Project Proposal

V. Kovalchuk, K. Salyer

November 2019

1   Goal

The goal of our project is to create a neural network which, depending on what dataset it is fed, will construct and send its own tweet from the account @ML-funsies1. We will design the neural network such that the training dataset can be easily changed: either an already constructed dataset such as this collection of 2015 New Year's Resolutions or construct our own using the Twitter API, Tweepy, to select and filter the tweets we want to feed into the neural network.

2   Process

2.1   Training Data

We will first try to create our own dataset by reading the most popular N tweets, filtering based off of a hashtag. This has more potential applications because learning what is being said about certain natural disasters can help in allocating resources [1]. We can specify the filtering hashtag and even location data in the code.

Another option available to us is to use any of a large number of free twitter datasets. This would require some labeling of data, but it would be interesting to try to train the neural network to reply to tweets. Using these, we can train our neural network to tweet a resolution or to tweet like Elon Musk.

2.2   Neural Network

We will have to use the Word2Vec approach in our neural network. Typically this is a group of shallow models that are trained to reconstruct linguistic contexts. We look forward to experimenting with numbers of layers, amount of data input, among other elements of the architecture. The most difficult part for us, however, will be interpreting the output of the neural network and having it craft a tweet. Since Word2Vec maps words to vectors, we can gauge how the training data affects the relationships that are learned by the neural network, and compare how those relationships vary with time or region. For example, a questions that we could explore is how the conversation about climate change has evolved over the years.

2.3 Tweeting

Once we have a trained model and its output tweet, we can have it automatically send this tweet on the account we created for this project: @MLfunsies1. In order to do this, we will have to use the Twitter API, Tweepy, which has pre-constructed methods that will make execution of this step fairly simple.

3 Conclusions

This project will evolve as we get started and learn about what types of questions are actually interesting and feasible to explore with our approach. For example, one exercise we discussed is comparing the output where the training data is all the same topic but from different years. It will also be interesting to think of other applications of this approach, given that we were inspired by the fact that using ML on Twitter data can solve real world problems like the allocation of resources in a natural disaster. Not only can we use it for the entertainment value of seeing what the neural network tweets, but also the neural network itself can be a foundation for solving big problems: something we look forward to exploring.

---

[1] D. Reynard and M. Shirgaokarb, "Harnessing the power of machine learning: Can Twitter data be useful in guiding resource allocation decisions during a natural disaster?" (2019), `doi:10.1016/j.trd.2019.03.002`.