GRAB C.H.A: A Web-based Interactive Application Exploring and Analyzing User and Customer Perceptions

Aaron Poh Weixin, Huang Huanyuan, Siow Chun Yen Jodie

Abstract— Social media has become intergral in the lives of everyone – young or old, consumers or business users. These days, brands can actively engage users by leveraging on social media platforms to strengthen their relationships with their customers, making them feel valued as a user. One practical way of doing this is for brands to practice social listening, to become more atuned to the demands and needs of customers to make better business decisions in the long run. In this study, we discuss how a company such as Grab can use text analysis to better understand and engage their end users.

1 Introduction

Following Grab's acquisition of Uber's business earlier this year, Grab is currently a monopoly in the ride-sharing market in Singapore, being the largest ride-hailing company here in the little red dot.

Despite its dominant presence, there still remains competition among the various services in the ride-hailing industry. Hence, despite the recent successful buy-out, it is imperative for Grab not to be complacent, but to continually strive to provide quality service for their users. Being proactive towards customers' feedback will help in ensuring that services provided are satisfactory, which will contribute towards healthy and sustainable bottom lines.

Consumers of Grab's services have been vocal in voicing their customer experiences on different platforms such as Twitter, Instagram and Reddit over the past years. Grab should leverage on such readily available data to better understand their customers and work towards improving their services based on relevant and constructive feedback.

2 Motivations and Objectives

Currently, Grab has been trying to develop a model to identify key topics of discussion amongst their customers as well as customers' sentiments regarding these topics. This will allow Grab to address inadequacies in their business practices, to better target their customers and to build a stronger brand image. Additionally, it can also help Grab to identify possible trends in the market to make better future business decisions to retain their market leadership.

Consequently, the objective of our visualisation is to help bridge the gap between analytics and Grab's business team, whereby business users are able to take control of the topic-modelling process to analyse online comments without requiring prior technical or coding knowledge. Specifically, our application allows Grab to better understand their customers and market trends using Latent Dirichlet Allocation (LDA) topic modelling.

Our web application aims to use topic modelling to provide Grab with the following capabilities:

1. To provide users with the freedom to upload their own csv files into the application to generate topic clusters for topic insights on a monthly basis
2. To allow business users to use the Topic Score Plot as a complementary tool with their existing business knowledge to determine ideal number of topics for text analysis
3. To allow users to visualise overall and individual topic insights as well as to deep dive into individual comments

4. To visualise the changes in topic salience amongst consumers over time

3 Related Works

Topic modelling with LDA has been commonly explored using different software programs. Due to the statistical capabilities of R as well as how we want to develop a web application, we decided to use the R environment – specifically R Studio and R Shiny to develop our web application. As such, we also referred to previous LDA projects done in the R environment.

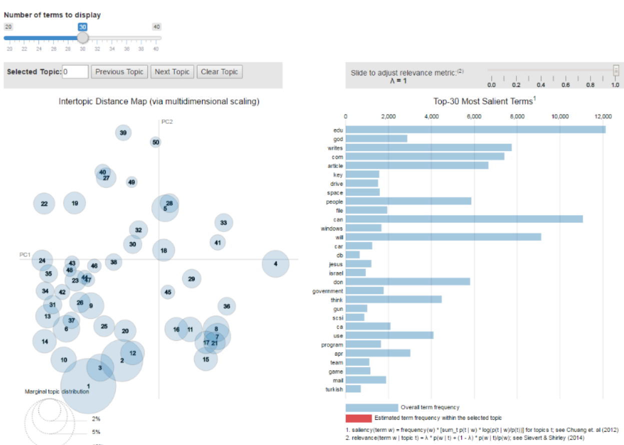

Fig 1. LDAvis in Shiny[1]

From this visualization done with LDAvis in R Shiny, it can be seen that LDA modelling allows users to sieve out terms that are relevant according to a relevance metric before clustering distinct terms into individual topics clusters for analysis. This allows users to get an overview of topic clusters in terms of frequency of terms in a cluster, frequency of terms relative to all terms in the database as well as how individual topics are related to other topics in the database.

R Shiny is commonly used in creating web applications. For LDA projects, bar graphs are commonly used to visualise top terms in individual clusters. This helps users to be able to quickly glance and identify the significant terms in each cluster. The use of bar graphs and colors increases the clarity of the visualisation.

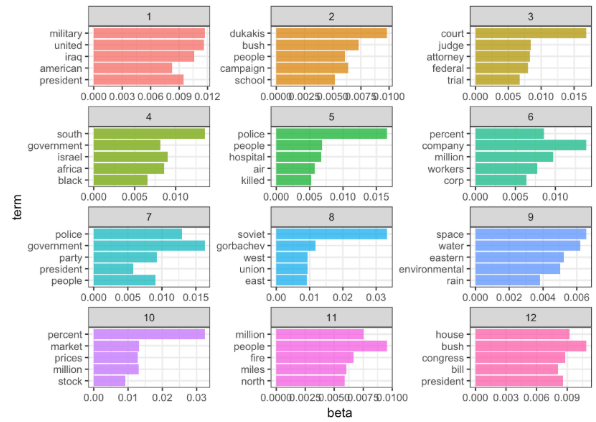

Figure 2. Top Topic Terms visualised using R Shiny[2]

4 User Interface Design

Our web application consists of two tabs – a 'Pre-Processing' tab to process and update data for the preparation and generation of LDA models as well as a 'Topic Output' tab which allows users to delve into the insights of individual topics and comments. On top of that, there is also a user guide tab explaining to new users how the application works. Our user interface is guided by the principal of being minimalistic while providing usability [3]. This is done by adopting a simple colour scheme, using 4 main colours of green, grey, red and white, as well as removing all extra elements that do not add much value to the application [4].

Upon entering the web application, the first tab that users will see is the 'Pre-Processing' tab, as seen below.

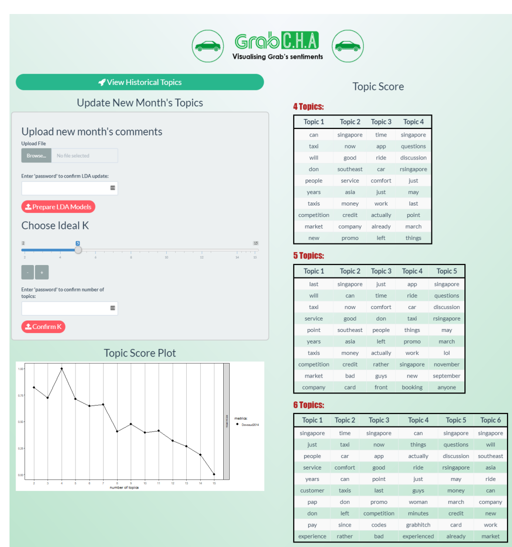

Figure 3: Pre-Processing Tab

Users will be given two options in the Pre-Processing tab – to choose to visualise past topic insights that have already been visualised, or to upload new data into the application so as to prepare LDA models to generate new topic insights with the updated data.

To visualise past topic insights, users have to click on ‘View Historical Topics’ to be redirected to the Topic Output tab.

When it comes to uploading new monthly data, users will have to add csv files from their computer systems into the application before LDA models can be prepared. During the preparation of the LDA models, a loading screen will appear and users will be able to see the system status progress in the generation of the LDA models, following user interface design best practices [5].

After the LDA models are prepared, the user has to select the ideal number of topic clusters (K) to be analysed. The top 10 terms of K-1, K and K+1 will be displayed on the right hand side of the visualisation (under ‘Topic Score’) as users toggle between K options using the ‘Choose Ideal K’ slider. The ‘Topic Score’ visualisation serves as a guide in helping users to choose the optimal number of topic clusters they wish to analyse based on the overview of topics.

Additionally, the ‘Topic Score Plot’ will also help business users in choosing the optimal number of topics to analyse. The ‘Topic Score Plot’ guides and recommends users to choose number of topic clusters (K) based on the Deveaud 2014 [6] scoring metric where the K with the highest score is preferred. At the same time, users are also able to use this in tandem with their existing business knowledge to make more informed decisions about the exact topic numbers to choose. For instance, if the business user is looking for more aggregated data, he can consider choosing 4 topics for analysis (the first peak in the graph), or he can choose between 5 and 7 topics for a more granular analysis of the data.

Understanding that confirmation dialogs can aid in reducing user errors [7], we have implemented safety measures in our application. Users are required to enter passwords before preparing new LDA models and before they select the number of topic clusters for analysis. On top of that, users are also prompted to confirm their decisions on the preparation of LDA models and on the number of topics for topic modelling before the application redirects them to the ‘Topic Output’ tab.

Next, our second tab is the ‘Topic Output’ tab as seen below.

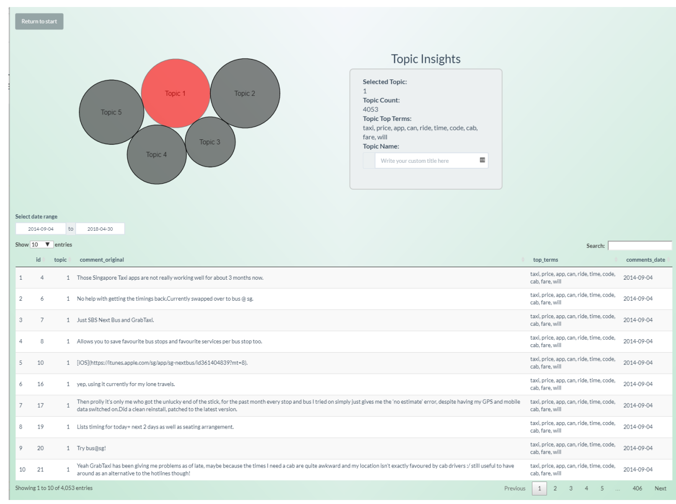

Figure 4: Topic Output Tab

In the ‘Topic Output’ tab, users will be able to see an overview of the topic models, where the size of the bubbles are representative of the relative size of a topic to the other topics selected. A tooltip containing more information about a topic such as the topic number and topic frequency will also appear when users hover over the topic bubbles.

Users can also reference ‘Topic Insights’ on the right side of the second visualisation to explore more in-depth information such as the top ten terms in a selected topic. Additionally, users are also given the interactivity function of entering topic names based on the top key words and individual comments in a topic cluster (located at the bottom of the tab). For instance, upon choosing and saving a topic name as ‘Ride Complaints’, the application is able to map future comments tagged to ‘Topic 1’ and recognise these new comments as ‘Ride Complaints’ as well.

Other interactivity functions included in this tab is the ability for users to select the range of dates for comments to be visualised, to see how salient topics may have changed over time. For instance, if topics analysed this month are largely similar to that of the previous month, it will be reflected by an increase in the topic count on the overall data, showing the growth and change in topic salience. On the other hand, if data visualised this month differs from that of the previous month, new topic bubbles will be added on the visualisation. This will be further explained in the data preparation section.

There is also a search function where users are able to filter out comments through searching for specific keywords. This will allow them to be able to sieve out individual comments to further analyse and understand what users are discussing about during that time period.

5 APPROACH

5.1 System Architecture

The backend of our LDA modelling is statistical in nature, where probabilities are used to determine the probability of a keyword belonging in a topic as well as the probability of a comment being in a topic. Probabilistic thresholds were also fixed on the backend to determine the threshold in which a keyword or a comment will belong to a particular topic. With such considerations in mind as well as wanting to create a web application, we have decided to design and develop our application in the R environment from end-to-end using R Studio and R Shiny.

5.2 Data Preparation

Before using LDA for topic modelling, we have to clean the raw data that was provided from Grab. The raw data consisted of approximately 43,000 comments scraped from different social media platforms such as Instagram, Reddit, Twitter and Google Playstore.

The data is first cleaned using the removal of irrelevant text such as punctuations and pronouns, before a list of stop words are removed from the comments. These stop words are determined by Grab, which include words such as go, also and the brand name of Grab etcetera. This will help to clean individual comments before remaining keywords will be modelled and clustered into individual topics. The comments are clustered into topics based on term-topic probabilities which is pre-determined back end.

Assuming that there are already previously loaded Topic Output data in the application, the new monthly uploaded data will go through the same cleaning process before the posterior method is applied to the new data. The posterior method will group new comments into the previous topics based on the term-topic probabilities if it exceeds a pre-determined probability threshold. Otherwise, a new LDA model will be used to cluster the remaining keywords into new topics to be analysed.

With this method, if keywords are repeated over time periods, we can expect to see an increase in topic count in a particular topic. This suggests that the topic is still a popular topic that is widely discussed by customers in the current time period, allowing Grab business users to visualise the salience in topics discussed over time.

6 FUTURE WORK

Our current application allows business users to visualise and understand customer comments based on text analysis using LDA modelling. To further improve on our application, we hope to incorporate sentiment scoring and sentiment analysis in our web application. This will allow business users to understand the overall sentiments of clusters in general as well as to understand the sentiment of individual comments collected from the various social media platforms. Eventually, this will allow Grab to be able to integrate topic sentiments on top of the topic modelling function on the application, providing more value-added services for business users.

ACKNOWLEDGMENTS

The authors wish to thank Professor Kam Tin Seong for his invaluable guidance and support.

REFERENCES

[1] Computing for the Social Sciences (2016). Text analysis: topic modelling. Retrieved November 15, 2018 from https://cfss.uchicago.edu/fall2016/text02.html

[2] StackOverflow (2016, Sep). Integrating LDAvis with Shiny. Retrieved November 15, 2018 from https://stackoverflow.com/questions/39237927/integrating-ldavis-with-shiny

[3] Yalanska, M. (n.d.). Lean and Mean: Power of Minimalism in UI Design. Retrieved November 15, 2018 from https://tubikstudio.com/lean-and-mean-power-of-minimalism-in-ui-design/

[4] Babich, N. (2016, Dec 29). Best Practices for Minimalist Design. Retrieved 22 Nov, 2018, from https://uxplanet.org/best-practices-for-minimalist-design-7af4a9b61ad7

[5] Wong, E. (Nov, 2018) User Interface Design Guidelines: 10 Rules of Thumb. Retrieved November 15, 2018 from https://www.interaction-design.org/literature/article/user-interface-design-guidelines-10-rules-of-thumb

[6] Deveaud, R., Sanjuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. Document Numérique, 17(1), 61-84. doi:10.3166/dn.17.1.61-84

[7] Nielsen, J. (2018, Feb 18). Confirmation Dialogs Can Prevent User Errors – If Not Overused. Retrieved November 15, 2018, from https://www.nngroup.com/articles/confirmation-dialog/