

NLP Codalab Competition.



Introduction

- SemEval-2020 Task 7: Assessing Humor in Edited News Headlines.
- Task- Rearession

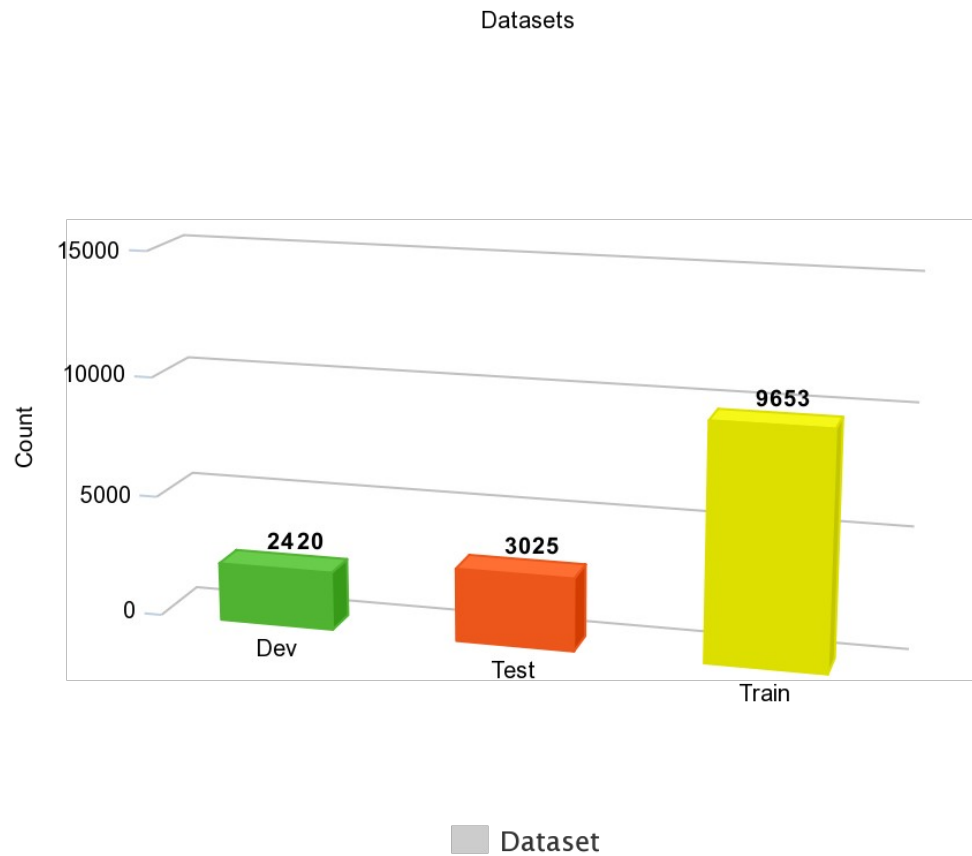
Original Headline	Substitute	Grade
Kushner to visit Mexico following latest Trump tirades	therapist	2.8
Hilllary Clinton Staffers Considered Campaign Slogan `Because It's Her Turn `	fault	2.8

- Organised by Nabil Hossain at the Unviesity of Rochest and John Krumm and Micahel Gamon at Microsoft Research AI.

Problem Definition

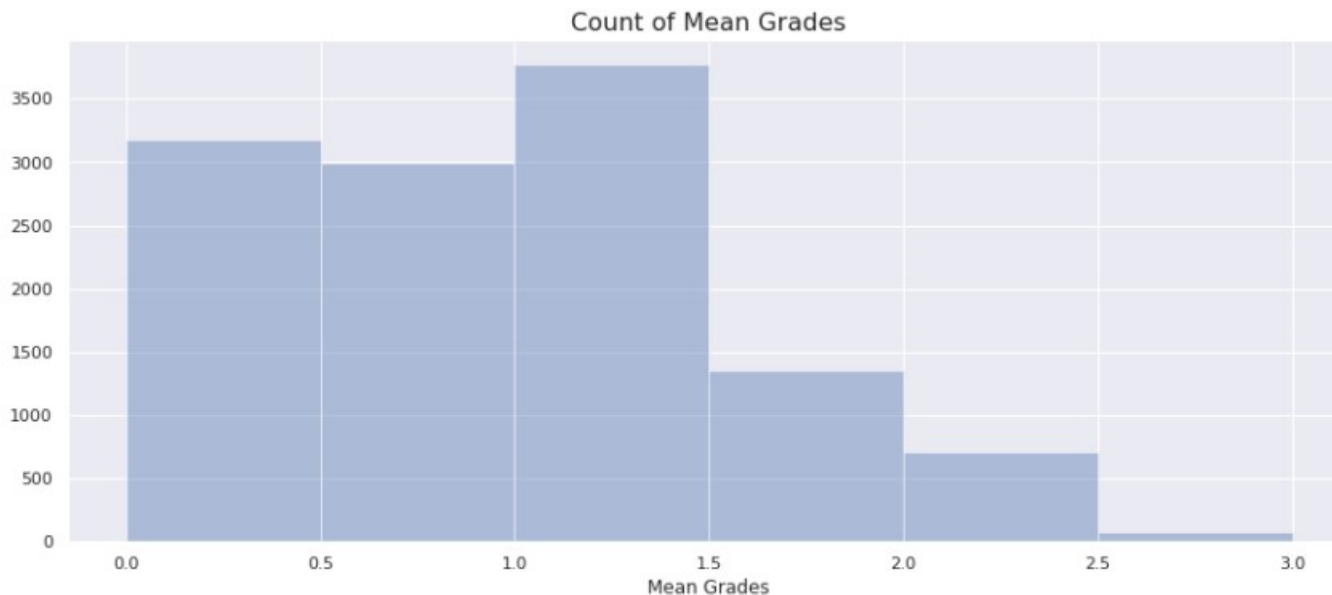
- Humor difficult task for humans to comprehend.
- News headlines will be substituted by new words.
- Humor subjective to different humans.
- Understand what factors lead to humor.
- Try to understand patterns that occur in humor.

Dataset Size



Dataset Exploration

	id	original	edit	grades	meanGrade
0	1723	Thousands of gay and bisexual <men/> convicted...	swans	22100	1.0
1	12736	Special <prosecutor/> appointed to Trump Russia	chef	21100	0.8
2	12274	Spanish police detain man and search Ripoll ad...	squad	21000	0.6
3	8823	N.Y. Times <reprimands/> reporter for sharing ...	applauds	32210	1.6



Pre-processing

- Input of the regressor - Use the finalized edited headline, by replacing the edited words.
- Use sentence pairs that combine edited headlines with the original text.
- Form - “From <original word> to <edited word>”.
- Add the “[SEP]” token to separate two sentences.

Training

- Train a BERT regressor with the preprocessed text as input and the mean grades as output.
- Load the pre-trained BERT tokenizer.
- Tokenize all the input text.
- Tokenize the text with added “[CLS]” and “[SEP]” tokens in the beginning.
- Padding to form 32-word long embeddings.

Training Contd..

- Tokenize text in the validation set in the same way.
- Torch DataLoader class used to create iterators.
- Load the pre-trained BERT model.
- Use BertForSequenceClassification class.
- Number of label=1.

Parameters

- Learning rate- $2e-5$
- adam epsilon of $1e-8$
- Epochs -2 to 4
- Seed Value to replicate- 42
- Validation score method- Root Mean Square Error

Results

```

===== Epoch 1 / 4 =====
Training...
/pytorch/torch/csrc/utils/python_arg_parser.cpp:756: UserWarning: This overload of add_ is deprecated:
  add_(Number alpha, Tensor other)
Consider using one of the following signatures instead:
  add_(Tensor other, *, Number alpha)
Batch   40 of   302. Elapsed: 0:13:54.
Batch   80 of   302. Elapsed: 0:27:28.
Batch  120 of   302. Elapsed: 0:41:04.
Batch  160 of   302. Elapsed: 0:54:39.
Batch  200 of   302. Elapsed: 1:08:03.
Batch  240 of   302. Elapsed: 1:21:25.
Batch  280 of   302. Elapsed: 1:34:57.

Average training loss: 0.34
Training epoch took: 1:42:18

Running Validation...
RMSE: 0.5434
Validation Loss: 0.30
Validation took: 0:07:45

===== Epoch 2 / 4 =====
Training...
Batch   40 of   302. Elapsed: 0:13:26.
Batch   80 of   302. Elapsed: 0:26:55.
Batch  120 of   302. Elapsed: 0:40:32.
Batch  160 of   302. Elapsed: 0:54:03.
Batch  200 of   302. Elapsed: 1:07:42.
Batch  240 of   302. Elapsed: 1:22:05.
Batch  280 of   302. Elapsed: 1:36:44.

Average training loss: 0.27
Training epoch took: 1:44:41

Running Validation...
RMSE: 0.5417
Validation Loss: 0.29
Validation took: 0:08:14
```

Conclusion

