2021

A Complete Text-Processing Pipeline for Business Performance Tracking

Minh Ngoc Dinh
RMIT University, Vietnam, minh.dinh4@rmit.edu.vn

Khanh Dinh Dang
RMIT University, Vietnam, s3618748@rmit.edu.vn

A Complete Text-Processing Pipeline for Business Performance Tracking

Full research paper

Minh Ngoc Dinh
School of Science, Engineering and Technology
RMIT University
Ho Chi Minh City, Vietnam
Email: minh.dinh4@rmit.edu.vn

Khanh Dinh Dang
School of Science, Engineering and Technology
RMIT University
Ho Chi Minh City, Vietnam
Email: s3618748@rmit.edu.vn

Abstract

Natural text processing is amongst the most researched domains because of its varied applications. However, most existing works focus on improving the performance of machine learning models instead of applying those models in practical business cases. We present a text processing pipeline that enables business users to identify business performance factors through sentiment analysis and opinion summarization of customer feedback. The pipeline performs fine-grained sentiment classification of customer comments, and the results are used for sentiment trend tracking process. The pipeline also performs topic modelling in which key aspects of customer comments are clustered using their co-relation scores. The results are used to produce abstractive opinion summarization. The proposed text processing pipeline is evaluated using two business cases in the food and retail domains. The performance of the sentiment analysis component is measured using mean absolute error (MAE) rate, root mean squared error (RMSE) rate, and coefficient of determination.

Keywords: Social listening, Machine learning, Sentiment analysis, Text summarization

1 Introduction

Nowadays, the more the business understands their customers, the more they will succeed. When a company or a brand knows what their customers want and think about their products or services, they can devise marketing and business strategies to boost their appearance and relevance in the market. Understanding customers has been done using methods such as surveying (paper-based and web-based), market research, and promotion campaigns. However, these methods are often time-consuming and can be an error-prone process when they are conducted manually. Survey data collected using these methods could be fake, or at the least, bias. But most importantly, these methods cannot scale and are likely to create delays in the subsequent market analysis process.

We take a famous case study, where a business failed to adapt to customers' opinions, as the motivation for this work. In the years leading up to 1985, Coca Cola's market shares suffered a significant loss due to the increase in demand for diet soft drinks and non-cola beverages (Great Ideas for Teaching Marketing 2020). To react, Coca-Cola introduced the new product "New Coke" while removing their flagship product from the market. However, taste aside, many consumers saw Coke as a cultural icon and were disappointed with its disappearance from the supermarket shelves. As a result, only 2.5 months later, Coca-Cola re-launched the original Coke formula, named as Coke Classic. This case study shows that misperceiving customers' opinion and their sentiment leads to harmful business strategies.

The rapid growth in portable device usage (smartphones, tablets, and laptops) has provided businesses with new interfaces to connect to their customers via websites and social media platforms. According to Esteban Ortiz-Ospina of https://ourworldindata.org/, there are 2.4 billion Facebook users, while other social media platforms including YouTube and WhatsApp also have more than one billion users each (Ortiz-Ospina 2019). As a result, a large number of reviews, comments, opinions, and textual feedback are generated continuously, presenting an endless source of business intelligence. Nevertheless, collecting customers' feedback and analyzing their sentiment is only part of the story. Companies need to do so in real-time in order to make informed decisions. This calls for text-processing system that can collect and process online reviews to identify which components (e.g., aspects) in those reviews that compose a given sentiment. The work in this paper employs the design science research approach and develops a text processing pipeline to analyze large corpus of text so customers' feedback and opinions can be quickly addressed to improve the business's performance. Our main contributions are:

- Exploring and evaluating a range of NLP models for analyzing Vietnamese text including:
  - A sentiment analysis engine to rate Vietnamese comments.
  - A topic model to extract key aspects from comments using their co-relation scores.
  - A text summarization model to summarize thousands of comments.
- Two case studies to demonstrate how key insights of business performance can be identified.

In the rest of the paper, section two presents the background and relevant work in several areas including sentiment analysis, topic modelling and text summarization. Section three proposes our text processing pipeline and its components including data collection and pre-processing, sentiment analysis, sentiment trend tracking, topic modelling, and opinion summarization. In section four, we present two use cases, one from McDonald and the other one with a product from the Vietnamese e-commerce platform Tiki[1]. We measure the accuracy rate of our sentiment analysis component and demonstrate benefits that the proposed text processing pipeline brings to business users. We conclude the paper in section five.

2 Background

2.1 Social Listening – A Market Review

In 2012, opinions about the Super Bowl XLVI were analyzed using the data gathered from social media platforms such as Twitter, Facebook, and blogs (Chumwatana and Chuaychoo 2017). In 2015, another method was proposed by Tse-Chuan Hsu to evaluate opinion research for the social network brand community by using social analysis. There are social analysis tools in the market including Mention and Awario for the international market (Baker 2020) and Dazikzak[2] for the Vietnamese market . These tools have their highlight features and some drawbacks. For example, Mention and Awario allow users to keep track of what the public has been saying through different forums (e.g., Amazon, Booking.com, Yelp,

[1] https://tiki.vn
[2] https://dazikzak.com

and Trip Advisor) and different social media platforms (e.g., Facebook, YouTube, and Twitter). They both focus on the statistics of factors such as the number of keywords mentioned based on location or timeline. However, these tools fail to deliver sentimental insights from the collected comments.

Social analysis arrived in Vietnam in 2015, but the up-taking has been quite low (only 1% of Vietnamese business explored social analysis solutions in 2018 (Business Woman Magazine 2020). In this market, the most popular tool, Dazikzak, provides different features for capturing feedback from different social media platforms. However, Dazikzak does not take the contents created through these activities such as comments, reviews and feedback as a data source which might omit many other key business performance insights. Reviewing existing commercial tools suggests several core features required of a social listening tool. They are listed below. However, these tools all lack functions for sentiment analysis.

- Keeping track of keywords mentions on social media.
- Comparison between users' business performances and competitors.
- Providing statistical reports based on collected data.

2.2 Sentiment Analysis

A sentiment analysis task classifies polarity of emotion (e.g., positive, negative, neutral) from texts such as user feedback and comments. Many NLP works are in this domain. Pang et al. survey different machine learning algorithms such as Naive Bayes, maximum entropy classification, and support vector machines for the sentiment classification problem given movie reviews (Pang, Lee, and Vaithyanathan 2002). Phan et al. present a method for analyzing the fuzzy sentiment phrases (FSPs) such as "relatively good", "not too delicious" and "not so bad", in Tweeter's tweets (Phan et al. 2019). The method achieves the recall rate of 77% and the precision rate of 75% for the detection step, while in the sentiment analysis step, the recall rate is 72% and precision is 71%. Naz et al. focus on experimenting with different combinations of N-gram and weighting schemes to find the best combination in terms of accuracy (Naz, Sharan, and Malik 2018). The authors also introduce the sentiment score vector of tweets as an external feature and analyzed its effects on SVM classifier performance. Regarding Vietnamese sentiment analysis, Kieu and Pham present a rule-based system which achieved F-measures of around 89% (Kieu and Pham 2010). Duyen et al. present an empirical study on Vietnamese sentiment analysis, where three machine learning (ML) algorithms including Support Vector Machines, Maximum Entropy Models, Naive Bayes, were studied with highest result of 76.8% accuracy (Duyen, Bach, and Phuong 2014). Overall, we found most recent research work remain at classifying sentiment polarity instead of achieving fine-grained sentiment score. Furthermore, they lack an evaluation on business applications.

2.3 Topic Modelling

A topic model utilizes statistical tools to identify abstract themes and text patterns, which are often presented as clusters of similar/recurring words, in a corpus of text (Blei 2012). By using topic models, businesses can extract the general consensus from a wealth of user textual data. Classical topic modelling techniques include Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA). While LDA assumes that each aspect of a domain topic will have various words related to it, and LDA techniques discover topics using the distribution of topics in documents over the distribution of words in the topic (Yue, Stuart, and Qiong 2017; Jelodar et al. 2019; Kang, Kim, and Kang 2019), LSA examines how words that are closer in meaning will occur in similar texts. As a result, by clustering texts using their semantic meaning, topics can be identified without the topic words present (Rehder et al. 1998; Kim, Park, and Lee 2020; Hasan, Sanyal, and Chaki 2018). On the neural network-based approaches, Attention-based Aspect Extraction uses word embeddings, which encodes word co-occurrence statistics, and attention mechanisms to identify significant words (He et al. 2017).

2.4 Text Summarization

Text summarization is an NLP domain that aims at creating summaries from multiple documents, for example, product reviews, or user feedback. Existing text summarization techniques are either extractive or abstractive. An extractive-based technique extracts important key terms, sentences, paragraphs from the original texts, with selection criteria are based on statistical and linguistic features of the sentences and concatenating them into shorter form. On the other hand, an abstractive summarization technique attempts to grasp the context of the original texts so that a generalized summary, which conveys information in a precise way, can be produced. Abstractive summarization is often preferred because it produces a text reflecting consensus opinion, especially when there are many texts (e.g., product reviews) involved in the process. As a result, we choose to develop an abstractive summarization solution in this work. We present some important works in this domain below.

Abstractive summarization can be generated using prior knowledge of the target text collection. Barzilay et al. identifies the common phrases across multiple sentences by using bottom-up local multi-sequence alignment (Barzilay and McKeown 2005). Similarly, Tanaka et al. developed a sentence fusion technique which inserts/substitutes common phrases to generate a summary of news broadcast through sentence revision process (Tanaka et al. 2009). There are also rule-based methods in which data extraction rules are used to select the most effective candidate aspects. For example, Genest et al. (Genest and Lapalme 2012) use predefined information-extraction rules to extract semantically related nouns and verbs, while Le et al. (Le and Le 2013) attempt abstractive text summarization using discourse rules, syntactical constraints and word graph. Recently, unsupervised modelling has become popular in summarization, especially because modern deep learning methods rely on large amounts of annotated data. Brazinskas et al. develop a generative model with a hierarchical variational autoencoder (VAE) model to capitalize on intuition and produce a text reflecting consensus opinions (Brazinskas, Lapata, and Titov 2019).

3 The Text Processing Pipeline

We develop a modular pipeline (Figure 1) so that different implementations of specific components can be swapped and tested flexibly. It also makes our text processing pipeline more scalable and robust for future implementations. In this pipeline, after data is collected by scraper bots (the implementation is not within the scope of this work), we proceed with pre-processing and analysis. Each step in the pipeline is presented below. Output from one step is input for the next steps according to Figure 1.

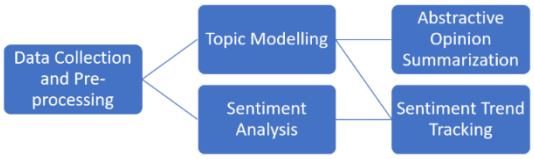

Figure 1: Text processing pipeline for business performance tracking

3.1 Data Pre-processing

Customers' comments may contain lots of different types of data such as words, numbers, emoji and punctuations which will make the accuracy of the machine learning model decreases if they are used for training directly. To clean the data, we first convert all characters to lower case and remove all punctuations from the text. We also remove all stop words such as 'và' (and), 'này' (this) etc, using a Vietnamese stop words list. We then apply text tokenizing process in order to split a sequence of strings into pieces such as words, terms, or phrases. Below are some example of pre-processing input/output. From this point onwards, for readability, we only include example user's comment and text in English.

- Input: *"Quán này sạch, đẹp, và nấu ngon quá." (The shop was clean, nice, and the food was tasty)*
- Output: *"quán sạch đẹp nấu ngon quá" (clean nice shop, tasty food)*

	MAE	**RMSE**
Count vectorization	2.45	3.12
N-gram (bigram) vectorization	2.11	3.05
TF-IDF	1.86	2.36

Table 1. Vectorization methods comparison (using the random forest regression model)

The cleaned text then goes through a process to convert non-numerical value into numerical features which can be fed into a machine-learning algorithm. In other words, we quantify the comments, a task that can be done using methods such as count vectorization, n-gram vectorizing and Term Frequency-Inverse document frequency (TF-IDF) (Brandão and Calixto 2019). If count vectorization only returns the word counts, N-gram vectorization generates a document-term matrix to present all combinations of adjacent terms for a given length. However, Term Frequency-Inverse Document Frequency (TF-IDF) uses the word frequency to determine which (combination of) words are important in the document (Brandão and Calixto 2019). The terms with high TF-IDF values are terms that appear in this document a lot, and a few in other documents. This helps flush out generic terms and hold terms of high importance (that document's keywords). Using the random forest regression model with each vectorization method to analyse the Vietnamese dataset mentioned above, we compare the performance of three different vectorization techniques. We use scale-dependent, error rate metrics including mean absolute error (MAE) and root mean squared error (RMSE) to measure the performance (Botchkarev 2019). Table 1 shows TF-IDF achieves better performance (lower error rate) and therefore was chosen for our pipeline.

3.2 Machine Learning Architecture for Sentiment Analysis

We use the cleaned data to train machine learning model to predict the rate of (or to rate) comments. While classification is a popular approach to map an input sentence to some pre-defined categories such as positive, negative and neutral, our pipeline performs deeper analysis using a regression model to give each comment a rating value in the scale from 0 to 10. We chose regression method because it enables quick retrain of the working model when given new labelled data. We collect 5,912 comments, labelled from 0-10, from Foody[3], to train our sentiment analysis model. The dataset is split into a training set and a testing set by using the K Cross-validation method, with K = 5. We trial a range of regression techniques including Random Forest Regression, Bayesian Ridge Regression, ElasticNet Regression, and Support Vector Regression (SVR) with either the polynomial kernel, or the radial basis function (RBF) kernel (Pedregosa et al. 2011). To further enhance the accuracy, we also apply ensemble learning with a Voting Regressor to combine some of the well-performed models (Pedregosa et al. 2011).

The performance of models in the ensemble is assessed using different metrics mentioned above including MAE, RMSE. We also measure the coefficient of determination value (R^2), which is a standard statistical metric to evaluate regression problems (Botchkarev 2019). Table 2 shows performance results for various regression techniques. Our tuning process helps us select three models including Random Forest Regression, Bayesian Ridge Regression and ElasticNet Regression.

Model	**MAE**	**RMSE**	**R^2**
Random forest	1.86	2.36	0.38
SVR polynomial	2.53	2.98	0.00
SVR RBF	1.87	2.35	0.38
Bayesian ridge	1.69	2.08	0.52
ElasticNet	1.71	2.08	0.51
Ensemble ML with Voting Regressor	1.69	2.08	0.52

Table 2. Machine learning models evaluations

3.3 Topic Modelling

Aspects in the form of key phrases can be extracted using unsupervised ML techniques in order to establish topics from the comment texts. With our pre-processed data into *bag-of-words* matrices, we develop a Latent Dirichlet Allocation (LDA) unsupervised model to cluster the words together based on the distribution of word co-occurrences.

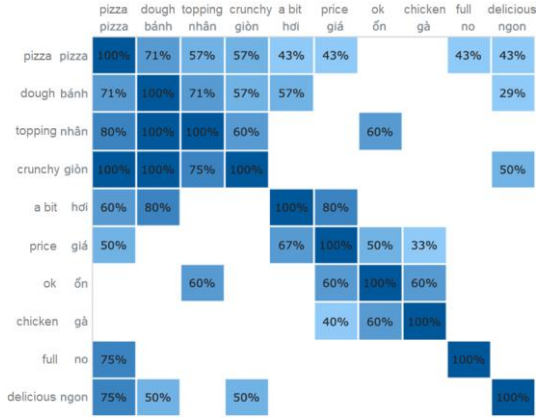

Figure 2: Co-occurrence matrix chart

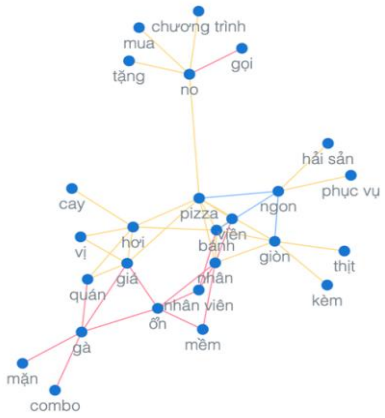

Figure 3: Network-of-words graph

While most common words/phrases express which topic needs to be investigated, establishing the keyword co-relation provides a deeper view on business's performance factors. Such co-relations can be visualized using the Co-occurrence Matrix, which displays the co-occurrences of top ten most frequent terms in the form of a heat map. As shown in Figure 2, by showing how likely a word in the x-axis appears simultaneously with a word in the y-axis, we can put terms together for a more coherent topic. For example, take row 2-column 1, we got "nhân"+"pizza" (pizza toppings).

Going beyond the correlation between two words, we also produce the Network-of-Words, in which not only frequent terms are displayed as a graph, their relationship is classified with the sentiment label

[3] https://www.foody.vn/

(i.e., negative-red, positive-blue, and neutral-yellow) (Figure 3). Using the force-directed algorithm, which minimizes the number of crossing edges and keeping all elements in the graph from not getting too close/far from each other, we use correlation to cluster common words (Kobourov 2012). Note that, the force-directed algorithm, which simulates the repulsive force between nodes, allows clusters of words to be an evolving system where node interaction update continuously with new data (Figure 4).

Steps	**Formula**
Calculate the effect of attractive forces between adjacent vertices	$f\_a\ (d) = d\hat{}2/k$
Calculate the effect of repulsive forces between all pairs of vertices	$f\_r\ (d) = -k\hat{}2/d$

Table 3. Force-directed process (d: distance between two vertices, k: optimal distance between vertices)

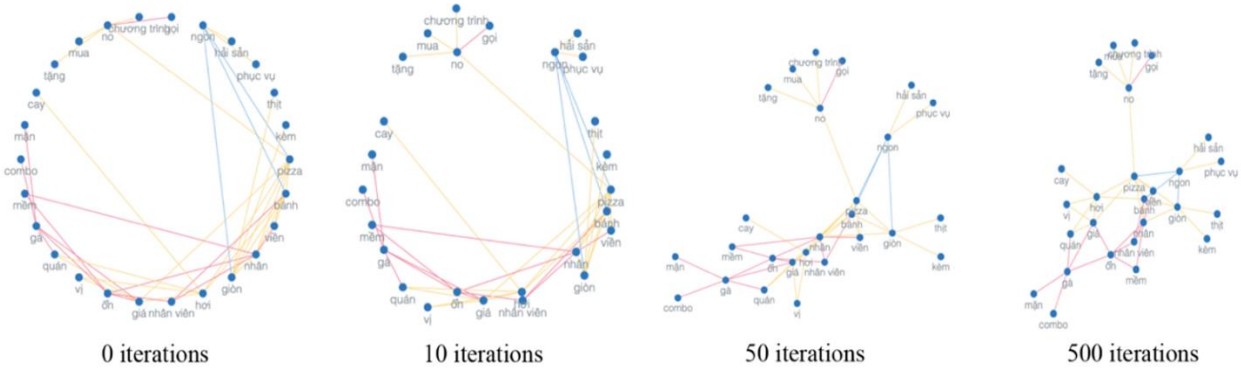

Figure 4: Force-directed graph with different iterations

3.4 Sentiment Trend Analysis

In this work, trend analysis is conducted by capturing the frequently mentioned topics plus assigning sentiment polarity to the stream of user activities. Using the sentiment classifier presented in section 3.2, we develop a sentiment trend analysis tool that enables business users to get answers for queries such as "when many positive/negative comments occur?", or "which topics associate with such a trend?". First, we classify incoming comments into three sentiment polarities. Comments are also clusters into topics, which allows us to map trending topics to given a sentiment label. Trend tracking can be done visually by displaying sentiment trends chart (Figure 5) on an hourly/daily basis. Overall, using this chart, a business owner can quickly spot performance degradation, and further identify an aspect that associates with a sentiment trend (e.g., a surge in negative comments). Such visualization aids enable business users to determine when and which parts of their business needs improvement.

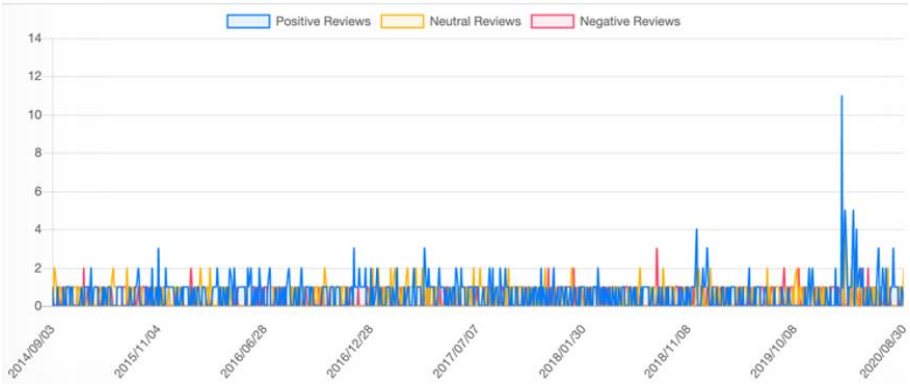

Figure 5: Sentiment trend chart

3.5 Opinion Summarization

To implement opinion summarization, we develop an unsupervised learning model using Variational Autoencoder (VAE) (Doersch 2016). An autoencoder is useful for summarization problem because its architecture resembles a bottleneck that ensures only the key details of a group of documents can go through and be part of the summary. However, this bottleneck could also filter fine-grain information and name entities such as technical information, product names, and menu items. In Figure 6, the Gated recurrent unit (GRU) encoder takes input vectors (r_1, ..., r_n) and performs dimensionality reduction to extract latent features (i.e., details that should remain in the summary). These features are combined

with the 'latent semantics' variable *c* to generate a set of semantics encoding variables z_i. The GRU decoder layer will turn these semantics encoding values into text that preserves the latent features of the input texts. We base our solution on (Brazinskas, Lapata, and Titov 2019) which implements an autoregressive model as part of the decoder take other reviews into account.

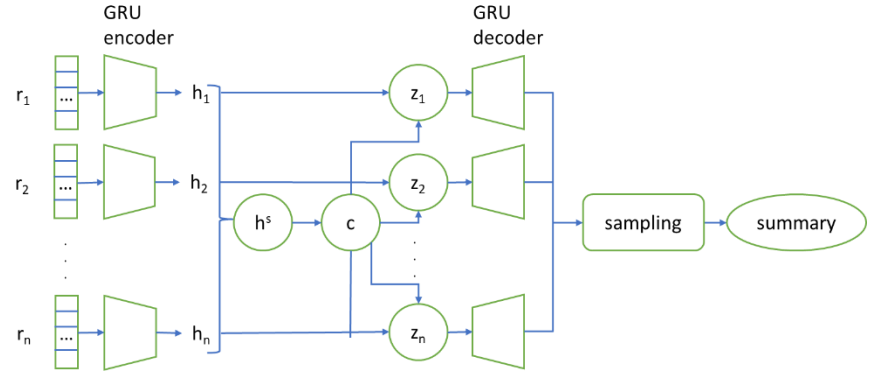

Figure 6: Text summarization workflow

4 Case Studies

To evaluate the proposed pipeline, we showcase two business cases in the food and retail domains including McDonald's and Tiki's products. In particular, we collect the online comments on their products and services and analyze them to retrieve insights of the business performance.

4.1 Use Case 1 - McDonald Opinions Analysis

McDonald's is well known for its hamburgers, cheeseburgers, and French fries. In Vietnam, McDonald's opened its first restaurant in 2014. In Ho Chi Minh city, by 2020, there are 23 restaurants. Our dataset includes 921 customers comments collected from 18 restaurants in Ho Chi Minh City, between 2014/03/09 to 2020/08/30, from the Foody website. We use our sentiment analysis model to predict the rating for all comments (0-10), with 10 is the most positive comment. In 921 comments, 400 comments are rated positive, 307 are neutral and 214 are negative. The average sentiment score for this dataset is 7.25 showing that McDonald's did not have any serious performance problems.

4.1.1 Tracking McDonald's performance

Figure 7 shows the sentiment trends of McDonald's comments between 2014 and 2020. Especially in the period 03/2020-06/2020, the number of positive comments raised sharply. Our topic modelling in combination with the sentiment trend analysis uncovers that many positive comments in that period of time involved the term "nhân-viên" (employees). As a result, employees' performance and employee management could be keys that resulted in customer satisfaction and improved the business performance. The hypothesis can be validated by an internal audit into how its employees have performed. For example, a change in management could lead to such an improvement.

Figure 7: Word-trending chart ("nhân-viên"-employee)

4.1.2 Topic Analysis

More analysis can be done by extracting top trending topics in the dataset. For example, one of the keywords found in the dataset was "ngon" (delicious) which presents a positive sentiment in the Food & Beverage industry. We are interested in identifying aspects that strongly associate with this sentiment keyword. For instance, a co-occurrence analysis on the top ten key terms found the term "ngon" follows the term "khoai" (potato) 71% of the time, a very high rate showing that customers love McDonald's fries. Another potential topic is "gà" (chicken), which also followed by "ngon" 67% of the time (in Figure 9).

Furthermore, the Network-of-Words was used to figure out which topics are more likely to co-occur. Figure 8 shows the word "gà" (chicken) most occurred with the word "khoai" (potato), "giòn" (crunchy), "miếng" (piece), "vị" (flavor) and "ngon" (delicious). In addition, the colours of the edges show that these comments were quite positive. Further topic analysis on the set of comments shows that "gà" is a favorite menu item, especially when it is sold with "khoai" (potato). This is an interesting insight as Mc Donald is famous for its beef hamburgers in the western context. However, deep-fried chicken burgers are more popular in Vietnam.

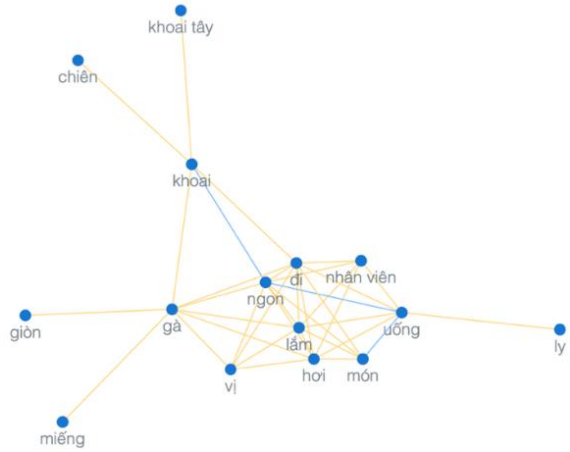

Figure 8: McDonald Network of Words

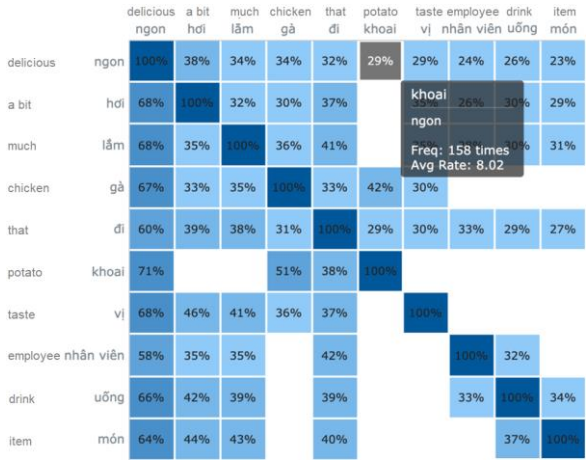

Figure 9: McDonald Co-occurrence matrix

4.1.3 Opinion Summarization

We train the model using a collection of 5,912 Vietnamese Foody[4] customer reviews then perform the opinion summarization task on the McDonald review dataset[5]. We group McDonald reviews using topics such as "gà" (chicken) and "nhân-viên" (employees) as focused above. Our ROUGE1 and ROUGE2 precision scores are 0.581 and 0.406 respectively. That presents a 40% overlap between our model generative summaries against the reference summary. This performance can be considered promising as abstractive summaries aims to not reuse much text from the original documents. We provide an example summary generated by the model along with several original reviews. Note that we only capture the relevant parts of the reviews and the generated summary in the example below.

Original reviews.

Deep-fried chicken not too oily, very crunchy, well marinated, delicious
Deep-fried chicken burger is not delicious..., ... Various dipping sauces are very delicious
Delicious chicken, hot, crunchy outside. Spicy chicken... I also like Mc for various dipping sauces
Better than the rumour. Chicken is delicious & crunchy, tasty beyond words...will come here often
Chicken nuggets are every nice, crunchy, hot, tender meat..., dip with chilli and mayone sauces...

Generated opinion summary.

Delicious chicken, hot, crunchy outside, stuffing with chicken wings ... Dipping chilli and mayone sauces are delicious and spicy...

The summary generates a concise global view given a sea of user comments. In the example above, business can confirm that chicken menu items, especially chicken wings are very popular in Vietnam, despite the business is more about beef patty hamburgers elsewhere in the world. Crunchy chicken pieces plus dipping sauces is the winning combination for McDonald in the Vietnam market.

[4] https://drive.google.com/file/d/1KWVNciIlaqNHDktoAJLqaG_Du3iBXc9_/view?usp=sharing
[5] https://drive.google.com/file/d/1kPMcoAfH6Ql4MYgQjirWD3zJc03xwLLF/view?usp=sharing

4.2 Tiki's Customer Service Analysis

Tiki is a Vietnamese e-commerce platform established in 2010. The dataset in this study contains 1,351 comments from 2018/06/19 to 2020/08/30 for the Xiaomi Gen 2S power bank collected from the Tiki website. We host the dataset here. Similar to the McDonald use case, we first apply our sentiment analysis model to rate all 1,351 comments[6]. We got 169 positives, 441 neutrals and 741 negatives with an average rating of 6.06.

4.2.1 Track Tiki's sale for Xiaomi power-bank product

The general sentiment trends of the Xiaomi power-bank sale at Tiki are shown in Figure 10, where the number of negative comments increased sharply around the end of 2019. We analyze this negative trend by looking at the comments involved the keyword "mua" (purchase) because the word-trending chart (Figure 11) shows many negative comments occurred in the same period. However, we could further explore the problem with a detailed topic analysis.

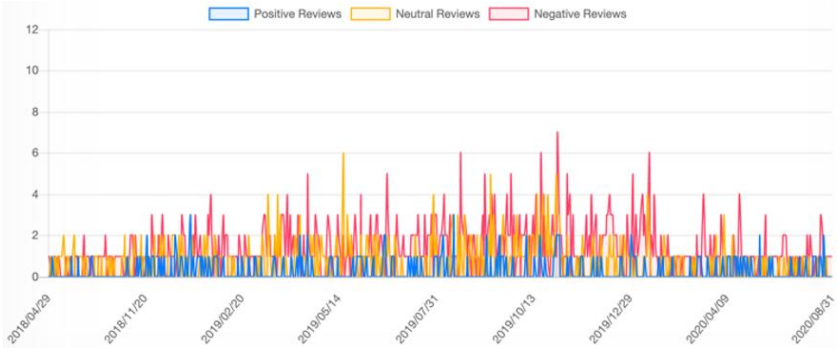

Figure 10: Tiki General sentiment trending chart

Figure 11: Tiki Word-trending chart ("mua" - purchase)

4.2.2 Topic Analysis

Beside words that relate to the nature of the product such as "sạc" (charging), and "pin" (battery), our LDA analysis shows some other most mentioned words: "giao" (deliver), "mua" (purchase), and "gói" (package). Furthermore, our sentiment labels show these key terms are among negative topics, which indicate that the purchasing and the delivery process receive concerns from customers.

The Co-occurrence matrix in Figure 13 shows that in 508 comments contain "hàng" (product), there are 131 comments (26%) contain the positive term "đẹp" (beautiful), and the average rate of these comments is 7.48. That indicates that most customers like the design of the Xiaomi power bank product. The Network-of-Words graph in Figure 12 delivers more insights on the issue with the purchasing process for this product. For example, the word "mua" (purchase) occurs mostly with words "giá" (price), "pin" (battery), "sạc" (charge), "sản phẩm"/"hàng" (product), and most of those phrases were labelled with low sentiment scores. So far, analysis of sentiment from customers' comments combining with the topic analysis capacity helps uncover problems with the purchasing process and the low quality of the product. Tiki could use these insights to improve the purchasing process for similar products in the future.

[6] https://drive.google.com/file/d/1R9MpqYYWmYSy-B-T3dp6GicfX-RM1szI/view?usp=sharing

4.3 Opinion Summarization

In this case study, we retrain our opinion summarization model using review data obtained from Tiki. We group Tiki product reviews using topics such as “sản phẩm”/“hàng” (product) and “giao” (delivery). The ROUGE scores present a 31% overlap between our model generative summaries against the reference summary, which is lower compared to the McDonald ROUGE scores. Studying the original review in detail shows that reviews from e-commerce platform are often more objective and logical than sentimental towards the products or services. For example, these reviews include product name & version, software/hardware information, battery life etc. Our summarization approach, while tries to remain abstractive, tends to filter out text tokens that are too specific in details. This drawback suggests extractive summarization could be more suitable for such a business case.

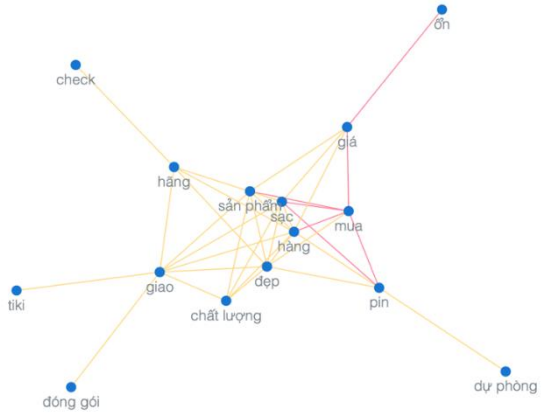

Figure 12: Tiki Network of Words

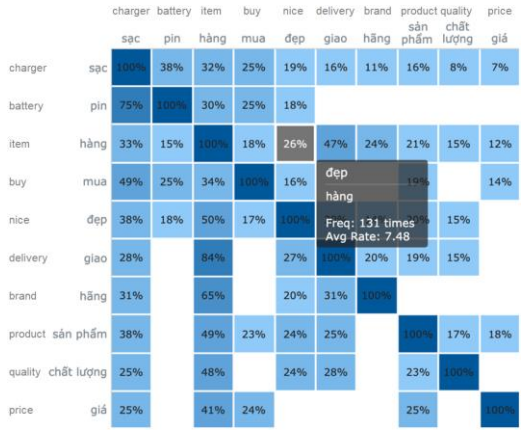

Figure 13: Tiki Co-occurrence matrix

5 Conclusion

Social listening is a key enabler for market research in the big data era. Being developed using AI as the foundation, social listening promises to bring significant benefits to businesses, especially those that focus on product and service delivery. In this paper, we present a complete social listening system in which large corpus of text can be recorded and processed automatically through topic and sentiment analysis. While e-commerce platforms often ask users to leave a rating (e.g., 5-star), many social media forums only record text comments without a rating. In this work, we present and integrate three core NLP technologies, along with performance evaluation results, to rate customer comments and generate overall summaries from those text. Once comments are rated, further text analysis can be done including identifying key aspects and their frequencies, finding clusters of keywords based on their co-relations, and finally trend-tracking. Overall, this work illustrates how a range of mature research ideas can be built into a practical technology. We present two business cases to show the potential of such social listening technology in terms of providing customer satisfaction and business performance insights.

6 References

Baker, Kristen. 2020. '15 Best Social Listening Tools to Monitor Mentions of Your Brand', Accessed 16/01/2021. https://blog.hubspot.com/service/social-listening-tools.

Barzilay, Regina, and Kathleen R. McKeown. 2005. 'Sentence fusion for multidocument news summarization', *Computational Linguistics*, 31: 297-328.

Blei, David M. 2012. 'Probabilistic topic models', *IEEE Signal Process. Mag.*, 27: 55-65.

Botchkarev, Alexei. 2019. 'Performance Metrics (Error Measures) in Machine Learning Regression', *Interdisciplinary Journal of Information, Knowledge, and Management*: 45-97.

Brandão, J. de Godoi, and W. P. Calixto. 2019. "N-Gram and TF-IDF for Feature Extraction on Opinion Mining of Tweets with SVM Classifier." In *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 1-5.

Brazinskas, Arthur, Mirella Lapata, and Ivan Titov. 2019. 'Unsupervised Multi-Document Opinion Summarization as Copycat-Review Generation', *ArXiv*, abs/1911.02247.

Business Woman Magazine. 2020. 'Nghịch lý 1% và sự ra đời của một công cụ lắng nghe mạng xã hội', Accessed 24/09/2020. https://nudoanhnhan.net/ngo-ngang-voi-nghich-ly-1-va-su-ra-doi-cua-mot-cong-cu-lang-nghe-mxh-uu-viet.html.

Chumwatana, T., and I. Chuaychoo. 2017. "Using social media listening technique for monitoring people's mentions from social media: A case study of Thai airline industry." In *2nd International Conference on Knowledge Engineering and Applications (ICKEA)*, 103-06. London.

Doersch, Carl. 2016. 'Tutorial on Variational Autoencoders', *ArXiv*, abs/1606.05908.

Duyen, N. T., N. X. Bach, and T. M. Phuong. 2014. "An empirical study on sentiment analysis for Vietnamese." In *2014 International Conference on Advanced Technologies for Communications (ATC 2014)*, 309-14.

Genest, Pierre-Etienne, and Guy Lapalme. 2012. "Fully Abstractive Approach to Guided Summarization." In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, 354–58. Jeju Island, Korea: Association for Computational Linguistics.

Great Ideas for Teaching Marketing. 2020. 'New Coke Case Study - Great Ideas for Teaching Marketing', Accessed 20/09/2020. https://www.greatideasforteachingmarketing.com/new-coke-case-study.

Hasan, H. M. M., F. Sanyal, and D. Chaki. 2018. "A Novel Approach to Extract Important Keywords from Documents Applying Latent Semantic Analysis." In *2018 10th International Conference on Knowledge and Smart Technology (KST)*, 117-22.

He, Ruidan, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2017. "An Unsupervised Neural Attention Model for Aspect Extraction." In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguitics*, 388-97. Vancouver, Canada: Association for Computational Linguistic.

Jelodar, Hamed, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. 2019. 'Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey', *Multimedia Tools and Applications*, 78: 15169-211.

Kang, Hee Jay, Changhee Kim, and Kyungtae Kang. 2019. 'Analysis of the Trends in Biochemical Research Using Latent Dirichlet Allocation (LDA)', *Processes*, 7: 379.

Kieu, Binh Thanh, and Son Bao Pham. 2010. "Sentiment Analysis for Vietnamese." In *Second International Conference on Knowledge and Systems Engineering (KSE)*, 152-57. Hanoi, Vietnam.

Kim, Suhyeon, Haecheong Park, and Junghye Lee. 2020. 'Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: A study on blockchain technology trend analysis', *Expert Systems with Applications*, 152: 113401.

Kobourov, S. 2012. 'Spring Embedders and Force Directed Graph Drawing Algorithms', *ArXiv*, abs/1201.3011.

Le, H. T., and T. M. Le. 2013. "An approach to abstractive text summarization." In *2013 International Conference on Soft Computing and Pattern Recognition (SoCPaR)*, 371-76.

Naz, S., A. Sharan, and N. Malik. 2018. "Sentiment Classification on Twitter Data Using Support Vector Machine." In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 676-79.

Ortiz-Ospina, Esteban. 2019. 'The rise of social media', Accessed 24/09/2020. https://ourworldindata.org/rise-of-social-media.

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. "Thumbs up? Sentiment Classification using Machine Learning Techniques." In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, 79-86.

Pedregosa, Fabian, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blonde, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, and David Cournapeau. 2011. 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research*, 12: 2825-30.

Phan, Huyen Trang, Ngoc Thanh Nguyen, Tran Van Cuong, and Dosam Hwang. 2019. "A Method for Detecting and Analyzing the Sentiment of Tweets Containing Fuzzy Sentiment Phrases." In *International Symposium on Innovations in Intelligent SysTems and Applications (INISTA)*, 1-6. Sofia, Bulgaria: IEEE.

Rehder, Bob, M. E. Schreiner, Michael B. W. Wolfe, Darrell Laham, Thomas K. Landauer, and Walter Kintsch. 1998. 'Using latent semantic analysis to assess knowledge: Some technical considerations', *Discourse Processes*, 25: 337-54.

Tanaka, Hideki, Akinori Kinoshita, Takeshi Kobayakawa, Tadashi Kumano, and Naoto Katoh. 2009. "Syntax-Driven Sentence Revision for Broadcast News Summarization." In *Proceedings of the 2009 Workshop on Language Generation and Summarisation*, 39-47. Suntec, Singapore: Association for Computational Linguistics.

Yue, Guo, J. Barnes Stuart, and Jia Qiong. 2017. 'Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation', *Tourism Management*, 59: 467-83.

Acknowledgements

This work has been supported by RMIT Vietnam's Thematic Research Fund (TRF 2021- Grant 3).

Copyright