Shahjalal University of Science and Technology
Department of Computer Science and Engineering

Aspect Based Sentiment Analysis On Ecommerce Product Reviews

MD. TARIFUL ISLAM

Reg. No.: 2018331042

4^{th} year, 1^{st} Semester

RAISA FAIROOZ

Reg. No.: 2018331050

4^{th} year, 1^{st} Semester

Department of Computer Science and Engineering

Submitted To

MOHAMMAD SHAHIDUR RAHMAN, PHD

Professor

Department of Computer Science and Engineering

18th October, 2023

Abstract

Sentiment analysis (SA) is a fundamental task in natural language processing (NLP). It involves the identification and classification of the sentiment (i.e., positive, negative, or neutral) expressed in a piece of text. Aspect-based sentiment analysis (ABSA) takes things a step further by identifying the sentiment of a specific aspect of a product or service, such as its price, quality, or customer service. ABSA is a challenging task, especially in languages like Bangla, where there is a lack of proper datasets and resources. In this project, we address this challenge by collecting a new dataset of Bangla, English, and Banglish e-commerce reviews. We then train a baseline LSTM and an Attention-based LSTM model for sentiment polarity finding, which is one of the four tasks of ABSA. We get F1 scores of 76.8% and 75.7% on these models respectively. The results of our experiments show that our model can achieve a competitive performance on the new dataset. This project has both business and academic implications. On the business side, ABSA can be used to improve the customer experience by identifying and addressing negative sentiments. On the academic side, this project can contribute to the development of new methods and resources for ABSA in Bangla and English.

Keywords: ABSA, LSTM, Attention-based LSTM

Contents

List of Tables

List of Figures

Chapter 1

Motivation

Sentiment analysis (SA) is a fundamental task in natural language processing (NLP). It involves the identification and classification of the sentiment (i.e., positive, negative, or neutral) expressed in a piece of text. The typical approach for SA focuses on sentence-level polarity. This means that the sentiment of the entire sentence is determined, regardless of the specific aspects that are being discussed. Aspect-based sentiment analysis (ABSA) is a task that takes things a step further by identifying the sentiment of a specific aspect of a product or service, such as its price, quality, or customer service. This is important because it allows us to understand the customer's opinion about the different aspects of a product, rather than just the overall sentiment of the review.

Reviews on e-commerce websites are a good channel for communicating with business owners and communities. These reviews often offer mixed sentiments. For example, the sentence "Product quality is decent but the packaging was really bad" gives a negative polarity when the sentiment is analyzed on a sentence level. However, the review for packaging overpowers the positive or neutral opinion about the actual product. Hence, with the typical approach of sentiment analysis, we often lose important relevant information.

ABSA is a growing research interest and has been gaining traction in recent years. The SemEval competition has been a major driver of this growth, as it has provided a common benchmark for researchers to compare their methods.

Extracting the aspect-based opinion is a crucial task. It has a significant impact on businesses.

Businesses can leverage user reviews to focus on the different aspects of the business. In the above example, it is important for the business to know which part of the business they need to improve and what works fine. In the above example, if the customer is unhappy with the packaging, the business can focus on improving the packaging. ABSA has a direct real-world impact on the business environment. Businesses can use ABSA to identify areas where they need to improve their products or services, target their marketing campaigns more effectively, improve customer satisfaction, build a better reputation.

Several approaches have been developed for performing ABSA in English and some other languages. But for Bangla, there is a lack of annotated datasets and corpus. Moreover, a majority of the e-commerce users in Bangladesh provide reviews in both Bangla and English as well as the romanization of Bangla using the English alphabet, commonly referred to as ”Banglish”.

The following four subtasks of ABSA was defined in SemEval 2014:

- Aspect term extraction: This task involves identifying the words or phrases in a sentence that refer to a specific aspect of a product or service. For example, in the sentence ”The food was delicious but the service was slow,” the aspect terms are ”food” and ”service.”
- Aspect term polarity: This task involves determining the sentiment polarity of the aspect terms. For example, the word ”delicious” is positive and the word ”slow” is negative.
- Aspect category extraction: This task involves identifying the higher-level category of the aspect terms. For example, the aspect terms ”food” and ”service” could be categorized as ”quality” or ”customer service.”
- Aspect category polarity: This task involves determining the sentiment polarity of the aspect categories. For example, the category ”quality” is typically positive and the category ”customer service” can be either positive or negative.

Chapter 2

Objectives

The project encompasses the following key objectives:

Aspect Term Polarity Extraction

The primary objective of this project is to address the second task within the subtask list, which involves performing aspect term polarity extraction. In this task, the project aims to design and implement a computational system capable of analyzing given sentences and identifying aspect terms within them. Subsequently, the system will determine the polarity associated with each identified aspect term, classifying it as positive, negative, or neutral.

Development of Multilingual Dataset

A vital undertaking of this project is the creation of a comprehensive dataset that spans three distinct linguistic forms: Bangla, English, and Banglish. This dataset will be meticulously curated to encompass sentences that reflect the diversity of linguistic usage, including instances of code-switching between Bangla and English.

Through the accomplishment of these objectives, the project seeks to contribute to the advancement of sentiment analysis within a multilingual and code-switched context, enabling a deeper understanding of sentiment expressions that transcend language boundaries.

Chapter 3

Related Work

Aspect-Based Sentiment Analysis (ABSA) has garnered attention across multiple languages, including initiatives such as SemEval's ABSA tasks held over consecutive years. The restaurant and laptop datasets that were used in SemEval 2014 [1] have established themselves as key benchmarks for ABSA evaluation. These datasets, derived from real-world reviews, offer valuable insights into sentiment analysis within specific aspects. However, their applicability beyond their original domains and languages may require careful consideration due to potential biases and limitations.

Limited work has been done in Bangla over the years. Due to the lack of proper datasets, advancements have been limited. But this is also a field with growing interest. In 2018, an annotated dataset for performing ABSA was published. It had two domains - namely, Cricket dataset and Restaurant dataset. [2] These datasets were annotated based on aspect category and their respective polarities. The cricket dataset was collected from Facebook comments regarding cricket. It had five categories - team, team management, batting, bowling, and others. The restaurant dataset was obtained by translating the original restaurant benchmark dataset. This particular dataset was used by several researchers for different ABSA tasks. They used three machine learning models - SVM, KNN and Random Forest for evaluation.

In 2020, Haque et al. used this dataset and trained two additional models - Linear regression and Naive Bayes and found best F1 score of 0.37.

Later in 2020, a new dataset was developed by collecting sentences from newspapers. The

aspect categories and polarities were tagged based on 4 categories - Politics, Sports, Religion and Others. [3] A CNN approach gave better accuracy than others, but the BiLSTM outperformed the CNN in all other measures - recall(78.33%), precision(80.46%) and F1 score of 79.38.

In 2021, Naim used the cricket and restaurant dataset and steered his experiment towards corresponding term extraction based on weight assignments on parts of sentences. [4] He employed CNN and found it to perform better than other methods.

Pretrained models like BERT was also employed for two of the tasks. [5] They also developed a dataset from Youtube comments of Bangladeshi users on five topics: Technology, Corruption, Sports, Politics, and Economy. BERT yielded an impressive f1 score of 0.97 for aspect extraction and 0.77 for sentiment analysis.

The available datasets primarily revolve around domains like news and, occasionally, YouTube comments. However, these datasets often struggle to adequately capture the spectrum of sentiments found in a single e-commerce review. Unlike stand-alone content, e-commerce reviews frequently express a blend of opinions regarding multiple aspects within a single piece of text. This complexity underscores the need for datasets tailored to e-commerce reviews, ensuring a comprehensive approach to aspect-based sentiment analysis within this domain. Moreover, all of the datasets available contain monolingual data only.

Chapter 4

Methodology

The work can be divided into two subtasks: collecting a dataset of reviews from the Daraz platform, followed by thorough text preprocessing to ensure data quality. The following step is to annotate the dataset with aspect words and give polarity labels in order to evaluate sentiment. We choose three polarity labels: positive, negative and neutral. As for the second subtask, we used two models: the Long Short-Term Memory (LSTM) model, recognized for its sequence handling capabilities, and an Attention-based LSTM model, which leverages attention mechanisms for enhanced performance.The process as a whole is depicted in Fig. 4.1

4.1 Data Collection

We used a systematic method to collect a representative sample of evaluations from the Daraz e-commerce website. Our data-gathering technique included two essential requirements that assured the reliability and relevancy of the reviews obtained. To begin, we chose products with the most reviews in order to represent a diverse range of user experiences. This criterion enabled us to collect a wide range of thoughts and views about the platform's numerous offerings. Second, in order to increase the legitimacy of the gathered reviews, we prioritized goods with the most sales.

We collected the reviews with three different forms of text: English, Bangla, and Banglish, in recognition of the multilingual character of interactions with customers. Through the extensive method, it was hoped to capture the complex nature of sentiment and aspect analysis in many languages and language mixtures. The Fig. 4.2 shows four different reviews of products purchased

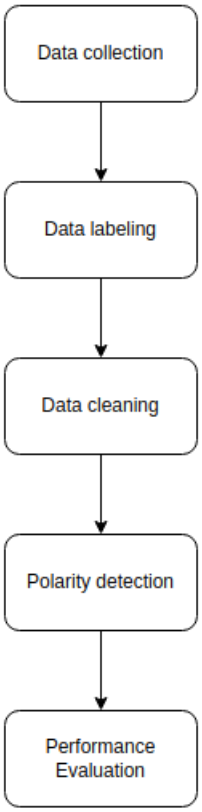

Figure 4.1: Aspect-Term polarity analysis process

from Daraz. The first review is in English and praises the product quality but criticizes the packaging. The second review is in Bangla and expresses mixed feelings about Daraz, the seller, and the coffee taste. The third review is in Banglish and discusses the quality and performance of earphones. The final review is a mixture of Bangla and English and praises Nescafe and Daraz Mall. Upon collecting the multilingual data, we engaged in a preliminary analysis to gain insights into the prevalent themes and terms within the dataset. To visually represent the frequency distribution of words, we generated a word cloud that encapsulated the most prominent terms across all languages. In Fig. 4.3 visualising the structure of the most prominent words from these three language forms.

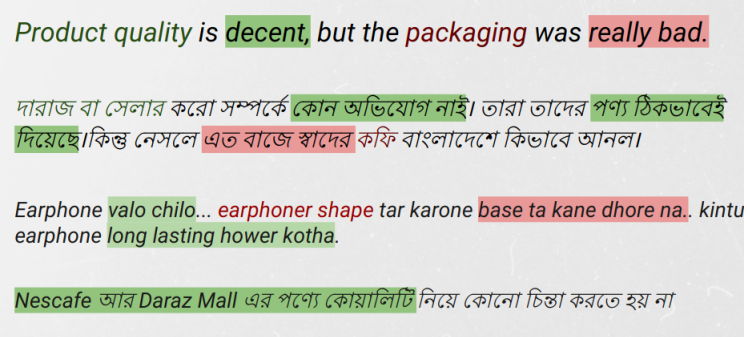

Figure 4.2: A diverse range of customer reviews on Daraz

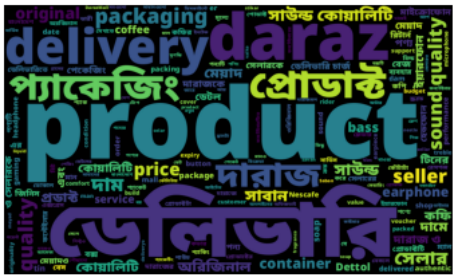

Figure 4.3: Representing the different forms of word appearances

4.2 Data Labeling

To provide our dataset with correct aspect term annotations and polarity categorizations, the data labeling procedure was carefully carried out. To begin with, multiline sentences inside reviews were manually transformed into single-line phrases to ensure coherence and consistency in the dataset. Following that, each review was individually annotated.

The initial step was to find the aspect phrase in each sentence. To extract the experience being addressed, a complete understanding of the background was necessary. After identifying the aspect phrase, the following step was to mark a suitable polarity category to express the feeling associated with that aspect. Our polarity categories included positive, negative, and neutral. In the final dataset, we annotated 970 reviews which are presented in three columns. One is for the reviews, the second one is for aspect-term and the final one is for polarity category. Fig. 4.4 contains a sample of our dataset.

1	SENTENCES	ASPECT TERM	POLARITY
2	Everything looks good and sound quality is also good but may not be the original QKZ. Because this box is not like the original.	looks	Positive
3	Everything looks good and sound quality is also good but may not be the original QKZ. Because this box is not like the original.	sound quality	Positive
4	Everything looks good and sound quality is also good but may not be the original QKZ. Because this box is not like the original.	originality	Negative
5	Everything looks good and sound quality is also good but may not be the original QKZ. Because this box is not like the original.	box	Negative
6	this is very good not bad and sound blocking also good.overall 9/10 I recommend .and the delivery rider is very very much much good .he is really very good person.and everything good.	sound blocking	Positive
7	this is very good not bad and sound blocking also good.overall 9/10 I recommend .and the delivery rider is very very much much good .he is really very good person.and everything good.	delivery rider	Positive
8			
9	আলহামদুলিল্লাহ খুবই সুন্দর প্রডাক্ট, সাউন্ড কোয়ালিটি তো মাস আল্লাহ, ইয়ারফোনটি আমার কাছে খুবই ভালো লেগেছে। ১০ এর মধ্যে আমি ১০ দিবো, মাত্র ২৪৬ টাকায় এই ইয়ারফোনটি পাইছি। চাইলে আপনারাও নিতে পারেন অনেক ভালো ইয়ারফোন।thank you so much darez..	সাউন্ড কোয়ালিটি	Positive
10	আলহামদুলিল্লাহ খুবই সুন্দর প্রডাক্ট, সাউন্ড কোয়ালিটি তো মাস আল্লাহ, ইয়ারফোনটি আমার কাছে খুবই ভালো লেগেছে। ১০ এর মধ্যে আমি ১০ দিবো, মাত্র ২৪৬ টাকায় এই ইয়ারফোনটি পাইছি। চাইলে আপনারাও নিতে পারেন অনেক ভালো ইয়ারফোন।thank you so much darez..	প্রডাক্ট	Positive

Figure 4.4: Sample of the labeled dataset

4.3 Model Training

We took advantage of the strengths of two cutting-edge models, Long Short-Term Memory (LSTM) and Attention-based LSTM, in pursuing the goal of precise polarity detection within the complex environment of aspect term extraction from Daraz product reviews. These models were chosen to effectively represent the complex nature of sentiment expression across languages due to the variety of linguistic forms included in our dataset, including English, Bangla, and Banglish.

LSTM: The Long Short-Term Memory (LSTM) model, a type of recurrent neural network, was employed to analyze sequential data in our aspect term extraction and sentiment analysis task. LSTMs are known for their ability to capture long-range dependencies in sequences, making them suitable for handling the contextual nature of text data. Our LSTM model consisted of interconnected memory cells that allowed the model to retain and selectively update information over varying time steps. The LSTM architecture shown in Fig. 4.5

Attention-based LSTM: The Attention-based LSTM model, a refinement of the traditional LSTM, was a pivotal component of our aspect term extraction and sentiment analysis. This model enhanced the LSTM's performance by dynamically focusing on different parts of the input text during processing. The attention mechanism allowed the model to weigh the importance of each word or token within a sentence, enabling it to give more attention to relevant information.

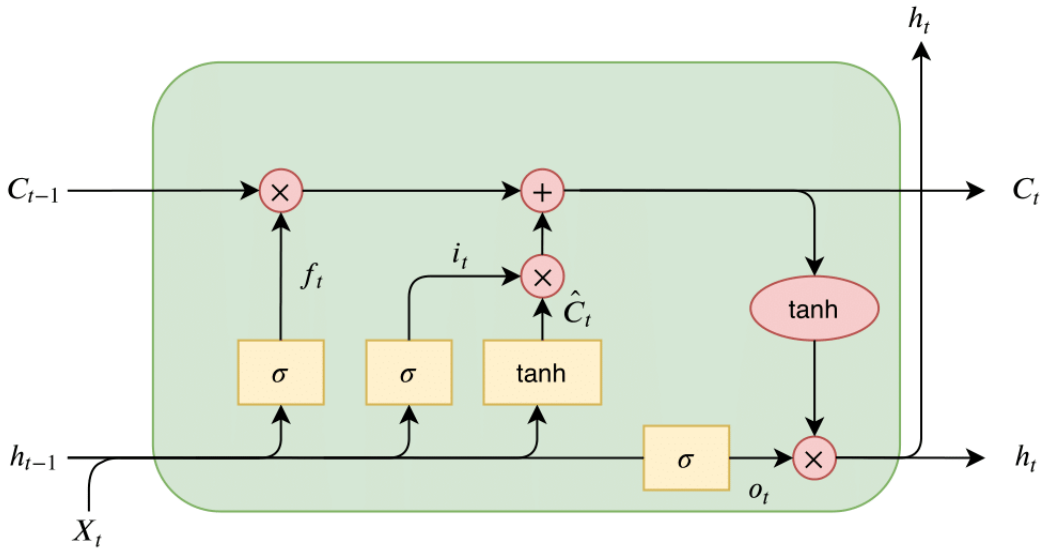

Figure 4.5: LSTM cell architecture, here Xt: input time step, ht: output, Ct: cell state, ft: forget gate, it: input gate, ot: output gate, $\hat{C}$t : internal cell state. Operations inside light red circle are pointwise.

Chapter 5

Results and Discussion

5.1 Preparing Environment

Deep learning-based experiments need powerful CPU and GPU. In order to conduct experiments, we used Kaggle notebook.

5.2 Result

After running the LSTM and Attention based LSTM we get the shown in the Table 5.1 accuracy 76.8% and 75.7% respectively.

Model	**Accuracy**
LSTM	76.8%
Attention-based LSTM	75.7%

Table 5.1: Model Accuracy

we calculated the loss value and accuracy during the training and testing process for the both model. After running the LSTM model for 15 epoch we got The lower graph on the Fig. 5.1 shows the accuracy of the train and test sets over the course of training. The accuracy of the train set starts out low and then increases over time. This is because the model is learning to make better predictions on the train set. The accuracy of the test set starts out higher than the accuracy of the train set, but then it plateaus or even decreases.

The upper graph on the 5.1 shows the loss of the train and test sets over the course of training.

The loss of the train set starts out high and then decreases over time. This is because the model is learning to make better predictions on the train set, which means that the difference between the predicted values and the actual values is getting smaller. The loss of the test set starts out high and then decreases over time, but it does not decrease as much as the loss of the train set. This is because the model is starting to overfit the train set and is not generalizing well to the test set.

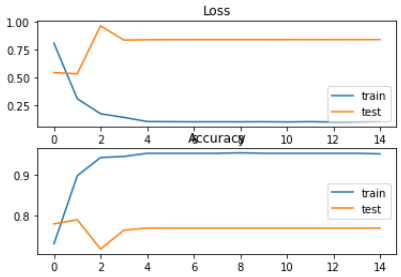

Figure 5.1: Loss and Acuracy curve during the training process of LSTM model

on the Attention-based LSTM part we run it on 5 epoch and The upper graph Fig. 5.2 shows the loss of the train and test sets over the course of training. The loss of the train set starts out high and then decreases over time. This is because the model is learning to make better predictions on the train set, which means that the difference between the predicted values and the actual values is getting smaller. The loss of the test set starts out high and then decreases over time, but it does not decrease as much as the loss of the train set. The lower graph 5.2 shows the accuracy of the train and test sets over the course of training. The accuracy of the train set starts out low and then increases over time. This is because the model is learning to make better predictions on the train set. The accuracy of the test set starts out higher than the accuracy of the train set, but then it plateaus or even decreases

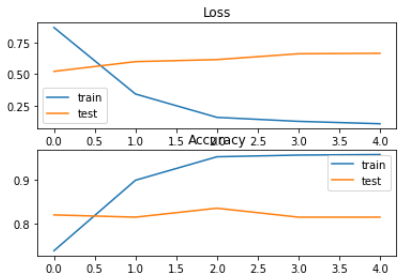

Figure 5.2: Loss and Acuracy curve during the training process of the Attention-based LSTM model

We can observe that the model exhibits overfitting tendencies. To address this, adjustments in hyperparameters are necessary, alongside an expansion of our dataset's size.

Chapter 6

Conclusion

In this report, we have presented a study on aspect-based sentiment analysis in Bangla. We collected a new dataset of Bangla, English, and Banglish ecommerce comments. We then trained a baseline LSTM model for sentiment polarity finding, which is one of the four tasks of ABSA. The results of our experiments show that our model can achieve a competitive performance on the new dataset.

Our study has several limitations. First, our dataset is relatively small. Second, our model is a simple baseline model. Third, we did not compare our model with other state-of-the-art models.

Despite these limitations, our study makes several contributions. First, we present a new dataset of Bangla, English, and Banglish ecommerce comments. Second, we show that a simple baseline model can achieve a competitive performance on this dataset. Third, we provide a baseline for future research on ABSA in Bangla.

In future work, we plan to collect a larger dataset of Bangla ecommerce comments. We also plan to explore more sophisticated models for ABSA in Bangla.

We believe that our work is a step towards the development of effective ABSA systems for Bangla. ABSA systems can be used by businesses to improve their products and services, as well as to better understand the needs of their customers.

Chapter 7

Future Work

Moving forward, several opportunities emerge for enhancing and extending the current study on aspect-based sentiment analysis (ABSA):

- **Aspect Category Extraction:** Expanding the analysis to encompass aspect category extraction will provide a more comprehensive understanding of sentiment dynamics by identifying higher-level thematic clusters.
- **Dataset Expansion:** Increasing the dataset's size and diversity will bolster model robustness and generalization capabilities, enabling more accurate sentiment analysis.
- **Data Augmentation:** Employing data augmentation techniques, such as synonym replacement and paraphrasing, can enrich training data and improve model adaptability.
- **Applying other Machine Learning and Deep learning Models:** Applying other models such as SVM, KNN has provided good performance for some related works. Also deep learning models can be implemented, as they have also shown competent accuracies for similar tasks.
- **Fine-Tuning Pretrained Models:** Fine-tuning pretrained language models like BERT and GPT can enhance performance, necessitating exploration of optimal architectures and domain-specific techniques.

References

[1] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. AL-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. M. Jiménez-Zafra, and G. Eryiğit, "SemEval-2016 task 5: Aspect based sentiment analysis," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. San Diego, California: Association for Computational Linguistics, Jun. 2016, pp. 19–30. [Online]. Available: https://aclanthology.org/S16-1002

[2] M. A. Rahman and E. Kumar Dey, "Datasets for aspect-based sentiment analysis in bangla and its baseline evaluation," *Data*, vol. 3, no. 2, p. 15, 2018.

[3] M. Ahmed Masum, S. Junayed Ahmed, A. Tasnim, and M. Saiful Islam, "Ban-absa: An aspect-based sentiment analysis dataset for bengali and its baseline evaluation," in *Proceedings of International Joint Conference on Advances in Computational Intelligence*, M. S. Uddin and J. C. Bansal, Eds. Singapore: Springer Singapore, 2021, pp. 385–395.

[4] F. A. Naim, "Bangla aspect-based sentiment analysis based on corresponding term extraction," in *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, 2021, pp. 65–69.

[5] M. Samia, A. Rajee, M. R. Hasan, M. Faruq, and P. Chandra Paul, "Aspect-based sentiment analysis for bengali text using bidirectional encoder representations from transformers (bert)," *International Journal of Advanced Computer Science and Applications*, vol. 13, 12 2022.