407 | 587 |
| ARTS-Rest | test | 1,953 | 473 | 1,104 |

Table 1: Statistics of datasets.

# O.O.D. & Adv. Result

- Largely improve robustness of target models

| Model | Lap | | | | Rest | | | |
|---|---|---|---|---|---|---|---|---|
| | O.O.D. | | Adv. | | O.O.D. | | Adv. | |
| | Acc. | F1 | Acc. | F1 | Acc. | F1 | Acc. | F1 |
| LSTM | 71.02 | 52.15 | 49.49 | 43.91 | 60.60 | 53.25 | 53.34 | 41.99 |
| w/ `pos-dp` | 71.48↑**0.46** | 50.98↓**1.17** | 50.74↑**1.25** | 44.38↑**0.47** | 63.39↑**2.79** | 58.57↑**5.32** | 53.57↑**0.23** | 42.11↑**0.12** |
| w/ `pos-wt` | 72.96↑**1.94** | 55.88↑**3.73** | 55.50↑**6.01** | 50.03↑**6.12** | 66.33↑**5.73** | 60.21↑**6.96** | 59.03↑**5.69** | 48.20↑**6.21** |
| LSTM-Attn | 71.61 | 53.61 | 51.33 | 46.11 | 62.85 | 54.97 | 58.45 | 49.65 |
| w/ `pos-dp` | 71.34↓**0.27** | 52.49↓**1.12** | 53.76↑**2.43** | 48.47↑**2.36** | 65.24↑**2.39** | 59.07↑**4.10** | 58.64↑**0.19** | 47.22↓**2.43** |
| w/ `pos-wt` | 72.84↑**1.23** | 56.18↑**2.57** | 58.53↑**7.20** | 53.54↑**7.43** | 68.90↑**6.05** | 64.48↑**9.51** | 64.80↑**6.35** | 55.34↑**5.69** |
| IAN | 72.09 | 54.44 | 52.91 | 47.54 | 63.82 | 55.20 | 57.75 | 48.12 |
| w/ `pos-dp` | 70.95↓**1.14** | 51.63↓**3.08** | 52.04↓**0.87** | 45.87↓**1.67** | 63.57↓**0.25** | 56.81↑**1.61** | 56.89↓**0.86** | 46.90↓**1.22** |
| w/ `pos-wt` | 72.86↑**0.77** | 54.88↑**0.44** | 56.03↑**3.12** | 50.30↑**2.76** | 62.45↓**1.37** | 55.95↑**0.75** | 63.49↑**5.74** | 54.04↑**5.92** |
| MemNet | 70.66 | 52.07 | 52.00 | 46.50 | 57.84 | 51.15 | 55.30 | 46.67 |
| w/ `pos-dp` | 69.93↓**0.73** | 53.37↑**1.30** | 53.54↑**1.54** | 47.93↑**1.43** | 61.94↑**4.10** | 54.49↑**3.34** | 57.31↑**2.01** | 45.23↓**1.44** |
| w/ `pos-wt` | 70.67↑**0.01** | 54.14↑**2.07** | 56.04↑**4.04** | 49.64↑**3.14** | 61.35↑**3.51** | 54.85↑**3.70** | 61.10↑**5.80** | 51.49↑**4.82** |
| AOA | 71.63 | 52.65 | 52.16 | 46.78 | 63.73 | 57.00 | 58.19 | 49.02 |
| w/ `pos-dp` | 72.30↓**0.67** | 53.73↑**1.08** | 53.56↑**1.40** | 48.18↑**1.40** | 65.33↑**1.60** | 58.31↑**1.31** | 56.24↓**1.95** | 45.63↓**3.39** |
| w/ `pos-wt` | 72.61↑**0.98** | 56.54↑**3.89** | 59.07↑**6.91** | 54.92↑**8.14** | 66.87↑**3.14** | 62.02↑**5.02** | 64.35↑**6.16** | 54.62↑**5.60** |
| RoBERTa | 83.16 | 72.99 | 73.57 | 69.26 | 77.62 | 71.34 | 79.08 | 71.79 |
| w/ `pos-dp` | 81.98↓**1.18** | 70.81↓**2.18** | 69.98↓**3.59** | 65.35↓**3.91** | 75.61↓**2.01** | 68.00↓**3.34** | 77.81↓**1.27** | 69.37↓**2.42** |
| w/ `pos-wt` | 83.43↑**0.27** | 74.08↑**1.09** | 75.72↑**2.15** | 72.09↑**2.83** | 79.40↑**1.78** | 74.44↑**3.10** | 79.47↑**0.39** | 73.10↑**1.31** |

Table 2: Robustness results (%). O.O.D. on Lap or Rest denotes a model is trained in current domain (Lap or Rest) and tested on another (Rest or Lap). Adv. denotes a model is trained in a domain and tested on its ARTS counterpart. Furthermore, w/ `pos-dp` means a model with position-biased dropout. w/ `pos-wt` means a model with position-biased weight. The small number next to each performance score indicates either performance improvement (↑) or drop (↓) compared with the original model without using position bias, and those highlighted in red are the best-performing ones among two variants.

# I.D. Result

- Does not harm I.D. performance

| Model | LAP I.D. | | REST I.D. | |
|---|---|---|---|---|
| | Acc. | F1 | Acc. | F1 |
| LSTM | 67.15 | 60.57 | 74.57 | 62.14 |
| w/ `pos-dp` | 67.34 | 60.27 | 74.23 | 61.55 |
| w/ `pos-wt` | 68.78 | 62.42 | 76.34 | 64.85 |

Table 3: I.D. results (%) of LSTM on LAP and REST.

# Case Study

- Case study on attention weights visualization
  - verifies the effectiveness

| w/ `pos-wt` | Example |
|---|---|
| ✘ | The price is reasonable although the quality is poor . |
| ✔ | The price is reasonable although the quality is poor . |
| ✘ | Awful food but the service was great ! |
| ✔ | Awful food but the service was great ! |

Table 4: **Case study.** The underlined words are aspects. The top two rows are O.O.D. examples, while the bottom two are Adv. examples. ✘ and ✔ refers to without and with `pos-wt` respectively.

# Conclusion

- To improve the robust of ASC models,  we propose a simple yet effective inductive bias that should be incorporated, that is, position bias.
- We proposed two mechanisms to capture position bias, namely ***position-biased weight*** and ***position-biased dropout***.
- The experimental results verify our hypothesis that position bias is beneficial for building more robust ASC models.

# The End

Thanks a lot.