

# Methods for building twitter-specific sentiment lexicon

## Using mutual information

[This page](#) has quite a bit of information regarding how the mutual information should be computed. I don't think there is any difference between multi-class mutual information and binary mutual information except that the entropy of the class labels is affected.

## Using Counts

I could use a simple method based upon counts of occurrences of words both in the negative and in the positive tweets with normalization for frequency of terms

## Using Feature Selection Methods

## Mutual Information

The issue with using mutual information for this is that, ideally, we would want a two tailed statistical test, while mutual information is a one tailed test. To fix this, there are two different strategies that I am going to try.

# Winner take all Mutual-Information

The winner take all mutual-information is the mutual information calculated for the positive class and the negative class, taking the larger of these two (negative mutual information will be multiplied by -1 to create a two tailed distribution of scores).

## Proportional Mutual-Information

In proportional mutual information, the mutual information for the negative class is subtracted from the mutual information for the positive class.

---

Revision #3

Created Mon, Sep 24, 2018 6:17 PM by [kenneth](#)

Updated Thu, Feb 6, 2020 4:53 PM by [kenneth](#)