SentiBERT: A Transferable Transformer-based Architecture for Compositional Sentiment Semantics

Da Yin\textsuperscript{1}, Tao Meng\textsuperscript{2}, Kai-Wei Chang\textsuperscript{2}

\textsuperscript{1}Peking University
\textsuperscript{2}University of California, Los Angeles
Motivation

- Sentiment composition is challenging.

Frenetic but **not really funny**.
Motivation

- How to encode sentiment composition in a contextual encoder?
- Can semantic composition learned from SST transfer to related tasks?

Better capture sentiment composition
Model

BERT

Sentiment Semantics Composition

Phrase Node Prediction
Model

- BERT
- Sentiment Semantics Composition
- Phrase Node Prediction

Layer 1: Attention to Tokens
- Layer 2: Attention to Children
Training Objectives

Masked LM

Phrase Annotation

Phrase Node Prediction

SentiBERT
Experiments

- Tasks:
  - SST-phrase
  - SST-5
  - SST-2, SST-3
  - Twitter Sentiment Analysis
  - Contextual Emotion Detection (EmoContext)
  - Emotion Intensity Classification (EmoInt)

- Evaluated under supervised learning protocol
- Test transferability
Experiments

- Results:

  - For sentiment semantic composition:

    - SST-phrase (Accuracy)
      - BERT
      - BERT w/ Tree-LSTM
      - SentiBERT
      - RoBERTa
      - SentiBERT w/ RoBERTa

    - SST-5 (Accuracy)
      - BERT
      - BERT w/ Tree-LSTM
      - SentiBERT
      - RoBERTa
      - SentiBERT w/ RoBERTa

More results and discussion are in the paper.
Experiments

- Results:
  - For transferability:
    
    | Model                | EmoInt (Pearson Correlation) | EmoContext (F1) |
    |----------------------|------------------------------|-----------------|
    | BERT                 | 0.645                        | 0.745           |
    | SentiBERT            | 0.65                         | 0.74            |
    | RoBERTa              | 0.655                        | 0.735           |
    | SentiBERT w/ RoBERTa| 0.665                        | 0.745           |

More results and discussion are in the paper.
Analysis -- Performance v.s. Sentiment Switch

- **Local difficulty**: the number of sentiment switches between a phrase and its children
- **Global difficulty**: the number of sentiment switches in the entire constituency tree

Diagram:

```
Frenetic  but  not
         / \          
      really  funny
```

- **positive**
- **neutral**
- **negative**
Analysis

- Results:

Global Difficulty

Local Difficulty

More results and discussion are in the paper
Case Study

More examples are in the paper
We present SentiBERT to better capture compositional sentiment semantics

SentiBERT can transfer the compositional sentiment semantics learned on SST to other related tasks

Thanks!
GitHub: https://github.com/WadeYin9712/SentiBERT