Smartphone Sentiment Analysis

Rhys Hewer

Contents:

Overview

Methodology

Findings
- Sentiment Categories
- Sentiment Comparison

Confidence
- Modelling Metrics
  - iPhone
  - Samsung Galaxy
- Analyst Insights
- Errors and Improvements
  - Data Sparseness
  - Sentiment Levels
  - Outliers

Implications

Overview

A project by Alert Analytics, a data analytics consulting organisation, on behalf of Helio, a smart phone and tablet app developer.

Helio is working with a government health agency to create a suite of smartphone medical apps for use by aid workers in developing countries. The government agency will be providing workers with technical support services, but need to limit the support to a single model of smartphone and operating system. To narrow this list down to one device, Helio has engaged Alert Analytics to conduct a broad-based web sentiment analysis to gain insight into the attitudes toward the devices.

Multiple devices were examined and two models chosen for in-depth analysis: iPhone & Samsung Galaxy.

Methodology

Common Crawl is an open repository of web crawl data (over 5 billion pages so far) that is stored on Amazon's Public Data Sets. This acted as our data source.

The general approach to collect the sentiment data was to count the words, associated with sentiment towards the smartphone in question, per web-page. The words were classified as having positive, unclear or negative sentiment based on predefined word-lists and were counted if they were in proximity of a mention of the smartphone.

As returning the results of 5 billion web-pages, many of whom would have no mention of the smartphone, would often be irrelevant, results were retained only if they had at least one mention of a phone or phone OS term, and at least one of the following terms were present within the web-page: review; critique; looks at; in depth; analysis; evaluate; evaluation; assess.

A small sample of the common crawl archive was targeted to create a sample of data, called the 'small matrix' which was then manually given a sentiment score based on a human review and interpretation of the webpage. This acted as the dataset to test and train the data.

The full common crawl archive was then mined for sentiment data to produce a large matrix to which the model developed on the small matrix was applied. It is the results of this that are discussed below.

Findings

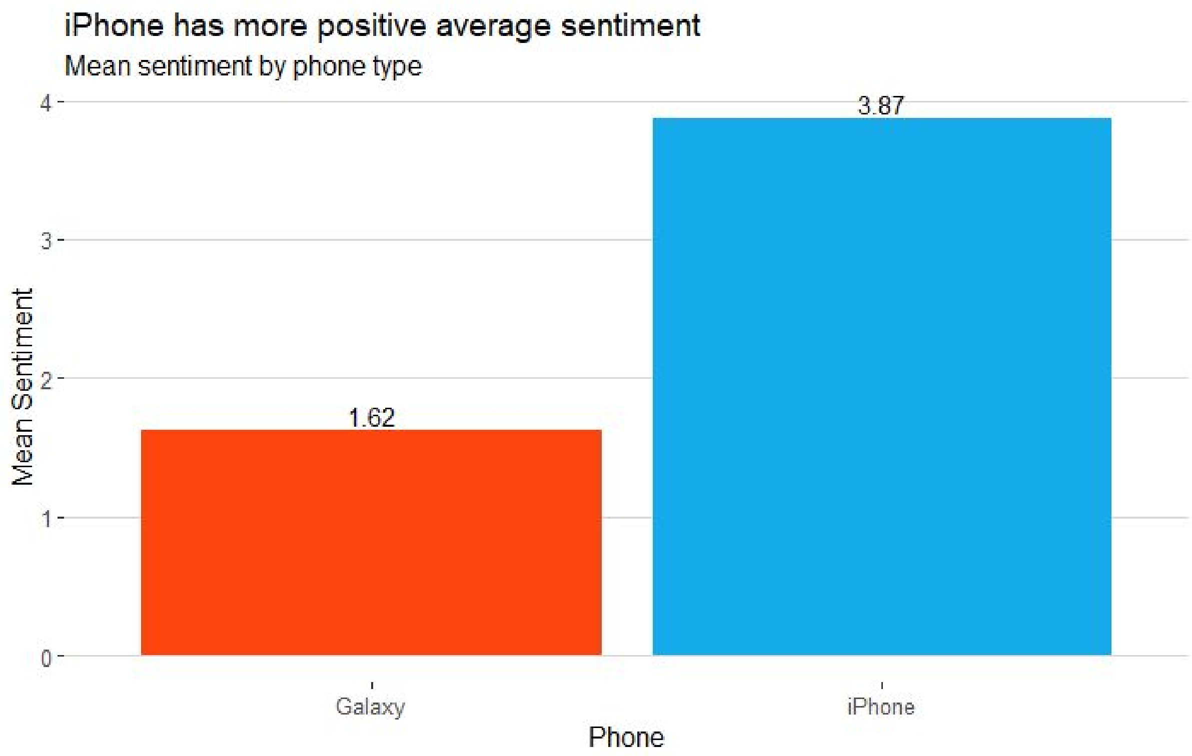
Sentiment Categories

Sentiment was originally recorded on a 6 point scale but was re-coded to a 4 point scale to allow for more focussed modelling. The 4 point scale used is:

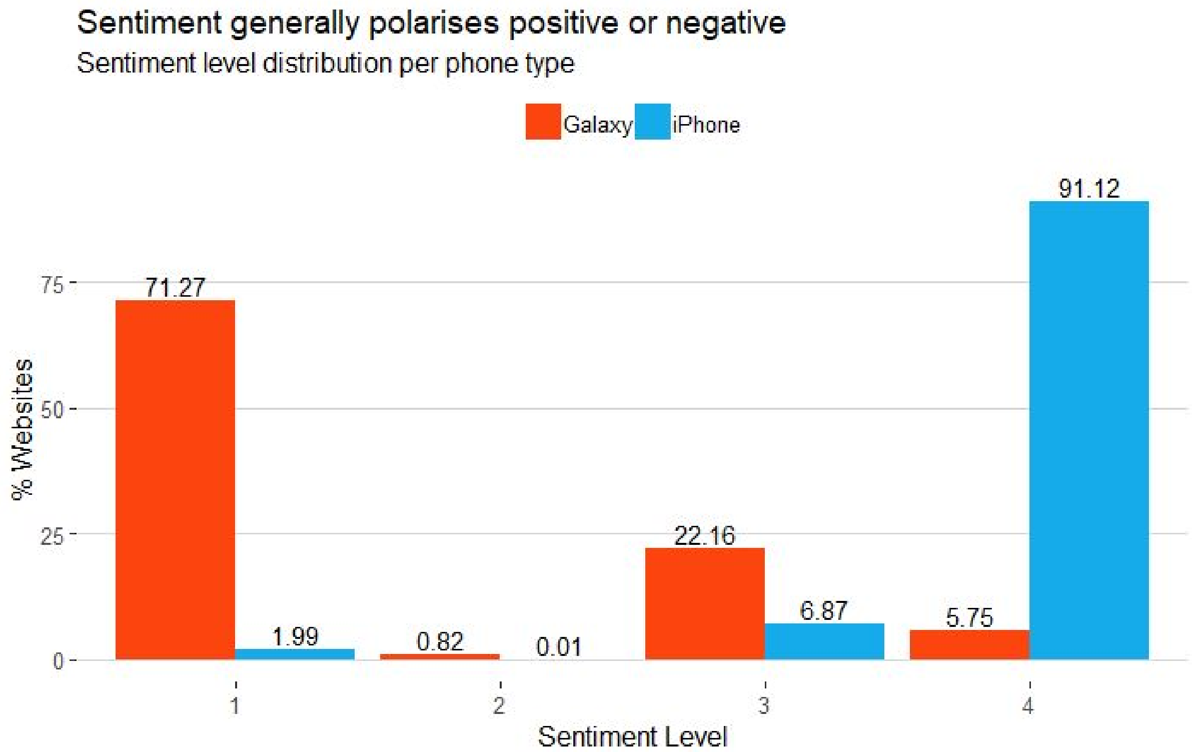
1. Negative
2. Somewhat negative
3. Somewhat positive
4. Positive

Sentiment Comparison

Of the two phones compared (iPhone & Samsung Galaxy), the iPhone had a clearly higher average sentiment, which is statistically significant.



Sentiment generally polarises towards *1 - Negative* or *4 - Positive*. Only 8.49% of total, combined, sentiment falls within the *2 - Somewhat Negative* or *3 - Somewhat Positive* zone.



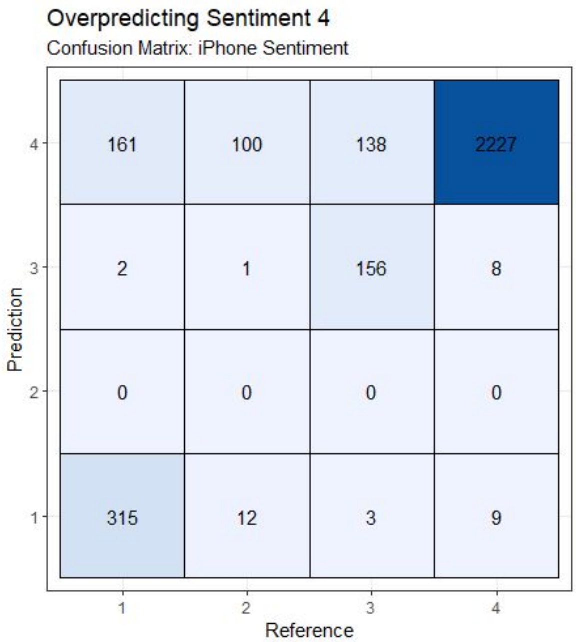
Confidence

Modelling Metrics

iPhone

Modelling iPhone sentiment gave the following results on the test set:

Accuracy Kappa
0.8614304 0.6331245

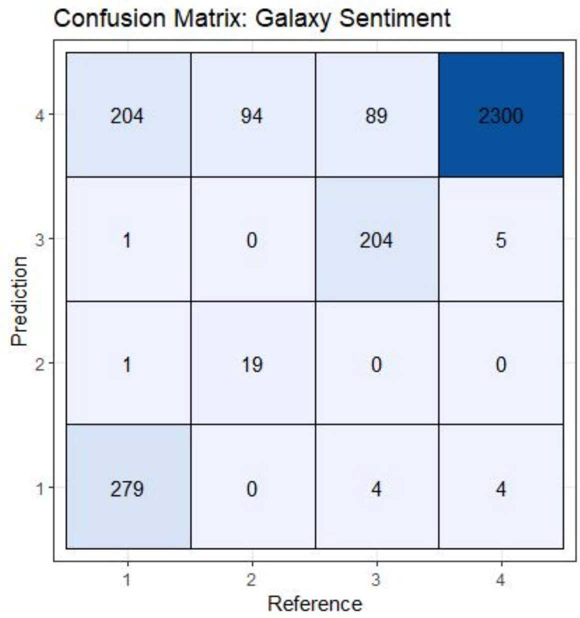


This means that over 86% of the sentiments were correctly predicted. The majority of those that were incorrect were predicted as sentiment 4.

Samsung Galaxy

Modelling Galaxy sentiment gave the following results on the test set:

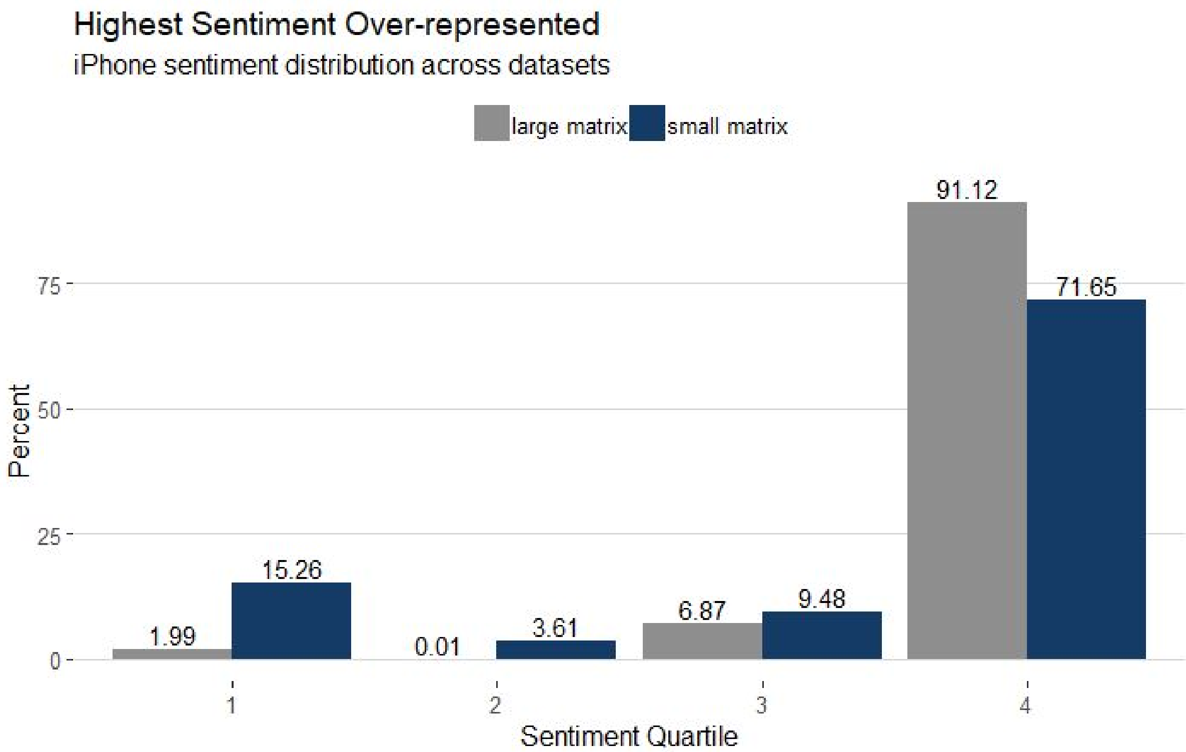
Accuracy Kappa
0.8745318 0.6661036



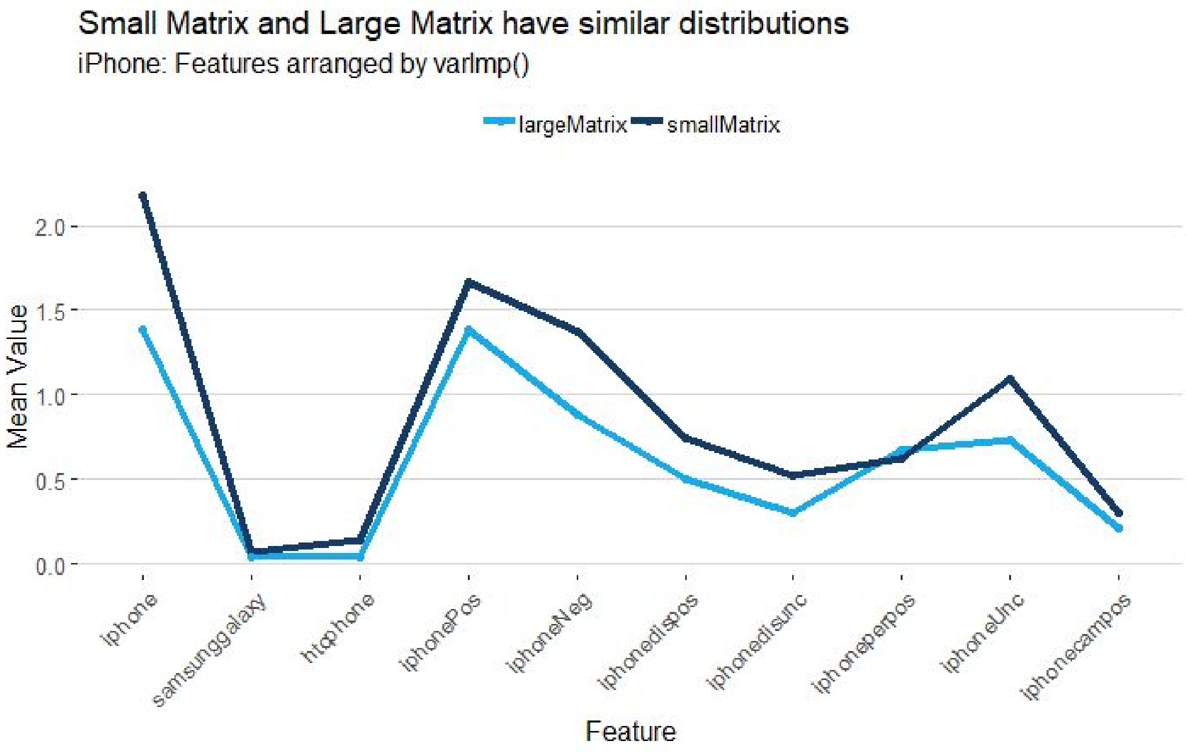
This means that over 87% of the sentiments were correctly predicted. The majority of those that were incorrect were predicted as sentiment 4.

Analyst Insights

The iPhone sentiment results applied to the Large Matrix (unknown sentiment data) have a similar distribution to the results of the Small Matrix (labelled sentiment data used to train model). Both show sentiment 4 as the significantly largest sentiment group.

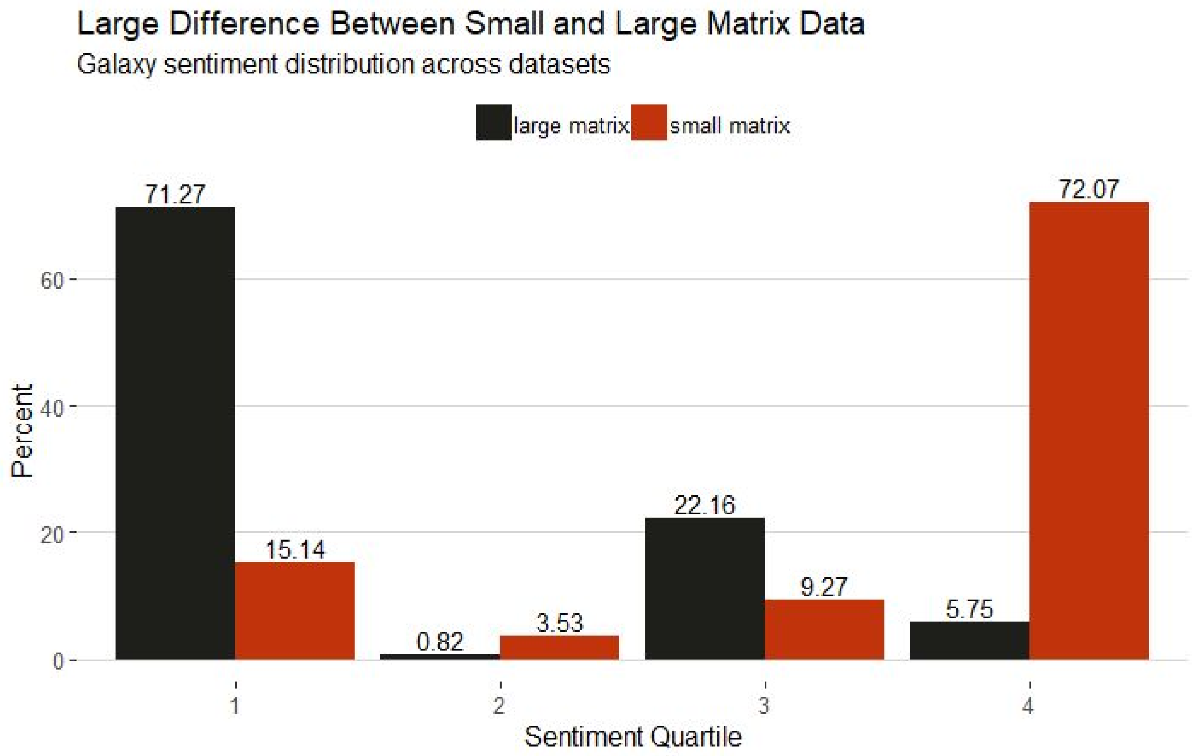


Comparing the means for the 10 features with the highest importance in the model we see that the general distribution of the small and large matrices are similar. This gives me confidence that the predictions of the large matrix sentiment are reasonable.

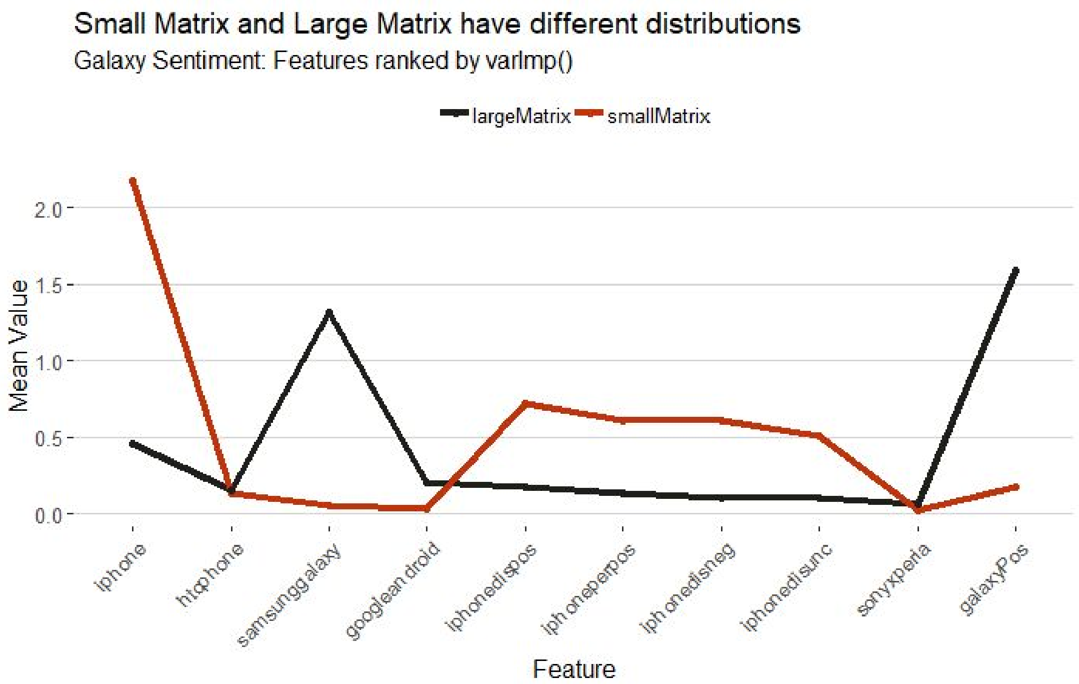


The model over-predicted sentiment 4 on the test set and I believe that this is likely to have occurred on the large matrix, meaning the 91% sentiment 4 prediction is probably marginally too high.

The galaxy sentiment results applied to the large matrix have the opposite distribution to those of the small matrix. In the small matrix 72% of sentiment is in quartile 4, in the large matrix 71% of sentiment is in quartile 1.



This is either a result of the model performing unsatisfactorily or, as I believe, of the data in the large matrix being substantially different from that of the small matrix.



The above diagram shows that the small matrix and large matrix have very different value distributions on the 10 most important features. This is a plausible explanation for why the sentiment quartiles were so different between the small matrix and large matrix.

That the features used by the model to predict sentiment are largely related to iPhone does cause concern and lead to doubt over the validity of the prediction.

To explore this further, the below table shows the median occurrences of positive and negative mentions. Median has been chosen to reduce the impact of the multiple outliers in the data. It clearly shows that iPhone has more positive that negative mentions and has a more positive profile than the galaxy. This supports the modelling results but does suggests an underlying issue with the data which I discuss in the next section.

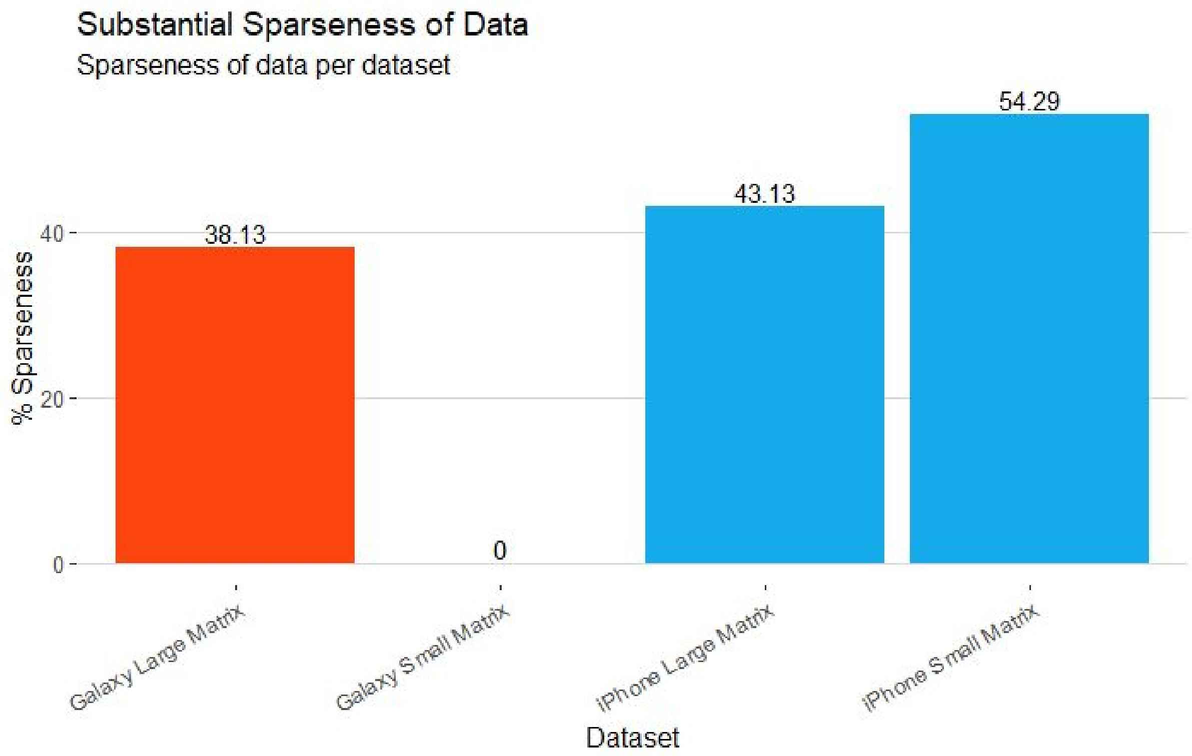
Mention	Median Occurrences
iPhone Positive	1.4
iPhone Negative	0
Galaxy Positive	0
Galaxy Negative	0

Errors and Improvements

Data Sparseness

The way in which the data was collected and the models were developed has led to inconsistencies in the results. Often, when the data was collected, there was a mention of the smartphone and one of the keywords (review; critique; looks at; in depth; analysis; evaluate; evaluation; assess) but very little other data.

I define sparseness as the lack of any information beyond a mention of the smartphone itself and one of the review keywords. This means that all of the remaining features in the data were unpopulated (e.g zero sentiment towards camera, OS etc). There is substantial sparseness in the data (Galaxy small matrix is an exception as there is an iphone entry for the vast majority of the web-pages visited. When iphone is excluded the sparseness becomes 54%)



This means that the models have limited information on which to build and are likely to be less accurate. The next time that the data is collated, I would recommend that it only be included if there is a minimum of 1 mention for all relevant features (e.g for camera, display, OS)

Sentiment Levels

In the next iteration of this project, I would also recommend that 3 sentiment levels are used: Positive, Neutral, Negative. Having humans code their interpretation of sentiment as the dependent variable does bring subjectivity into the modelling. Reducing the number of sentiment levels should make it easier to contain the subjectivity.

Outliers

In the small matrix there were significant outliers (e.g total positive mentions on a web-page ranging from 0 to 237). In the next iteration of the project I would experiment with transforming the data to minimise the potential skewing effects of these outliers. This could be through logging or binning the data.

Implications

Our situation is that we are confident with the iPhone predictions, whilst accepting that it is likely over-predicting positive sentiment, and we have concerns about the validity of the Samsung Galaxy sentiment predictions.

Either our concerns about the Samsung Galaxy sentiment predictions are correct, and the model has underestimated positive sentiment or our concerns are incorrect and the sentiment towards the Samsung Galaxy truly is significantly negative. If there is positive sentiment towards the Samsung Galaxy then we are faced with known positive sentiment towards the iPhone and potential positive sentiment towards the Samsung Galaxy. Alternatively there is negative sentiment towards the Samsung Galaxy.

In either of these cases, I would choose to move forward with the iPhone. I believe that the indications that the sentiment analysis has given us are broadly representative and would recommend the iPhone based on the sentiment analysis conducted.