A Neural Framework for Analyzing the Tweets of dril

Grant Storey
Cornell University
gjs224@cornell.edu

Abstract

Natural Language Processing is a major field within computer science research, but by definition it limits itself to common human language. Many interesting linguistic phenomena can be found in unnatural, absurd language, especially of the sort common in the twitterverse. In this work, we present the first scholarly study of the tweets of Twitter user dril. We gather a new dataset containing the text of dril tweets and associated absurdity rankings collected from mechanical turkers. Then, we suggest state-of-the-art neural techniques to perform the given classification task on these tweets. We report promising experimental results that suggest wider value for this specific task.

1 Introduction

Vast scholarly time and effort has been focused on solving various problems in Natural Language Processing, the sub-field of Computer Science which focuses on natural language processing. In spite of this, very little attention has been paid to linguistic corpora which are not strictly natural language but are quite similar to it: pig latin, gibberish, random yelling, 8-year-olds on Youtube, and wild Twitter accounts. Our work seeks to begin remedying this gap.

For our first task, we hope to analyze the absurdity of tweets from the Twitter user dril. No existing corpora concerning these tweets exist, so we utilize mechanical turkers to create a new dataset and curate it for optimal usability.

We then utilize modern neural network techniques for text analysis to create a model for performing the analysis task given a tweet. Using this model on our dataset gives promising results.

Our contributions are threefold:

1. We collect and curate a new dataset for analyzing the absurdity of dril Tweets.
2. We propose a state-of-the-art neural model for performing this analysis task.
3. We analyze applications for this research and directions for future work.

2 Related Work

Some prior work in Emoji language, including Barbieri et al.'s skip-gram semantic model(Barbieri et al., 2016) but also other analyses(Miller et al., 2016; Kelly and Watts, 2015) is tangentially related in analyzing slightly different forms of communication found on social media, often detached from natural human language.

Parker's work on echos of Cthulu's primal bellows seeks to provide structure to unnatural language, but is mostly focused on waveform analysis and is primarily concerned with non-human language (Parker, 1999). Our work, in contrast, seeks to study language originating from this planet.

Lastly, studies on train conductor announcements (Yarini and Fubata, 2002) and blackout frat boy heckling (Johnson, 2011) present analyses of edge cases of human language, but these works focus more on translating less coherent language into natural language rather than analyzing the alternative languages in their own right.

3 Data

To the best of our knowledge, there is no prior work on analyzing absurd tweets. In order to make this task feasible, we collect a new dataset of dril's tweets, the Dril Tweet Foundational Dataset (DTF Dataset). We gathered all 7519 of dril's tweets from the first tweet on September 15th, 2008 to

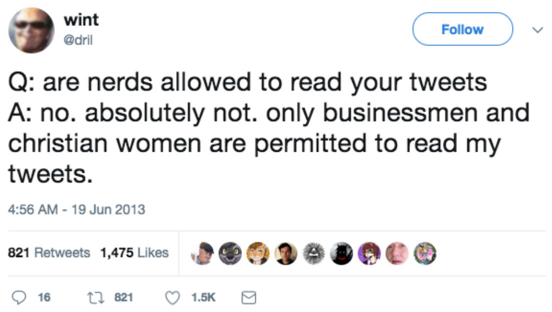

Figure 1: An example of one of dril's tweets.

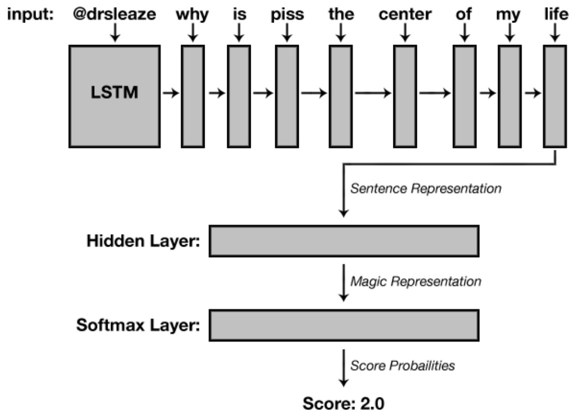

Figure 2: Illustration of the model architecture.

12:30 PM (EST) on March 4th, 2018. An example tweet is presented in Figure 1.

We then gathered mechanical turkers to perform our labeling by harassing them via Facebook Messenger. Turkers were asked to rate tweets on an absurdity scale of 1-5, with 1 being reasonably normal and 5 being completely unhinged.

For each score from 1 to 5, users were given an example tweet with that score to calibrate their analysis. They were then presented with 50 tweets and 1 check (to ensure they were paying attention). These 50 initial tweets were the same for each user. After completing the 50 tweets, users could optionally continue and rate more random tweets in batches of 50.

A total of 5 users performed the task over the course of 10 days. All users passed every attention task, and 1 user completed more than the initial 50 tweets (they rated 108 tweets total, including the initial 50).

After gathering the data, we assigned a score to each rated tweet by choosing the median score provided to it. The Cronbach's alpha for the dataset was 0.8009, where a score above .8 represents "good" quality agreement, showing that the amount of inter-annotator agreement was quite "good".

We then tokenized the text of each tweet so that punctuation was treated as its own token. Each tokenized tweet is paired with its median absurdity score across all raters.

Finally, we divided our 108 data points into Training, Validation, and Test sets of size 88/10/10 respectively. The DTF Dataset can be found at `http://www.grantstorey.com/SIGSEGV/2018/DTF.zip`.

4 Method

Our model is designed to solve the task at hand: given a sequence of tokens representing the tweet, predict an absurdity score from the set $\{1, 2, 3, 4, 5\}$. We begin with preprocessing, feed the tokens through an LSTM, and then choose the most likely label for the given tweet.

4.1 Preprocessing

We analyze the entire data set, and replace every token that appears only once with the special token <UNK> (unknown). 240 tokens appear at least twice and are mapped to a unique embedding, while the remaining 820 tokens (which appear once each) are mapped to the UNK embedding. We do not use pre-trained embeddings because many of the words used are misspellings or nonsense not present in any pre-trained embeddings.

4.2 Neural Model

See Figure 2 for an illustration of the model. Following on prior work in Natural Language Processing, we use a Recurrent Neural Network with Long Short-Term Memory (Hochreiter and Schmidhuber, 1997) for analysis of the text. We first extract the 50-dimensional embedding associated with each of the tokens in the tweet. These embeddings are fed into a two-layer LSTM to generate a representation of the sentence. This representation is passed through a hidden layer with 128 dimensions to generate 5 output dimensions, which are in turn passed through a SOFTMAX layer to generate probabilities for each of the possible labels. The label with highest probability is chosen, and the associated negative log-loss is calcu-

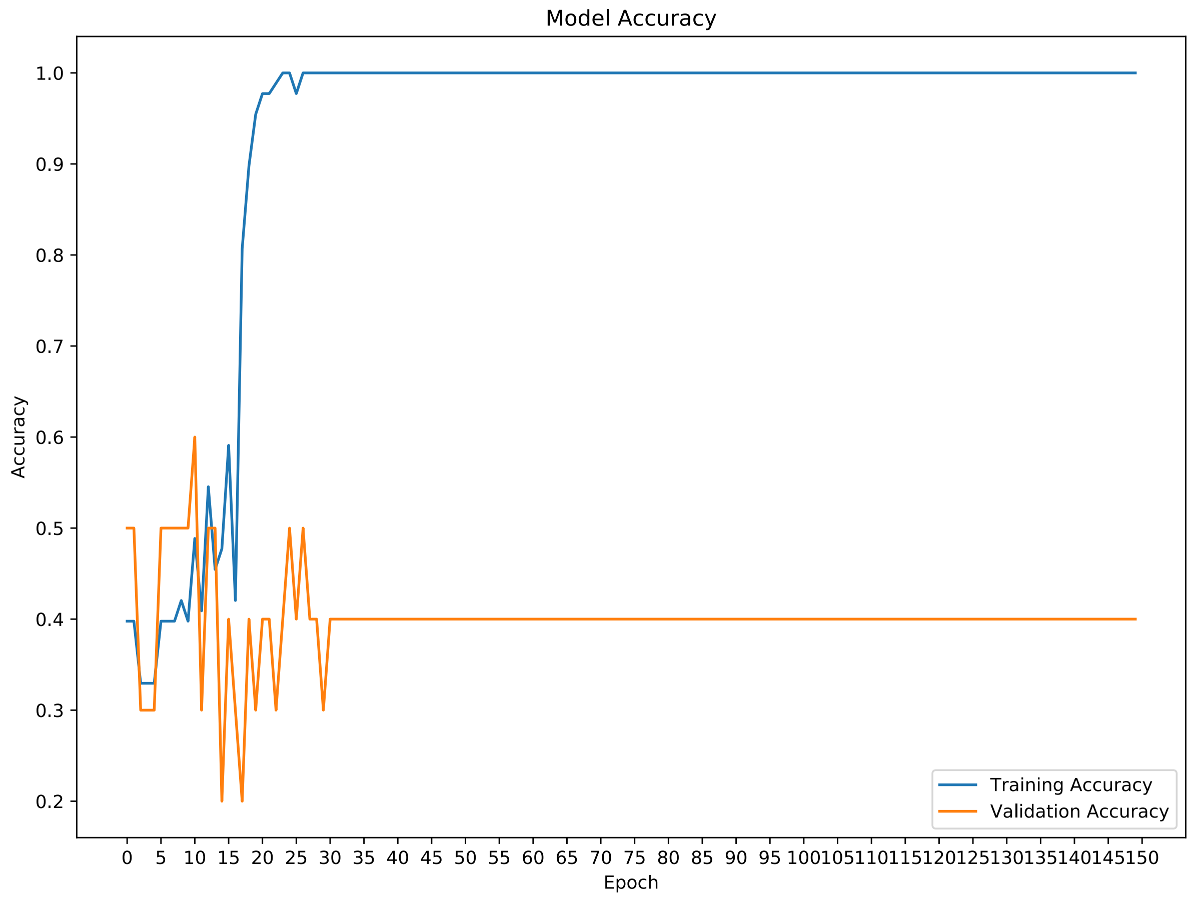

Figure 3: Training and validation accuracy by epoch.

lated and back-propagated through the network. In general, the model can be described by the following equations:

$$A = P(p_a + r_e * n * t)/l(y)$$

$$I = N_e * E_d$$

$$E_{qu} = A^t + I_{0,N} * \vec{s}$$

$$L_i[k] \pm \epsilon = T(H_i * \vec{s})$$

4.3 Parameters

We trained our model for 150 epochs, with a batch size of 1, and beam search with beam size large enough to obliterate Alderaan in one shot and crush any hope for the rebellion with our display of unstoppable military might.

4.4 Model Choice

We examined the model's performance on the validation set at each epoch, and chose the model from the last epoch with 100% training accuracy before validation accuracy began to go down - Epoch 26 in this case. See Figure 3 to see the training and validation error by epoch.

Model	Validation Acc	Test Acc
Majority-Class Baseline	**50%**	**50%**
Our Model	50%	30%

Table 1: Comparison on the Validation and Test sets of the performance of a baseline and our model.

5 Results and Discussion

The results for our model on the validation and test set is compared to the performance of the majority-class baseline (which always chooses the most common label in the training data) in Table 1.

The majority-class baseline clearly outperforms our model. Nevertheless, we consider this to be a positive result. The failure of state-of-the-art methods to solve this problem show that it is a **difficult** task, and likely unsolvable by computational methods.

In fact, this difficult is quite promising, as this task (or other similar tasks) could be used to prevent dangerous artificial intelligences from posing a threat to humanity by distracting them with unsolvable problems. Other problems in a similar class include convincing young children to sit still for an entire day, determining why kids love Cinnamon Toast Crunch, and finding meaning in an uncaring universe.

6 Conclusion

We set out to tackle a new task: determining the absurdity of dril tweets. We collected a robust dataset to allow evaluation of models on this task. Our proposed model is then applied to this task and found to have worse performance than a simple baseline. We use this result to successfully argue that this task can stand between humanity and the robot apocalypse by presenting an unsolvable task for potentially dangerous Artificial Intelligences.

Acknowledgments

We would like to thank the brave souls who contributed to our Tweet dataset for their valuable contributions. We would also like to thank the world for providing us with dril, a national treasure. This work was generously funded by grant EM-45209 from Elon Musk's Artificial Stupidity initiative, with additional funding provided through Aperture Labs' W.H.E.A.T.L.E.Y. Program.

References

Francesco Barbieri, Francesco Ronzano, and Horacio Saggion. 2016. What does this emoji mean? a vector space skip-gram model for twitter emojis. In *LREC*.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9(8):1735–1780.

Chad Johnson. 2011. *Becoming Bro: English from Fraternity to Main Street*, volume 1. Harvard University Press, Cambridge, MA.

Ryan Kelly and Leon Watts. 2015. Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships. *Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances, and Design* .

Hannah Miller, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen, and Brent Hecht. 2016. Blissfully happy or ready to fight: Varying interpretations of emoji. *Proceedings of ICWSM* 2016.

Smyth Parker. 1999. Interpreting sounds of darkness with complex analysis. *Studies in Necromancy, Voidwalking, Shadowcasting, and Other Forbidden Arts* .

Sheila Yarini and Alice Fubata. 2002. Hrmnrm hmr next stop frmhrm: Translating new york subway announcements into english. In *Trains! Trains! Trains!*.