

Minimum Viable Product for NLP Project - Cards Against Humanity

(caution: some contents are explicit and inappropriate for children under 18 years of age)

Belle Peng | 08/2018

Domain Description and Motivation

"Unlike most of the party games you've played before, Cards Against Humanity is as despicable and awkward as you and your friends." - YouTube. The game is simple. Each round, one player asks a question from a Black Card, and everyone else answers with their funniest White Card from their hand of 10 cards. Cards Against Humanity (CAH) lit up numerous parties at my house. When brainstorming for an idea for a Natural Language Processing project, it got me thinking - what makes CAH so irresistibly funny? What if I can create a joke-generator to generate an Answer card (white card) to a randomly chosen Question card (black card)? In addition to the joke-generator, I also want to do topic modeling in finding structure or clusters in the White and Black cards amongst themselves. Potentially find funny Answer cards for each Question card within a given humor type.

Data

I can download a .json file of the Base set and all 7 of the Expansion Sets from this [site] [<https://crhallberg.com/cah/>]. For each set I have access to all of its Answer cards (white cards) and Question cards (black cards). I have a total of 1155 Answer cards and 295 Question cards.

Some example Questions are as follows:

- "Why can't I sleep at night?"
- "I got 99 problems but _ ain't one."
- "What's a girl's best friend?"
- "What's that smell?"
- 'This is the way the world ends / This is the way the world ends / Not with a bang but with _.'
- "What is Batman's guilty pleasure?"
- 'TSA guidelines now prohibit _ on airplanes.'
- 'What ended my last relationship?'
- "MTV's new reality show features eight washed-up celebrities living with _."
- 'I drink to forget _.'

Some example Answers are:

- 'Coat hanger abortions.'
- 'Man meat.'
- 'Autocannibalism.'

- 'Vigorous jazz hands.'
- 'Flightless birds.'
- 'Pictures of boobs.'
- 'Doing the right thing.'
- 'The violation of our most basic human rights.'
- 'Viagra®.'
- 'Self-loathing.'

Known Unknowns

The biggest limitation of this project is that the answer cards can be rather short and also my limited knowledge in NLP will mean I get to experiment with the NLP techniques. So far I am able to parse through the data and use word vectorizer to vectorize it. I am working on implementing Word2Vec. I am also not entirely clear how to use RNN and Topic Modeling in this project yet, but some unsupervised technique for clustering the white and black cards are in my pipeline.