02/22/2015 Mengfan Cheng

Is News A Downer?

1. Introduction

There has been assumption that hard news is usually about bad news. However, no data has been collected to test whether the hypothesis is true. Now we will analyze tweets collected from the twitter accounts of 10 media that have been considered to produce "hard news" to test the assumption. **The null hypothesis in our analysis is that the percentage of negative news is not significantly greater than that of other news.**

These twitter accounts include New York Times (@nytimes), Financial Times (@FinancialTimes), NPR (@nprnews), CNN (@CNN), New Yorker (@NewYorker), AFP (@AFP), Wall Street Journal (@WSJ), Associated Press (@AP), Guardian (@guardian), Reuters (@Reuters). 6000 most recent tweets are collected in total with 600 tweets from each account. The analysis process is as follows. First, each tweet is split into single words. Then each word is tested and determined to be positive, negative or neutral (sentiment testing method is adopted from Dr. Pablo Barbera's social media workshop). Finally we get the total percentage of positive words, negative words and neutral words from each account. The output data is shown in Table 1 and further in graphs in Appendix. The mean percentage is 24% for positive words, 26% for negative words and 50% for neutral words.

Twitter Account	Positive%	Negative%	Neutral%
New York Times	23%	27%	50%
Financial Times	17%	20%	63%
NPR	40%	6%	54%
CNN	27%	30%	43%
New Yorker	28%	26%	46%
AFP	17%	32%	51%
Wall Street Journal	27%	20%	53%
Associated Press	19%	35%	46%
Guardian	24%	30%	46%
Reuters	22%	34%	44%
Mean	**24%**	**26%**	**50%**

Table 1

2. Methodology

Judging from the means we get from the output, it seems that there is no significant difference between the percentage of positive words and negative words, and that the percentage of neutral words is much greater than both that of positive words

and negative words. Yet to test the null hypothesis, we need to run t-test for two hypotheses.

I. *Test whether the percentage of negative words is equal to that of positive words.*[1]

H_0: There is no statistically significant difference between the percentage of negative words and the percentage of positive words.
H_1: There is statistically significant difference between the percentage of negative words and the percentage of positive words.

The calculated t-statistic is -0.3502 and the p-value is 0.7342, which means that we fail to reject the null hypothesis.

II. *Test whether the percentage of negative words is smaller than that of positive words.*[2]

H_0: There is no statistically significant difference between the percentage of negative words and the percentage of positive words.
H_1: The percentage of negative words is not significantly smaller than that of neutral words.

The calculated t-statistic is -5.5826 and the p-value is 0.0002, which means that we are able to reject the null hypothesis at the 0.01 alpha level.

3. Conclusion

Judging from the data we collected from these tweets, the percentage of negative words is not greater than the percentage of positive words. And the percentage of negative words is significantly less than the percentage of neutral words. We might be able to reject the hypothesis that hard news is about bad news; yet there are some problems need to be addressed before we draw the conclusion.

Firstly, only 10 observations are made in this test, which is too small a sample size to draw any justified conclusion. Secondly, these tweets are the most recent 600 tweets from each account, which means that they only cover a relatively short period of time. Thirdly, there is still doubt about whether it is legit to test the hypothesis using only data from tweets instead of data from print or online articles and TV news. Fourthly, the result might be different if a different coding of sentiment is adopted or if different twitter accounts are selected.

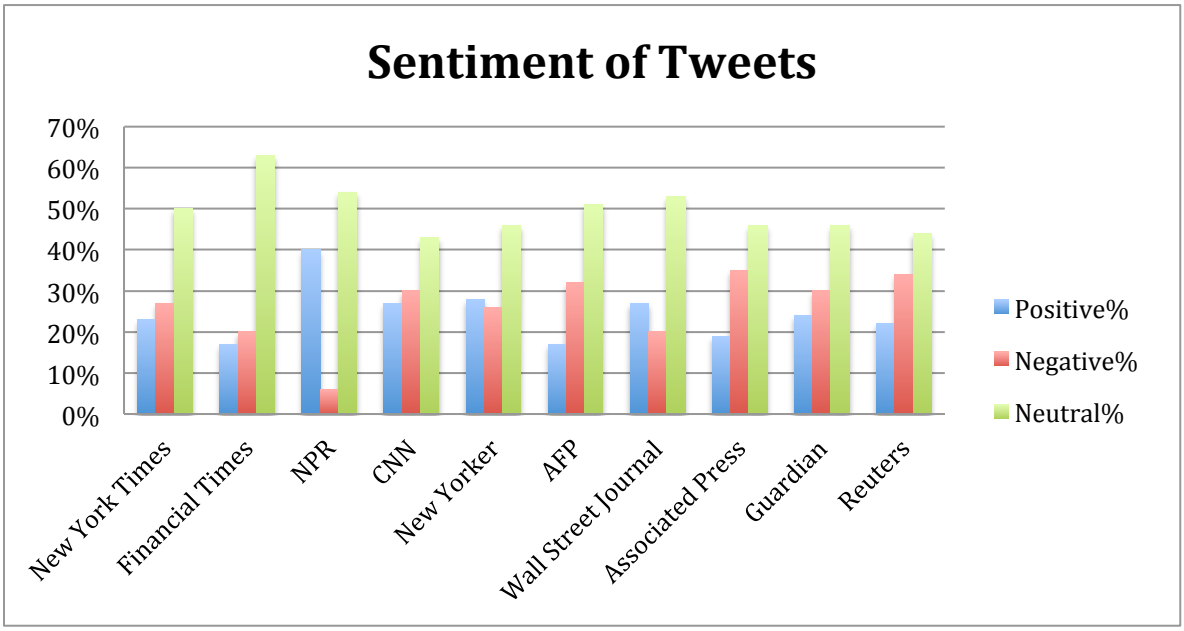
---

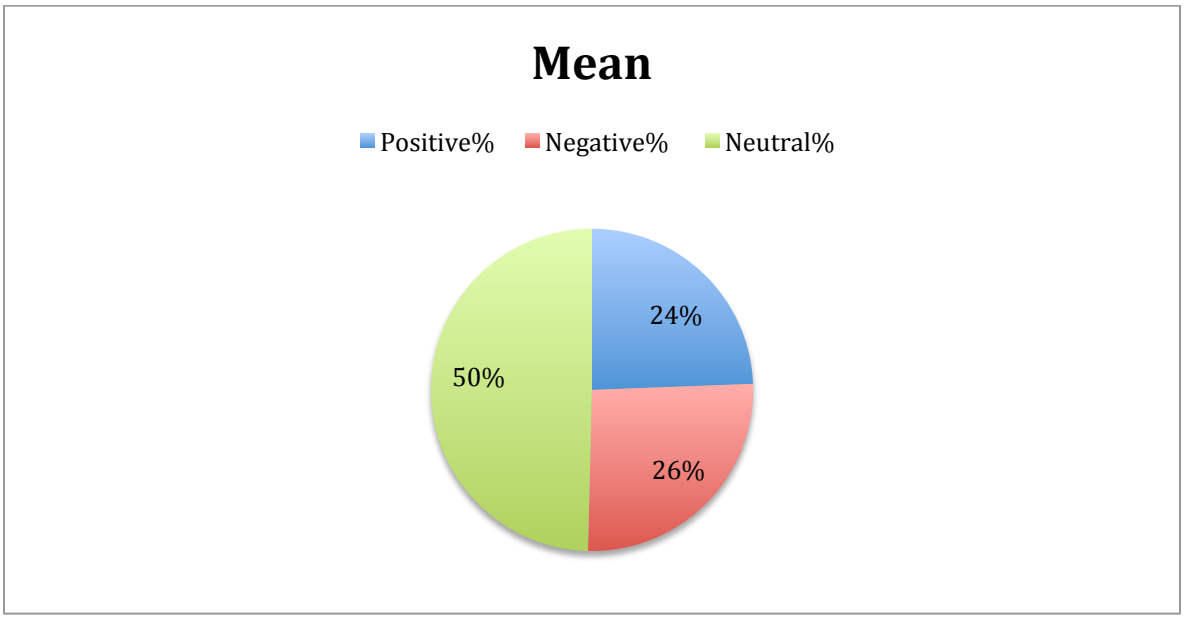
[1] A two-tail test is adopted because we do not have a strong prediction of the direction of difference.

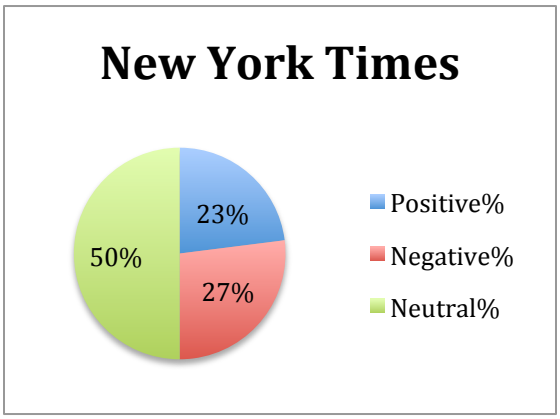
[2] A one-tail test is adopted because we have a strong prediction of the direction of difference from the output.

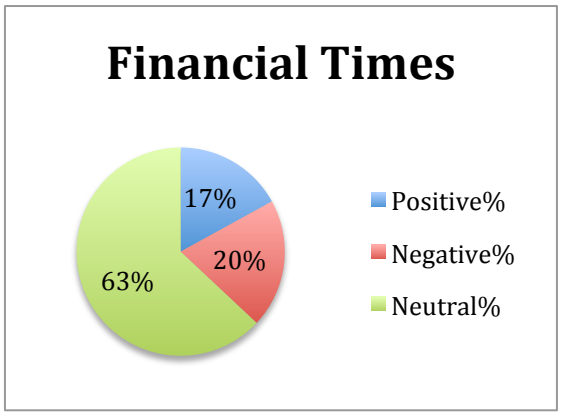
02/22/2015 Mengfan Cheng

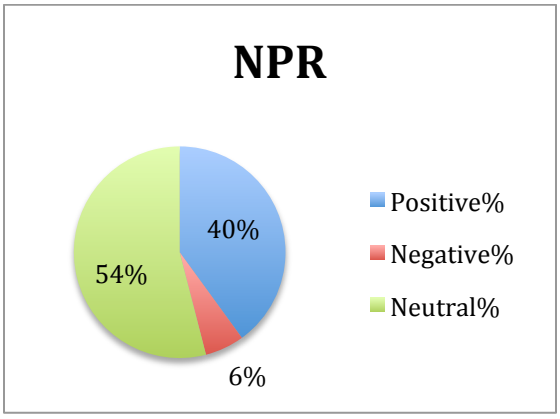
APPENDIX

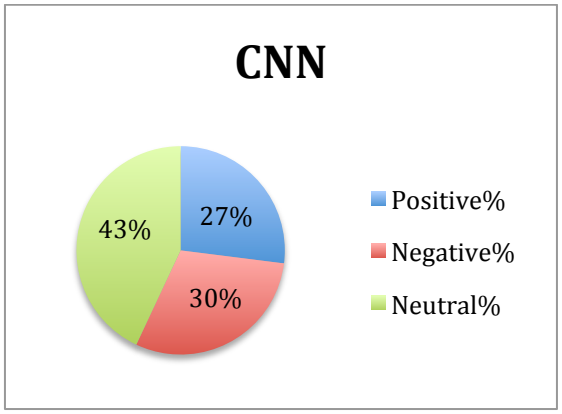


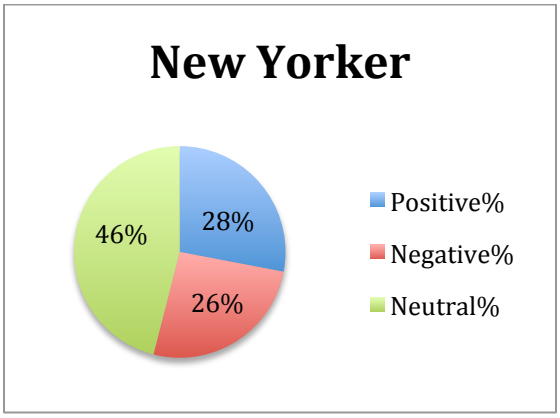


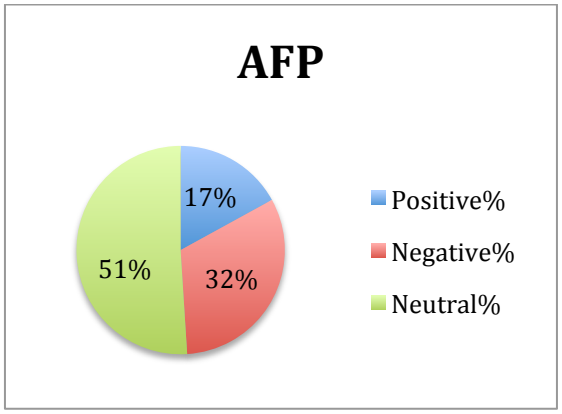

New York Times
23%
27%
50%
Positive%
Negative%
Neutral%


Financial Times
17%
20%
63%
Positive%
Negative%
Neutral%


NPR
40%
6%
54%
Positive%
Negative%
Neutral%


CNN
27%
30%
43%
Positive%
Negative%
Neutral%


New Yorker
28%
26%
46%
Positive%
Negative%
Neutral%


AFP
17%
32%
51%
Positive%
Negative%
Neutral%

02/22/2015 Mengfan Cheng