ML Joke Bot

CS 487 Project

Spring 2020

C. Baker | M. Dickson | L. Mendoza

Motivations:

For thousands of years, we solved problems directly through conversation, chatbots are a throwback to this simpler time. Everyone loves laughing. When people laugh, it doesn't just lighten their load mentally, it actually induces physical changes in their body such as reducing stress, easing tension and stimulates many organs. But humor is generally thought of as a uniquely human trait, so we thought it would be interesting to see if we could use a machine learning algorithm to create unique jokes and to gauge whether people actually find them funny. Specialists see humor as the final frontier for artificial intelligence, because it requires mastery of sophisticated functions like self-awareness, empathy, spontaneity, and linguistic subtlety. If we can create a good joke bot, then we can make progress in making computers and robots pioneering their own comedic stylings by learning humor in humans. Machines are here to kill all right, but kill us of laughter.

Definition of Problem:

Training an ML algorithm to author its own jokes.

- Our algorithm should recognize jokes.
- Output a joke in text format when the program is run.
- Gather information regarding whether people find it amusing.

Solution:

As we recognize the needs for our machine learning algorithm, we plan to scope the work into the following development process:

1. Prepare Data : The creation of an ontology using a larger dataset of jokes from kaggle.
2. Pre-Processing : Add grammar into the machine learning algorithm so that our joke bot can understand jokes from the dataset.
3. Creating a **deep learning** model to train our bot.
4. Track our progress of our model accuracy. (If the jokes are funny or not)
5. Add it to an application/GUI
6. Test our deep learning joke bot, if the jokes aren't funny, we need to add more datasets to our bot.
7. Deploy bot.

Data:

Our bot will initially be run on a dataset called "shortjokes" which we downloaded from https://www.kaggle.com/abhinavmoudgil95/short-jokes. One issue with this dataset is that many of the jokes contain racial slurs and inappropriate content for a school project so we will need to preprocess this dataset to remove any jokes which contain offensive or inappropriate content. The dataset currently contains 65,000 jokes which will be enough to train our bot even after certain jokes are removed.

Any additional datasets will be added to the report at a later date.