Smart-Phone Twitter Analysis

Mahdi Azarafrooz, Alireza Louni and Ebrahim Safavi
Depatment of Electrical and Computer Engineering
Stevens Institute of Technology

I. INTRODUCTION AND MOTIVATION

The idea of big-data has recently found a great importance in the market analysis. Users' expectations change over the time and the market planner should be able to keep up with the latest changes. Big-data has come to the picture as a valuable source of information due to the large scale information it carries that can lead to a realistic picture of the existing needs.

A business analyst then can use these information to come up with the optimized strategic plans and fill the market gaps in the most reliable ways. But the main challenge would be to find different approaches to turn these large scale information to meaningful data. Some of these approaches has practiced in our midterm project including linguistic, co-occurrence frequency and information spread analysis.

II. LINGUISTIC ANALYSIS

At the beginning of our analysis, we tried to pull out our desired data from the twitter platform. We used Python-Twitter package for the data acquisition purpose. For each phone item, we pulled tweet data for a time duration of 12 hours. The next step after data acquisition is data cleaning. This is of more importance for the Galaxy smart-phone, due to the existence of a chocolate company with the Galaxy brand name.

After the data cleaning step, we had approximately 6000 tweets per phone. In this part we have applied the linguistic analysis to our collected data. The main idea of linguistic analysis is determining common word categories among tweets. At the first linguistic analysis step, we omitted meaningless words like pronouns and junctions and etc. In the training phase, the top 100 frequent words have been determined. We then came up with five different main categories including software, hardware, marketing, appearance and competitors, where we modeled the market as game with these five major players.

Based on each appearance of each category, we have drawn the perceptional map which provides us with the insight over popular features of each products.
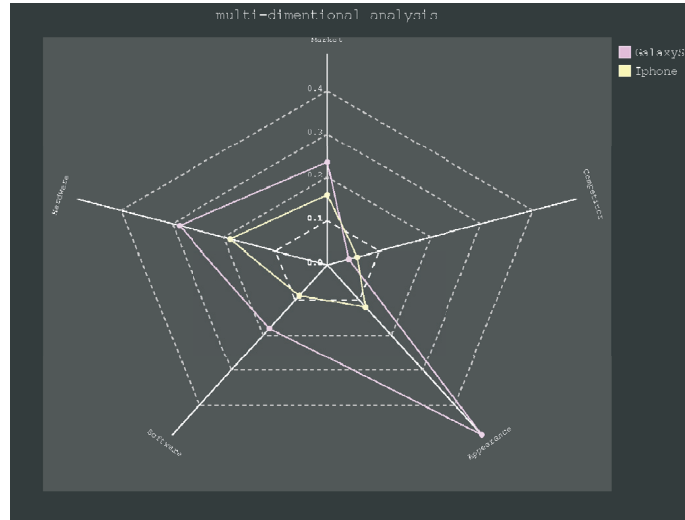

Fig. 1.   The smart-phone game perceptional map based on the tweet linguistic analysis

III. CO-OCCURRENCE FREQUENCY ANALYSIS

In this part we have analyzed tweets using frequency of co-occurrence with "Galaxy". It has been observed that some words appear in tweets with the same value of co-occurrence frequency. For example, "eyebased" and "scrolling" are appeared in tweets about Galaxy with the same value of co-occurrence frequency, 0.81, and "features" with relatively close value, 0.21. Using this method, we can have deeper insights about how costumers think of products (rather than separate and independent word analysis). This observation can be seen with more details in the graph visualization. Vertices of the graph are keywords that show high values of re-occurrency with the word "Galaxy. There is an edge between any pair of keywords if the mutual

co-occurrence frequency of them in each tweet is greater than a specific threshold amount. It can be seen that for large values of threshold, the graph is broken into different clusters and some isolated nodes in which each cluster points to a hot topic about "Galaxy" among twitter members.

IV. EXTRACTING RETWEET ORIGINS FOR GALAXY4S

An understanding of information flow in social networks has many applications. One relevant application is the marketing impact on social media. This sort of analyzes can help the market designer to take advantage of the situation to maximize his profit by widespread advertisements. In this part, as the simplest step of analyzing of information propagation, we wanted to extract the originating source for Galaxy4S from retweets.

Retweet paltform of tweeter has evolved several times. We have used the current platform to extract the origin of a retweet about GalaxyS4 with convetions such as "RT@users" users or "via@user". We looked at retweet count field value of tweets. If the value is more than zero we look for the text of the tweet to find the referred twitter URL. We also had intended to do the same analyze for iPhone to get a grasp at the effect of graph of retweets on the success of a specific product. However this try failed for the iPhone and a more comprehensive analyis of Galaxy due the failure of python-twiter package at the last stage of our simulations. This part was not discussed on the presentation for the aforementioned reasons. However the python code has attached in the corresponding folder.