Master Thesis

For Programme Applied Data Science

Topic Modeling on Online Reviews

Chenjun Zhang

c.zhang5@students.uu.nl

1214837

18th May, 2022

Graduate School of Natural Sciences

Dr.Rense Corten

Abstract

Online markets became progressively dominate in today’s society, accompanied with this wide spread of online market, online rating systems emerged. Online reviews and rating system can provide useful information for both the customers and the product or service suppliers. This project focused on the suppliers’ perspective towards the online reviews with data from a dutch website Werkspot.nl which provides plumber service. Implemented text mining and LDA topic model to the reviews, the project explored the concerns of the homeowners and subsequently proposed several methods that can help the professionals improve themselves to obtain high rating scores and more working opportunities.

Keywords:online reviews; customer behavior; topic modeling

Content

1 Introduction

Online market is increasingly significant in current society. It provided a brand new way to connect product suppliers and the customers. Both customers and product suppliers earn benefit from plenty of opportunity online market provides.The appearance of this new form of trade created a market which demands less time and human to complete a trade (Tadelis, S., 2016). Additionally, online information translation lead to a new type of word-of mouth:an electronic word of mouth, in where people are able to start easy conversation on the internet.

Review systems is significant for the online market, also it is an obvious difference from the traditional market,online consumer review actually built a new product or service channel with popularity and significance (Chen, Y., & Xie, J., 2008). In the traditional physical market people can only know the quality of products or efficiency of service through people they know and their own previous experience. Online review and rating system provide another way to learn about the product or service they decide to purchase based on electronic word of mouth. Therefore, reviews of the product or service can provide more complete and comprehensive information about the service or product.

Normally, customers of online platform provides service always paid more attention to the reviews of the service suppliers, since customers are able to see almost all of the reviews and rating scores of all of the service suppliers. Because of that situation, the customers can choose the service suppliers conform their expectation,which made them become the first beneficiary from the online review and rating system. Despite the fact that the feedback of the consumers can also help service suppliers to learn more about the needs of consumers and then increase their competition in the online marketplace, the service suppliers,however, can only receive limited number of reviews of themselves,and it is hard for them to improve their service through these limited feedback. When the service suppliers can understand the concerns of the customers, they can adjust themselves towards a better direction, where they could have more working opportunities and then they can even offer a higher price towards specific service.

Therefore understanding the concerns of the customers is significant and necessary for the suppliers as well. In the present study,I conducted the analysis on the dataset retrieved from the ‘Werkspot.nl’ website. I mainly implemented topic modeling(LDA) on the dataset and tried to obtain the most interpretable and correlated results on different size of data. I then connect the results with the rating scores to find out the concerns of the consumers and in more details, to explore which kinds of behavior or features will lead to a better experience.

2 Theoretical Background

2.1 Electronic Word of Mouth

The consumers will discuss their experience of the products or service in their daily dialogues. Because of the emergence of internet, the social media became increasingly significant in people’s daily life,and it also expands the social circle of the people which infers that the information can spread faster and wider. Additionally the pictures and messages presented on social media also provide a brand new channel for people to acquire information. The emergence of internet and the widely use of social media therefore brought up a new form of word of mouth:electronic word of mouth which are more infectious than the traditional word of mouth.

Consumers will use online platform such as social media to share their opinions about the products or service they consumed (Gupta, P., and Harris, J., 2010). Consumers consider the feedback of previous customers as a more reliable method to help them decide whether or not should they purchase the products or service. Online reviews of the platform will sure effect the final decision of the customers.

2.2 Customer Behavior

Along with the widely use of internet,online business started to offer more convenience compared to traditional physical stores (Rita, P., Oliveira, T., & Farisa, A., 2019). For one thing, customers can complete the whole process of purchasing only through the internet. Furthermore,the online business enlarge the choice of the customers, for example, the sellers of ebay can show their products to the whole world, and that definitely increase the probability of selling out the products. Another example is Werkspot, homeowners are able to find the professionals they prefer all over the Netherlands,the platform obviously improve efficiency to solve the problem happened in the house.

When companies can learn about the customer behavior,they are capable to decide an optimal price for certain products or can come up with a clearer insight on marketing strategies. Especially in the online marketplace,because on one hand, a successful marketing or recommendation systems will definitely improve the sales of the products or the usage of the website. On the other hand, it is easier to analyze the customer behavior in e-commerce environment. For example, it is possible for entrepreneurs or online platform to trace back the purchase history of the consumers, and their view history in certain situations.

2.3 Topic Modeling

Topic models such as Latent Dirichlet Allocation (Blei, D. M., Ng, A. Y., & Jordan, M. I., 2003)is aim to search for the semantic similarity that behind certain texts. Because human can discover the hidden topics of texts easily, however, it is hard for a program [8]. In the 1980, topic modeling was initially developed and branched off from the subject area of ‘generative probabilistic modeling’ (Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W., 2016). This type of modeling assumes there exists a relationship between observed variables and unobserved parameters within the data in the dataset (Steyvers, M., & Griffiths, T., 2007). The development of the concepts of topic modeling arose from the need to briefly describe elements in a large dataset and without disturbing the statistical relationships. Topic modeling is trying to create algorithm so that the program can determine the topics as well. And it is especially important when the aim of the project is to classify the large amount of texts into clusters based on their similarities.

Latent semantic indexing (LSI) (Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R., 1990) is the initial model of topic model, it is not a probabilistic model (Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W., 2016), therefore it is not an actual topic model. Probabilistic latent semantic analysis (PLSA) (Hofmann, T., 2001) was proposed by Hofmann and is a genuine topic model were built after the LSI model. Based on the previous models, LDA model was created,and it is a more complete and common used model until now.

Topic modeling can be conducted on both long texts such as newspaper and short texts like tweets or review. Normally,topic model such as LDA has a satisfied results. However there is still some limitations of the model. The most significant one is that the algorithm cannot model the topic correlation, for example,the word ‘gene’ will be more likely in the group of words related to disease,not the word ‘X-ray astronomy’ (Blei, D., & Lafferty, J., 2006).

2.4 Research Question

The research question of the project is to find the concerns of the consumers,in other word to explore the customer behavior of the people, since customer can be the one to purchase either products or service. The decision of the customer is affected by the needs and preference of themselves (Applebaum, W., 1951). In this project ,most customers(homeowners in this circumstance) will search for the professionals when they really need them. Therefore, it is interesting to explore their preference on the professionals. From the dataset, the reviews are the most relevant variable when it comes to the preference of the customers. Because most of the time people will write a review when their concerned features appeared on the performance of professionals.

For instance,when some people care more about the communication with the professionals, and they happened to have an efficient conversation with the professionals during the whole process, they will probably write a review such as ‘easy to communicate’.More importantly, the customer behavior is particularly significant for the professionals which are the service suppliers in this project, since they do not have access to all of the customer. This leads to the research question ‘What do homeowners expect from the plumbing work and which features of plumbers are related to levels of satisfication?’

The results of the analysis can help the professionals adjust themselves based on the concerns of homeowners. For example, if people care more about the price, the professionals can adjust their price accordingly. Therefore, both professionals and consumers can have a better experience during the whole working process.

Additionally, in the present study, I will conduct the topic modeling on more correlated short text data which is the reviews of the customers from the website. Firstly, I will analyze the original data,after that I split the raw data into 3 more related subsets based on sentiment(unsatisfied,satisfied and very satisfied), then I did analysis on each subset. Previous project always conducted the model on the processed original data, and I did a comparison during the analysis process which is between the original dataset and classified dataset. Through this kind of analysis I hope to explore whether more related and correlated texts can help improve the performance of the model. Especially to find out whether shorter and specific data can lead to a correlated results from the topic model.

3 Data

3.1 Brief Introduction of Data

The data analyzed in the project is collected from the ‘Werkspot.nl’ website. The website is a platform that connects homeowners and professionals on housework. Homeowners can search for certain professionals on the website and they can choose which one to start a conversation with based on their personal thoughts and the reviews or ratings scores presented on the website. I focused on one kind of housework, and therefore I chose the reviews and ratings cores of plumbers. The data includes more than 20,000 reviews and rating scores over 12 years. And the final collected dataset consist the ID of the professionals, the title of the review, the full review, the star, the date for each review and the reviewer ID(some reviewers are anonymous). All the reviews in the data set are in dutch. And for each professional the reviews are sorted by the date from the most recent one. In the project I aim to discover the concern of the homeowners form the present data therefore my central point is the rating stars and the reviews.

3.2 Process the Data

As for the preprocessing of the data, firstly, I checked whether there existed any missing or duplicated values in the dataset. The missing value exists in different columns and if I just deleted all of the missing values there will be a great loss of data. Therefore I only focused on the reviews and star which are the relevant variable in the analysis afterwards to reduce the loss of the original data. After this process the the dataset has 22052 records left. The original dataset consists both the title of the reviews and the full reviews in the column.Since the reviews will not be long texts,I just used all the text in the full reviews and built a new dataframe of all the reviews contain full texts, and I also added a column named ‘stars’ which is the attribute ‘star’ in the original data set,for a better illustration, I also reset the index of the new dataset.

When the new data set was created,I started to process the review. For the analysis afterwords, all the reviews need to be tokenized and I used Spacy to do the tokenization. Spacy is a free open-source library for Nature Language Processing (https://spacy.io/), and it includes pipeline that supports different language concludes dutch. After tokenization, I also deleted the stop words and the punctuation because these tokenization will not be relevant to concerns of homeowners, furthermore all the word are inverted to lower case for convenience.

4 Method

To answer the research questions, I first split the original dataset into 3 subsets and then conducted analysis on each of them.To split the dataset, I first checked the distributions of the rating scores in the original dataset .

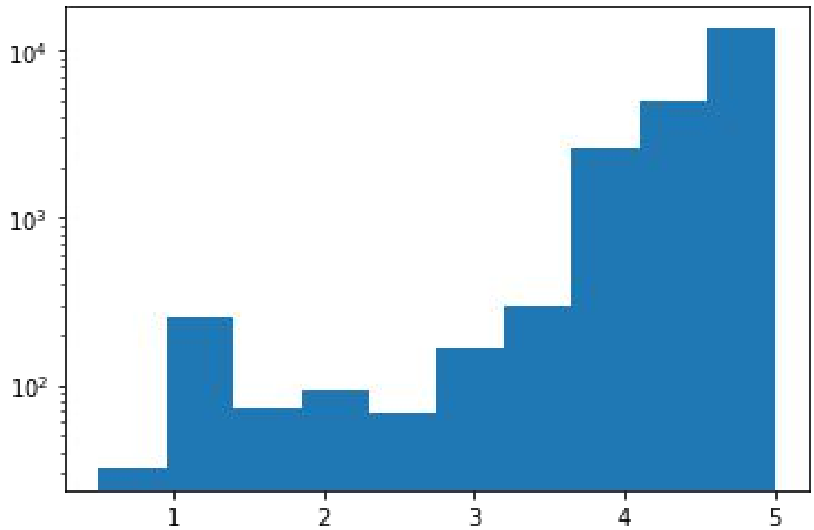

Figure 1: number of reviews in distributed in different rating scores

As the figure 1 exhibits ,the original data is noticeably imbalanced. More than 80 percent of the data are distributed in the range of 4.5-5. It will be less meaningful to classify the data based on all of the ten rating scores, since the records with the rating score below 4 are only a small part of the whole data. Therefore I decided to classify the data as 3 parts,first part is the data with a rating score below 3 and I defined this part as ‘unsatisfied’, this classification indicates when people’s problems were not solved or they were angry about some aspects during the process; the second part which is ‘satisfied’ includes data with rating score between 3 and 4 suggests people were satisfied during the whole experience,however people probably not pleased with some aspects during the process, in this parts two margin values(three and four) are included; and the left part of the data was then defined as ‘very satisfied’, which basically can represent people are totally consider this process as a perfect experience.

To retrieve the topics from all of the texts in the project, I decided to use the LDA topic modeling. LDA is now one of the most popular probabilistic text modeling techniques,it allows the hidden parameter or semantic in text mining to explain the similar observed variables (Tong, Z., & Zhang, H., 2016). Additionally, some previous projects, for example to use the LDA model to detect related issues from historical newspaper to assist historical research (Yang, T. I., Torget, A., & Mihalcea, R., 2011) and implement LDA model on user’s comments on social media (Ramamonjisoa, D., 2014)prove that LDA model can produce a reasonable result on the review data in this project, therefore I chose to implement the LDA model in the analysis process.

For the used texts in the analysis afterwards, I tried all of the words in the reviews first.

Figure 2: Wordcloud of all the texts in the original dataset

Regarding figure 2, words such as ‘goed’, ’klus’, ’kwam’, ‘snel’ etc are more frequent than the other words. Obviously, the results of all the words in reviews also contain adjectives such as ‘good’ ‘fast’ ‘great’ etc. With all the adjectives, it is hard to find out and summarize the concerns of the homeowners from this results include adjectives, therefore I only selected nouns to train the LDA model in subsequent analysis. Moreover, I ruled out the common first names in the texts, for example ’Gert’, ’Michale’, ’Dennis’ etc, since the first names are not relevant to the customer behavior, and they will affect the results of the analysis probably.

I used package Genism (https://radimrehurek.com/gensim/) to achieve LDA topic model on the data, Genism is an open source to train large-scale semantic models. Genism model required several parameters. First, I defined the limitation of all the review words. The model will only perform on word that appear at least two times ,and the probability of one certain word exists in the all of the word should be less than 0.85. Therefore, the left word will be more helpful as for the topic detection, otherwise the results may contain some meaningless but frequent mentioned words. Subsequently, I got all the vocabularies through the function ‘dictionary’, so that every word connect to its integer ID. Then bag-of-word(BOW) corpus were created based on the dictionary built before.

Because topic modeling is an unsupervised learning and it gives no guarantee on the interpretability of the outputs (Röder, M., Both, A., & Hinneburg, A., 2015), and coherence measure is a method to evaluate whether the results of topic model are understandable. Therefore I calculated the coherence of the model on different number of topics from 1 to 50 with a coherence measure based on sliding window and cosine similarity to find the optimize number of topics trained in the final model(Röder, M., Both, A., & Hinneburg, A., 2015).

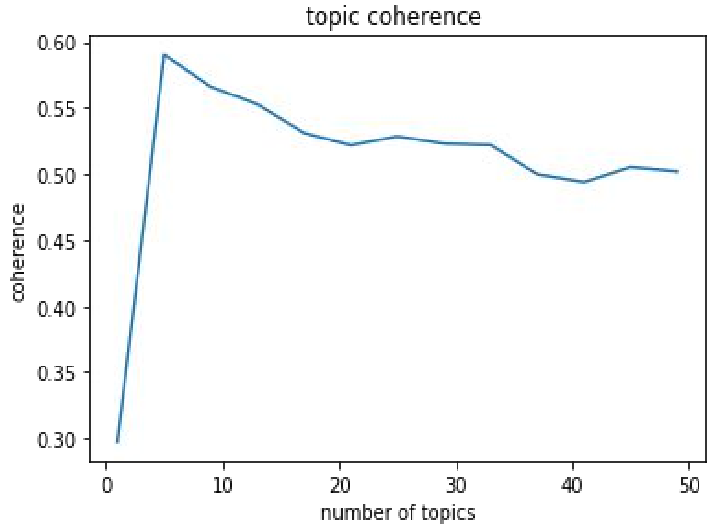

Figure 3:Coherence of the LDA model

From the results of the plot, I decided to train the LDA model with 6 topics, 10 optimized intervals which is the option to allow some topics to be more prominent than others(https://radimrehurek.com/gensim/) and 1000 iterations.

As for the three subsets, I used the same package Spacy and the same selection to process the text data. After that I implemented the topic model on each of them still also with the parameter: 'min_df = 3, max_df = 0.85', because even the dataset with the fewest records still has 523 reviews, and it is still fine to only consider the words appear at least 3 times.

5 Results

Figure 4: Wordcloud of the result of all the reviews

The results reflected that topics related to price and types or places of the plumbers' work are the main topics of all the reviews, since words 'vakman', 'price', 'klus', and 'keuken' are more frequent than the other words. However, the wordcloud only shows limited information of the results. Therefore, table below can illustrate the results extensively.

Table 1:Results of all review texts

Topic		Explanation
1	probleem kraan afvoer keuken wim heer verstopping lekkage vaatwasser muur	Reflects different classification of plumber work, which includes leakage,drain and kitchen
2	ketel cv installatie paul radiator radiatoren henk onderhoud service airco	Mainly tells about different types of the plumber problems, including heating, installing and maintenance

3	werk klus prijs vakman tijd aanrader prima communicatie netjes afspraken	The topics discusses the price, efficiency and the appointment or agreement of the whole process
4	werkzaamheden klus vakman zaken bedrijf werk klussen afspraken ervaring resultaat	The topics reflects experience during the working process, and there exits some noise words
5	klus dag afspraak contact offerte opdracht keer prijs dagen werkspot	These topics reflect the agreement between homeowners and the professionals
6	badkamer huis vloer resultaat mannen keuken toilet tegels plafond team	These topics mainly describes different types of plumbers' work

From the explanation of table 1, topic 1 and 2 are both reflects the types of plumbers' work, topic 3 is about the communication or whether two sides can reach an agreement, topic 4 and 6 talk about the result or the professional level, topic 5 is about the price. Although we obtained the main topics which most of them can be used to infer the concerns of the homeowners, it is also necessary to explore which kinds of behavior or features during the process will lead to more satisfaction.Foremost, the table below describes the topics among the reviews showed unsatisfactory.

Table 2: Results of the 'unsatisfied' subset

Topic	Explanation

1	probleem lekkage klus aantal maanden contact keer fout zaken malen	These topics reflects mistakes and accidents happened
2	werk communicatie klus ervaring prijs review keer geld dagen werkspot	The topics mainly describes the agreement should have been reached between homeowners and professionals
3	euro monteur bedrijf vosse installatie leidingen kosten vakman btw begin	These topics reflects price and some types of plumber's work
4	badkamer tegels huis vakman geld werk toilet vloer deur water	The topics mainly describes the places where the homeowners need plumbers
5	uur bedrijf heer factuur kraan tijd werkzaamheden prijs loodgieter rekening	The topics are mainly about the price or the bill of the work
6	afspraak klus dag contact offerte meneer keer reactie tijd afspraken	These topics reflects contact and appointment during the work process

As the table 2 demonstrates, topic 3 and 4 are all about some specific places which need to work on, such as bathroom and some types of plumbers' work, for instance,installation. Topics 1 mentioned there exists mistakes during the whole working experience. Topic 2, 5 and 6 mentioned price and appointment which should be included in the agreement reached before.

Table 3:Results of 'satisfied' subset

Topic		Explanation
1	montage water team gasleiding rekening dingen rest uitleg deel muur	Some noise words, left topics are about the specific places need to be be repaired
2	vervanging plekke buizen klusjesman ervaring buren garantie mensen zand sterren	Most are the noise words, few words related to the types of work
3	totaal hal koffie afzuiging minuten staat lof dienstverlening communicatie plek	The topics describe the communication between the homeowners and the plumbers
4	werk klus prijs vakman tijd afspraak afspraken communicatie ervaring man	The topics reflects the communication and the appointment happened between the process
5	dingen schutting stopcontact lampen draad stopcontacten afzuigkap staat slag regenpijpen	Most of the topics are noise words,and it is hard for me to find out the correlation among them
6	plafonds minpunt mening verwarming afzuiging oplossing wensen badkamer inspectie renovatie	The topics describe the different types of plumbers' work and the different places as well

The table 3 is the results from the 'satisfied' subset, it shows that topic 1, 2 and 6 describes the types and places of plumber work, topic 3 and 4 reflects the communication and the appointment between the plumbers and the homeowners.

Table 4:Results of the 'very satisfied' subset

Topic		Explanation
1	klus tijd man vakman afspraak werk klussen keer werkzaamheden prijs	Reflects the agreement between the homeowners and professionals
2	badkamer resultaat huis toilet keuken vloer mannen verbouwing team planning	The topics are about the places of the plumbers' work
3	werkzaamheden zaken ketel advies installatie radiator verstand offerte kennis uitleg	The topics describes types of plumbing work
4	kraan keuken verstopping afvoer lekkage probleem vaatwasser leiding water buitenkraan,	The topics are related to a specific types of plumber work:water piping
5	werk prijs afspraken communicatie aanrader vakman ervaring service tijd super	The topics are about the communication and agreements between homeowners and plumbers
6	badkamer huis toilet resultaat vloer keuken mannen verbouwing team planning	The topics are related to different places which need plumbers

From the results of the table 4,topic 2, 4 and 6 are about different types or places of plumbers' work such as heating and water piping in kitchens, and toilets,topic 1, 3 and 5 are all about communication and agreements between the plumbers and the homeowners.

In summary, topics related to mistakes only appear in the the 'unsatisfied' group; the topics describes types and places of plumbers' work occur in all of the three subset;

the topics related to appointment emerge more in the ‘unsatisfied’ and ‘very satisfied‘ groups; topics about the communication pop up in the ‘satisfied’ and ‘very satisfied’ subsets, especially in the latter one.

6 Conclusion

6.1 The Performance of Topic Modeling

I have conducted the LDA model on 4 dataset, one of them is the processed original dataset, and the other three are the subsets based on rating scores of the original dataset. Most of the topics are easy to understand and clearly provide summarized information from a bunch of datasets, furthermore most of the topics retrieved from model accord to the common sense of these review. However, the results contain noise words in almost every dataset except the original data, moreover even if the results are interpretable, they have barely no correlation with each other.

Compared the original dataset with the three subsets,the ‘unsatisfied’ and ‘satisfied’ subset actually performed worse than the processed original dataset. I would attribute this situation of two subsets to the small amount of the data, and therefore this infers that LDA model probably are more helpful when the dataset contain more information. The ‘very satisfied’ dataset which includes more than 80 percent of the original data has a best performance among all of the four dataset however, two groups of topics retrieved from the model are more correlated than the results from other subsets. This conclusion implies that when a large dataset is more correlated or have been classified before, the LDA model can perform better and offer more helpful information.

6.2 The Concerns of Customers

Through the results of the original dataset, homeowners care about the efficiency and the whether the price is reasonable, besides, the communication is also important. All of the results conclude some features customers will value during the working process of the plumbers, however, these are summarized concerns and it is still unclear which kinds of behaviors will actually lead to a better feedback of customers which will help the professionals improve and adjust themselves.

Therefore I will conclude the results from the three subsets. Foremost, from the topics of unsatisfied subset, I found that price appears as a common mentioned topic by the homeowners. Besides, unpunctuality will lead to unsatisfied reviews. Furthermore, in the unsatisfied subset, there will exists some mistakes during the working process.When it comes to the satisfied reviews,more than half of the results are all about the different places in the house which need plumbers to fix and also kinds of plumbers' work such as water leakage and heating system, which implies that when the work can be done efficiently, homeowners will be satisfied about the experience. However, high efficiency itself is not enough for customers to leave a high rating scores above four. Finally, the results retrieved from the 'very satisfied 'subsets infers that homeowners are willing to rate high scores when the plumbers can not only efficiently finished the work but also have successfully reached and obeyed the agreement of the price and the appointed time with the homeowners, moreover, they communicated well with the homeowners during the whole process.

In conclusion, the price are significant, since an unreasonable price can easily lead to a low rating score. Additionally, when the appointed time is decided between the homeowners and professionals, it is better for plumbers to obey the time, otherwise it will lead to a disappointed feedback. Moreover, communication is significant during the process, a neat and tidy conversation would more likely lead to high rating scores which will help the plumbers to have more customers in the future. Overall, when plumbers hope to improve their rating scores and acquire more working opportunities from the homeowners, they can offer a reasonable price conform to their ability and the market, then come to the place at the appointed time which was decided in the agreement with the homeowners before. Furthermore, communications happened during the working process will always be the key to high rating scores, therefore the plumbers can try to communicate more to the homeowners.

7 Discussion

7.1 Limitations of the Project:

The analysis of the project have mainly three limitations.Foremost, LDA model itself has some problem on the correlation of the results. Form the results of the analysis, when the original data is large and related, the results will be more correlated, however this method are not practical in some situations.

Secondly, the data is imbalanced because the high rating scores data constitute more than 80 percent of the data. This kind of imbalance is probably due to conformity, which implicitly influences people's behavior to group norms(Liu, Y., Cao, X., & Yu, Y., 2016). In this project conformity happens when customers still leave high rating scores even if they are unsatisfied because the previous customers mainly leave high feedback. Situation like this will obviously affect the classification I made based on the rating scores, since there probably exist unsatisfied reviews with high rating scores and I classified them into the wrong subset.

The third limitations in the project is about the reviews language in the original dataset. Werkspot is a dutch website and therefore most of the reviews are recorded in dutch(I have asked native speakers to help translate the results retrieved from the results). In the preprocessing of the text data, I implemented an open source 'Spacy' package to exclude all the stop words and select only nuns for the subsequent analysis. Because dutch is not widely used as English,the package probably cannot accurately ruled out all the stop words and selected only nouns,which will influence the result of the analysis afterwards.

7.2 Future Work

Future work can be related to revise the LDA model to improve the correlation issue,so that the results of the model can be more informative. Besides, it is also interesting to explore the conformity in online platforms ,and to explore how and in at which level this phenomenon can influence the online rating systems. Since online rating systems are widely exist in today’s society,from shopping website to movie website, it is significant to find out the authentication of these ratings.

References

[1]Applebaum, W. (1951). Studying customer behavior in retail stores. Journal of marketing, 16(2), 172-178.

[2]Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

[3]Blei, D., & Lafferty, J. (2006). Correlated topic models. Advances in neural information processing systems, 18, 147.

[4]Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. Management science, 54(3), 477-491.

[5]Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. Journal of the American society for information science, 41(6), 391-407.

[6]Gupta, P., and Harris, J. (2010). How e-WOM recommendations influence product consideration and quality of choice: a motivation to process information perspective. J. Bus. Res. 63, 1041–1049.

[7]Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. SpringerPlus, 5(1), 1-22.

[8]Liu, Y., Cao, X., & Yu, Y. (2016, September). Are you influenced by others when rating? Improve rating prediction by conformity modeling. In Proceedings of the 10th ACM conference on recommender systems (pp. 269-272).

[9]Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. Machine learning, 42(1), 177-196.

[10]https://radimrehurek.com/gensim/(online)

[11]https://spacy.io/(online)

[12]Ramamonjisoa, D. (2014, March). Topic modeling on users's comments. In 2014 third ICT international student project conference (ICT-ISPC) (pp. 177-180). IEEE.

[13]Rita, P., Oliveira, T., & Farisa, A. (2019). The impact of e-service quality and customer satisfaction on customer behavior in online shopping. Heliyon, 5(10), e02690.

[14]Röder, M., Both, A., & Hinneburg, A. (2015, February). Exploring the space of topic coherence measures. In Proceedings of the eighth ACM international conference on Web search and data mining (pp. 399-408).
[15]Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. In Handbook of latent semantic analysis (pp. 439-460). Psychology Press.[8]A review of topic modeling methods (IkeVayanskya,Sathish A.P.Kumar)
[16]Tadelis, S. (2016). Reputation and feedback systems in online platform markets. Annual Review of Economics, 8, 321-340.
[17]Tong, Z., & Zhang, H. (2016, May). A text mining research based on LDA topic modelling. In International conference on computer science, engineering and information technology (pp. 201-210).
[18]Yang, T. I., Torget, A., & Mihalcea, R. (2011, June). Topic modeling on historical newspapers. In Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (pp. 96-104).