THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS
xkcd: 1838

An Intro to BERT

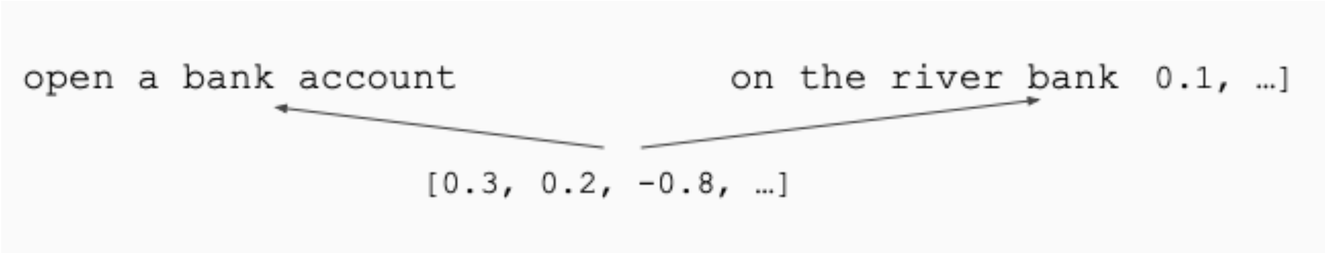
Fine-tuning Google's SOTA NLP model to classify movie reviews

Steven Newton
29.04.21

Making a deep contextual language model

Contextual embedding: “word embeddings are the basis of deep learning for NLP”*

Contextual pre-trained word embeddings (ELMo)



Contextual pre-training: “word masking is the crux of BERT”*

```
the [MASK] store called, they’re running out of [MASK]
```

*Jacob Devlin, co-author of *BERT: Pre-training of Deep Bidirectional Transformers* (2019)

DATA

Large Movie Review Dataset

ai.stanford.edu

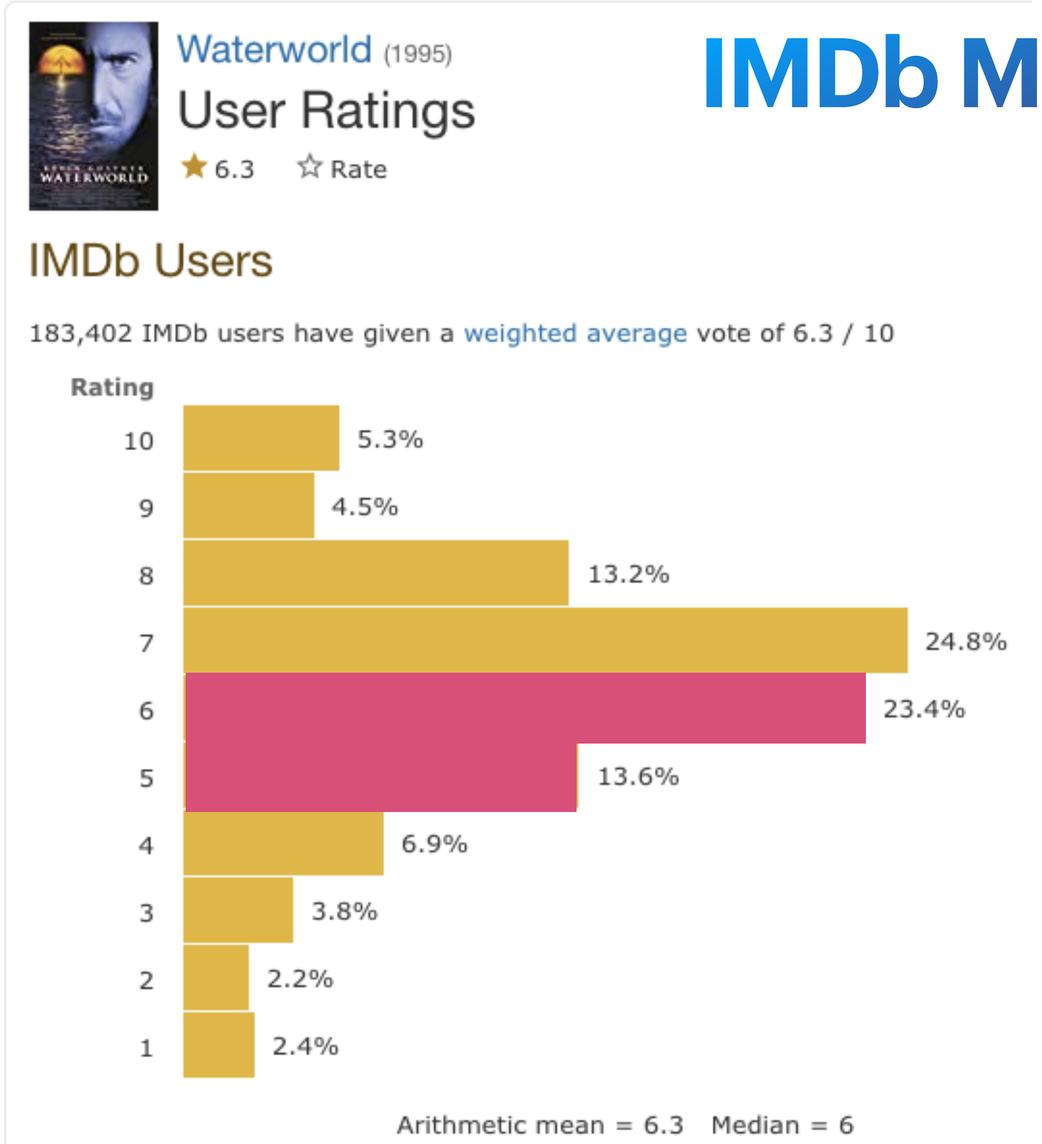
IMDb

ENVIRONMENT

python

MODELING

IMDb Movie Reviews



50,000 polar movie reviews

Binary sentiment labels

- positive
- negative

Source: Andrew L. Maas, et al. (2011).
Learning Word Vectors for Sentiment Analysis.

Results

89 percent accuracy after 20 minutes training on 1 GPU

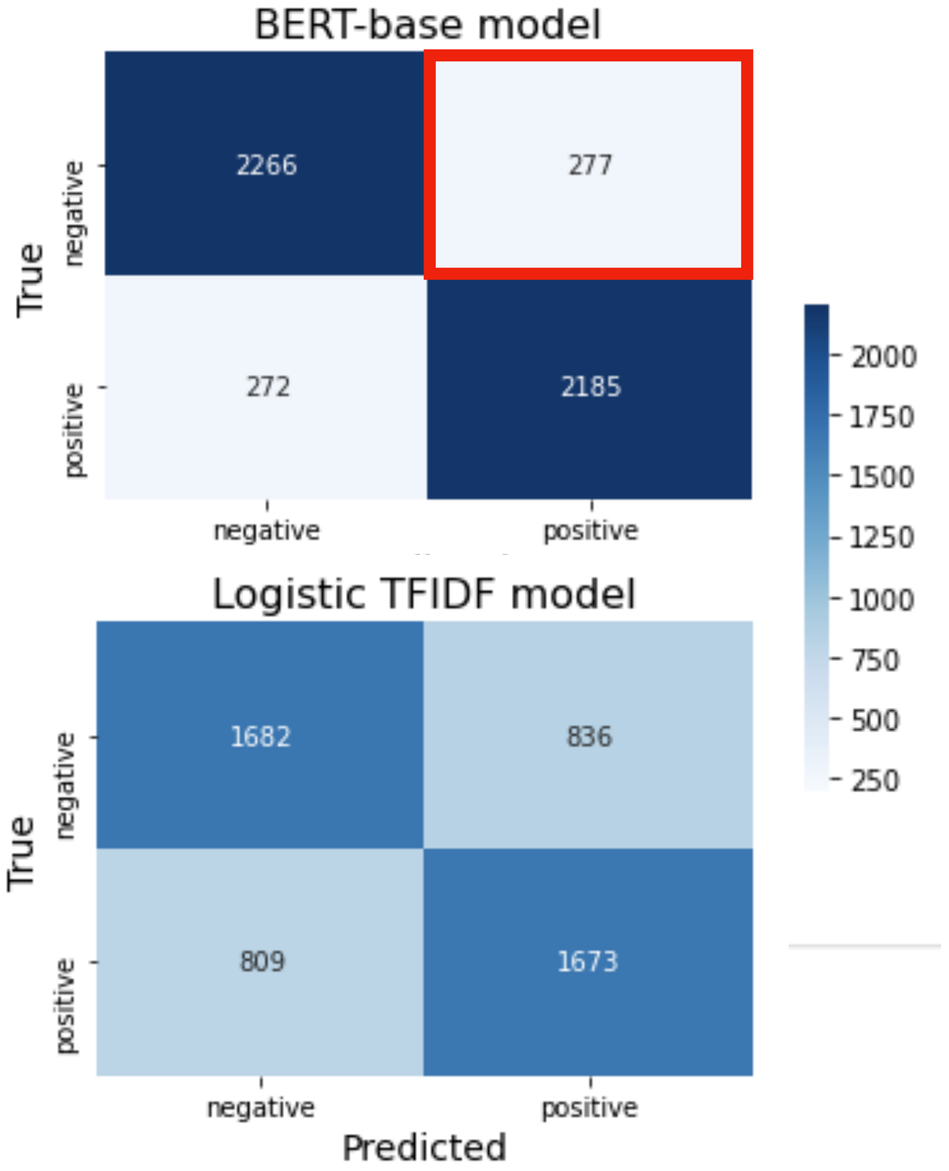
89%

Further gains are CTI

Ten percent of reviews truncated by 512 token input limit

- Transformer-XL
- BigBird

67%



NEGATIVE predicted as positive

model attention

Ambiguous - While the new film is gorgeous to view and the soundtrack is lovely, it is not Jane Austen's *Pride & Prejudice*...

human attention

human attention

Trolling - I only watched 15 minutes of this film...the concept alone is so original. It is so brilliant so original so GREAT. ENJOY! 3 out of 10

model attention

machines don’t pick up on sarcasm

Steven Newton

github.com/konstanzer/bert
steven.j36@protonmail.com

Any sufficiently advanced technology is indistinguishable from magic.

-Clarke's Third Law