**Mutual Information (MI)**

MI measures the mutual dependence of two random variables.

The higher it is, the more dependent the two random variables are with each other.

Its value is always positive.

The equation of MI:



**Pointwise Mutual Information**



If x and y are independent, MI is 0.

Let w1 = amazing, f(w1) = 2825

w2 = movie, f(w2) = 2657

w3= bad, f(w3) =

f(w1,w2) = 50

N = 67063111

**Steps:**

1. From the labelled reviews, Find the positive and negative terms based on the frequency of occurrence.

Positive\_terms = 'good', 'nice', ‘best’, 'great', 'awesome', 'outstanding', 'fantastic', 'terrific', ':)', ':-)', 'like', 'love'

Negative\_terms =  'bad', 'terrible', 'crap', 'useless', 'hate', ':(', ':-('

2. Find the PMI of the test review terms with all the positive and negative terms.

Eg: He is my best friend.

2.1 After preprocessing, we have the unigrams: ‘best’, ‘friend’

2.2 Since ‘best’ comes under positive term, SO(best) = positive

2.3 Find the PMI of ‘friend’ with all the positive and negative terms. Consider 3 positive and 3 negative terms.

N = 104

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Good | Nice | Best | Bad | Terrible | Crap | friend |
| f(w1) | 20 | 15 | 16 | 12 | 10 | 13 | 12 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| f(w1,w2) | Good | Nice | Best | Bad | Terrible | Crap |
| friend | 5 | 3 | 2 | 1 | 0 | 0 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **w1** | **w2** | **f(w1)** | **f(w2)** | **f(w1,w2)** | **N** | **p(w1,w2)** | **p(w1)** | **P(w2)** | **Nr/Dr** | **PMI** | **SO** |  |
| good | friend | 20 | 12 | 5 | 104 | 0.048077 | 0.192308 | 0.115385 | 2.166667 | 1.115477 |  |  |
| nice | friend | 15 | 12 | 3 | 104 | 0.028846 | 0.144231 | 0.115385 | 1.733333 | 0.793549 |  |  |
| best | friend | 16 | 12 | 2 | 104 | 0.019231 | 0.153846 | 0.115385 | 1.083333 | 0.115477 | 2.024504 | ∑\_(r ∈ V^+) |
| bad | friend | 12 | 12 | 1 | 104 | 0.009615 | 0.115385 | 0.115385 | 0.722222 | -0.46949 | -0.46949 | ∑\_(r ∈ V^-) |
| terrible | friend | 10 | 12 | 0 | 104 | 0 | 0.096154 | 0.115385 | 0 | 0 |  |  |
| crap | friend | 13 | 12 | 0 | 104 | 0 | 0.125 | 0.115385 | 0 | 0 |  |  |
|  |  |  |  |  |  |  |  |  |  |  | **2.493989** |  |

3. Using V+ and a vocabulary of positive terms and V- for the negative ones, the Semantic Orientation of a term t is hence defined as:

Since the SO(friend) is 2.493989, the review is positive.