Sentiment Analysis via Google's BERT

A Galvanize Capstone Project
Chase Middleman
github.com/cmiddleman/BERT_sentiment_analysis

Two Tasks

- Positive / negative sentiment classification for Amazon product reviews.

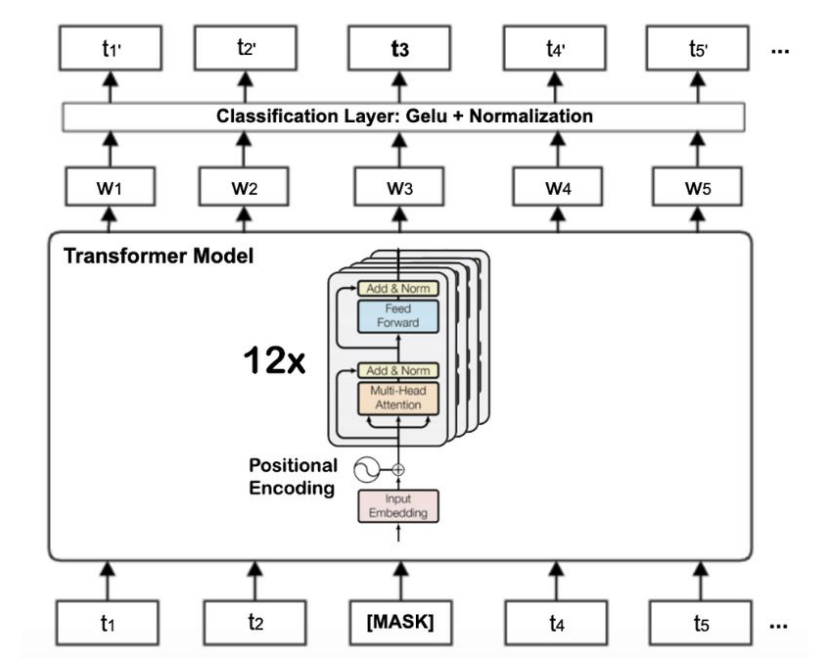
Two Tasks

- Positive / negative sentiment classification for Amazon product reviews.

- Chat analysis of the popular livestream platform twitch.

What is BERT?

- B(idirectional) E(ncoder) R(epresentations) from T(ransformers)
- Masked language model
  - Learns by filling in the ____
- Trained on a *large* corpus of words
  - 2.5 billion words from wikipedia
  - 800 million words from various literary works
- Original BERT paper



Amazon review sentiment

Amazon review sentiment

The data

Amazon review sentiment

The data

- Four million Amazon.com product reviews from kaggle.com
- Labeled positive or negative
- Balanced

Amazon review sentiment

Training

Amazon review sentiment

Training

- Slow
  - Amount of data
  - BERT layers are unfrozen (100+ million trainable parameters)

Amazon review sentiment

Training

- Slow
  - Amount of data
  - BERT layers are unfrozen (100+ million trainable parameters)
- Training time per epoch
  - CPU - more than two days
  - GPU - 20 hours

Amazon review sentiment

Training

- Slow
  - Amount of data
  - BERT layers are unfrozen (100+ million trainable parameters)
- Training time per epoch
  - CPU - more than two days
  - GPU - 20 hours
  - TPU - 2.5 hours
    - Gave some issues with pickling trained model for production
    - Work arounds were found

Amazon review sentiment

Results

Amazon review sentiment

Results

- Baseline BERT - 90% accuracy
  - HuggingFace’s out of box transformer

Amazon review sentiment

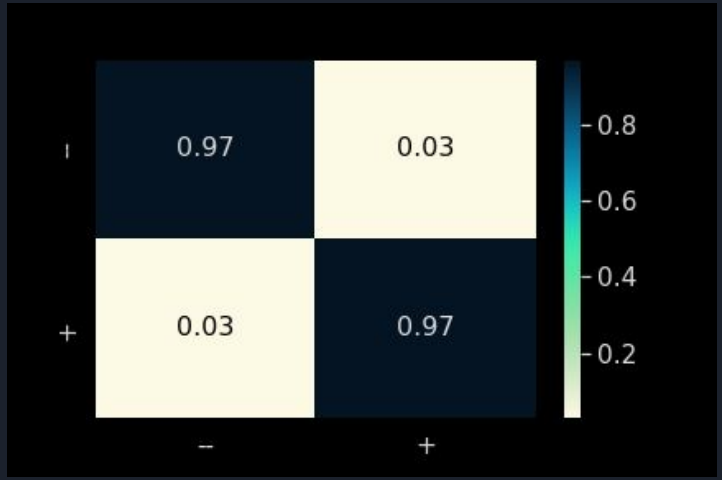
Results

- Baseline BERT - 90% accuracy
  - HuggingFace’s out of box transformer
- Our task specific BERT - 97% accuracy
  - Three epochs

Amazon review sentiment

Results

- Baseline BERT - 90% accuracy
  - HuggingFace's out of box transformer
- Our task specific BERT - 97% accuracy
  - Three epochs
- Symmetric confusion matrix



Amazon review sentiment

Results

- Baseline BERT - 90% accuracy
  - HuggingFace's out of box transformer
- Our task specific BERT - 97% accuracy
  - Three epochs
- Symmetric confusion matrix
- Some misclassifications

review	label
KIND 126: EK GAAN VIR DIE KIND BOLIVE MAAK DAN MOT HY NIE HUIS TO GAAN NEE EK WILL DIE KIND SPEEL DAN	+
cute toy: I had one whenever I was a kid, and I loved it. I recently ordered this doll for my sister.	--
Returned: I apparently ordered an incorrect size as the filter did not fit my car thus I returned it!	+
Great movie: I liked the movie a lot. I saw it several years ago and it's one that I would see again.	--
Just Fein: Not only were the blades cheeper than the store bought blades, they seemed to last longer.	+

Twitch Chat Emote Analysis

Twitch Chat Emote Analysis

Twitch Chat Emote Analysis

FeelsStrongMan I BELIEVE FeelsStrongMa
BELIEVE FeelsStrongMan I BELIEVE

zlulomegalulz: just fall

Michw12: hmm

Zenyeeta_: 🙏

lelon_musk: POGGERS

Randooooooooooom: PepeLaugh

serpantshreader: PauseChamp

SpeedOMike:

bvcs_:

glomaro: PepeLaugh

CodeFraud: nymnCorn

XAfterHoursO: GachiPls 2.38

Twitch Chat Emote Analysis

- xQc

Twitch Chat Emote Analysis

- xQc
- > 60,000 viewers on average

Twitch Chat Emote Analysis

- Chat
  - Text
  - Emotes

Twitch Chat Emote Analysis

- Chat
  - Text
  - Emotes
- Prior work tends to separate -> classify sentiment -> combine

Twitch Chat Emote Analysis

Fundamental Assumption and Methodology

The collective sentiment of twitch chat is directly captured by emotes, and that on average the sentiment of chat text can be captured by the emotes they are surrounded by.

Twitch Chat Emote Analysis

Fundamental Assumption and Methodology

The collective sentiment of twitch chat is directly captured by emotes, and that on average the sentiment of chat text can be captured by the emotes they are surrounded by.

- Turns task into supervised learning problem
- Label comments by surrounding emotes
    - Weighted by difference in time stamps

Twitch Chat Emote Analysis

Results and Applications

Twitch Chat Emote Analysis

Results and Applications

- Similar training paradigm
- One million chat comments from xQc’s streams

Twitch Chat Emote Analysis

Results and Applications

- Similar training paradigm
- One million chat comments from xQc's streams
- Compared average emote sentiment of 500 comments (~40 sec.) from watching a coronavirus vaccine debate video to average of random sample.

	WutFace	Kappa	BatChest	LUL	xqcCursed	DansGame	xqcL	xqcHead	:)	xqcT
vaccine debate	0.133	0.142	0.121	0.203	0.084	0.053	0.086	0.038	0.050	0.089
random sample	0.126	0.140	0.113	0.210	0.071	0.083	0.078	0.052	0.053	0.073

Twitch Chat Emote Analysis

Results and Applications

- Similar training paradigm
- One million chat comments from xQc's streams
- Compared average emote sentiment of 500 comments (~40 sec.) from watching a coronavirus vaccine debate video to average of random sample.

	WutFace	Kappa	BatChest	LUL	xqcCursed	DansGame	xqcL	xqcHead	:)	xqcT
vaccine debate	0.133	0.142	0.121	0.203	0.084	0.053	0.086	0.038	0.050	0.089
random sample	0.126	0.140	0.113	0.210	0.071	0.083	0.078	0.052	0.053	0.073

It seems feasible for this to be applied to streamers reading sponsored ads to gauge how their chat is reacting when compared to typical content. Definitely more hidden information waiting to be uncovered!

Demo Time!

- Special thanks to:
  - My instructors Skylar and Alex
  - My classmates and the rest of the Galvanize staff team
  - My family
  - My roommate Tyler
  - The homies from R2N1