Funniness in Edited News Headlines

Group:
Rohith Nomula - IMT2016072
Bukka Nikhil Sai - IMT2016091
Arunaav Reddy - IMT2016108

Contents

Overview:

1. Problem Statement
2. Dataset
3. Architecture
4. Training Procedure
5. Comparisons
6. Future Scope

Problem Statement

Task-1:

Regression

Given the original and the edited headline, the participant is required to predict the mean funniness of the edited headline.

Grade	Meaning
0	Not Funny
1	Slightly Funny
2	Moderately Funny
3	Funny

Task-2:

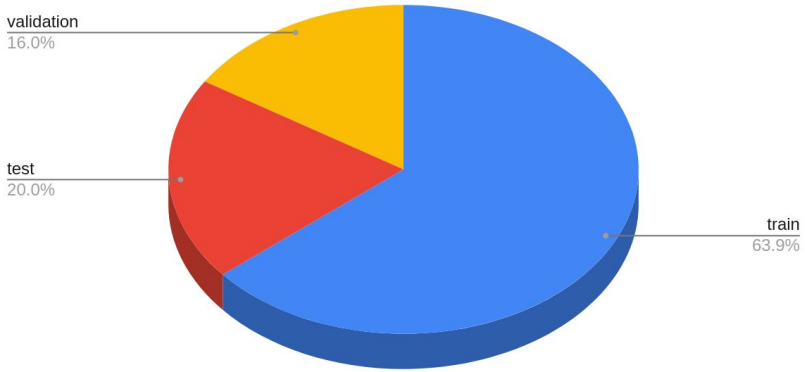
Predict the funnier of the two edited headlines:

Given the original headline and two edited versions, the participant has to predict which edited version is the funnier of the two.

Original Headline	Substitute	Grade
Kushner to visit **Mexico** following latest Trump tirades	**therapist**	2.8
Hilllary Clinton Staffers Considered Campaign Slogan `Because It's Her **Turn**'	**fault**	2.8
The Latest: BBC cuts **ties** with Myanmar TV station	**pies**	1.8
Oklahoma isn't **working**. Can anyone fix this failing American state?	**okay**	0.0
4 **soldiers** killed in Nagorno-Karabakh fighting: Officials	**rabbits**	0.0

Dataset

- The data set is called **Humicroedit** which was newly released, for research in computational humor.
- It is an openly available dataset.
- Each data point consists of a regular news headlines paired with versions of the same headline with some simple replacements which were made to make them funny.
- 5k headlines with 3 edits per headline. Each graded with 5 judges.
- The Ground truth is the average of all the 5 judges score
- Training - 9652, Testing - 3024 & Validation - 2419
- The categories of the headline are **World News and Politics.**

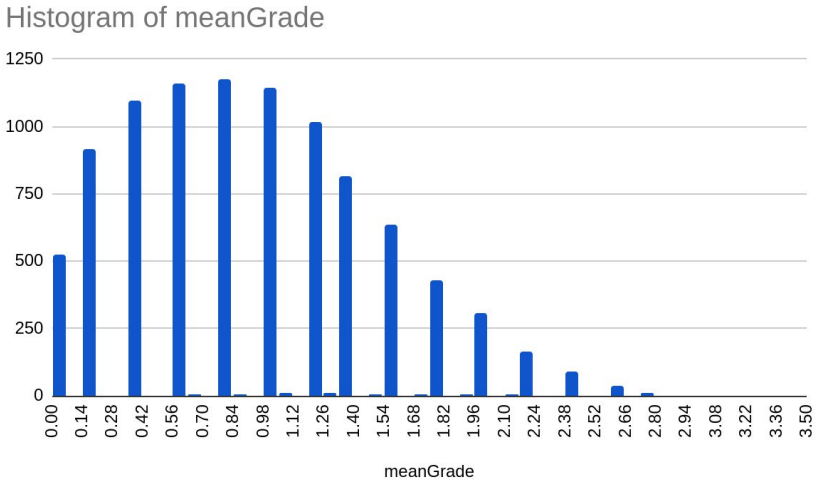


Pre-Processing

- Junk Values of mean grade other then 0-3 were removed
- The sentences are padded to a size of 80
- Each input sequence is tokenized.
- BERT has a constraint on the maximum length of a sequence after tokenizing and the maximum sequence length after tokenization is 512.
- BERT expects input data in a specific format, with special tokens to mark the beginning ([CLS]) and separation/end of sentences ([SEP]).
- We tokenized our text into tokens that correspond to BERT’s vocabulary using Bert Tokenizer.

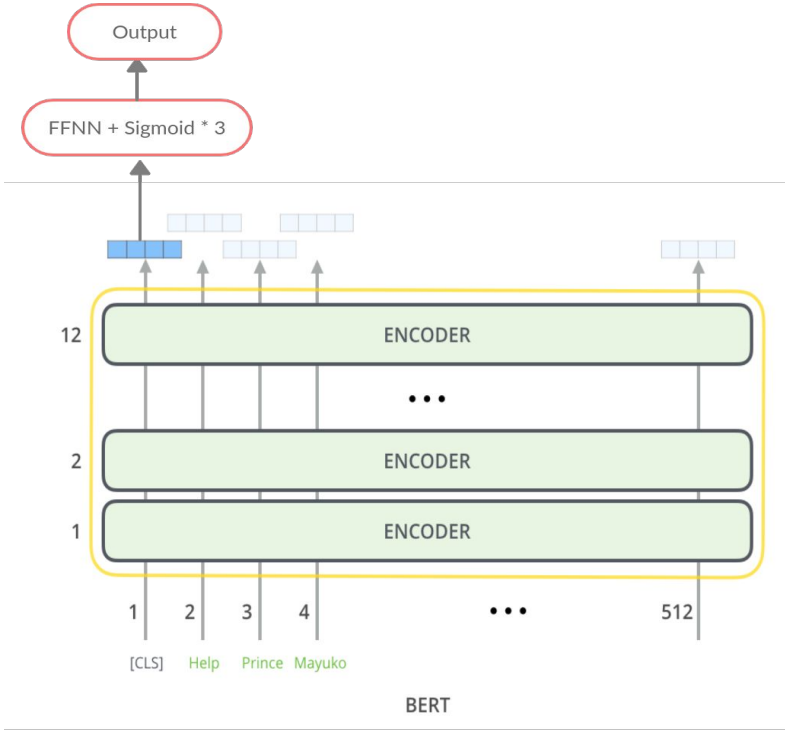
Models

- Linear Model - Linear Regression
- LSTM
  - Single Layer LSTM
- Bert
  - Fine Tuned Bert for the task



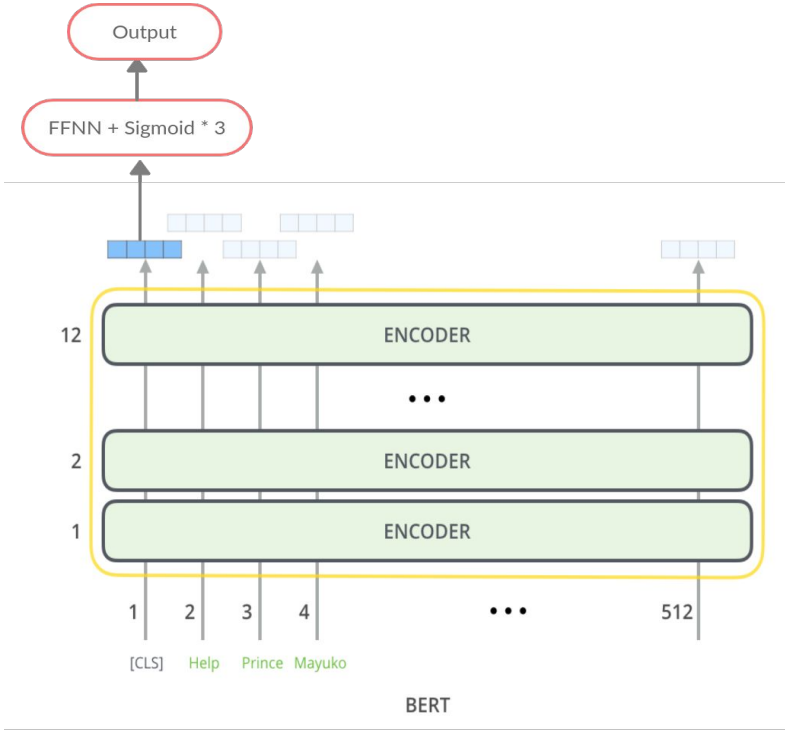
Architecture

- We used pytorch interface provided by Hugging face to fine tuned BERT base model.
- It has 12 encoding layers and by default the Maximum number of tokens you can input is 512.
- The output of [CLS] is sent to Feed Forward neural network + sigmoid layer.



Architecture

- Training Parameters
  - batch_size = 50
  - epochs = 30
  - hidden_units = 768
  - learning_rate = 0.005
  - optimizer = Adam
  - loss function = RMSE



LeaderBoard

Sub-Task 1 Results (ignore when viewing results for sub-task 2)

#	User	Entries	Date of Last Entry	Team Name	RMSE ▲	RMSE@10 ▲	RMSE@20 ▲	RMSE@30 ▲	RMSE@40 ▲
1	oyx	1	03/23/20		0.50157 (1)	0.80027 (34)	0.68195 (36)	0.60493 (36)	0.54541 (36)
2	Pramodith	20	04/26/20	lmml	0.51695 (2)	0.82450 (33)	0.69790 (35)	0.61897 (35)	0.56023 (35)
3	ZiyangLin	17	07/18/20		0.52054 (3)	0.84302 (30)	0.71518 (31)	0.63145 (34)	0.56844 (34)
4	ttzhang	1	03/12/20	ECNU7	0.52187 (4)	0.85769 (26)	0.72911 (29)	0.64118 (30)	0.57365 (33)
5	Ferryman	1	03/09/20		0.52776 (5)	0.83953 (31)	0.72031 (30)	0.63722 (31)	0.57503 (32)
6	rtdesetty	50	04/27/20	TCS Research	0.53259 (6)	0.86191 (25)	0.73269 (25)	0.64732 (27)	0.58267 (29)
7	YuWang17	7	07/18/20		0.53283 (7)	0.85332 (29)	0.73201 (26)	0.64697 (28)	0.58192 (30)
8	docekal	2	07/08/20	BUT-FIT	0.53336 (8)	0.82605 (32)	0.71130 (32)	0.63497 (32)	0.57611 (31)
9	Farah	3	03/24/20	JUST_Farah	0.53396 (9)	0.90910 (21)	0.76645 (22)	0.66720 (23)	0.59243 (24)
10	theoliao	3	04/26/20		0.53466 (10)	0.85477 (27)	0.73183 (28)	0.64650 (29)	0.58343 (28)
11	Mjason	1	03/11/20	Lunex	0.53518 (11)	0.87481 (24)	0.74673 (24)	0.65606 (24)	0.58769 (27)
12	kdehumor	13	07/07/20	KdeHumor	0.54084 (12)	0.85344 (28)	0.73185 (27)	0.64800 (26)	0.58798 (26)
13	Buhscitu	5	03/27/20		0.54558 (13)	0.90077 (23)	0.76379 (23)	0.67021 (22)	0.59988 (23)
14	cseligson	12	04/21/20		0.55019 (14)	0.90634 (22)	0.77340 (21)	0.67619 (21)	0.60613 (22)
15	Jiajun	5	03/23/20		0.55087 (15)	0.91140 (20)	0.77551 (20)	0.67825 (20)	0.60654 (21)
16	aniton	15	04/30/20	SO	0.55298 (16)	0.78974 (36)	0.70163 (33)	0.63314 (33)	0.58838 (25)
17	RasalasJau	2	07/18/20		0.56802 (17)	0.97283 (19)	0.81985 (19)	0.71247 (19)	0.63176 (19)
18	rabbitsheep	31	06/25/20		0.57225 (18)	0.97577 (17)	0.82360 (17)	0.71729 (18)	0.63639 (18)
19	sgallon	5	07/10/20		0.57454 (19)	0.98609 (11)	0.83059 (14)	0.72249 (15)	0.63986 (16)
20	VMAtm	82	06/07/20		0.57470 (20)	0.98548 (12)	0.83020 (16)	0.72239 (16)	0.63991 (15)
21	juliaive	2	06/02/20		0.57471 (21)	0.98616 (10)	0.83064 (13)	0.72266 (14)	0.64004 (14)
22	yanghaocsg	2	03/10/20	HWMT_Squad	0.57471 (21)	0.98616 (10)	0.83064 (13)	0.72266 (14)	0.64004 (14)
23	lorettaxux	6	07/17/20		0.57519 (22)	0.97458 (18)	0.82352 (18)	0.71763 (17)	0.63810 (17)
24	Snow_MingXueLiu	12	06/24/20		0.57682 (23)	0.98933 (9)	0.83208 (12)	0.72462 (13)	0.64207 (13)
25	shallowZzz	4	07/09/20		0.57975 (24)	0.98476 (13)	0.83047 (15)	0.72488 (12)	0.64380 (12)
26	jam	22	05/16/20	smash	0.58655 (25)	0.79045 (35)	0.70129 (34)	0.65027 (25)	0.61401 (20)
27	hokyeejau	16	07/09/20		0.58945 (26)	0.98461 (14)	0.83951 (10)	0.73426 (10)	0.65327 (11)
28	Murphy-MIAOYuanzhi	16	06/25/20		0.59155 (27)	1.00096 (8)	0.84610 (9)	0.73657 (9)	0.65500 (10)
29	arunaav	5	07/19/20		0.59512 (28)	0.97886 (16)	0.83249 (11)	0.73160 (11)	0.65514 (9)

Sub-Task 2 Results (ignore when viewing results for sub-task 1)

#	User	Entries	Date of Last Entry	Team Name	Accuracy ▲	Reward ▲
1	oyx	1	03/23/20		0.67237 (1)	0.30015 (1)
2	ttzhang	1	03/12/20	ECNU7	0.64384 (2)	0.25076 (2)
3	docekal	3	07/08/20	BUT-FIT	0.64041 (3)	0.25038 (3)
4	Farah	2	03/24/20	JUST_Farah	0.62633 (4)	0.21438 (5)
5	YuWang17	9	07/18/20		0.61986 (5)	0.21553 (4)
6	ChengyiZhao	8	07/16/20		0.60388 (6)	0.18143 (6)
7	Ferryman	1	03/09/20		0.60274 (7)	0.17709 (7)
8	sid.mahurkar	18	04/15/20	LRG	0.58714 (8)	0.14186 (9)
9	rtdesetty	5	04/25/20	TCS Research	0.58029 (9)	0.14323 (8)
10	RasalasJau	6	07/16/20		0.57801 (10)	0.13508 (11)
11	jam	31	05/15/20	smash	0.57496 (11)	0.13668 (10)
12	Warren	2	03/11/20	Warren	0.55365 (12)	0.09452 (12)
13	ZiyangLin	2	07/15/20		0.51788 (13)	0.01796 (17)
14	Snow_MingXueLiu	4	07/15/20		0.51104 (14)	0.02002 (14)
15	uir	1	03/13/20		0.51027 (15)	0.04132 (13)
16	Murphy-MIAOYuanzhi	23	07/16/20		0.50989 (16)	0.01986 (15)
17	Murphy-MIAO-YuanZhi	2	07/16/20		0.50951 (17)	0.01956 (16)
18	lorettaxux	2	07/16/20		0.46651 (18)	-0.05084 (18)
19	arunaav	4	07/19/20		0.45662 (19)	-0.06629 (19)

Results

#	User	Entries	Date of Last Entry	Team Name	RMSE ▲	RMSE@10 ▲	RMSE@20 ▲	RMSE@30 ▲	RMSE@40 ▲
1	HonoMi	7	03/02/20	Hitachi	0.49725 (1)	0.77914 (48)	0.66726 (48)	0.59322 (48)	0.53814 (48)
2	alonzorz	2	03/07/20	Amobee	0.50726 (2)	0.81093 (44)	0.69345 (46)	0.61210 (47)	0.55204 (47)
3	jtoma	13	03/08/20	YNU-HPCC	0.51737 (3)	0.81704 (43)	0.70441 (45)	0.62222 (45)	0.56280 (46)

Task 1 - Bert	0.59512	0.97885	0.83248	0.73160	0.65513

Results

#	User	Entries	Date of Last Entry	Team Name	Accuracy ▲	Reward ▲
1	oyx	1	03/23/20		0.67237 (1)	0.30015 (1)
2	ttzhang	1	03/12/20	ECNU7	0.64384 (2)	0.25076 (2)
3	docekal	3	07/08/20	BUT-FIT	0.64041 (3)	0.25038 (3)

Task 2 - Bert	0.45662	0.06629

Limitations

- Each epoch takes about 30 min. We were unable to run more than 18 epochs as colab was resetting the state of the GPU.
- Dataset is more skewed to the Non funny side data

Future Scope

- Task 1 can be improved more by running more epochs.
- Task 2 can be improved by building models which takes the context difference of both the edited text into consideration.

Thank you!