Initializing Convolutional Filters with Semantic Features for Text Classification

Shen Li\textsuperscript{1,2} \hspace{1cm} Zhe Zhao\textsuperscript{4,5}
shen@mail.bnu.edu.cn \hspace{1cm} helloworld@ruc.edu.cn

Tao Liu\textsuperscript{4,5} \hspace{1cm} Renfen Hu\textsuperscript{3,†} \hspace{1cm} Xiaoyong Du\textsuperscript{4,5}
tliu@ruc.edu.cn \hspace{1cm} irishere@mail.bnu.edu.cn \hspace{1cm} duyong@ruc.edu.cn

\textsuperscript{1} Institute of Chinese Information Processing, Beijing Normal University
\textsuperscript{2} UltraPower-BNU Joint Laboratory for Artificial Intelligence, Beijing Normal University
\textsuperscript{3} College of Chinese Language and Culture, Beijing Normal University
\textsuperscript{4} School of Information, Renmin University of China
\textsuperscript{5} Key Laboratory of Data Engineering and Knowledge Engineering, MOE

Appendix

The crux of our initialization technique is n-gram selection, which assists neural networks to extract important n-gram features at the beginning of the training process. In the following tables, we illustrate those selected n-grams of different classes and datasets to understand our technique intuitively. Since all of MR, SST-1, SST-2, CR, and MPQA are sentiment classification datasets, we only report the selected n-grams of SST-1 (Table 1). N-grams selected by our method in SUBJ and TREC are shown in Table 2 and Table 3.

| Class     | Very Positive | Positive | Neutral | Negative | Very Negative |
|-----------|---------------|----------|---------|----------|---------------|
| Unigram   | standout      | heartening | kin     | choppiness | flopped       |
|           | perfection    | virtuosic | reworked | woozy     | indescrimably |
|           | releases      | affectionately | michelle | meager     | atrocity      |
| Bigram    | mesmerizing   | with raw  | man vs  | left slightly | definitely meaningless |
|           | music         | remarkable about | kin ’s | been conjured | wasted nearly |
|           | satisfying evenings | this much | the sides | ridiculous wig | is meaningless |
| Trigram   | best films of | grounded in an | even one word | conjured up only | devoid of substance |
|           | making it one | enjoyable and satisfying | pleasant enough and | difficult to fathom | is definitely meaningless |
|           | enjoyable and satisfying | | than to receive | dumbed down approach | with this silly |

Table 1: Examples of the selected n-grams in SST-1 dataset. The results are self-explanatory. There are five classes in SST-1 dataset. The polarities of n-grams selected from very positive texts to very negative texts change smoothly. Adjectives with positive sentiment are easily selected in positive texts, e.g. “enjoyable”, “beautiful” and “satisfying”. Obviously, n-grams indicating negative emotions are more likely to be selected in negative texts such as “wasted” and “meaningless”.

| Class     | Subjective | Objective |
|-----------|------------|-----------|
| Unigram   | amusing    | discovers |
|           | laughs     | 233       |
|           | i          | decide    |
|           | entertaining | boyfriend |
| Bigram    | movie that | his father |
|           | it does    | him to    |
|           | but it     | he finds  |
|           | the performances | where he |
| Trigram   | but it’s   | is the story |
|           | a movie that | the help of |
|           | if you ‘re | falls in love |
|           | it’s not   | in order to |

Table 2: Examples of the selected n-grams in Subj dataset. We can observe that adjectives such as “amusing” and “entertaining” are more likely to be selected in subjective reviews, and neutral words such as “his” and “him” are more likely to be selected in objective reviews.
Table 3: Examples of the selected n-grams in TREC dataset. Strong indicators of question types are selected by NB weights. For example, “acronym”, “stand for”, and “the abbreviation of” are selected for the abbreviation question type. The n-grams that are related to entities’ attributes such as “disease” and “animal” are selected for the entity question type. Human’s actions (e.g. “portrayed”, “who invented”) are selected for the human question type. “what country” and “what city” possess large NB weights in questions about location. In questions of the number type, “how many” and “when was” are selected.