

# Project 4: Humor evaluation: judging funniess level of the dialogue in The Big Bang Theory

Hongfei Li(hl963), Shibo Zang(sz428)

---

## 1. Problem or Task

Commedy TV shows are playing a very special part of people's life nowadays. The key to the success of commedy Tv shows is the level of funnies of the themselves. Our task is to use natural language processing technics to evaluate comedy script's funnies level. We may evaluate the funnies level on a scale of 10.

One example from a popular show Big Bang Theory is the following

*Raj: I don't like bugs, okay. They freak me out.*

*Sheldon: Interesting. You're afraid of insects and women. Ladybugs must render you catatonic.*

Clearly the second sentence is the laugh point. Our system will look into the context of the laugh point and may give the laugh point a funniess level of 8/10.

The reason we choose to investigate in this is because we love watching Big Bang Theory and it is the TV series that is making the most money. We want to look into something that can be potentially useful and practical.

## 2. General Approach or System Architecture

The main task of our project is to use NLP technics to evaluate comedy script's funnies level. The general idea we propose to solve this task is in some sense similar to Professor Cristian Danescu-Niculescu-Mizil's work on judging politeness of corpus. We would like to break our task into several subtasks:

1.. Humor Strategy Classifcation:

Following examples illustrate the type of linguistic elements that underlie humor.

---

Example	Strategy
"What do you use to talk to an elephant? An elly-phone."	Phonological Similarity
"MIT stands for Mythical Ilogy."	Acronym multiple sense
"Infants don't enjoy infancy like adults do adultery."	Pun
"Of all the things I lost, I miss my mind the most."	Human centric vocabulary
"Money can't buy your friends, but you do get a better class of enemy"	Negative Orientation
"It was so cold last winter that I saw a lawyer with his hands in his own pockets."	Professional Communities

## 2.. Predicting Funniess Level

We plan to apply machine learning algorithms to this task. Mainly there are three categories that needs to be considered: not funny, neutral, and funny. And with classification algorithm, we could first train the training dataset produced by turkers' annotation using bag of word (BOW) classifier or linguistically (Ling.) informed classifier. The BOW classifier is based on Support Vector Machine using a unigram feature representation. And the linguistically informed classifier is an SVM using the linguistic feature.

The reason we decide to divide training dataset

## 3. Data and Data Annotation

We will get all transcripts of every epsiode of a commedy TV show such as Big Bang Theory, which can be found at <https://bigbangtrans.wordpress.com>. Mark all laugh point with a label as well as a funny score associated with a label <lp,9> using amazon's mechanical turk (9 means this laugh point has a funnies of 9 out 10). Since The score is subjective score, we may have 5 or more people rate the laugh point and take the average of them. Each turk is going to read through the transcript of one episode and rate the laughing points.

For example, for the following transcript which is the first couple of lines in season 1 episode 2 of Big Bang Theory. The label are added after the turk's work.

*Leonard: There you go, Pad Thai, no peanuts.*

*Howard: But does it have peanut oil?*

*Leonard: Uh, I'm not sure, everyone keep an eye on Howard in case he starts to swell up. <lp,8>*

*Sheldon: Since it's not bee season, you can have my epinephrine. <lp, 7>*

*Raj: Are there any chopsticks?*

*Sheldon: You don't need chopsticks, this is Thai food.*

*Leonard: Here we go.*

*Sheldon: Thailand has had the fork since the latter half of the nineteenth century. Interestingly they don't actually put the fork in their mouth, they use it to put the food on a spoon which then goes into their mouth. <lp,3>*

#### **4. Methods and System Development**

#### **5. Implementation**

#### **6. Evaluation**

#### **7. Anything else**

#### **8. Individual Member Contribution**