UNDERSTANDING SMART LOCKER USER BEHAVIOR THROUGH TWITTER

Colette Malyack and Pius Egbelu
New Jersey Institute of Technology, USA

ABSTRACT
Understanding smart locker sentiment and use is an area of increasing interest for package delivery organizations. Applications of this data could result in cost savings through route optimization and increased placement of smart locker technology. However, there has been little effort applied to gathering information related to public sentiment on this topic. Therefore, we gather and analyze Twitter data related to smart lockers to determine if there is change in sentiment over time and if socialization is greater in certain regions or communities. This analysis is performed through multiple statistical analyses, linear regression, time series decomposition, and logistic regression. Some significant findings indicate that socialization of tweets related to smart lockers increased over time and socialization is greatest in the more densely populated continent of Asia. Future studies are encouraged to continue analysis of data related to smart lockers based on population density, as this could provide marketing and delivery optimization improvements to decrease cost without decreasing customer sentiment or service.

KEYWORDS
Smart Locker, Microblogging Sites (MS), Statistical Analysis, Time Series Decomposition, Logistic Regression

1. INTRODUCTION

Movement of goods through package delivery services has become a requirement of modern life. With Amazon and other online retailers providing the ability to purchase goods and have those goods arrive at a customer’s home within twenty-four hours, more people than ever have been shipping packages. From Thanksgiving 2019 to New Year’s Day 2020, USPS is expecting to ship an average of 20.5 million packages per day (USPS, 2019). While customers do not usually think about the process behind package delivery, companies like UPS and USPS (United States Postal Service) prepare all year to ensure customers are satisfied with their package delivery experience, especially during the holiday season.

Smart lockers provide a new opportunity to ensure customers have a positive delivery experience while reducing the number of stops necessary for delivery of those packages. Adopting smart locker technology can assist companies in what is known as the “last mile delivery problem”, where delivering a package from the hub to a pickup location could account for up to 53% of the total shipping cost (Dolan, 2018). Reducing this cost would benefit these companies and their customers, who could see a reduction in shipping costs or incentives to use smart lockers. These lockers also provide other benefits for customers such as avoiding the need for repeat delivery calls and protection from package theft.

However, getting customers to use smart lockers requires gaining an understanding of how they view this new delivery option. Understanding customers and their use of available services is a well-known aspect of all companies that provide a service or product, including package delivery. Customers have multiple options for package delivery ranging from requiring a signature on delivery to informing companies of a preferred location on their property to drop the package. The launching of Amazon’s smart lockers in the early 2010s allowed a growing population access to shared and secure package storage, with lockers currently in over 900 cities and towns across the United States (CBS San Francisco, 2012; Holsenbeck, 2018). However, delivery optimization algorithms do not currently consider customer preference for shared pickup locations and lockers at this time.
Lack of customer sentiment incorporation in these models is likely due to lack of access to information relating to the topic. The gap could be closed through the use of microblogging sites (MS), such as Twitter, where customers publically share opinions and knowledge on a wide range of topics. These posts, known as “tweets” in the case of Twitter, can allow companies to gain real-world insights into how their products are being viewed and where to focus marketing strategies. In the case of package delivery, this could be used to optimize the use of smart locker locations to decrease driver stops and increase customer satisfaction. As with all optimization and model development, it is first necessary to analyze the applicability of a data point prior to considering adding the attribute to a model. The intent of this research is to examine the potential role of MS in understanding customer sentiment for smart lockers.

2. LITERATURE REVIEW

To our knowledge, there has been no previous work done on the analysis of customer sentiment through MS relating to the use of smart lockers. However, there have been many previous accomplishments relating to the incorporation of customer sentiment through MS in other industries. Several studies have focused on the effects of MS on marketing campaigns ranging from the stock market to product sales. In one study by Mittal and Goel, it was found that Twitter data allows for the collection of public sentiment through natural language processing, before showing a correlation between certain moods and stock values (Mittal & Goel, 2012). A finding that is further defended by Li et al, who reveal that stock markets are sensitive to public sentiment relating to online blogs and news postings (Li et al., 2014). These articles reveal the effect MS can have on financial markets and how public sentiment can result in changes to those values. Another area of interest for public sentiment is marketing products to consumers. One study, conducted by Gaikar and Marakarkandy, attempted to determine the effect of valance and volume in MS reviews on intention to purchase a product. Results indicated that “valence of the review, attitude towards the product and prior product knowledge was found to significantly predict purchase intention of the product” (Gaikar & Marakarkandy, 2015). Even purchasing an experience MS data can be applied to understand marketing campaign impacts. This is shown in the study by Philander and Zhong, where multiple presence of resorts on MS were compared to show the effect of promotions and campaigns showing positive associations (Philander & Zhong, 2016).

3. RESEARCH QUESTIONS

With the goal of attaining an understanding of public sentiment for smart lockers, we conducted an evaluation through the use of the microblogging site (MS) known as Twitter. Socialization involving smart lockers was applied for this analysis. For the purposes of this analysis socialization was defined as the conversation occurring on the MS, this was quantified by the number of retweets, favorites, mentions, and hashtags relating to smart lockers. Data based on this definition allowed us to obtain an understanding of the conversations occurring on the MS and the number of potential participants in those conversations. To accomplish our goal of obtaining an understanding of socialization related to smart lockers through MS we aimed to look at the data historically and in real-time, with the additional application of sentiment analysis for tweet content, to analyze the following hypotheses.

- **H1**: Tweets (postings) related to smart lockers have increased in socialization over time (count, retweets, favorites, mention use, and hashtag use).
- **H2**: Tweets (postings) with positive sentiment related to smart lockers have increased in socialization over time (retweets, favorites, mention use, and hashtag use).
- **H3**: Tweets (postings) related to smart lockers have greater socialization (count, retweets, favorites, mention use, and hashtag use) in different continents.
4. SAMPLING

First it was necessary to choose the terms that would be applied to query the tweets. For the purposes of this study we queried for hashtags “#amazonhub”, “#amazonhublocker”, “#amazonlocker”, “#deliverylocker”, “#packagelocker”, and “#smartlocker”. The hashtags “#smarthub”, “#packagehub”, and “#deliveryhub” were purposefully excluded to avoid gathering tweets related to package sorting hubs or computer packages instead of smart lockers for package pickup.

For each term it was then necessary to access the historical data through Twitter using the GetOldTweets python package (Henrique, n.d.). This package uses queries to search Twitter through word search, username, hashtags, etc. In our case we used the query search and returned all tweets for each term for each month from June 2018 to November 2019. The given time period was chosen since smart lockers reached over 70 cities in June 2018 (Holsenbeck, 2018). For each term we searched for the full hashtag above in the text of the tweets, no spelling changes or abbreviations were applied to avoid potential bias.

More recent data containing location information was attained through the use of the Twitter API and rtweet package (Kearney et al., 2020). Since the Twitter API limits the historical period that can be queried, tweets from December 2nd, 2019 to January 6th, 2020 were collected, resulting in 1,616 tweets. This time period was intentionally chosen to include Cyber Monday and the following days since this time period marks an increase in online shopping, and therefore package deliveries, for the holiday season. Additionally, it includes twelve days after Christmas to allow for returns and late package delivery. This makes it more likely tweets relating to delivery and smart lockers will occur. We searched for the full hashtag above with no abbreviations.

Tweets with locations were also attained through the Twitter API. Since the Twitter API limits the historical time period that can be queried we were able to attain 143 unique tweets for the above time period. These tweets were analyzed by hand to determine location based on the location and language of the tweet text, where location is considered continent (Europe, North America, Asia, etc.). Since a majority of the tweets were found to be neutral and informative, sentiment was dropped.

5. METHODOLOGY

Raw tweets from Twitter were obtained historically for all tweets mentioning each hashtag and sentiment analysis was performed. This sentiment analysis was done through the use of the TextBlob python package (Loria et al., 2020). Output from the analysis was applied to categorize the text of the tweet as “Positive”, “Neutral”, or “Negative” based on natural language processing. Once this was accomplished the data was aggregated for analysis.

5.1 Statistical Analysis of Historical Twitter Data

In order to examine whether the data elements changed over time and test our first two hypotheses we initially applied a time series decomposition with a seasonality of six months. This time period was chosen since we did not have the minimum of two full years of data to make up the required two period for decomposition. Additionally, we fit a regression model to the data, where time was the predictor and the results from each term were considered separate observations. Once this was completed an F-Test was performed to determine if change over time was significant.

5.2 Statistical Analysis of Geographical Twitter Data

In testing our third hypothesis, we examined whether the location based data was normally distributed, we employed the Shapiro-Wilke Test and tests of skewness and kurtosis. Since the results of these tests indicated that none of the variables had strong probability of being normally distributed, we applied the non-parametric Mann Whitney U (Wilcoxon Ranked Sum) Test in order to make comparisons between locations. Finally, we also created a logistic regression model to determine which factors had the most significant relationship to location.
6. VARIABLE DEFINITION AND RESULTS

Data was aggregated to include aggregated variables with sentiment labels from historical Twitter data and geographical Twitter data. Here we present a breakdown of results for historical and geographical based Twitter data analysis. The importance of these results is discussed later in the Discussion section.

6.1 Time Series Decomposition

As part of evaluating trend over the given time period for our hypotheses an additive time series decomposition was completed on the historical Twitter attributes. For this decomposition a seasonality of tweet counts over six months was chosen to ensure a minimum of two periods would be available to provide the trend breakdown. Results show that for a majority of attributes (count of retweets, favorites, etc.) we see a spike and then a decrease, this can be seen in Figure 1. Interestingly, the trend for neutral related behaviors was shown to increase in the final months based on the decomposition in Figure 2.

![Figure 1. Time Series Decomposition for Number of Tweets per Month (shows time series similar to other attributes (retweets, favorites, etc.))](image1.png)

![Figure 2. Time Series Decomposition for Number of Neutral Retweets per Month](image2.png)
6.2 Linear Model over Time

To confirm if the observed trends should be considered statistically significant and officially prove or disprove our first two hypotheses, we fit a regression model to the data. As can be seen in Figure 3, all attributes (count of retweets, favorites, etc.) were summed for the month and month was used as the predictor variable. This means for each attribute (count of retweets, favorites, etc.) 17 data points were used to fit a linear model to the data.

F-Tests to determine whether or not \( \beta_1=0 \) for each regression line were then performed. Only the slope for Neutral Hashtags regression line showed \( \beta_1 \neq 0 \) at the \( \alpha=0.1 \) significance level, with a \( \beta_1 \) value of 12.663. The F-score of 3.086 was compared to the F-statistic of 3.048 to attain this conclusion.

6.3 Tests of Normality

The four independent variables counting retweets, mentions, favorites, and hashtags from geographical Twitter data were tested for normality. The results showed that none of the variables had a normal distribution, all were skewed to the right. This can be further illustrated in Figure 4, which shows a histogram of the values for number of hashtags in each tweet. Normality results can be found in Table 1.
Table 1. Shapiro-Wilke, Skewness, and Kurtosis Test Results

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Shapiro Wilke Test (P-Value)</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retweet Count</td>
<td>&lt;2.2e-16</td>
<td>4.465</td>
<td>24.839</td>
<td>Non-Normal</td>
</tr>
<tr>
<td>Favorite Count</td>
<td>&lt;2.2e-16</td>
<td>4.130</td>
<td>20.272</td>
<td>Non-Normal</td>
</tr>
<tr>
<td>Mention Count</td>
<td>&lt;2.2e-16</td>
<td>4.693</td>
<td>28.329</td>
<td>Non-Normal</td>
</tr>
<tr>
<td>Hashtag Count</td>
<td>&lt;2.2e-15</td>
<td>1.966</td>
<td>6.646</td>
<td>Non-Normal</td>
</tr>
</tbody>
</table>

6.4 Univariate Analysis

In our third hypothesis, for the purposes of comparison we consider tweets created in different continents. Only results from Asia compares to tweets created outside of Asia were found to be significant. According to the Mann Whitney U (Wilcoxon Ranked Sum) Test two attributes showed a significant difference between the populations at the $\alpha=0.01$ significance level. It is found that favorite count is greater and mention count is smaller for each tweet in Asia with p-values of 5.76e-4 and 9.065e-4, respectively. All univariate analysis results are in Table 2.

Table 2. Univariate Analysis

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Mean in Asia</th>
<th>Mean Outside Asia</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retweet Count</td>
<td>0.388</td>
<td>0.270</td>
<td>0.962</td>
</tr>
<tr>
<td>Favorite Count</td>
<td>2.24</td>
<td>0.54</td>
<td>5.76e-4 (Significant)</td>
</tr>
<tr>
<td>Mention Count</td>
<td>0.025</td>
<td>0.302</td>
<td>9.065e-4 (Significant)</td>
</tr>
<tr>
<td>Hashtag Count</td>
<td>2.91</td>
<td>3.76</td>
<td>0.391</td>
</tr>
</tbody>
</table>

6.5 Logistic Regression Results

In order to complete logistic regression and prove or disprove our third hypothesis it was first necessary to gain an understanding of the variable distribution and covariance. During the course of this evaluation it was necessary to consider this for the geographical Twitter data. A correlation between the number of retweets and the number of favorites was seen in the results for Asia vs. other continents. However, this makes sense when considered that many Twitter users will favorite and then retweet a tweet so their followers can see it. After controlling for covariant variables that are correlated the data was run through a series of 15 logistic regression models to determine which attributes could be considered significant. These results did show that hashtag and mention count is a significant variable in determining location when comparing tweets created in Asia vs. outside of Asia. Of the 15 models created, 6 of the models identified mention count as significant and 4 of the models identified hashtag count as significant, both at the $\alpha=0.01$ level, with p-values ranging from 0.003 to 0.009. This showed that for every mention and hashtag count decrease there are greater odds that the tweet was created in Asia. Results for 1 the 6 models can be found in Tables 3 as an example.

Table 3. Logistic Regression Analysis. (Significant Model Example)

<table>
<thead>
<tr>
<th>Attribute Included</th>
<th>P-Value</th>
<th>Adjusted Odds Ratio</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retweet Count</td>
<td>0.483</td>
<td>0.75</td>
<td>-0.293</td>
</tr>
<tr>
<td>Favorite Count</td>
<td>0.044</td>
<td>1.53</td>
<td>0.425</td>
</tr>
<tr>
<td>Hashtag Count</td>
<td>0.009 (Significant)</td>
<td>0.77</td>
<td>-0.259</td>
</tr>
<tr>
<td>Mention Count</td>
<td>0.003 (Significant)</td>
<td>0.09</td>
<td>-2.354</td>
</tr>
</tbody>
</table>
7. DISCUSSION

In the linear analysis over time performed on historical Twitter data our findings indicated that hashtags in neutral tweets relating to smart lockers had increased. This lends some weight to our first hypothesis that socialization amongst tweets related to this package storage option have increased. However, it should be noted that hashtags are often used as an attempt to increase the socialization of a tweet, similar to mentions. Retweets and favorites show socialization between individuals since they indicate that others are actually viewing the tweet. Therefore, the results of the time series and linear analysis provide limited support for our first hypothesis. Unfortunately, no support is provided for our second hypothesis since no significant change over time were found for positive tweets.

For our third hypothesis relating to socialization in countries in Asia there is support. Favorites and mentions are shown to increase in Asia through the Mann Whitney U Test, indicating that more individuals are directing tweets to others and reading tweets related to smart lockers in Asia. Retweets not showing a substantial change could be the result of individuals not wanting to share tweets to their followers, since tweets relating to this topic could differ drastically from the theme of a user’s Twitter account and result in decreased follower count. It is interesting that in the linear regression models mentions and hashtags appear as significant at the $\alpha=0.01$ level, while favorites do not. However, favorites do show significance at the $\alpha=0.05$ level. This makes sense since it shows a connection between attempts to socialize a tweet (mentions and hashtags) and actual socialization of that tweet (favorites and retweets).

Additionally, it could reasonable that socialization of smart locker tweets would increase in Asia, since this correlates to population density. Currently Asia has the highest population density, with 257.8 people per square mile as of May 2019. This is significantly higher than the other continents, with the next runner up being Europe with 187.7 people per square mile (Rosenberg, 2019). Individuals living in densely populated areas would have more exposure and interest in smart lockers, since smart lockers offer an easily accessed pickup point with security measures in place. Additionally, there are more individuals within the given area that would participate in the conversation. Therefore, both attempts and actual socialization of tweets related to smart lockers would increase for the region.

8. LIMITATIONS

There is a lack of literature related to this topic, which limits the available information and best practices to be applied to the data. Previous research would allow for comparison of results and additional insights into the change over time and in continents. However, this also allows us to fill in a gap in the literature and provide information for future studies.

The main limitation of this study is the sample size. Due to the limits on querying historical data through the Twitter API there were limited results with location data available. Even for those available, location data had to be understood and categorized for continents by hand. However, this is a typical limitation for studies examining Twitter data. Since this data is based entirely on public conversations and topics that will attain the most socialization, the number of tweets is subject to popular discussions at the time.

9. CONCLUSION

Through this research we have provided a review of Twitter data related to smart lockers. Historically we reviewed data from June 2018 to November 2019, which did show a small increase in socialization overall providing some support to our first hypothesis. Unfortunately, the results did not support our second hypothesis that tweets identified to have positive sentiment would increase in socialization over time. Location based data was also considered for December 2nd, 2019 to January 6th, 2020, revealing an increase in socialization in the densely populated continent of Asia. This supported our third hypothesis that socialization would increase in this region.
Future studies in this area should continue to focus on the concept of popularity in regions. Increased socialization in densely populated regions is reasonable since those individuals would have more use for smart lockers, but that will require further investigation to prove. Additional applications for this research is in delivery optimization models, since knowledge of where smart lockers would be best utilized relies on individual sentiment towards the lockers and their use. Alternatively, marketing campaigns could apply the data to focus on areas with low sentiment and increase smart locker use. Either application could then be applied with the goal of saving delivery resources through the increased use of smart lockers.

REFERENCES


