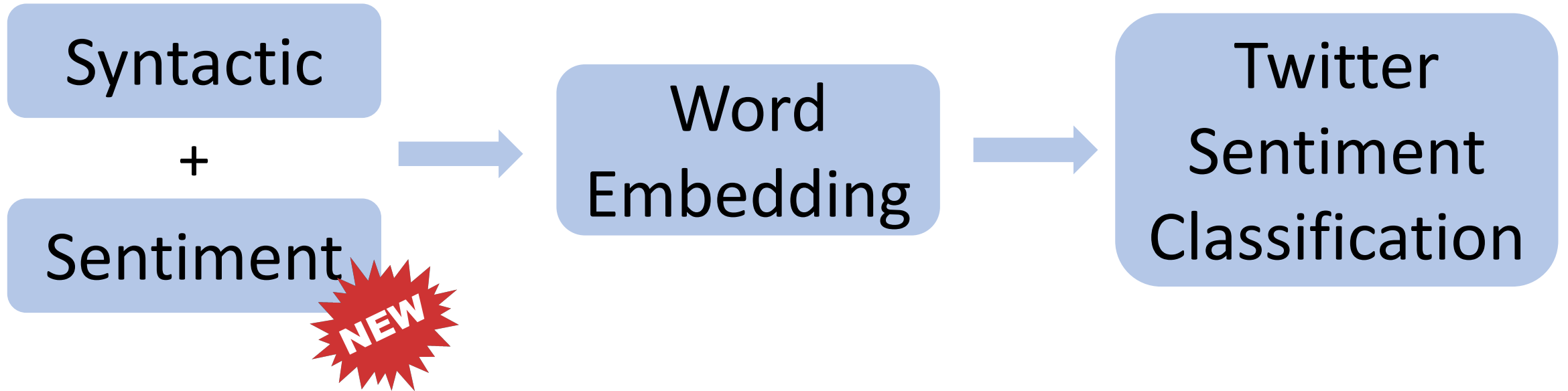


Learning Sentiment-Specific Word Embedding (SSWE) for Twitter Sentiment Classification

Chang Shu 04/30/2020



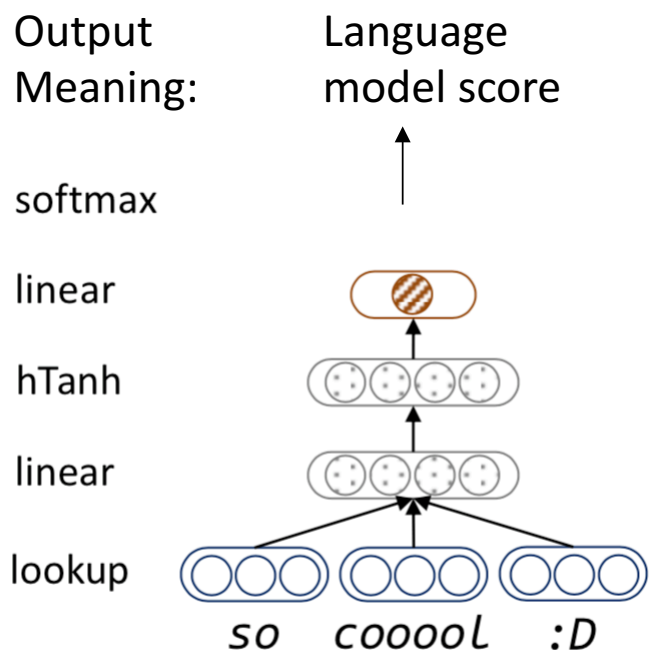
Algorithms comparison

Model	Syntactic	Sentiment
C&W	YES	NO
SSWE(h)	NO	YES
SSWE(r)	NO	YES
SSWE(u)	YES	YES

* Sentiment-Specific
Word Embedding
(SSWE)

The authors' approaches, will be introduced in the coming slides

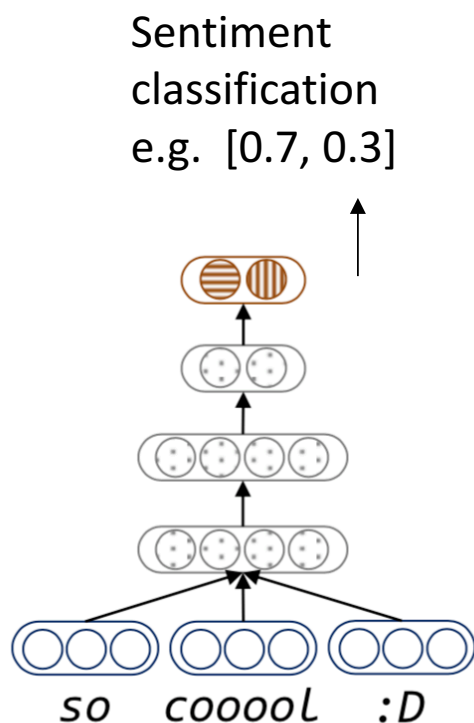
C&W



Input: One-hot for **pairs** of text
(original & corrupted)

Example of model input

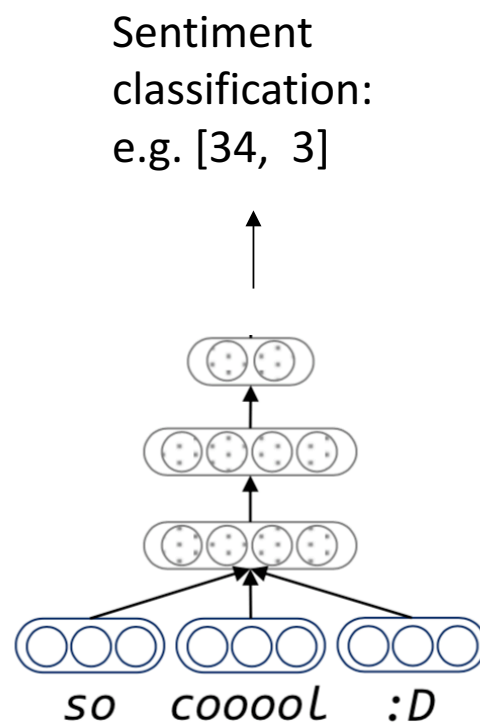
SSWE(h)



One-hot for slide of text

Example of model input

SSWE(r)

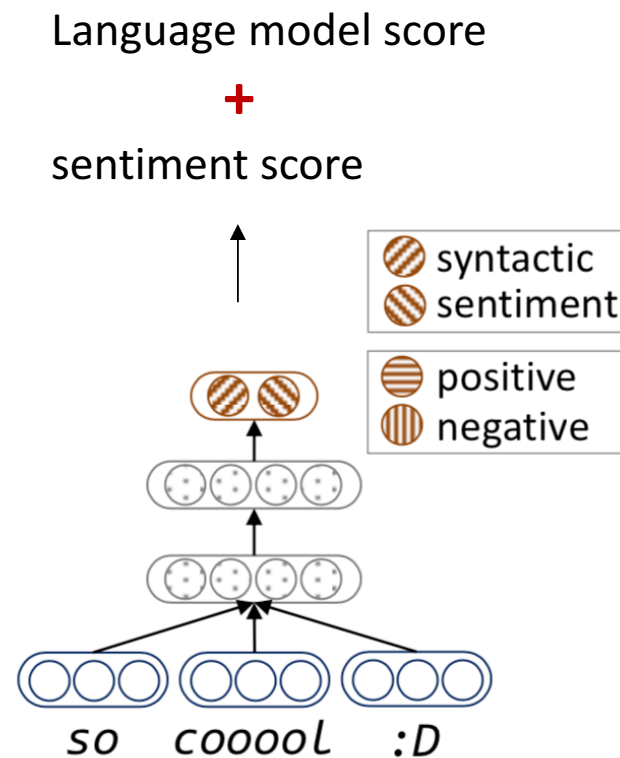


The same as SSWE(h)

Example of model input

SSWE(u)

Best



The same as C&W

Example of model input

Dataset

- Context

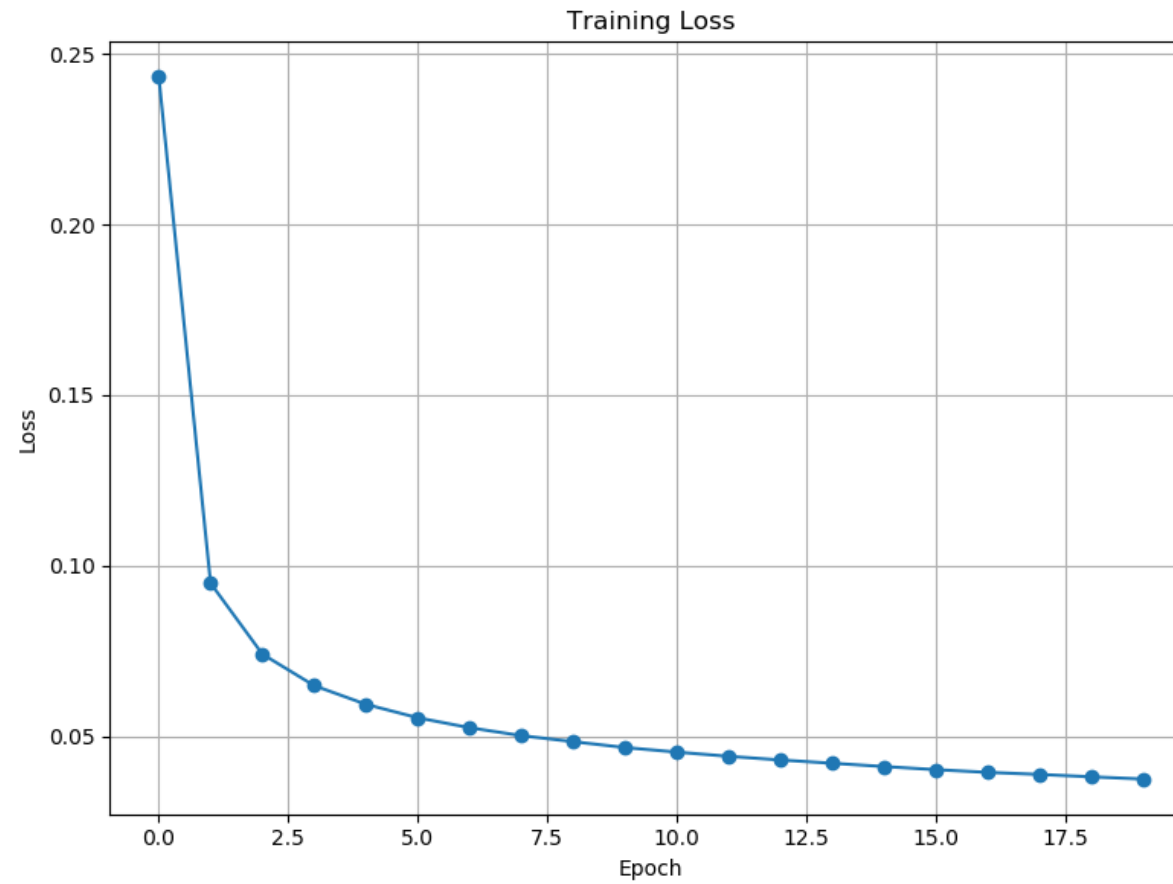
- Not from the paper. This is the sentiment140 dataset. It contains 1,600,000 tweets extracted using the twitter api . The tweets have been annotated (0 = negative, 4 = positive) and they can be used to detect sentiment . After removing sentences which length are shorter than 3, I obtained a training set with 1,214,748 samples and a validation set with 303,687 samples.

- It contains the following 6 fields:

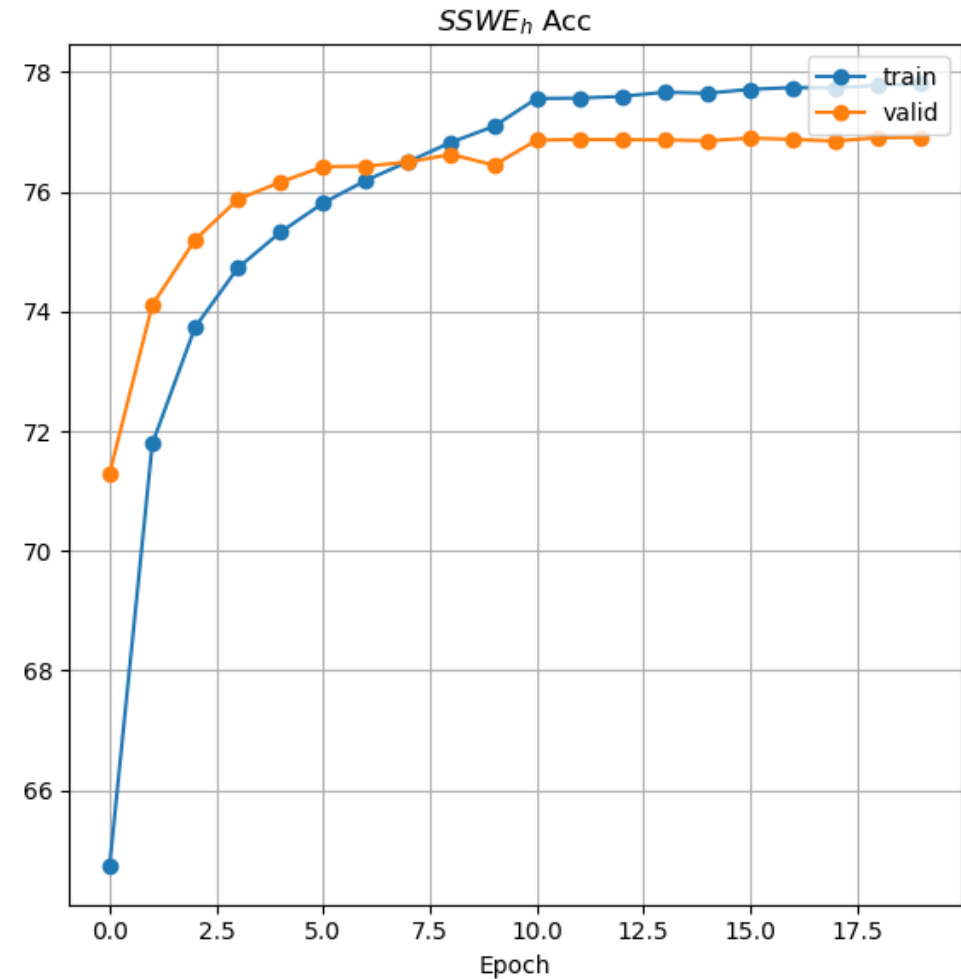
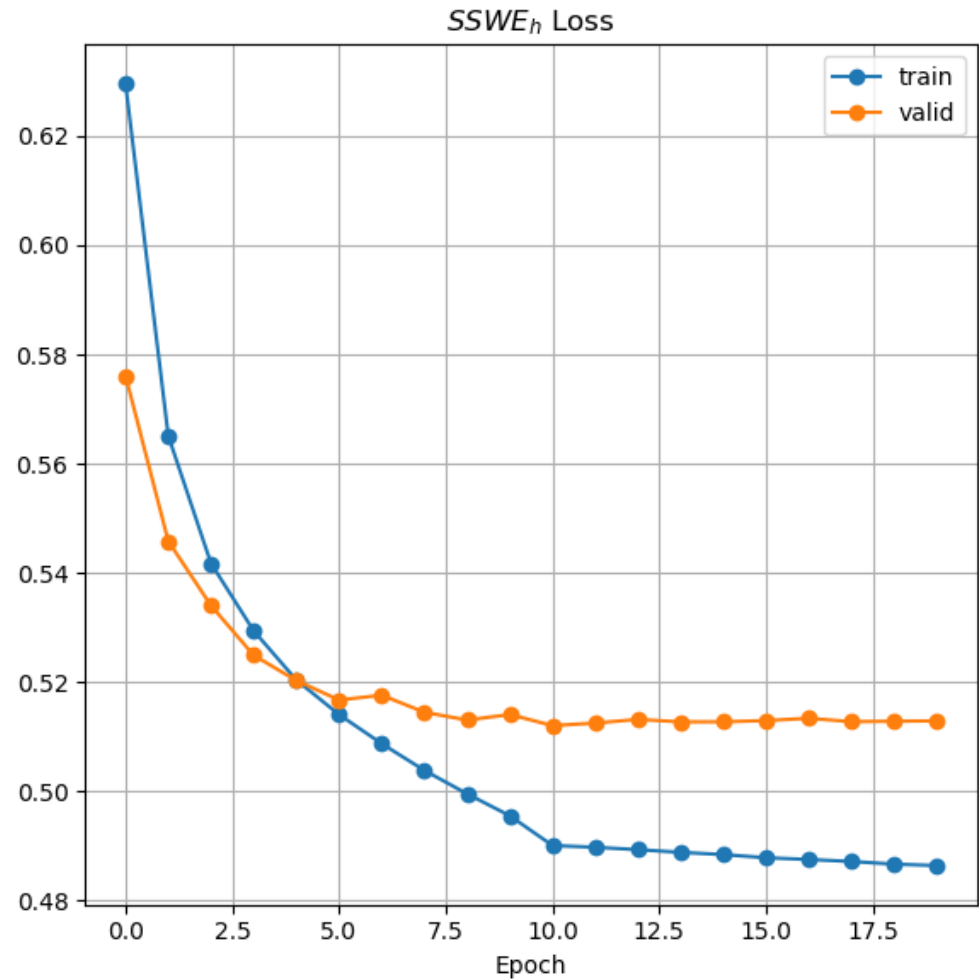
- target: the polarity of the tweet (0 = negative, 2 = neutral, 4 = positive)
- ids: The id of the tweet (2087)
- date: the date of the tweet (*Sat May 16 23:58:44 UTC 2009*)
- flag: The query (*/lyx*). If there is no query, then this value is NO_QUERY.
- user: the user that tweeted (*robotickilldozr*)
- text: the text of the tweet (*Lyx is cool*)

2	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	@kenichan I lived many times for the ball. Managed to save 50% The rest go out of bounds
3	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire

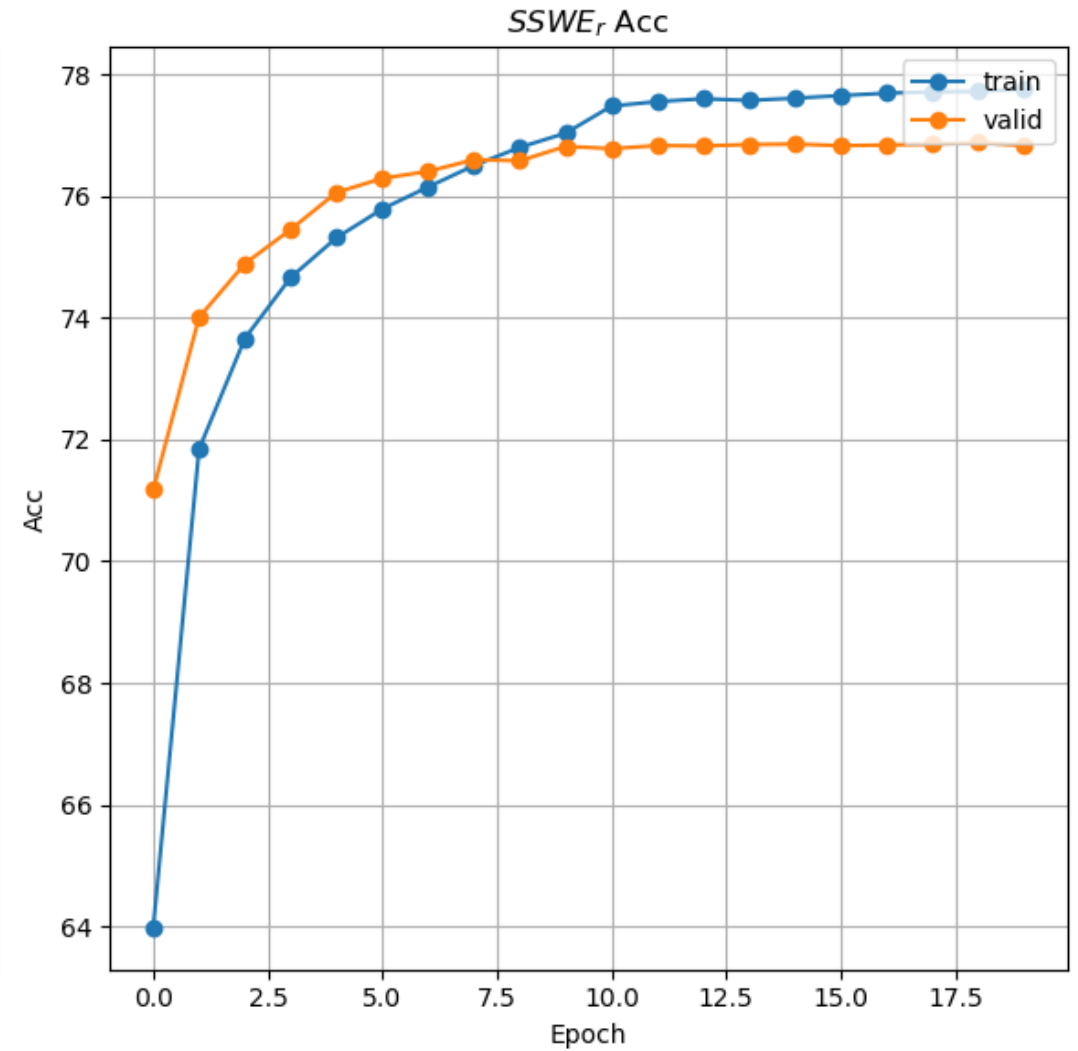
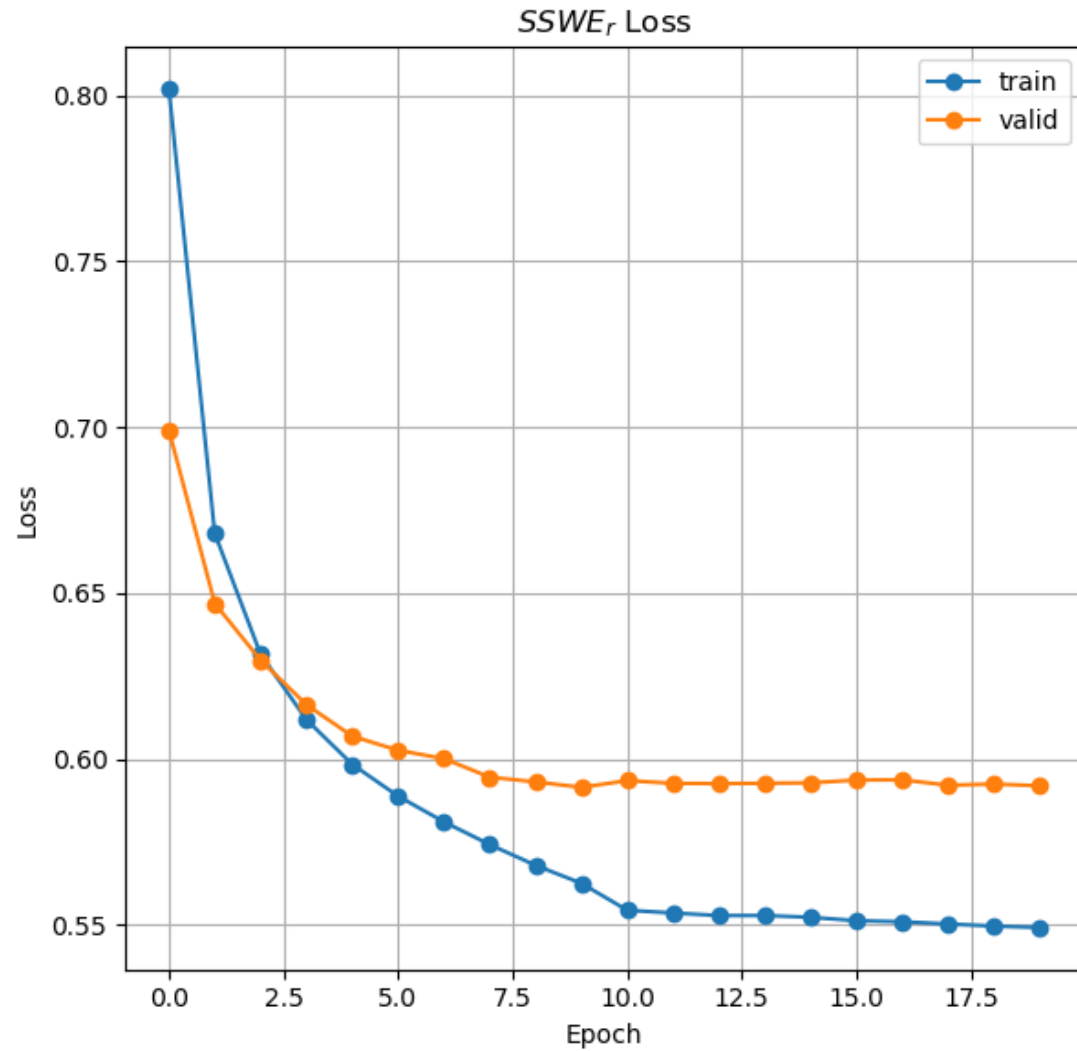
Model result: C&W - loss



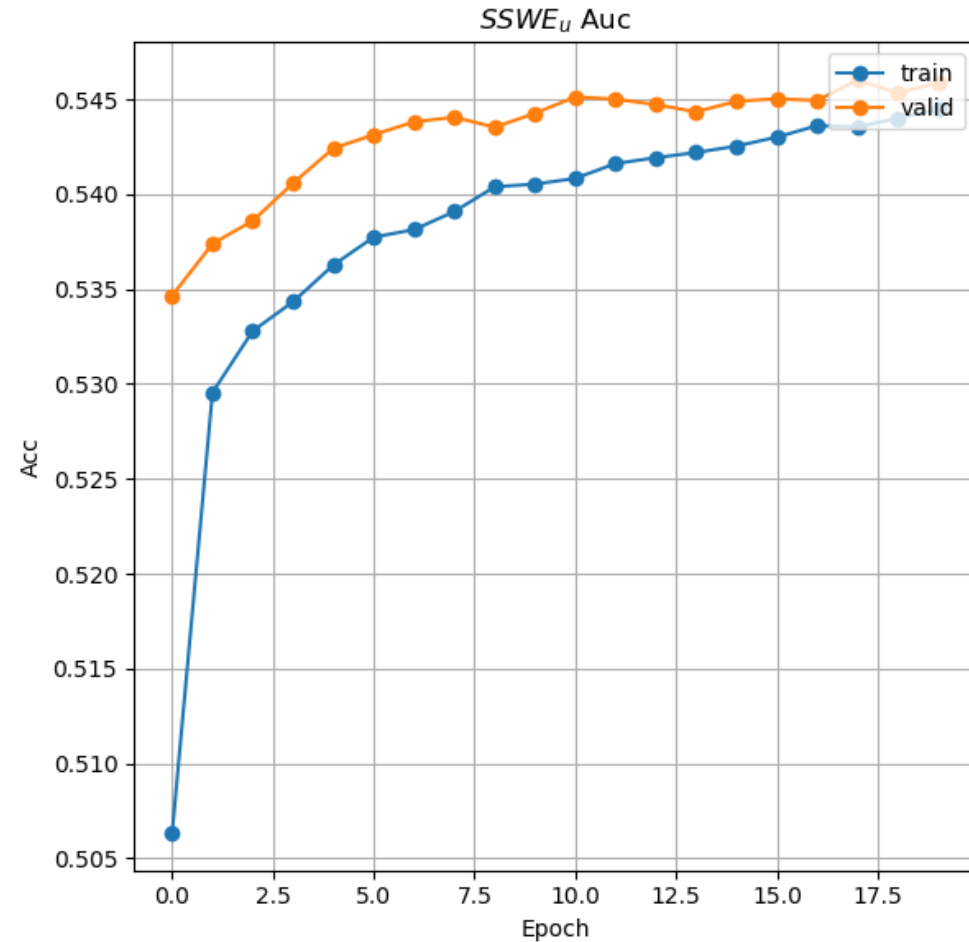
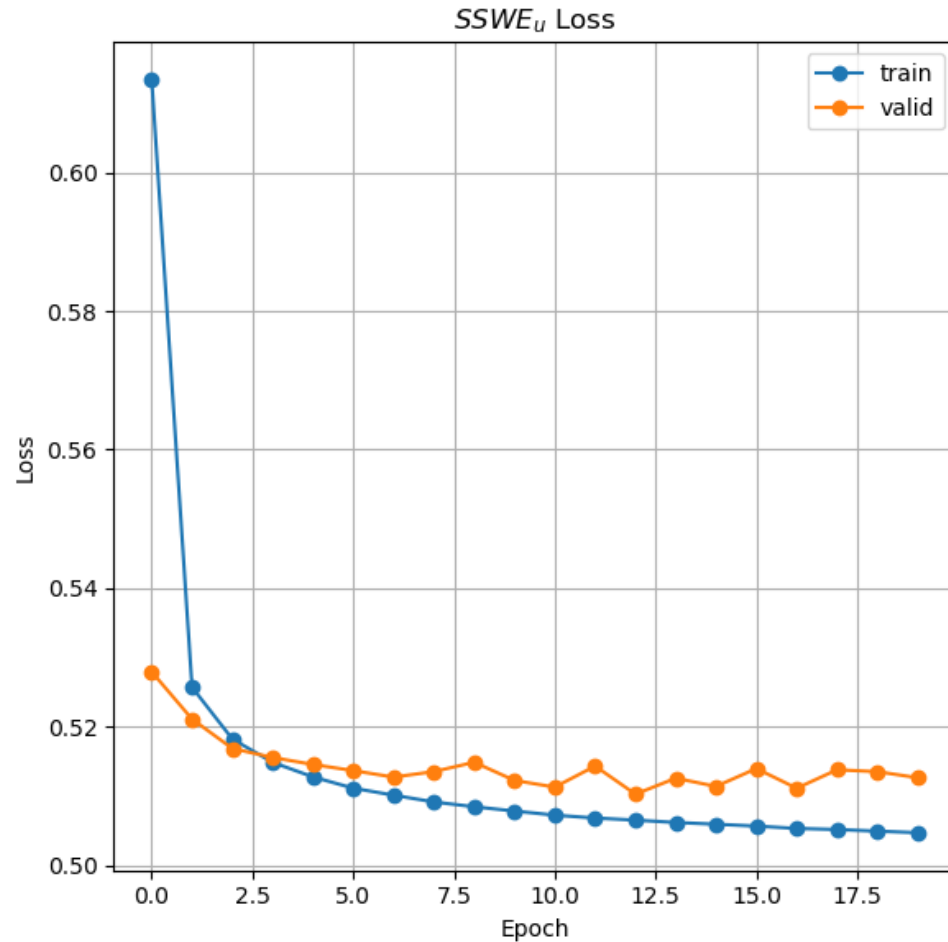
Model result: SSWE(h) – loss & Accuracy



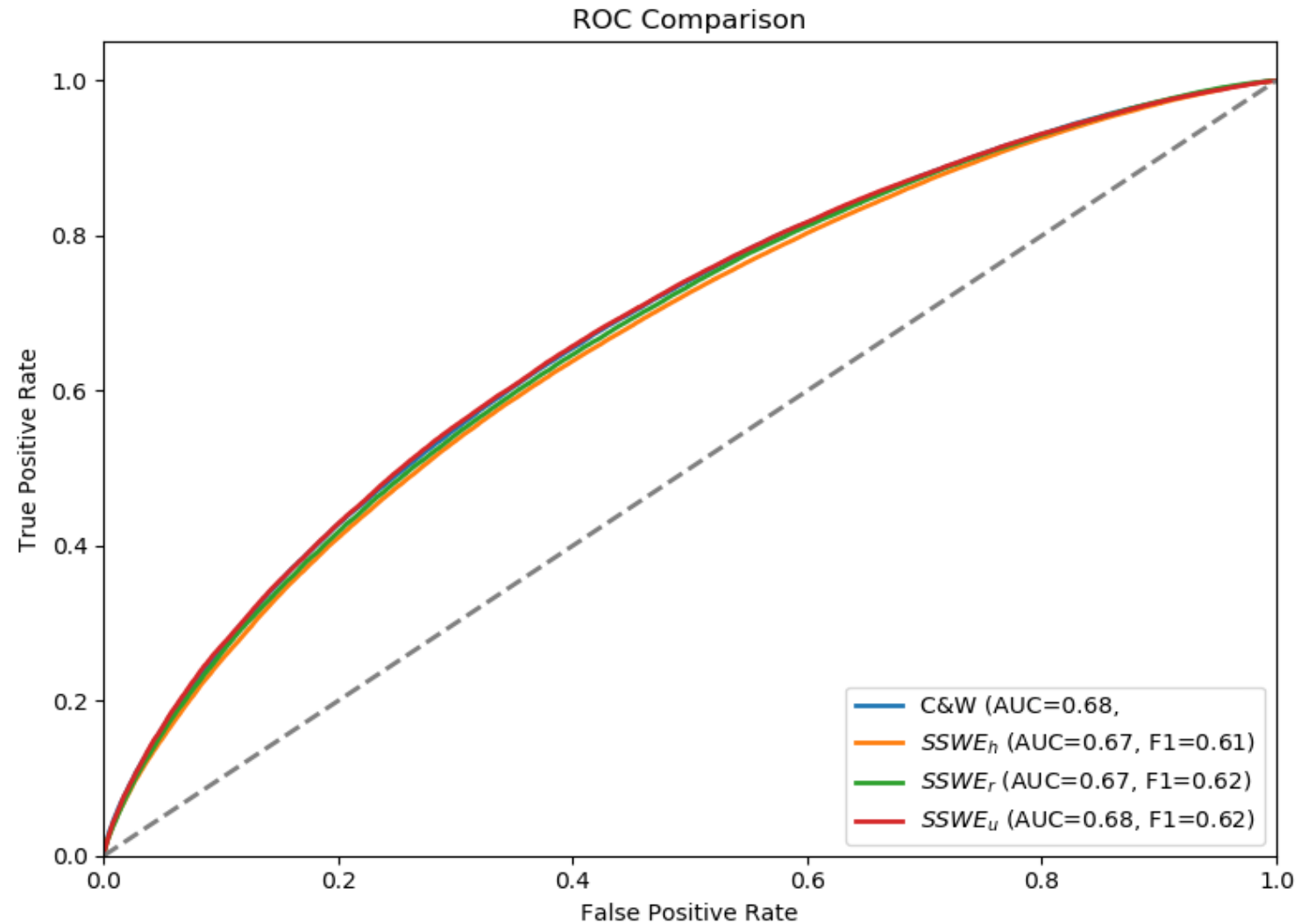
Model result: SSWE(r) – loss & Accuracy



Model result: SSWE(u) – loss & Auc



Model result: AUC & macro-F1



Limitation

- Only use uni-gram
- Just trained 20 epochs
- Haven't use Twitter-specific tokenization

Live Demo

Thank you and stay safe!