

TWEET SENTIMENT PHRASE EXTRACTION

by Aditya Rustagi

Aim & Methodology

- Pretrained BERT from Hugging Face pretrained on combination of MLM and NSP tasks.
- Fine tuning on question answering task
 - Question —> the sentiment given in the dataset(positive/negative/neutral)
 - Context —> the text/tweet given in the dataset
 - Answer —> the phrase causing sentiment
- Extract location of answer, train using cross entropy loss
- Performance Metric: As a final check of performance we use jaccard score to see how close the predicted extracted sentiment is to the given selected text

DATA

Unique Text ID	Tweet	Extracted Tweet	Sentiment
cb774db0d1	Sooo SAD I will miss you here in San Diego!!!	Sooo SAD	negative
6e0c6d75b1	2am feedings for the baby are fun when he is all smiles and coos	fun	positive
e050245fbd	Both of you	Both of you	neutral

27481 data points

Training : 80% (22000)

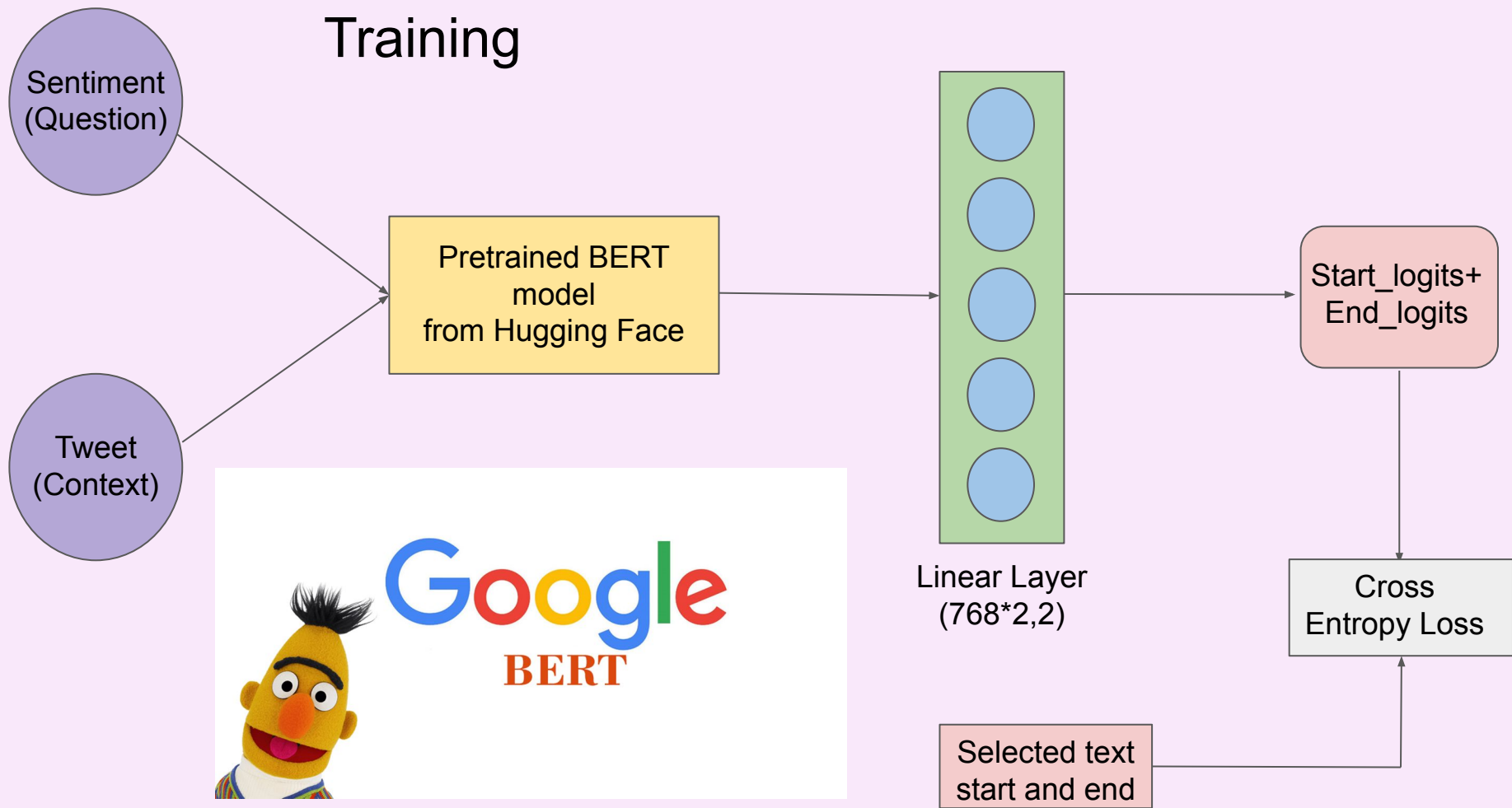
Validation 20% (6481)

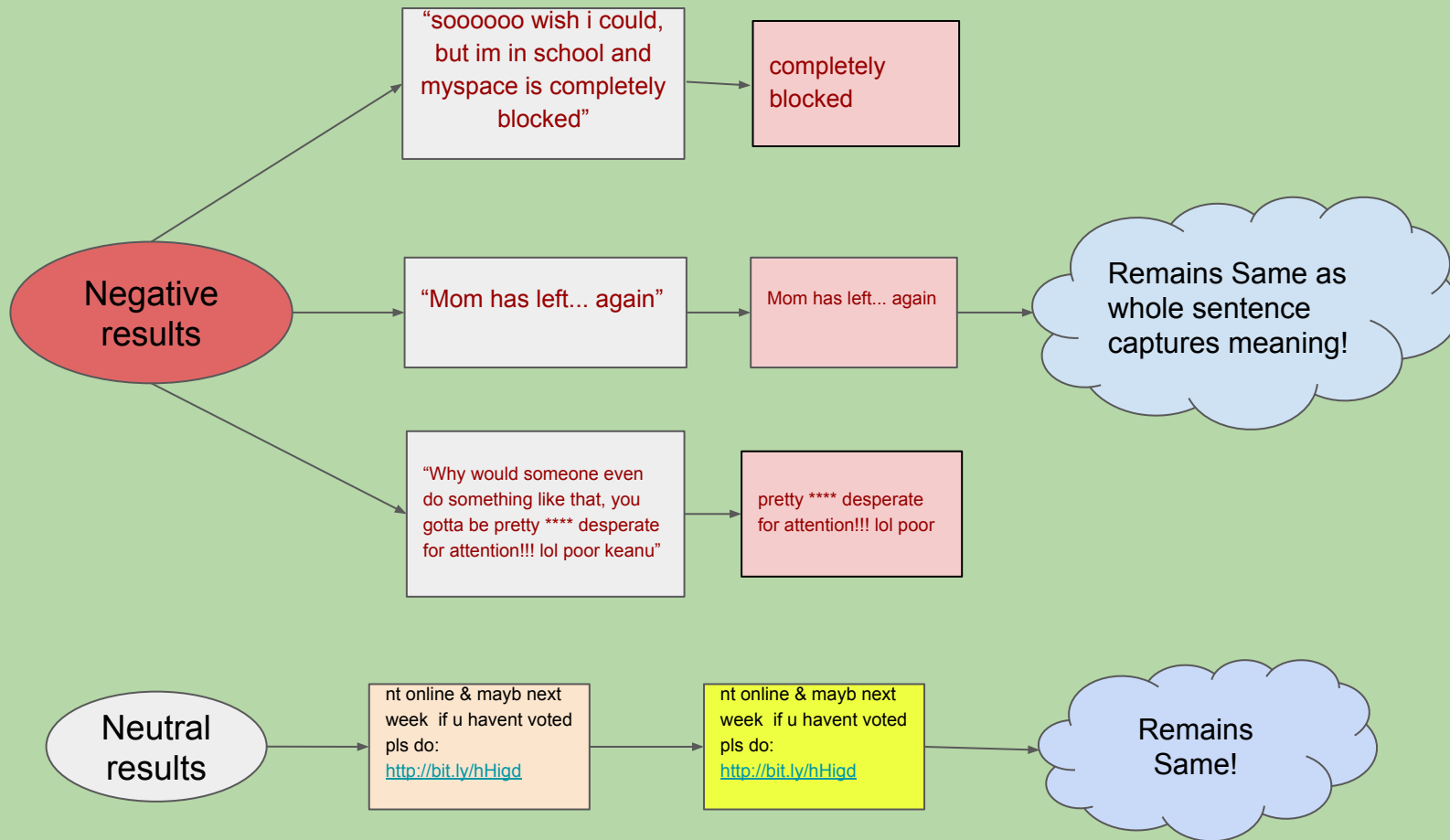
Testing : 3534

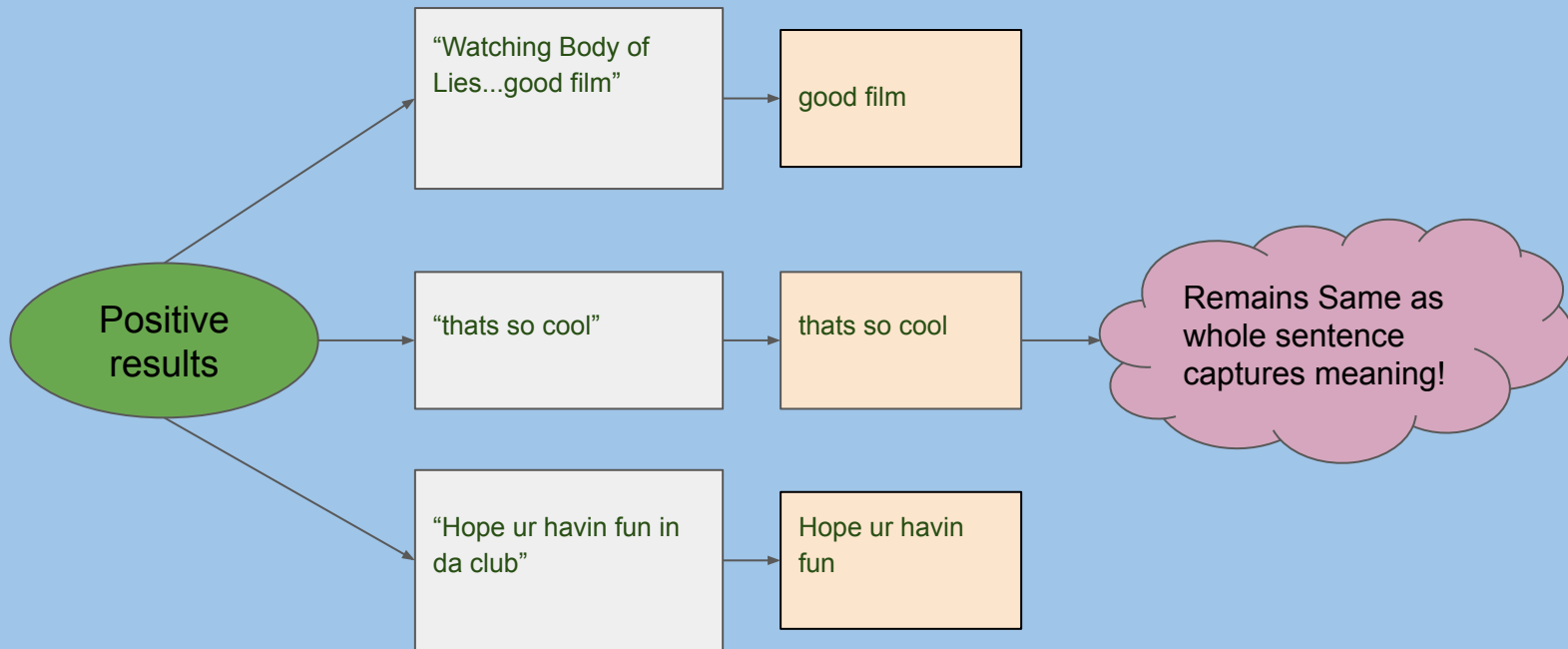
Processing

- 1) **Get index of the desired text (start+end)**
- 2) **Create character level 0-1 encoding (because tokens might break our word into smaller words eg. youtube -> 'you', 'tube')**
- 3) **Extract ID's and offsets (remove ['CLS'] and ['SEP'])**
- 4) **Find the indexes that we are interested in (those that contain at least 1)**
- 5) **Prepare for BERT input (QA model)**
- 6) **Padding**

Training







textID	text	sentiment	selected_text
187dea47db	Last session of the day http://twitpic.com/67ezh	neutral	Last session of the day http://twitpic.com/67ezh
96d74cb729	Shanghai is also really exciting (precisely -- skyscrapers galore). Good tweeps in China: (SH) (BJ).	positive	Good
eee518ae67	Recession hit Veronique Branquinho, she has to quit her company, such a shame!	negative	such a shame!
01082688c6	happy bday!	positive	happy bday!
33987a8ee5	http://twitpic.com/4w75p - I like it!!	positive	I like it!!
726e501993	that's great!! weee!! visitors!	positive	that's great!!
261932614e	I THINK EVERYONE HATES ME ON HERE lol	negative	HATES
afa11da83f	soooooo wish i could, but im in school and myspace is completely blocked	negative	completely blocked
e64208b4ef	and within a short time of the last clue all of them	neutral	and within a short time of the last clue all of them
37bcad24ca	What did you get? My day is alright.. haven't done anything yet. leaving soon to my stepsister though!	neutral	What did you get? My day is alright.. haven't done anything yet. leaving soon to my stepsister though!
24c92644a4	My bike was put on hold...should have known that.... argh total bumner	negative	argh total bumner
43b390b336	I checked. We didn't win	neutral	I checked. We didn't win
69d6b5d93e	.. and you're on twitter! Did the tavern bore you that much?	neutral	.. and you're on twitter! Did the tavern bore you that much?
5c1e0b61a1	I'm in VA for the weekend, my youngest son turns 2 tomorrow.....it makes me kinda sad, he is getting so big, check out my twipics	negative	sad,
504e45d9d9	Its coming out the socket I feel like my phones hole is not a virgin. That's how loose it is... :{	negative	not a virgin.
ae93ad52a0	So hot today =_= don't like it and i hate my new timetable, having such a bad week	negative	i hate
9fce30159a	Miss you	negative	Miss you
00d5195223	Cramps . . .	negative	Cramps . . .
33f19050cf	you guys didn't say hi or answer my questions yesterday but nice songs.	positive	nice songs.
17718b3c23	I'm going into a spiritual stagnentation, its exploding my ego!. I now realise, i'm not all that great. and i'm ok with that.	neutral	i'm going into a spiritual stagnentation, its exploding my ego!. I now realise, i'm not all that great. and i'm ok with that.
9ef44428d0	Stupid storm. No river for us tonight	negative	Stupid storm.
be634ebeb0	My dead grandpa pays more attention to me than you do	negative	My dead grandpa pays more attention to me than you do
3dcf4f7e13	... need retail therapy, bad. AHHH.....gimme money geebus	negative	bad.
10ef04109b	about to go to sleep	neutral	about to go to sleep
8be365118e	you are lame go make me breakfast!!	negative	you are lame
c50bdd4567	thats so cool	positive	thats so cool
334954f215	hey peoples, dont you just hate being grounded haha, im just sat eating an apple and watching death note (some anime)	neutral	hey peoples, dont you just hate being grounded haha, im just sat eating an apple and watching death note (some anime)
b783916431	Huh, another ScarePoint coding Sunday	neutral	Huh, another ScarePoint coding Sunday
1fa8e6ad66	look who I found just for you ---> http://twitter.com/DJT2009	positive	look who I found just for you ---> http://twitter.com/DJT2009
be38b29042	No AC, the fan doesnt swing our way ... we are sweating it out on a hot humid day	negative	sweating it out on a hot humid day
95701e8ed9	guess that depends on if you want to be on the jury	neutral	guess that depends on if you want to be on the jury
4dd58ecd2c	resorted to eating Mickey Ds ALONE.	neutral	resorted to eating Mickey Ds ALONE.
cf553cccd5	There is a faux gothy chick looking at me, sorry I am not going to camden and I like pop-punk and jimmy eat world	neutral	There is a faux gothy chick looking at me, sorry I am not going to camden and I like pop-punk and jimmy eat world
55c69c2d52	did he ask for your Twitter ID? Your sun sign?	neutral	did he ask for your Twitter ID? Your sun sign?
9b1a52cc02	where'd you go!	neutral	where'd you go!

loss_avg:	0.6849843155009201	Jaccard_avg:	0.851290744028521
loss_avg:	0.6846546315037094	Jaccard_avg:	0.8513249623594633
loss_avg:	0.6844640764112244	Jaccard_avg:	0.8513216751365743
loss_avg:	0.6833294685207196	Jaccard_avg:	0.851426941793303
loss_avg:	0.6839765336896692	Jaccard_avg:	0.8513422557777608
loss_avg:	0.6836463865611249	Jaccard_avg:	0.8513767079237267
loss_avg:	0.6830667539461125	Jaccard_avg:	0.8515529154354519
loss_avg:	0.6825007332109772	Jaccard_avg:	0.8517066001501855
loss_avg:	0.6824516021153506	Jaccard_avg:	0.8517386055116567
loss_avg:	0.6827482451791288	Jaccard_avg:	0.8517521583946952
loss_avg:	0.6824875242180295	Jaccard_avg:	0.851832054893286
loss_avg:	0.6822703174877445	Jaccard_avg:	0.8518688084616745
loss_avg:	0.6815787037502635	Jaccard_avg:	0.8519534396576545