

Building Large-Scale Twitter-Specific Sentiment Lexicon: A Representation Learning Approach

Duyu Tang¹, Furu Wei², Bing Qin¹, Ming Zhou², Ting Liu¹

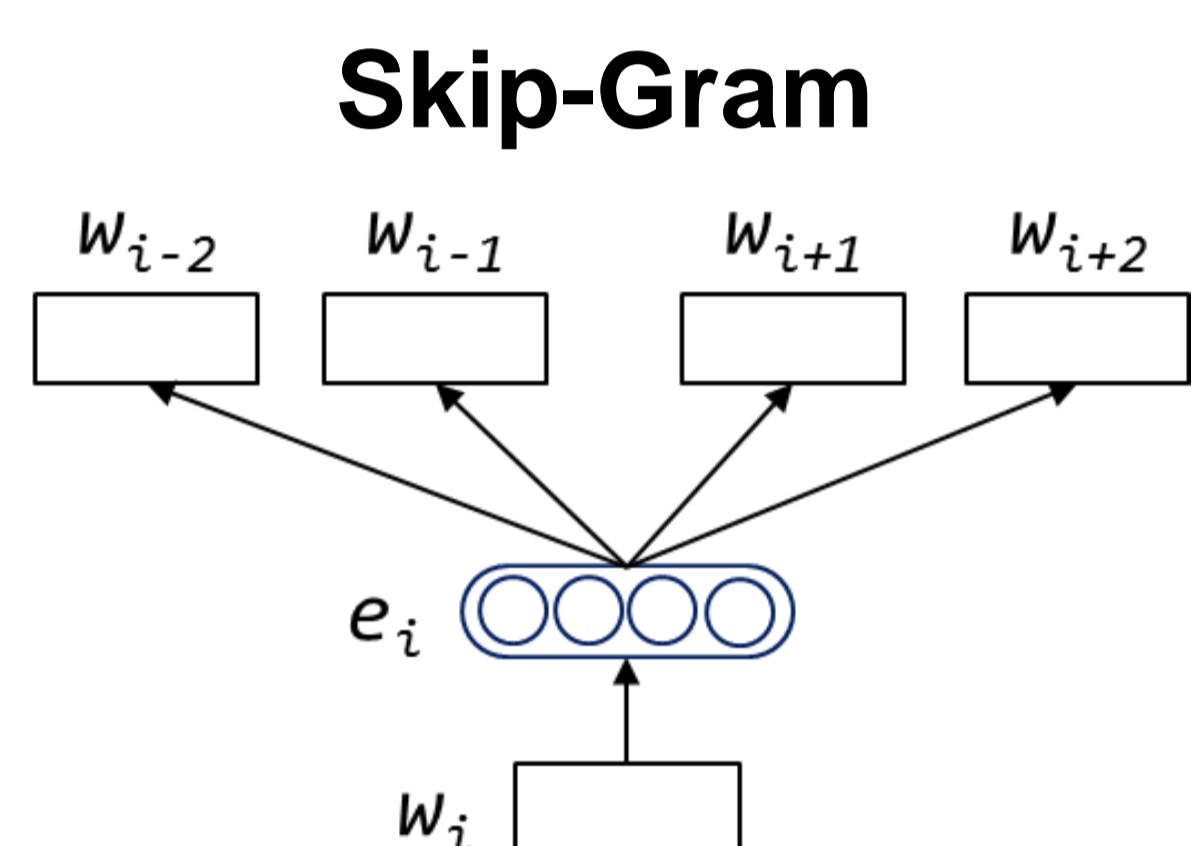
¹ Harbin Institute of Technology, Harbin, China

² Microsoft Research, Beijing, China

1. Introduction

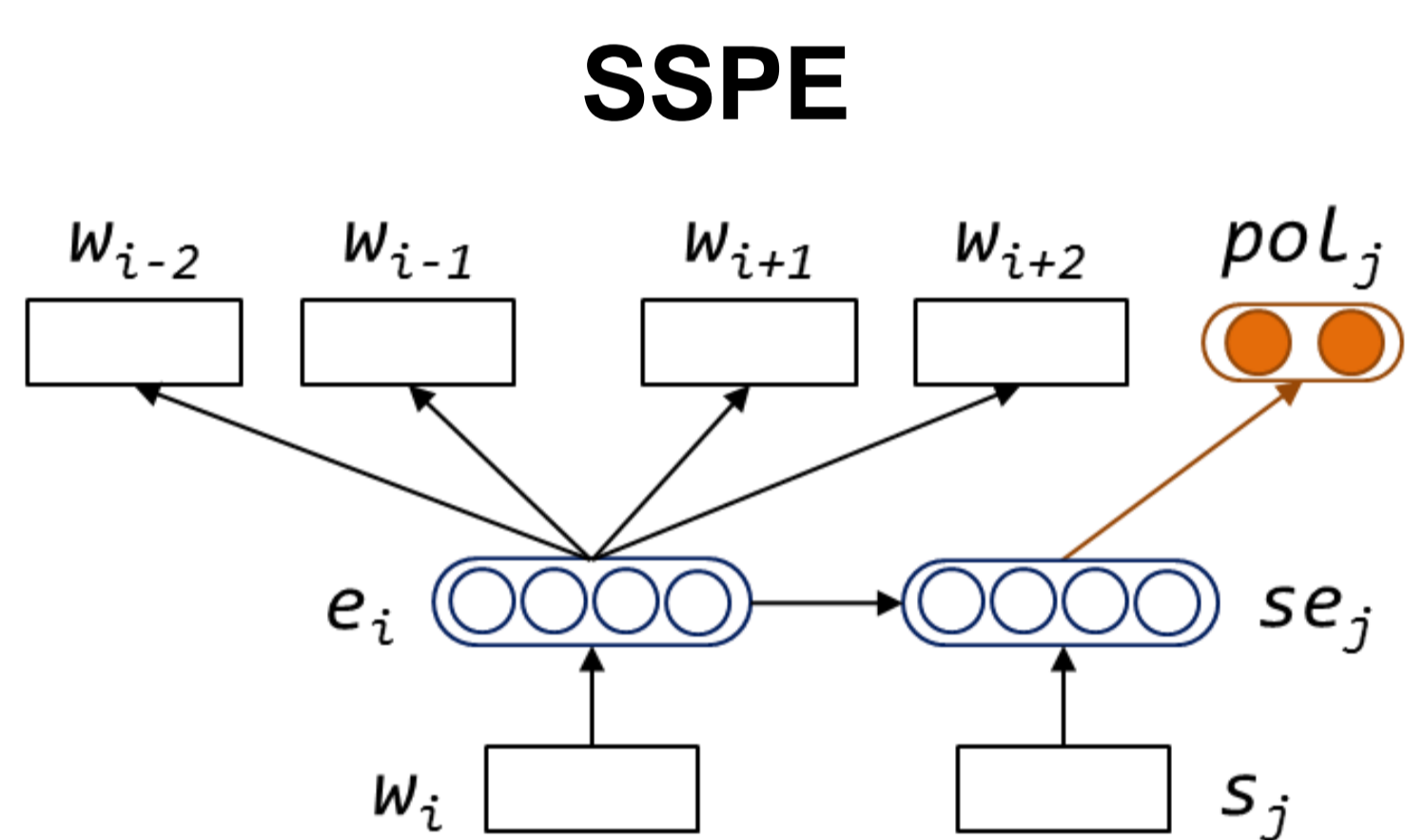
- Sentiment lexicon is important for sentiment analysis.
- In this work, we build large-scale sentiment lexicon within a classification framework.
- Two challenges:
 - Develop feature representations of phrases
 - Get training data with minor manual annotations

3. Sentiment-Specific Phrase Embedding



- Predict the **context words** of the given phrase

$$\frac{1}{T} \sum_{i=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{i+j} | e_i)$$

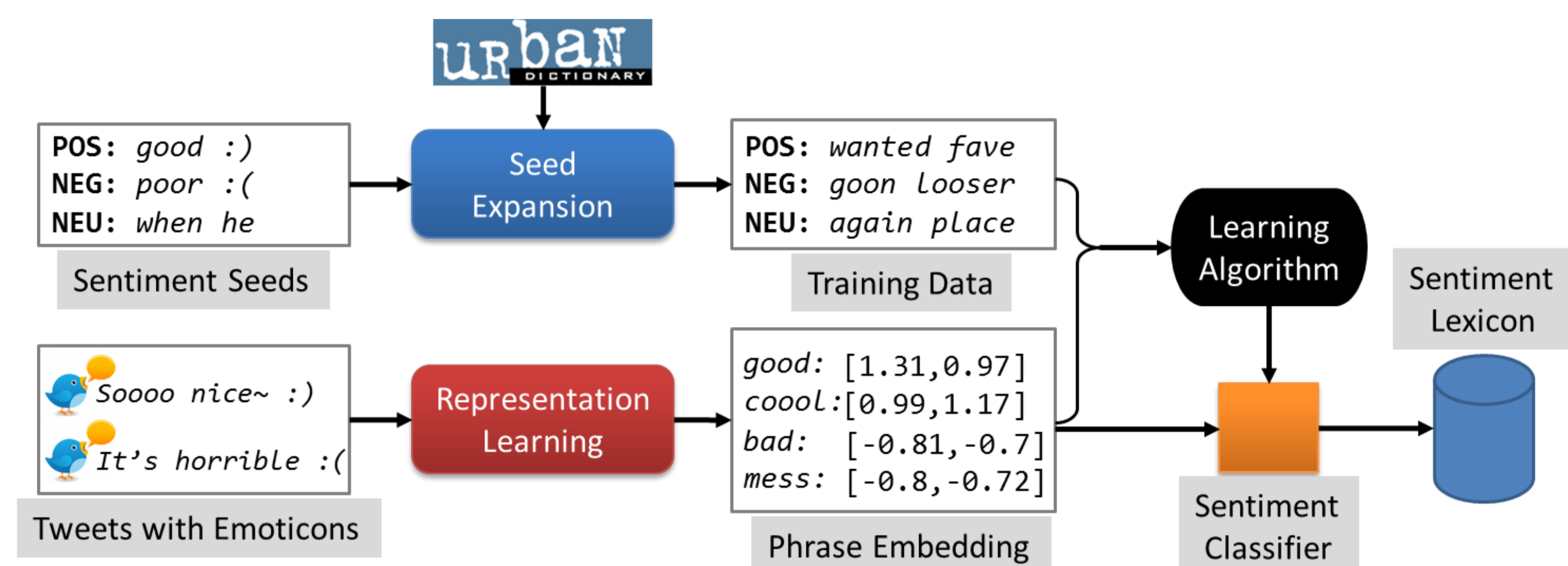


- Predict the **context words** of phrases and the **polarity labels** of sentences

$$\alpha \cdot f_{syntactic} + (1 - \alpha) \cdot \frac{1}{S} \sum_{j=1}^S \log p(pol_j | se_j)$$

2. Our Approach

- Learning phrase feature representation
 - Sentiment-specific phrase embedding (**Part 3**)
- Getting more training data
 - Seed expansion with Urban Dictionary (**Part 4**)



4. Seed Expansion with Urban Dictionary

- Use the similar words in Urban Dictionary to expand a small size of sentiment seeds
- We build a KNN classifier to select expanded words with high confidence.

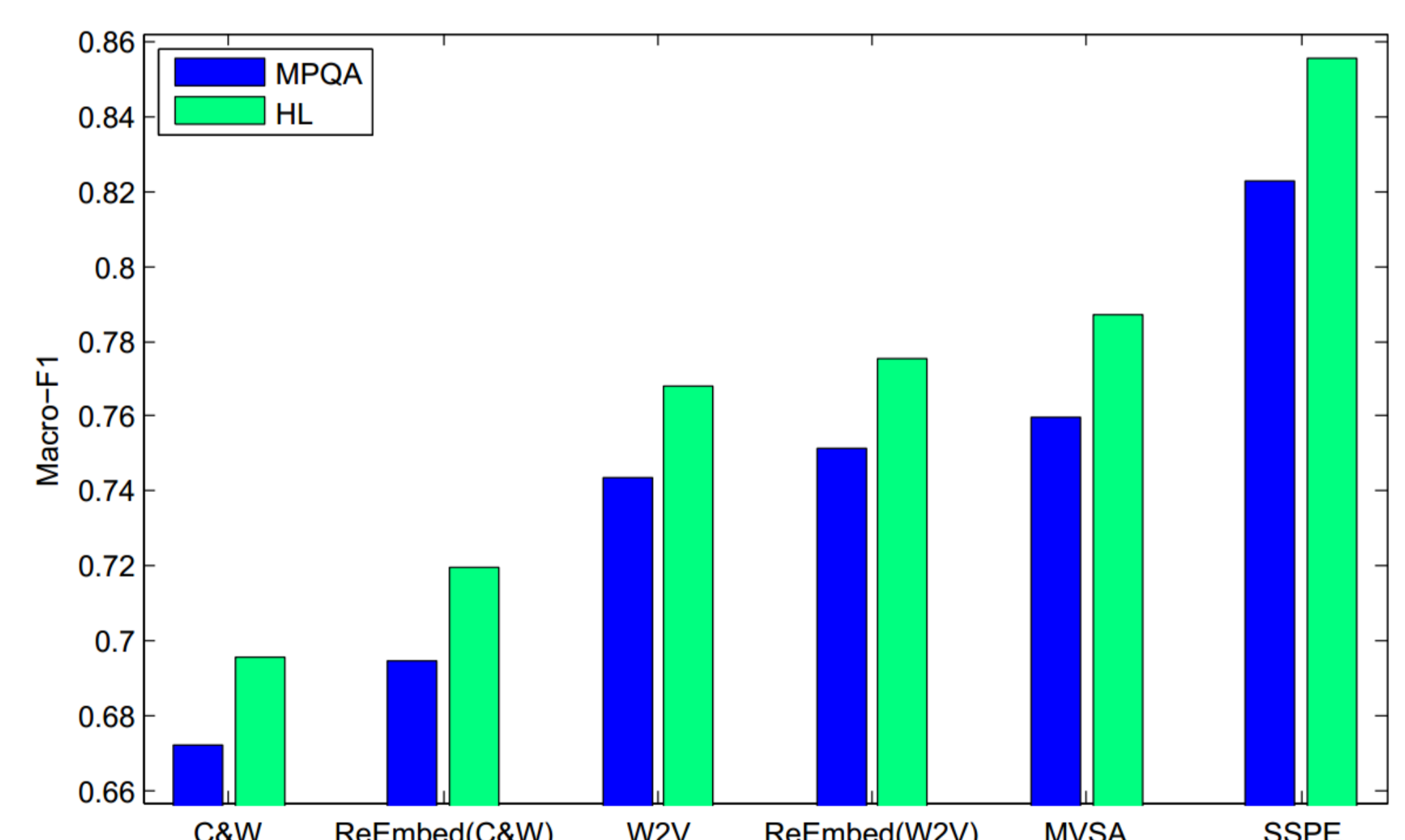


5. Experiment

- The statistical information of sentiment lexicons

Lexicon	Positive	Negative	Total
HL	2,006	4,780	6,786
MPQA	2,301	4,150	6,451
NRC-Emotion	2,231	3,324	5,555
TS-Lex	178,781	168,845	347,626
HashtagLex	216,791	153,869	370,660
Sentiment140Lex	480,008	260,158	740,166

- Evaluate SSPE through sentiment classification of lexicons



- Apply sentiment lexicon for Twitter sentiment classification

Lexicon	Unique	Appended
HL	60.49	79.40
MPQA	59.15	76.54
NRC-Emotion	54.81	76.79
HashtagLex	65.30	76.67
Sentiment140Lex	72.51	80.68
TS-Lex	78.07	82.36

Lexicon	Unique
Seed	57.92
Expand	60.69
Lexicon(seed)	74.64
TS-Lex	78.07

- Unique: Only use the lexicon feature
- Appended: Use feature combination

- Effect of alpha in SSPE for Twitter sentiment classification

