

Prediction of Price Fluctuation of Used Cellphones on Amazon by Twitter

Tonghe Zhan, Qingyuan Wen

Introduction

Nowadays, with the rapid update of cellphones, more and more people who change their cellphones frequently will to experience different kinds of cellphones. So it is really a saving way for them to go to the used items market. Amazon provides a good platform for those kinds of people. It has a 3rd party used market for them. And our project focuses on the used cellphone from 3rd party on Amazon. People want to buy the cellphones at a best time, so they will to know the price fluctuation about this cellphone. We think it is a quite meaningful project, and it is really has a high practical applying value. So we want to begin our research on this area, and there are many different kinds of cellphones, we choose Samsung galaxy S i9000 as a representative. We want to use sentimental classification to find out whether we can predict the price of this cellphone by twitter, and find out if there are any correlation between twitter and second hand cellphone. We assume that the comments of cellphones on Twitter will affect the price of the used cellphone from 3rd party on Amazon.

Data

In our project, there are two mainly kinds of data. One is the raw dataset comes from professor Aron Culotta. There are more than 10,000 tweets in the raw dataset. They are the tweets contain the words: “galaxy s”, “i9000” from September 1 to December 31, in 2011. Because the year of the cellphone released is 2010, we choose collecting the data one year later, because we think this period has the best value for us to do the research on second hand market. The other one is the price fluctuation of this cellphone in same period of the tweets above. We find an Amazon price tracker “Keepa” to collect information. And because there are many types of Samsung Galaxy S, so we just choose two representatives of it which sells well in 2011. Then we record the data manually.

Methods

Price Tracker

At first, we are going to use an Amazon price tracker named Camelizer, the one we have mentioned in our proposal. But we find that it only can provide a trend of the price of a certain cellphone without exact price in each day. So we find another Amazon price tracker website --Keepa.com, which will show the price in each day of this cellphone exactly and it is more convenient for us to record data.

Sentiment Analysis

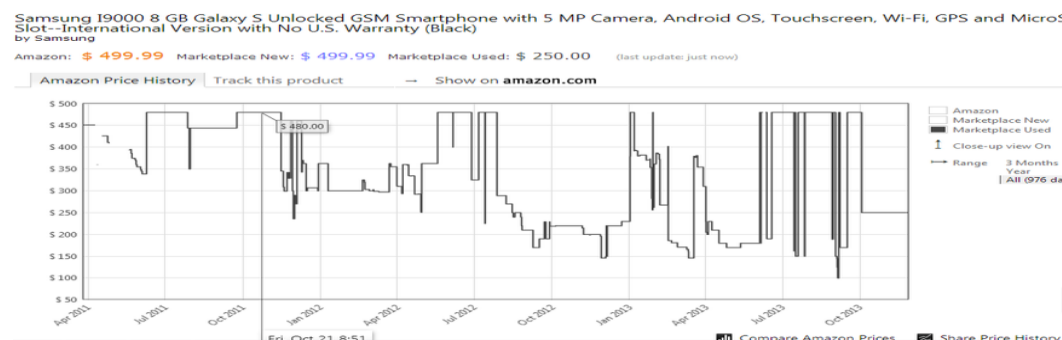
We dealt with our sentiment analysis by applying Alex Davies' word list in order to see if a simple approach is sufficient enough to correlate to price change. For this, we use a pre-generated word list of about 5,000 common words with two log probability of 'happy' and 'sad' associated with the respective words. Because the probability of 'happy' and it of 'sad' equals 1, we only used the probability of 'happy'.

The word list application works as follows. First, each tweet is tokenized into a word list. The function in the application will divide the tweets into individual words with whitespace. Next, it is going to look up each token's log-probability in the word list. Then, we used the log-probability of 'happy' of each word or even emoticon to calculate the value of 'happy' of the tweet. Finally, we combine the tweets in same days into an average daily sentiment value.

Experiment

We choose the Samsung Galaxy S (I9000) as a representative. To make the dataset more reliable, we combine the price trend of two different models of Galaxy S into one cart.

The following images are the price fluctuation trend of the Samsung Galaxy S from keepa.com. If we move the mouse in different positions, the price and the time will change automatically.



First model: Samsung I9000 8 GB Galaxy S Unlocked Smartphone.



Second model: Samsung I9000 8 GB Galaxy S Unlocked Smartphone.

Then we use Excel to handle those data, and take an average of the two prices in each day.

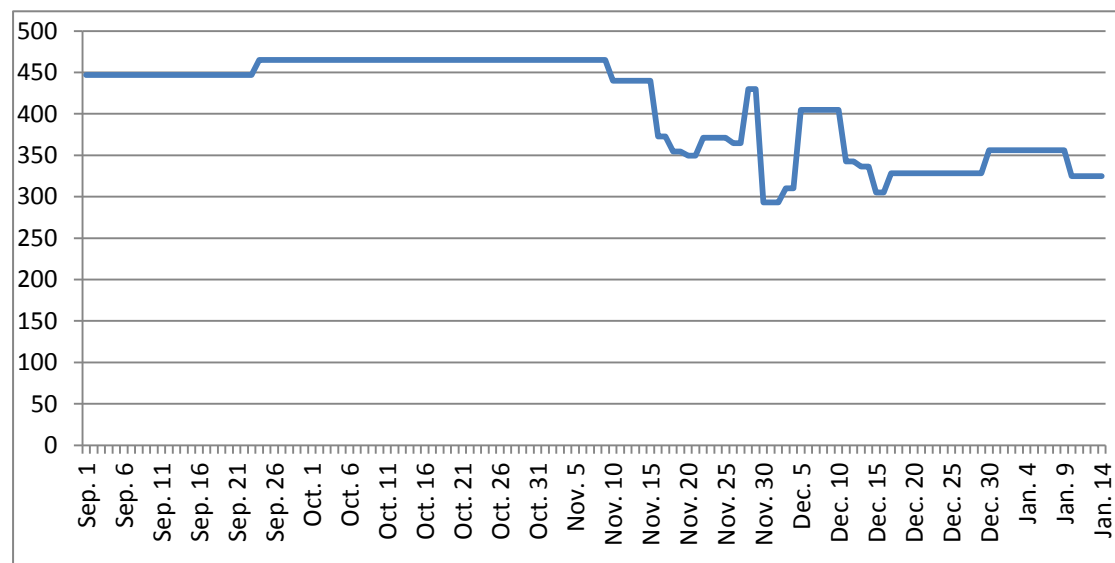


Figure 1: The average price fluctuation of i9000 combined with 2 models.

In the other side, we also handle raw data from Twitter. Because we are new to Python, we cannot handle complex data with Python. According to this problem, we used Java to deal with the data.

Filter

Because of lots of noise, the raw data should be filtered to obtain the tweets are meaningful and useful. For example, we eliminate those are talking about 'galaxy s ii' or 'galaxy s 2' which also contain 'galaxy s', and other situations, such as, in a tweet, user actually talk about galaxy s 2, but they will mention galaxy s just as a comparative object. After filtering the dataset, we get 3369 tweets which really talk about this sort of cell phone. Then we sort the data by date and calculate the average probability of 'happy' with the word list for each day as we can see in Figure 2.

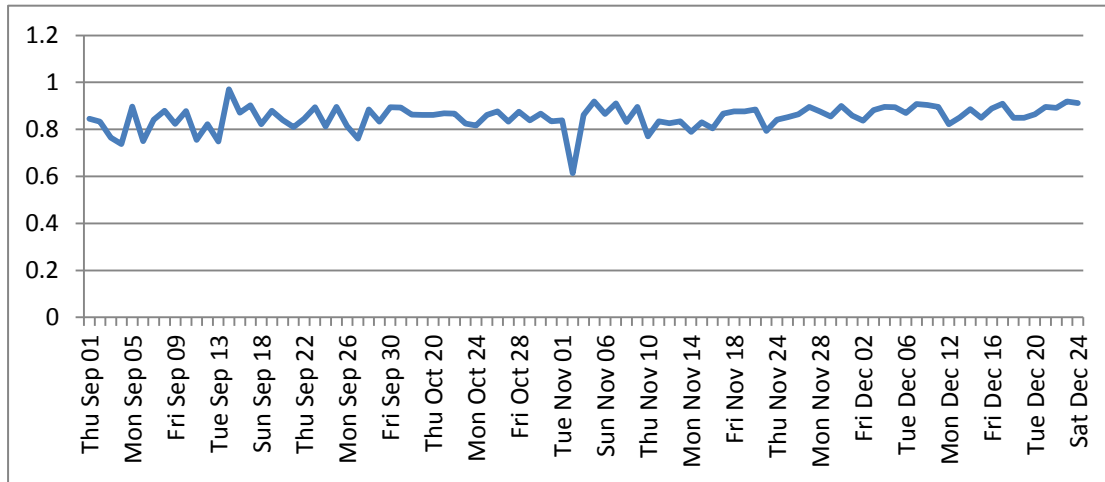


Figure 2: the average probability of 'happy' for each day.

Factor 'X'

As the Figure 2 shows, we can hardly find any useful information to analysis the relationship between the price of this phone and evaluation given by tweets. Then we came up an idea that the tweet's sentiment will not change by time. Because if somebody thinks it is a good cell phone, he will always think so. But the price will decrease on the whole as time goes by. So we suppose a factor 'X' that represents the influence of time.

We select a long range time to calculate the factor 'X'. It represents the decreasing of the influence of a tweet which is talking about an outdated cell phone. Then we import fact 'X' to the method and acquire the Figure 3.

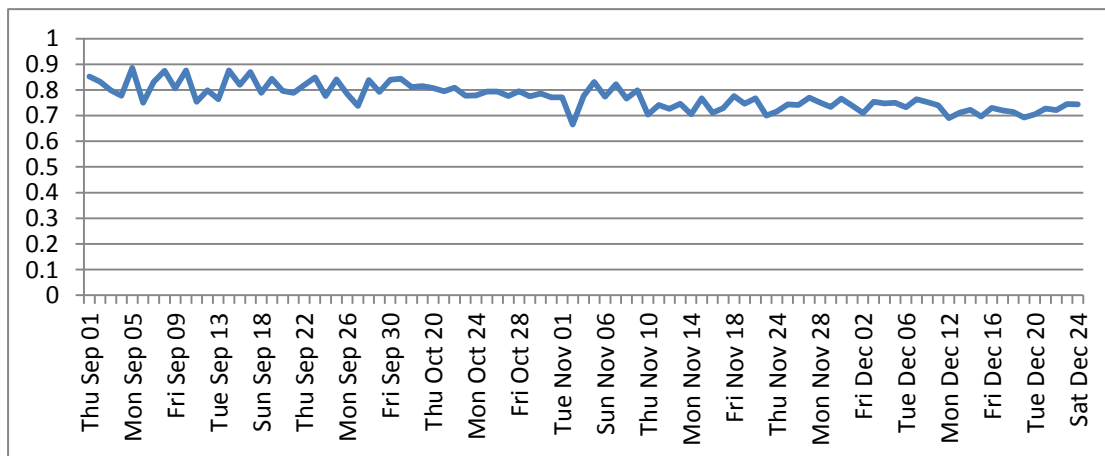


Figure 3: the result after added factor 'X'

But it still changes frequently and hard to analysis. To make it more smooth, we combine 3 days into 1 point. So we get the final result shown with Figure 4.

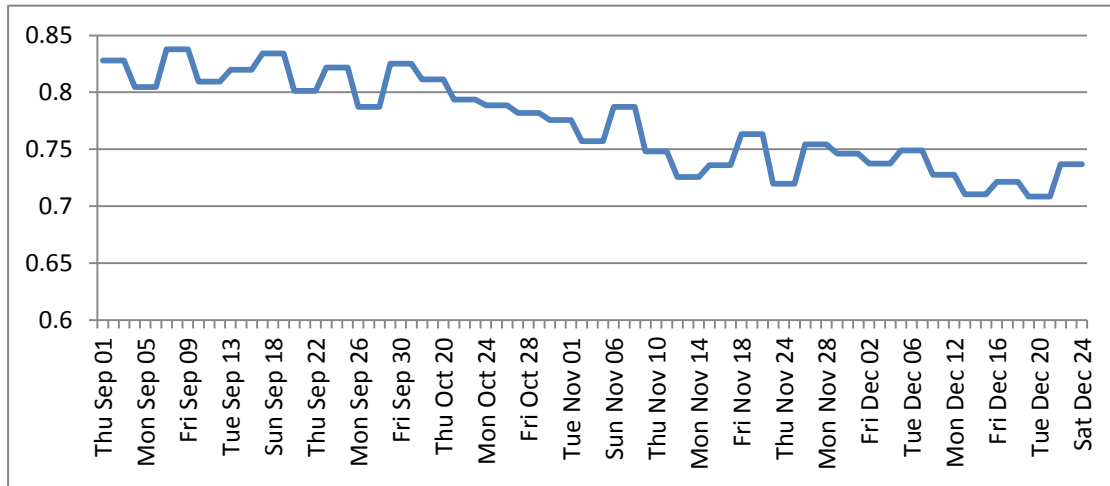


Figure 4: final result

Result

If we just focus on Figure 1, it may be hard for us to find the law. At beginning, the price of the cellphone does not change frequently, and in Nov. 10th, the price of it begins to decrease and the price changes more frequently than before. And about Figure 2, it shows that the “happy” probability starts to decline in Oct. 21th, it changed frequently, but after Oct. 21th, it never reaches 0.8. It proves that, the positive comment on this cellphone decreases generally. And when we put these two figures together, we can see that the trend of these two line charts are all decrease in a certain time, the interval of them both begins to decrease just about 20 days.

The change coincides with the reality, when the positive comments to a product decrease generally, the price of this product will not decline immediately. It will take some time for the second hand market receiving those information and have reactions. So, after 20 days, the price of this product declines, it is really reasonable. To sum up, our conclusion is that the price fluctuation of a cellphone in 3rd party used market on Amazon will be affected by the trend of the evaluation about the cellphone on twitter.

Future work

We still have a lot of future work to do, to make the project more perfect.

First of all, we should find some cellphones in other brands to test our method, our result. We choose Samsung cellphone as representative this time, because it is a famous brand in the world and all people in different countries trust the quality of Samsung and would like to buy the products of Samsung. So we will get a lot of data about it, but there are many other good cellphones, so whether they all have the same law or not, we should have a further test on it.

Secondly, this time our data all come from between Sep. 1 to Dec. 31, about four months, we think it is enough for us to do the research, but if we want a perfect result, we can still larger the dataset to make our result more accurate and trustable.

Thirdly, we can do more detailed work, such as, the period of our data come from, are there any other new cellphone released leading to the change of the price of the cellphone we focus on. Because new cellphones will attract people's eyes, and the new one must have updates compare to the cellphone in our experiment, so in people's tweets, the "happy" probability of it will decrease obviously, there is no doubt that it will have a impact on the price of this cellphone, and this influence may affect the result.

Related paper

1. Ray Chen, Marius Lazer: *Sentiment Analysis of Twitter Feeds for the Prediction of Stock Market Movement*
2. Alex Davies. *A Word List for Sentiment Analysis of Twitter*.
<http://alexdavies.net/2011/10/word-lists-for-sentiment-analysis-of-twitter/>