Comparison with bae_garg_2019

Bert-base-uncased-yelp-polarity model:

Our Attack

```
| Attack Results                 |        |
+-------------------------------+--------+
| Number of successful attacks: | 21     |
| Number of failed attacks:     | 4      |
| Number of skipped attacks:    | 0      |
| Original accuracy:            | 100.0% |
| Accuracy under attack:        | 16.0%  |
| Attack success rate:          | 84.0%  |
| Average perturbed word %:     | 12.51% |
| Average num. words per input: | 133.36 |
| Avg num queries:              | 238.84 |
+-------------------------------+--------+
```
textattack: Attack time: 252.68779397010803s

bae_garg

```
| Attack Results                 |        |
+-------------------------------+--------+
| Number of successful attacks: | 11     |
| Number of failed attacks:     | 14     |
| Number of skipped attacks:    | 0      |
| Original accuracy:            | 100.0% |
| Accuracy under attack:        | 56.0%  |
| Attack success rate:          | 44.0%  |
| Average perturbed word %:     | 6.44%  |
| Average num. words per input: | 133.36 |
| Avg num queries:              | 314.88 |
+-------------------------------+--------+
```
textattack: Attack time: 764.2062220573425s

Bert-base-uncased-imdb model:

Our Attack

```
| Attack Results                 |        |
+-------------------------------+--------+
| Number of successful attacks: | 21     |
| Number of failed attacks:     | 4      |
| Number of skipped attacks:    | 2      |
| Original accuracy:            | 92.59% |
| Accuracy under attack:        | 14.81% |
| Attack success rate:          | 84.0%  |
| Average perturbed word %:     | 5.29%  |
| Average num. words per input: | 248.96 |
| Avg num queries:              | 419.72 |
+-------------------------------+--------+
```
textattack: Attack time: 268.29629945755005s

bae_garg

```
| Attack Results                 |        |
+-------------------------------+--------+
| Number of successful attacks: | 12     |
| Number of failed attacks:     | 13     |
| Number of skipped attacks:    | 2      |
| Original accuracy:            | 92.59% |
| Accuracy under attack:        | 48.15% |
| Attack success rate:          | 48.0%  |
| Average perturbed word %:     | 2.97%  |
| Average num. words per input: | 248.96 |
| Avg num queries:              | 561.52 |
+-------------------------------+--------+
```
textattack: Attack time: 1153.5890488624573s

Albert-Base-v2-yelp-polarity model:

<u>Our Attack</u>   <u>bae_garg</u>

```
| Attack Results              |        |   | Attack Results              |        |
+----------------------------+--------+   +----------------------------+--------+
| Number of successful attacks: | 17    |   | Number of successful attacks: | 11    |
| Number of failed attacks:   | 8      |   | Number of failed attacks:   | 14     |
| Number of skipped attacks:  | 0      |   | Number of skipped attacks:  | 0      |
| Original accuracy:          | 100.0% |   | Original accuracy:          | 100.0% |
| Accuracy under attack:      | 32.0%  |   | Accuracy under attack:      | 56.0%  |
| Attack success rate:        | 68.0%  |   | Attack success rate:        | 44.0%  |
| Average perturbed word %:   | 8.19%  |   | Average perturbed word %:   | 4.72%  |
| Average num. words per input: | 133.36 |   | Average num. words per input: | 133.36 |
| Avg num queries:            | 242.28 |   | Avg num queries:            | 287.24 |
+----------------------------+--------+   +----------------------------+--------+
```
textattack: Attack time: 146.80171513557434s **textattack**: Attack time: 438.28860211372375s

Albert-Base-v2-imdb model:

<u>Our Attack</u>   <u>bae_garg</u>

```
| Attack Results              |        |   | Attack Results              |        |
+----------------------------+--------+   +----------------------------+--------+
| Number of successful attacks: | 22    |   | Number of successful attacks: | 16    |
| Number of failed attacks:   | 3      |   | Number of failed attacks:   | 9      |
| Number of skipped attacks:  | 3      |   | Number of skipped attacks:  | 3      |
| Original accuracy:          | 89.29% |   | Original accuracy:          | 89.29% |
| Accuracy under attack:      | 10.71% |   | Accuracy under attack:      | 32.14% |
| Attack success rate:        | 88.0%  |   | Attack success rate:        | 64.0%  |
| Average perturbed word %:   | 5.2%   |   | Average perturbed word %:   | 2.73%  |
| Average num. words per input: | 244.71 |   | Average num. words per input: | 244.71 |
| Avg num queries:            | 434.92 |   | Avg num queries:            | 512.32 |
+----------------------------+--------+   +----------------------------+--------+
```
textattack: Attack time: 584.6124613285065s **textattack**: Attack time: 1051.5286781787872s