Amazon vs Walmart

Shuopeng Wu

Abstract

Twitter messages are able to provide a good reflection of public attitude in aggregation. In the project, I want to capture the users' attitude towards sales in Holiday season by examining the words frequencies patterns, the relationships between each brand and keywords, the estimated locations towards each in U.S. Additionally, I examine the effectiveness of various machine learning techniques on providing a positive or negative sentiment on a tweet corpus.

Introduction

Thanksgiving Day is a traditional holiday in the United States. For the Americans, are not only to share the happy time with their family, but also to go shopping to reward their hard work for the whole year. Therefore, companies wishing to attract the attention of the customers have to continually improve the promotional products and/or service in order to ensure a good business relationship with customers. Last year on Black Friday, consumers spent $12 billion at real stores and $1.9 billion online. As a retailer, you can't ignore a weekend that can potentially earn 60 percent of your annual income.

As a social media website, twitter is a very famous social media network where users can post and share messages, links, pictures and so on. The aim of this project is to compare between Walmart and Amazon according to the Twitter users' attitude. This thanksgiving social media battle will influence the companies' social status and the reputations. These are really crucial for a retail company.

Preprocessing

(a)Keywords Selection

I selected keywords related to amazon and Walmart from @amazon @walmart #amazon #walmart amazon amazon's amazondeals walmart walmart's. But I did not put very specific keywords in both brands because it can put extra weight on specific area, for example, the keyword " amazonpayments" could generate greater percent of frequency words on payment problems than the general "amazon" itself.

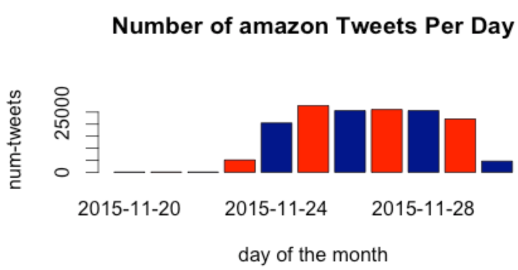
(b) Date Selection

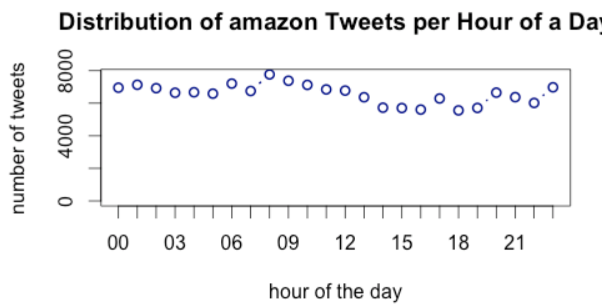
Because the analysis is about the Thanksgiving period, so the time frame is from Nov 20^{th} to Nov 30^{th} (10 days in total). So the Black Friday and the Cyber Monday are all in the frame as well.

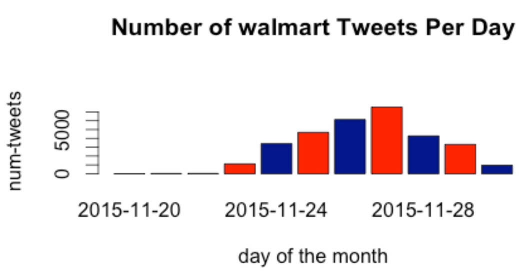
Influence Analysis

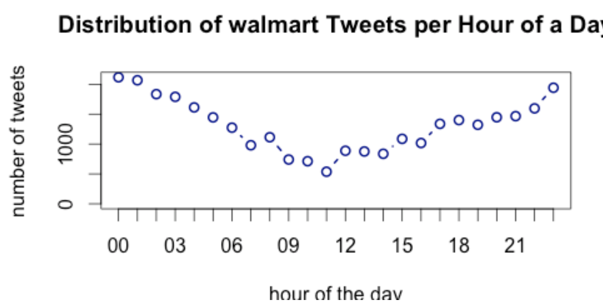
From the tweets per day, Amazon definitely was more active in tweeter than Walmart. From statistic result, the tweets generated around Amazon is about five time more than those on Walmart. And from the bar plot, we could also see the attention on Amazon was very high and stable from Nov 24th to Nov 29th. On the contrary, the amount of tweets on Walmart increased until Nov 27th (peak time), then it slowly decreased.

```
[1] "Amazon tweets per day in Thanksgiving"
2015-11-20 2015-11-21 2015-11-22 2015-11-23 2015-11-24 2015-11-25 2015-11-26 2015-11-27 2015-11-28 2015-11-29 2015-11-30
        76         93         86       5119      20514      27722      25634      26106      25636      22131       4628
[1] "Walmart tweets per day in Thanksgiving"
2015-11-20 2015-11-21 2015-11-22 2015-11-23 2015-11-24 2015-11-25 2015-11-26 2015-11-27 2015-11-28 2015-11-29 2015-11-30
        10         14         22       1113       3403       4674       6133       7545       4273       3309        951
```





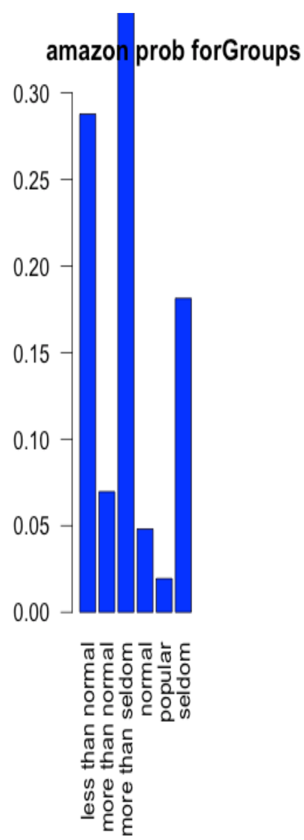


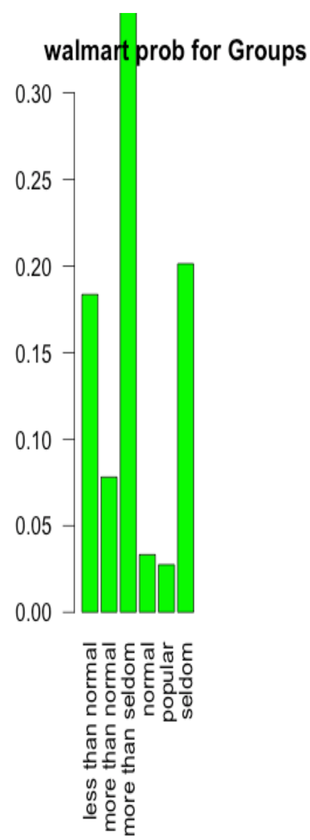


As we also interested in which time of a day in those period, those two brands were active on tweeter. The results are very interesting. The Amazon got tweets all the time without any significant peak time. However, Walmart got the least attention in the afternoon compared to the other times. It is understandable that in the vocation night, people tweets about their holiday shopping, while in the afternoon they are just too busy together.

I want to know what kind of people were tweeting those brands during the Thanksgiving period. Let's assume:

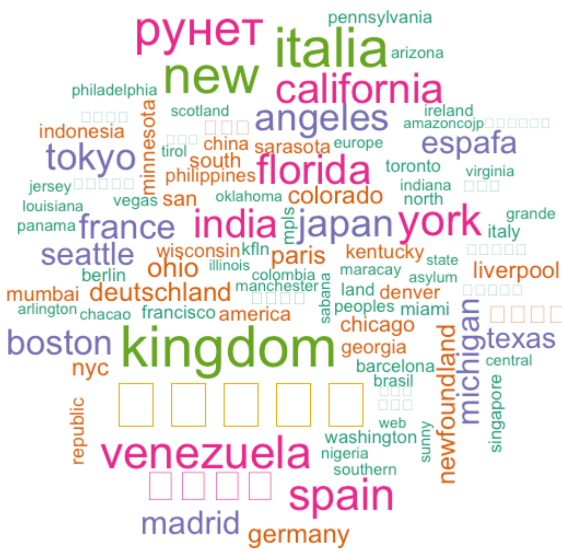
Numbers of Twitter followers In Range	User in the Group
[,100]	Seldom
[100,1000]	More than seldom
[1000,5000]	Less than Normal
[5000,10000]	Normal
[10000,50000]	More than Normal
[50000]	Popular





The results are also impressive. For both Walmart and amazon, people tweeted them the most are those in the "more than seldom" group range. The follower two most tweeted group are the "less than Normal" and the "seldom" user groups. So I say, in general, people who do not use twitter everyday are willing to check and tweet on the retail shopping experiences during the Thanksgiving period.

And meanwhile, we are also interested in which area in U.S. are more interested in those retail brands. However, the result had lots of noise because clearly people around the world were talking about Walmart and amazon. (left one is for Amazon, and the right one is for Walmart)



Even though there are some differences in the users' location between Amazon and Walmart. But overall, popular city/state in U.S. has more shopping needs than the rural one. For example, we can see California, Los Angeles , New York, Chicago, Florida in both clouds. The other reason, I think, is both Walmart and Amazon are popular in U.S. Even though Walmart is a physical store and Amazon is an online store, they all have inventory distribution center all over the place.

Sentimental Analysis

After building the topic model and the LDA model to deep dig what were people talked around amazon and Walmart during the period, we could see those words, like (‘black Friday’ ,’deal’, ‘sales’ , ‘cyber’, ‘thansgiving’ ,’got’), showed as hot topics for both brands.

However, the top topics on amazon seem more positive. The words like “save”, “like”,”good”,”thanks”,”want” and so on. I did notice the amazon’s prime service is very prompted during the rush shopping period. And its products/services as “kindle”, “fire”(as kindle fire), “itunes”,“video”, “apple” are tweeted a lot. And yes, there were sales on those stuff during the period.

On the contrary, the topics in Walmart model are more neutral or even negative, we got words as “don't”, “fuck”,”someone”. And one interesting thing in Walmart topic models is that its competitors, (“amazon” “target”) were also frequently mentioned. I don’t think it is a good sign. When people consider it is a sale or good deal, they will not be that calm to compare between competitors.

(Left is the amazon topic model, and the right is the Walmart topic model.)

And yes, thanksgiving sales topic changes little by little every day during the period. However, I did not see any surprising growing topic as the day past. Below is the amazon word clouds form day 26 to day 29 (from left to right, then top down).

We generate a dictionary with count occurrence of all words. Then we track each text in tweets with a feature vector over the dictionary words we just generated. It is a 'Bag of words' approach, which ignores the grammar and orders of words, but works on the multiplicity. However, we only have 158318 in our words dictionary for amazon and even less in Walmart sample as 31682. I got more than Training Error more than 45% in the Naïve Bayes. So I am not going to use this method. I use the rule-based sentiment rules defined in vader package to evaluated our sentimental values. Because this well defined package is based on sensitive of both polarity ad intensities in social media contexts (especially its frequent word dictionary was built on tweet text). I instead use this in python (code uploaded in the project). Specifically, the vader package predefined quantifications positive and negative probabilities for each words in each sentence. It is an advanced version of our term-frequency analysis, as the degree adverbs impact the intensity. For example, after we tokenized the text, ['very','good'] increases the positive probability of just ['good'].

The result of this sentiment analysis is also very convincing. The negative score of Walmart is pretty high, and the total sentimental score of Walmart is much worse than Amazon. And yes, most words are neutral because we did not rule out the links as well as some code like words.

We get:

Brand	Negative	Neutral	Positive	Aggregate Score
Amazon	0.01190798898419	0.9522555805404427	0.03430239138948185	0.04125391048396074
Walmart	0.03374244050249352	0.9230682406413827	0.04284259200808061	0.01863040212107906

Conclusion

Because both tweet sentiment analysis and the stock market are actually reflecting public attitude in aggregation, we could use the stock market value during the same timeline as a measure of our evaluation success.

It's not surprising that Amazon did the best in the past Black Friday period. From our word frequency report, we found out the appearance of keyword "deal" co- occurrence with "amazon" more than with other brands. Given by USA today, the profit is estimated to grow 261% in the fourth quarter. As Walmart have the more negative score in our sentimental analysis, it is also reasonable to find out its stock price went done compared Amazon during that period (graph above from yahoo finance). Those extremely cases are able to match the reality world. However, this textual analysis result could be substantially improved for more sensitive analysis.

R Code Other Results Not Included Yet:

(a)

```
[1] "Amazon Percentage of tweets with different number of follower"

less than normal more than normal more than seldom          normal         popular          seldom
           45395            10987            62112            7582            3062           28607
[1] "Walmart Percentage of tweets with different number of follower"

less than normal more than normal more than seldom          normal         popular          seldom
            5773             2455            14978            1046             864            6331
```

(b)

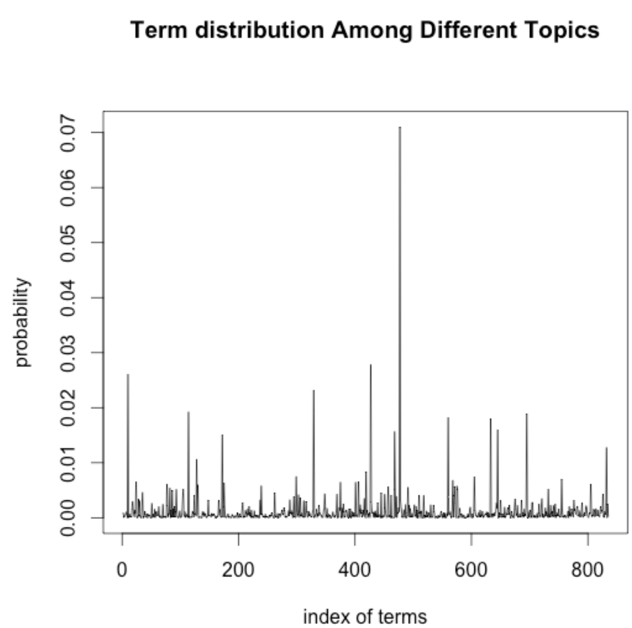
Below the first part is amazon location, the the second part is for Walmart.

```
  愛知名古屋     kingdom      italia         new       руннет   venezuela       spain        york    愛知一宮  california     florida       india
   7084.185    5654.403    5466.554    5126.409    4710.833    4364.819    4214.008    4163.878    4131.494    3869.040    3853.701    3688.280
      japan     angeles       tokyo      boston      france      madrid      espafa     seattle    michigan       texas        ohio    愛知江南
   3513.686    3439.613    3398.883    3110.235    3032.401    2770.717    2729.556    2705.359    2675.761    2371.510    2186.836    2054.436
deutschland     germany newfoundland      paris    colorado         nyc   liverpool     chicago         san      日本国       south   minnesota
   2048.449    2035.260    1987.546    1961.005    1888.460    1760.207    1700.016    1512.756    1505.797    1488.847    1469.697    1399.325
     mumbai     america   indonesia    kentucky    sarasota      denver     georgia   wisconsin    republic philippines       china    岐阜川島
   1372.919    1332.197    1314.714    1311.711    1292.499    1285.641    1275.732    1240.500    1219.570    1210.341    1203.969    1174.561
      miami       italy
   1152.944    1143.730
     toronto   california          new        texas      florida      angeles      chicago         ohio       dallas         york
   1780.4533    1133.0221    1108.5418    1054.5907     999.7776     898.5382     797.4577     767.3646     746.7981     733.4249
         nyc    arlington          san       brasil        vegas      atlanta      georgia      indiana    tennessee   washington
    662.4536     606.6650     570.3483     563.0844     523.9971     522.7816     519.4714     492.0264     491.9727     477.6320
    kentucky     michigan        south       kansas      ontario     oklahoma  philadelphia     illinois      alabama      houston
    468.1635     460.5090     447.2874     438.3784     432.3215     430.4264     421.9389     419.7870     414.9579     410.6671
      puerto         rico          bay       denver       jersey      america        north      calgary      seattle    cleveland
    406.4594     401.9142     388.3495     380.0822     365.8620     365.5944     350.4976     347.1722     345.1424     344.5497
      brazil    wisconsin         utah     virginia     maryland     colorado        state progressiveland pennsylvania    oregon
    342.2044     340.6813     334.9426     328.8029     325.8488     311.4020     311.0993     310.5182     296.3382     294.8553
```

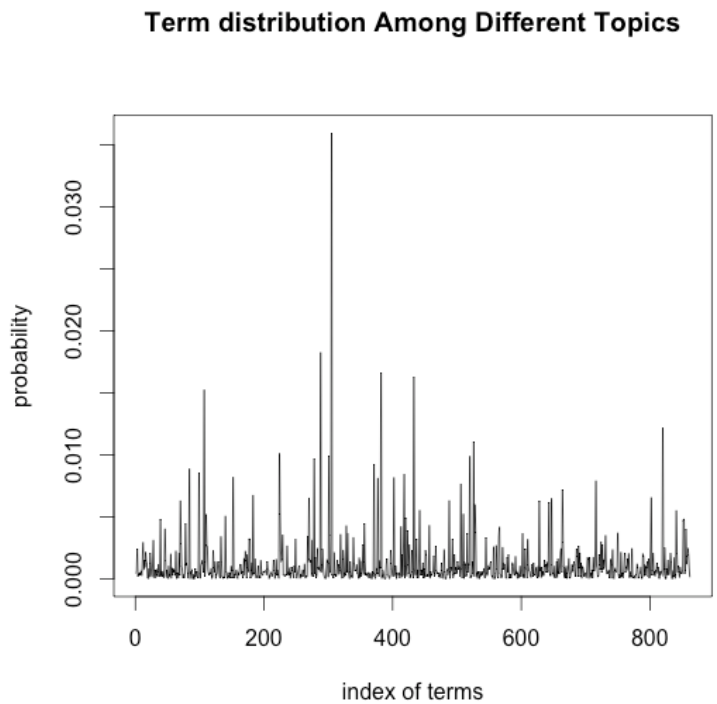
(c)

```
[1] "15 most freqterms for 10 topics in amazon"
      Topic 1       Topic 2       Topic 3    Topic 4       Topic 5       Topic 6     Topic 7       Topic 8   Topic 9       Topic 10
 [1,] "black"       "fire"        "man"      "get"         "sale"        "black"     "gift"        "just"    "best"        "blackfriday"
 [2,] "will"        "kindle"      "one"      "free"        "amazons"     "friday"    "card"        "prime"   "prime"       "friday"
 [3,] "new"         "goldbox"     "new"      "itunes"      "book"        "now"       "just"        "deal"    "free"        "win"
 [4,] "get"         "new"         "ads"      "black"       "memory"      "amp"       "today"       "amazons" "blackfriday" "deals"
 [5,] "friday"      "save"        "nazi"     "friday"      "includes"    "deals"     "itunes"      "friday"  "deals"       "get"
 [6,] "deals"       "now"         "high"     "deals"       "ram"         "get"       "get"         "list"    "new"         "per"
 [7,] "days"        "book"        "just"     "amazons"     "cyber"       "use"       "fire"        "best"    "bezos"       "black"
 [8,] "blackfriday" "kindlefire"  "castle"   "book"        "monday"      "account"   "deal"        "black"   "get"         "amazons"
 [9,] "delivery"    "dec"         "subway"   "just"        "air"         "can"       "win"         "like"    "friday"      "itunes"
[10,] "like"        "see"         "amp"      "blackfriday" "drone"       "kindle"    "enter"       "get"     "jeff"        "just"
[11,] "price"       "tea"         "hiring"   "fire"        "friday"      "flipkart"  "via"         "cyber"   "now"         "cloud"
[12,] "buy"         "show"        "free"     "kindle"      "blackfriday" "looking"   "blackfriday" "can"     "video"       "amp"
[13,] "need"        "select"      "software" "buy"         "yep"         "think"     "giveaway"    "wish"    "ich"         "month"
[14,] "blue"        "blackfriday" "apple"    "sales"       "taking"      "holiday"   "love"        "monday"  "book"        "prime"
[15,] "shepard"     "stick"       "engineer" "can"         "check"       "bought"    "free"        "listed"  "hey"         "buy"
[1] "15 most freqterms for 10 topics in walmart"
      Topic 1       Topic 2  Topic 3  Topic 4  Topic 5       Topic 6       Topic 7  Topic 8        Topic 9           Topic 10
 [1,] "black"       "get"    "friday" "think"  "black"       "just"        "get"    "just"         "friday"          "just"
 [2,] "friday"      "black"  "black"  "just"   "going"       "one"         "people" "thanksgiving" "amp"             "like"
 [3,] "get"         "dont"   "deals"  "friday" "got"         "people"      "like"   "deals"        "get"             "people"
 [4,] "like"        "one"    "like"   "like"   "like"        "lol"         "black"  "people"       "going"           "time"
 [5,] "now"         "need"   "best"   "time"   "target"      "target"      "now"    "workers"      "buy"             "dont"
 [6,] "buy"         "buy"    "going"  "dont"   "friday"      "got"         "lol"    "amp"          "sendhallmark"    "now"
 [7,] "time"        "friday" "people" "can"    "mom"         "blackfriday" "deals"  "friday"       "thesimpleparent" "got"
 [8,] "target"      "can"    "work"   "buy"    "thats"       "friday"      "want"   "work"         "someone"         "shopping"
 [9,] "went"        "people" "today"  "turkey" "need"        "see"         "went"   "mom"          "now"             "friday"
[10,] "hate"        "see"    "cant"   "lol"    "saw"         "get"         "going"  "like"         "like"            "lol"
[11,] "shopping"    "going"  "kohls"  "know"   "now"         "dont"        "day"    "blackfriday"  "right"           "right"
[12,] "match"       "store"  "just"   "night"  "get"         "can"         "just"   "really"       "people"          "going"
[13,] "blackfriday" "ive"    "went"   "got"    "price"       "black"       "fuck"   "want"         "lol"             "man"
[14,] "cant"        "amp"    "lol"    "last"   "thanksgiving" "need"       "back"   "best"         "deals"           "never"
[15,] "price"       "just"   "camera" "black"  "people"      "today"       "open"   "open"         "crazy"           "probably"
```

(d) For Amazon



For Walmart


Term distribution Among Different Topics
probability
0.000
0.010
0.020
0.030
0
200
400
600
800
index of terms