Summarization and Sentiment Analysis of Reviews

Aayush Indrapratap Singh
North Carolina State University
asingh48@ncsu.edu

Bhavya Omprakash Agrawal
North Carolina State University
bagrawa@ncsu.edu

Diksha Paliwal
North Carolina State University
dpaliwa@ncsu.edu

1 Background and Introduction

Visits to online businesses have surged rapidly as internet access has expanded to remote regions. As a general tendency, people seek suggestions and feedback from the prior customers to determine whether or not to acquire goods online. This project will contain a method for providing users with concise and accurate product reviews.

The project's goal is to create a model that accurately summarizes reviews, which will be useful for both customers and sellers who can use the data to enhance their services and products. It is also taking into consideration two kinds of people, one sample who gives a low rating with a negative review and others who give a low rating for an overall positive review. We intend to overcome this disparity with our summarizing and sentiment extraction.

Furthermore, reading long reviews requires more time and effort, but these long reviews contain the maximum information. Most individuals simply overlook these helpful reviews due to their length. Because of this, the buyer might not know about all the aspects including the positive as well as the negative points, that they should be considering, before buying the product. People will be able to get the information more quickly if they were provided with the summarized reviews.

Doing this would allow the customer to read less data but still gain the most important information needed to decide on the product. It would also improve productivity by speeding up the surfing process of the user.

2 Related Work

Silva et. Al.[4] have discussed that the similarity between two words can be calculated by measuring the distance between the text, a similarity value zero will be assigned if the two words occur concurrently. Different similarity values will be assigned based on how far apart the words are. The semantic similarity i.e. the likeliness of words having similar meaning, is calculated based on the probability of the words occurring together. If the frequency of words occurring together is high, a higher similarity value is assigned to the words. A variation to this can be using semantic similarity based on normalized distance, which is determined by comparing the word with a group of keywords.

Ordonez et. al.[2] proposed a method of detecting the relevance of words with its neighbouring text and collecting them in a semantic area as a matrix. The values of the matrix would express the relationship between the corresponding words in rows and columns. In this method, a weight is assigned to each element based on the similarity with its neighbours. Similarity of words is obtained by employing values to the respective vector of words and calculating the relatedness by applying the cosine measure. The calculation is based on the word location and similarity.

Jain et. al.[1] discuss abstractive text summarization using Structured based approaches and Semantic based approaches. Structured based approaches extract the most important information from text using templates, extraction rules and structures like trees and ontology. Whereas semantic based approaches capture the concept in the text and forms the relation between these concepts. Some of the semantic based approaches explained:

- Multimodal Semantic Model: It captures the concepts of the text and finds the relation between different concepts and are presented as new sentences.
- Information Item based method: In this method the summary is generated using the abstract representation of the original text. The abstract representation is nothing more than a text's smallest information item, the information item.
- Semantic Graph based method: In this method the summary is generated by creating a semantic graph called rich semantic graph (RSG) on the original text and then condensing the semantic graph to create the final abstractive summary.

Vaswani et. al. [5] introduced the use of attention mechanism along with encoder - decoder model.The attention mechanism extracts information from the all past encoder states, which allows the decoder to assign weights according to all states rather than just the last state.

Xu Yun et al. [6] from Stanford University applied existing supervised learning algorithms such as perceptron algorithm, naive bayes and supporting vector machine to predict a review's rating on Yelp's rating dataset. They used hold out cross validation using 70% data as the training data and 30% data as the testing data. The author used different classifiers to determine the precision and recall values.

Callen Rain [3] proposed extending the current work in the field of natural language processing. Naive Bayesian and decision list classifiers were used to classify a given review as positive or negative. Deep-learning neural networks are also popular in the area of sentiment analysis.

3 Method

3.1 Encoder-Decoder LSTM seq2seq model

The Encoder-Decoder architecture is mainly used to solve the sequence-to-sequence (Seq2Seq) problems where the input and output sequences are of different length and type. The objective is to build a text summarizer where the input is a long sequence of words (in a text body), and the output is a short summary (which is a sequence as well). So, we can model this as a Many-to-Many Seq2Seq problem.

There are two major components of a Seq2Seq model:

1. Encoder
2. Decoder

The Encoder and Decoder Models are developed using LSTM Layers. Long Short Term Memory (LSTM) layers help process time-series data like sentences by maintaining states and propagating information through the layer. These are also capable of capturing long term dependencies by overcoming the problem of vanishing gradient.

3.1.1 Encoder

An Encoder LSTM model reads the entire input sequence wherein, at each time step, one word is fed into the encoder. It then processes the information at every time step and captures the contextual information present in the input sequence. The diagram 1 illustrates this process:

The hidden state (hi) and cell state (ci) of the last time step are used to initialize the decoder. Remember, this is because the encoder and decoder are two different sets of the LSTM architecture.

3.1.2 Decoder

The decoder is also an LSTM network which reads the entire target sequence word-by- word and predicts the same sequence offset by one time step. The decoder is trained to predict the next word in the sequence given the previous word.

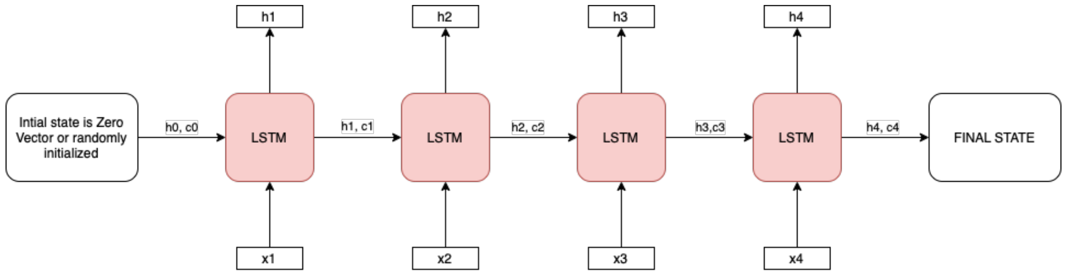

Figure 1: Encoder Model

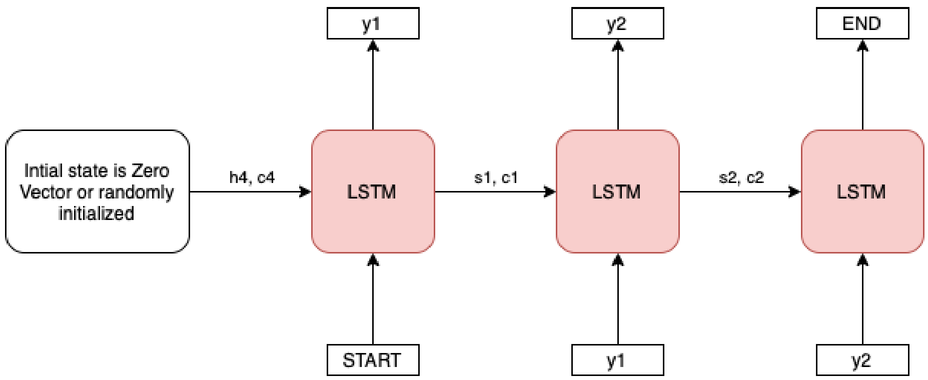

Figure 2: Decoder Model

<start> and <end> are the special tokens which are added to the target sequence before feeding it into the decoder. The target sequence is unknown while decoding the test sequence. So, we start predicting the target sequence by passing the first word into the decoder which would be always the <start> token. And the <end> token signals the end of the sentence.

3.1.3 Attention Layer

It is well-known that a few words contribute more to the meaning of the paragraph than other words do. This can be reasoned as follow: The adjectives and action verbs generally hold more meaning in product reviews than prepositions or other part of speeches. The goal of summarization is to capture the essence which is highly hidden in these part-of-speeches.

This raises a demand of focusing on keywords in the review, which can be achieved by using an Attention Layer. The Attention Layer utilizes the decoder outputs and encoder outputs to attend over all positions in the input sequences and helps it understand the context over the training period.

The attention mechanism aids in the memorizing of long source texts. It generates shortcuts between the context vector and the full source input, rather than creating a single context vector out of the encoder's previous concealed state. For each output element, the weights of these shortcut links can be changed. The context vector has access to the complete input sequence, thus it remembers the previous states. The context vector learns and controls the alignment between the source and destination.

3.2 Sentiment Analysis

Sentiment analysis models detect polarity within a text (e.g. a positive or negative opinion), whether it's a whole document, paragraph, sentence, or clause.

Understanding people's emotions is essential for businesses since customers are able to express their thoughts and feelings more openly than ever before. By automatically analyzing customer feedback, from survey responses to social media conversations, brands are able to listen attentively to their customers, and tailor products and services to meet their needs. The service and the product review's polarity is the rating the user provides for that review. The Good Reviews are those with rating 5 stars and 4 stars, and Bad Reviews are those with rating 3 stars, 2 stars and 1 star. Finally, when a feature sentiment is extracted the sentiment phrase is sent to a polarizer method, this method basically returns 1 if the phrase is a positive sentiment else 0 if the phrase is a negative sentiment.

4 Plan & Experimental Setup

4.1 Plan

The Primary goal of this project is to generate concise and meaningful summaries of the reviews. Secondary goal is to extract sentiment from the reviews.

Apart from these goals, the project would aim to test the hypothesis that the Title column in the data is not an accurate summary representation of the review, and thus accuracy tested on this column and the predicted summary should be low.

4.2 Dataset & Pre-Processing

The Project utilizes the Amazon Reviews Dataset from Kaggle (link to dataset) which contains the rating, title and reviews of approximately 3 million amazon customer reviews in the training set and approximately 600,000 amazon customer reviews in the testing set.

	Rating	Title	Review
0	5	Inspiring	I hope a lot of people hear this cd. We need m...
1	5	The best soundtrack ever to anything.	I'm reading a lot of reviews saying that this ...
2	4	Chrono Cross OST	The music of Yasunori Misuda is without questi...
3	5	Too good to be true	Probably the greatest soundtrack in history! U...
4	5	There's a reason for the price	There's a reason this CD is so expensive, even...
...	...	...	...
2999994	1	Don't do it!!	The high chair looks great when it first comes...
2999995	2	Looks nice, low functionality	I have used this highchair for 2 kids now and ...
2999996	2	compact, but hard to clean	We have a small house, and really wanted two o...
2999997	3	Hard to clean!	I agree with everyone else who says this chair...
2999998	1	what is it saying?	not sure what this book is supposed to be. It ...

2999999 rows × 3 columns

Figure 3: Training Dataset

	Rating	Title	Review
0	4	Surprisingly delightful	This is a fast read filled with unexpected hum...
1	2	Works, but not as advertised	I bought one of these chargers..the instructio...
2	2	Oh dear	I was excited to find a book ostensibly about ...
3	2	Incorrect disc!	I am a big JVC fan, but I do not like this mod...
4	2	Incorrect Disc	I love the style of this, but after a couple y...
...	...	...	...
649994	5	Pretty Cool!	We got it for our mom's birthday. She LOVES it...
649995	5	great cd	this cd is very good. i especially love "cats ...
649996	2	An interesting look into Boston's comedy clubs	This was a good documentary on the history of ...
649997	5	Du vol...pour les cowboys!	Avez-vous déjà vu un CD double et un DVD avec ...
649998	4	A Companion Read To GUNS, GERMS, AND STEEL	If you like books that offer explanations for ...

649999 rows × 3 columns

Figure 4: Testing Dataset

The project is utilizing several techniques mentioned below to preprocess the data into the format needed for summarization and sentiment analysis.

1. **Convert all text to lowercase**: The lower() method returns the lowercase string from the given string. It converts all uppercase characters to lowercase. So that the input is uniform.
2. **Removing all punctuation**: Format words and removing unwanted characters like ,$,&, etc, since these are not relevant to the meaning of the sentence.
3. **Removing stop words**: Removing unnecessary words by using English stop words imported from spacy-de library. Stopwords include any word in the sentence which do not provide any meaningful insight to the project's data analysis.
4. **Word Embeddings** : It is the process of representing words as real valued vectors such that words with similar meaning have similar encoding. There were two possible approaches to implement word embeddings : Numberbatch embedding and GloVe embedding, it was observed that GloVe embedding produced better results for the input dataset.
5. **Tagging** : Adding start and end tokens in the sentences to mark the start and ends of sentences.
6. **Tokenization**: Tokenization is the process of tokenizing or splitting a string, text into a list of tokens. It is done in order to split the input sentences into word tokens.
7. **Text to sequence**: It is the process of converting input sentences in to sequence of numbers which can be fed to the model.
8. **Padding**: The process of making the length of all the input sequences equal is called padding. As the model requires the inputs to be of same shape and size, padding is necessary.

4.3 Exploratory Data Analysis

Based on the Data Visualization shown below, it can be inferred that the data is very balanced and proportionate.

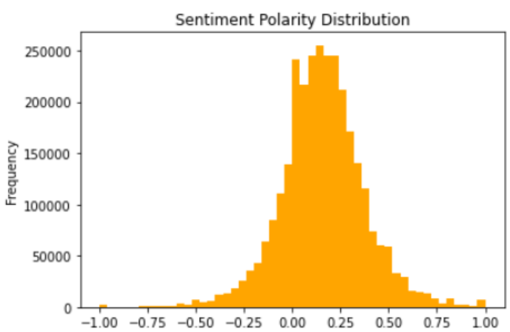

Figure 5: Sentiment Polarity Distribution

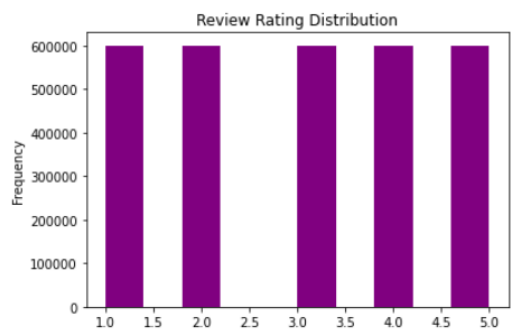

Figure 6: Review Rating Distribution

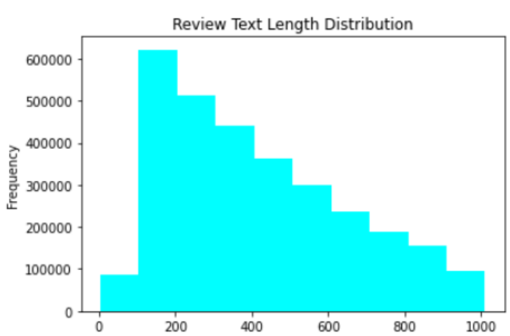

Figure 7: Review Text Length Distribution

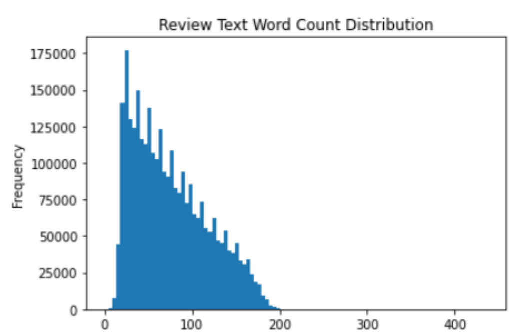

Figure 8: Review Text Word Count Distribution

4.4 Experimental Design

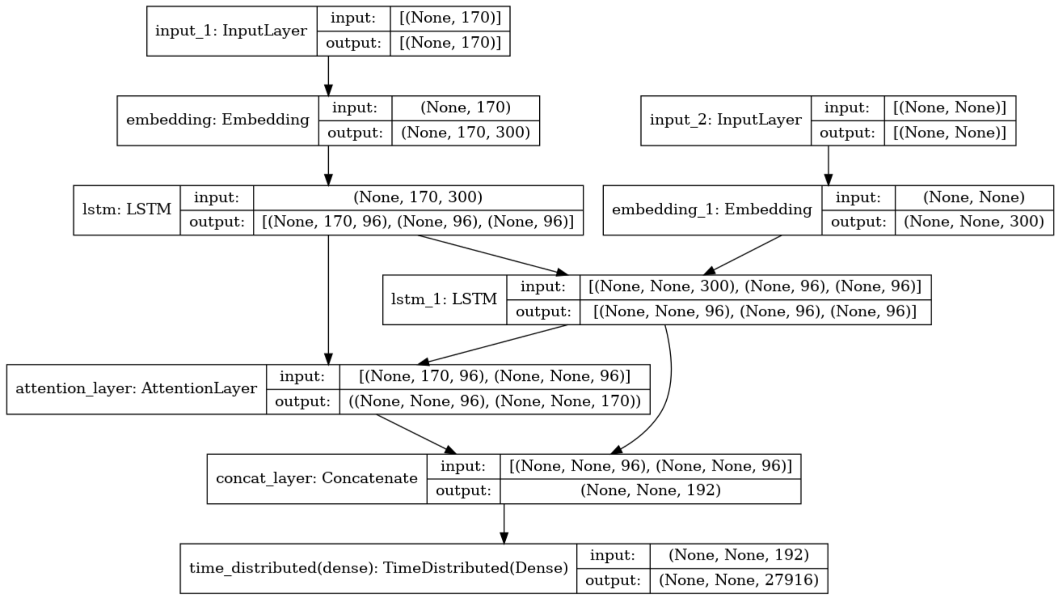

Figure 9: Summarization Model

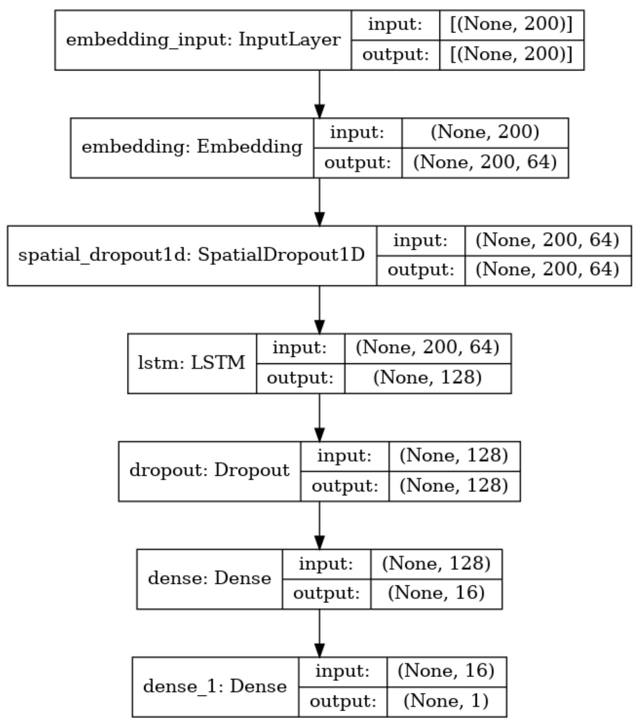

Figure 10: Sentiment Model

Summarization Model is a Seq2Seq Model developed using Embedding Layer, LSTM Layer, Attention Layer and Time Distributed Dense Layer as show in Figure 9. The Embedding layer is used to convert input words into corresponding Vector using the Word Embeddings as weights of the layer.

The Outputs of Encoder and Decoder model are fed to the attention layer and then feed forwarded to Dense layers for the Training Phase. For the Inference Phase this Dense Layer acts as the Context

Phase. The input review is fed to encoder and the decoder layer is made to output the Summary by reversing the states and links.

Sentiment Analysis Model is a Shallow Neural Network with similar layers as Summarization Model, however this is a simple feed forward network for binary classification. The architecture can be seen in Figure 10

Hypothesis Testing: To test the previously mentioned hypothesis, accuracy of the predicted summaries is tested using ROUGE - N metric, which tests for overlap between between N-gram reference sentence i.e. the Title column and the predicted summaries. This ROUGE - N metric score should be low to prove our hypothesis right. And BLUE score is calculated to evaluate the summarization model.

5 Results

5.1 Sample Outputs

Review 1: let preface review saying fan patricia cornwell scarpetta series book however books series disappointed mystery features cast characters scarpetta mysteries however crazy sadist gault appears caricature person one formulaic clue another becomes unbelievable protagonist kay scarpetta shop worn new characters introduced never developed marino politically incorrect chain smoking cop real endearing character book supposedly mystery book drops clues sky line fantasy

Original summary: disappointing not up to other books in this series
Predicted summary: not as good as the first
Predicted sentiment: Negative

Review 2: although enjoy patricia cornwell style and would loved read another books cannot stand better thou militaristic amoral atheistic adulterous lascivious feminism that promoted book mother and sister feminists go church every week care children husbands remain unconditionally faithful spouses carry gently loving most importantly desire equal rights not superiority men

Original summary: adulterous amoral atheistic feminist saves the day
Predicted summary: a must read
Predicted sentiment: Positive

Review 3: one disappointing purchases made amazon book one long myopic libertarian apology

Original summary: no lessons learned
Predicted summary: not what i expected.
Predicted sentiment: Negative

5.2 Outcome

Since abstractive summarization was performed, the output was expected to be newly generated sentences from the input reviews which was achieved and can be seen in the output sample below. ROUGE- N score of 0.26 was calculated for the predicted summaries which indicates that our hypothesis is correct.

BLEU Score: Bilingual Evaluation Understudy Score is a metric score that evaluates a generated sentence to a reference sentence. It takes into account how many of the words present in the reference sentence are present in the generated sentence. Thus, to evaluate the summarization model BLUE score was the best choice.

BLEU Score achieved for the above summarization model : **58.76**

To measure the accuracy of the sentiment analysis model, the predicted sentiment was compared to the Rating of the review, where a Rating <= 3 implied a Negative sentiment and a Rating > 3 implied a Positive sentiment. Accuracy is the chosen metric for this model, as equal significance is required for the positive as well as the negative Sentiment Prediction.

Accuracy of above sentiment analysis model : **90.8 %**

6 Conclusions

Abstractive summarization model was expected to output a meaningful summary of the input reviews, which was achieved, whereas if extractive summarization was used, the output summaries would not have been able to capture the meaning of the text as well as abstractive ones. BLEU score is a good measure to evaluate the abstractive summarization model but human verification is still a much better way to evaluate the results. To improve the model, training could be done using proper summaries of the review rather than using the title because the title column doesn't necessarily capture the summary of the review.

Sentiment Analysis model is able to capture the correct sentiment of 90.8 % of the reviews when it is trained using the Rating column as the polarity decider, with Rating = 3 as the threshold. A survey done over 500 reviews showed that many reviews with low ratings have positive sentiment overall and reviews with high ratings have negative sentiment, and since we are using Ratings to validate and verify our predicted sentiment, the low accuracy of the model is explained.

For future scope, instead of summarizing individual reviews, we could output the clubbed summary of all reviews for a particular product.

7 Github

https://github.ncsu.edu/asingh48/engr-ALDA-fall2021-P26

References

[1] 2021. *ICMLC 2021: 2021 13th International Conference on Machine Learning and Computing* (Shenzhen, China). Association for Computing Machinery, New York, NY, USA.

[2] Carlos Ordonez, Yiqun Zhang, and S. Lennart Johnsson. 2018. Scalable machine learning computing a data summarization matrix with a parallel array DBMS. *Distributed and Parallel Databases* (2018), 1–22.

[3] Callen Rain. 2013. Sentiment analysis in amazon reviews using probabilistic machine learning. *Swarthmore College* (2013).

[4] Gabriel Silva, Rafael Ferreira, Rafael Dueire Lins, Luciano Cabral, Hilário Oliveira, Steven J. Simske, and Marcelo Riss. 2015. Automatic Text Document Summarization Based on Machine Learning. In *Proceedings of the 2015 ACM Symposium on Document Engineering* (Lausanne, Switzerland) *(DocEng '15)*. Association for Computing Machinery, New York, NY, USA, 191–194. `https://doi.org/10.1145/2682571.2797099`

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. `https://arxiv.org/pdf/1706.03762.pdf`

[6] Yun Xu, Wu; Xinhui, and Wang; Qinxia. 2015. Sentiment Analysis of Yelp's Ratings Based on Text Reviews. (2015).