2019 Winter NLP

Project 1 – Twitter Analysis

Overview

- Three main types of analysis
 - Award Recognition
 - Host Identification
 - Award Association Tasks
 - Nominees, Winner, Presenters

Award Recognition

Multiple award names analyzed to determine common Part of Speech (POS) Structure*



Tweets containing the word "best" are identified



Using POS tagging and chunking, possible substrings are identified



Strings starting with "best" are taken, and the final subset is determined by implementing a frequency cutoff

Although performance is sporadic, this method is generalized to many different award shows – which many other methods may not have

^{*} This is shown in awards_methodology_explain.ipybn, and is done in a way that is very generalizable.

Host Identification

Tweets containing the word "host" are identified



People KB

Names are identified using all capital Bigrams and Trigrams



Most common name is extracted as definite host, and cutoff for cohost frequency is used to determine possibility for two hosts

Performs well on 2013 and 2015 and accounts for a single host and cohosts

Award Association Tasks - Nominees, Winner, Presenters

