BLACKLIGHT (machine learning)

Training & Testing

🟥 1. data

🟩 2. vectorize

🟩 3. model

Application

☑ 1. preprocess

□ 2. apply model

🟥 3. "comment"

Data:

☑ 1. categorize sentences

🟥 2. parse sentences to phrases

🟩 3. categorize phrases

size: 150,000 ?

Vectorize:

~~1. Count Vectorizer~~

2. ngram vectorizer (tfidf)

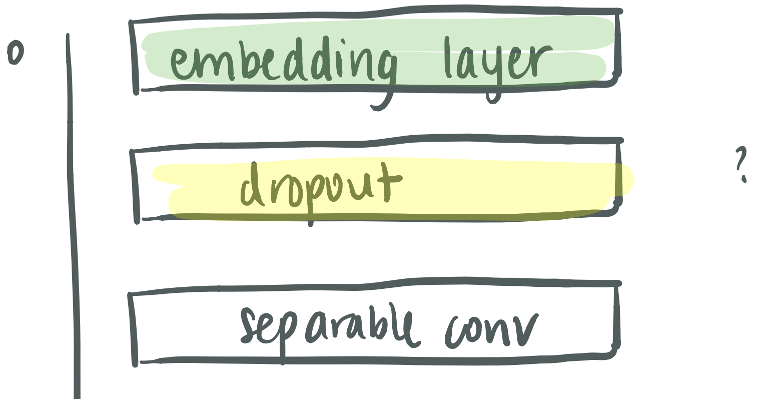
3. Hashing vectorizer

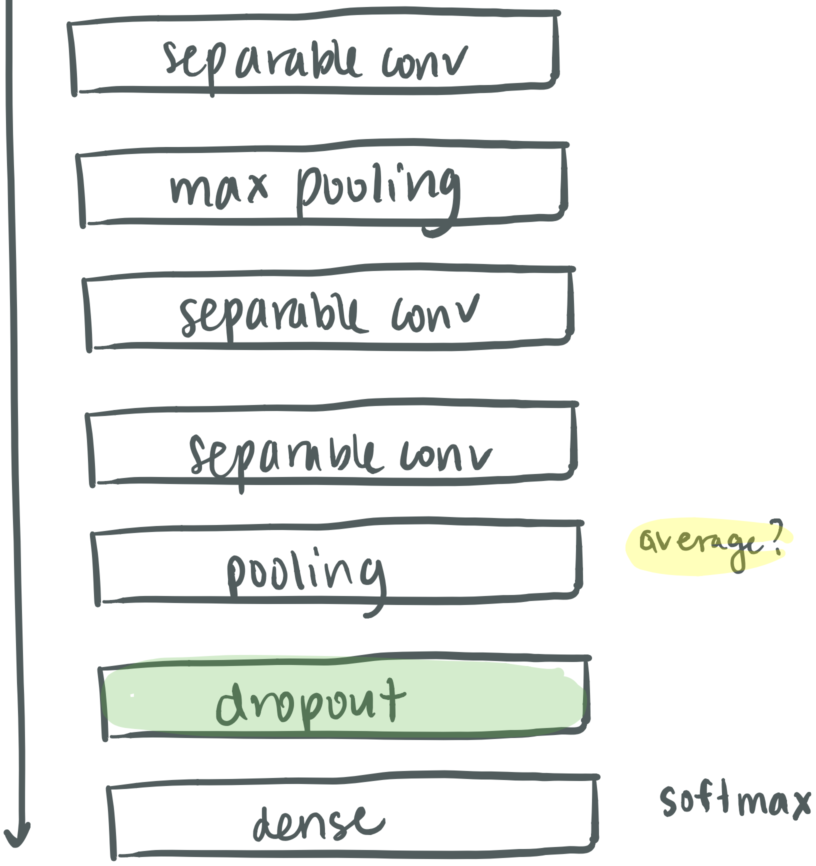
4. sequence vectorizer (.texts_to_sequences)

model:

(i don't think there's a lot of value in making a model if we can only analyze/test it w/ bad training data?)

Separable CNN:





deep pyramid CNN ?

CNN w/ batch normalization & ReLU activation?

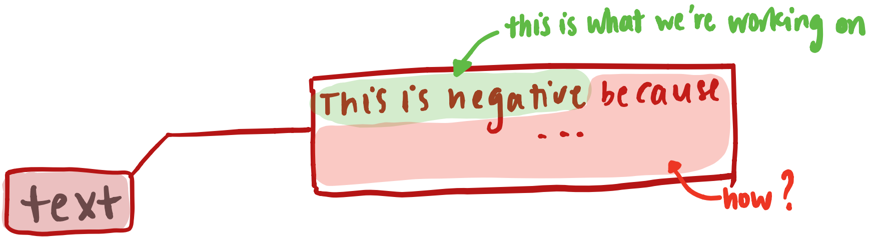
---

preprocessing: lemmatize ?

- ☑ load file
- ☑ clean text (whitespace, case)
- ☑ focus on sentences w/ relevant words/phrases

---

annotate:


this is what we're working on
This is negative because
...
how ?
text