Towards Automatic Sentiment-based Topic Phrase Generation

Raj Ratn Pranesh,[1] Ambesh Shekhar, [1] Sumit Kumar [1]
[1] Birla Institute of Technology, Mesra India
raj.ratn18@gmail.com, ambesh.sinha@gmail.com, sumit.atlancey@gmail.com

Abstract

For obtaining a comprehensive understanding and knowledge of customers'expectations and demands, analysis of user-generated online product and service reviews is very important. Utilizing crucial insights from customers' feedback for developing business strategies can significantly improve product quality. In this paper, we have proposed a pretrained language model based encoder-decoder framework which generates a topic that describes the concise meaning of a customer's review text based on its corresponding polarity. We performed a comparative performance analysis of three language models, namely, BERT, ALBERT and GPT2 for the topic generation task on a dataset containing 8,124 customer reviewers along with topics and it's associates sentiment. In our experiment, we found that GPT2 model outperformed the other two models by achieving lower perplexity of 1.82.

Introduction

With the ever-growing consumer goods and services industry, in order to stay above the competitions, it has become very important for companies to precisely understand the customer's demands and expectations. There exists a huge user-generated data on the Internet that helps potential customers in deciding by taking past customers opinion under consideration. Hence, it becomes very essential for companies to pinpoint custom's issues and address it in a quick and efficient manner.

From the managerial and operating point of view, the majority focus is mostly given toward the understanding of customer review's sentiment, and along with that finding a key driver that triggered a customer's positive or negative sentiment. Considerable amount of work have been done on understanding customer's opinion sentiment (Buche, Chandak, and Zadgaonkar 2013) (Asghar et al. 2014). Manually generating customer analytic are prone to misinterpretations and, at the same time, it is very time consuming, exhaustive and inefficient. Various customer analytics software has been developed that uncovers granular insights about service/product-customer relationship which helps in developing business strategies focusing on solving customer-centric issues. In this paper, we proposed an encoder-decoder based method for generating minute details about a customer's opinion from a given review text and sentiment. We utilized pretrained language models finetuned on our dataset for developing our topic generation models and systematically investigated their performances. We utilized multiple automatic model performance evaluation metrics which are widely used for the task of machine translation. In our analysis, we observed that the GPT2(Radford et al. 2019) model performed superior as compared to BERT(Devlin et al. 2018) and ALBERT(Lan et al. 2019) models. To the best of our knowledge, this paper presents very first work that utilizes customer's review and sentiment for generating sentiment triggering topics.

Methodology

In this section, we presented a detailed discussion about the following: (i) the dataset used in the experiment, and (ii) the proposed encoder-decoder model architecture.

Dataset

We utilized a publicly available Amazon.com product review dataset[1] consisting of customer's feedback over variety of Amazon products. Based on the customer's review and rating, we manually annotated(a team of 5 undergraduates) 8,124 reviewers into two sentiment classes(positive and negative) and also assigned corresponding topic. For example, (a) review: "*The product was delivered on time with perfect packaging*", sentiment: *Positive*, topic: *Delivery* (b)review: "*Disappointing customer service. Zero assistance was given by the customer care representative*", sentiment: *Negative*, topic: *Customer Service*. Out of 8,124 reviews, 4,678 reviews were labelled positive and 3446 reviews were labelled negative. .We divided the dataset into following 3 parts: (i) train dataset(80%), (ii) validation data(10%) and (iii) test data(10%)

Proposed Model Architecture

The figure 1 presents a schematic diagram of encoder-decoder model. During the training, encoder receives input in the form of `[CLS]`sentiment`[SEP]`review`[SEP]`, which is then passed through the sequential encoders layers.

[1]https://www.kaggle.com/datafiniti/consumer-reviews-of-amazon-products

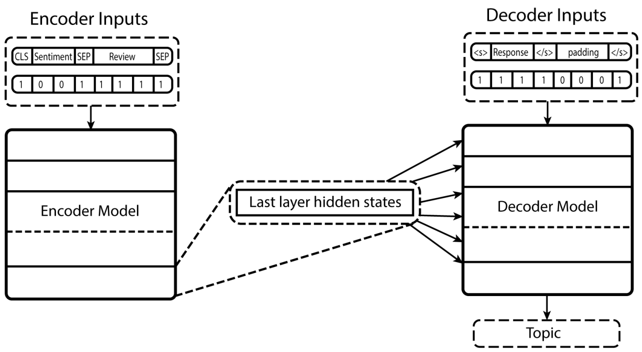

Figure 1: Our proposed language model based encoder-decoder framework architecture.

	ALBERT	**BERT**	**GPT2**
Perplexity	6.11	2.34	**1.82**
BLEU-2	4.87%	7.51%	**8.02%**
BLEU-4	3.01%	8.12%	**9.94%**
METEOR	1.21%	5.75%	**6.71%**

Table 1: Performance score of various models.

The hidden state output of the final encoder layer was then supplied to all the decoder layers. The decoder then generates the topic by predicting the next word in for the output sequence. The encoder-decoder models(BERT-BERT, GPT2-GPT2 and ALBERT-ALBERT) were jointly trained. Models were fine-tuned for 10 epochs with learning rate = 1e-4 and cross entropy loss. The input and output max token length was set at 400 and 4 respectively. The model's hyperparameters were fine-tuned using validation dataset.

Result

We used three automatic metrics evaluation methods, perplexity, BLEU-*n* (Papineni et al. 2002) and METEOR (Lavie and Agarwal 2007). BLEU-*n* and METEOR are popular machine translation evaluation metrics whereas perplexity is used for measuring generated responses smoothness and quality. All the scores were calculated on the test dataset.

As we can see in the table 1, GPT2 reported the best perplexity score of **1.82**, whereas for BERT and ALBERT it was **2.34** and **6.11**. This shows that the GPT2 model's output quality was superior to the other two models. On BLEU-*n* and METEOR metrics, the GPT2 model achieved higher scores, closely followed by BERT model. Overall, GPT2 performed best on all the performance benchmarks, followed by BERT and thirdly ALBERT. Figure 2 shows the perplexity graph for all the 3 models.

Conclusion

In this paper, we presented a noble approach for sentiment-based topic generation using customers reviews. We created a custom dataset for our experiment and explored various language models for designing encoder-decoder model for the topic generation task. In our analysis, we found that the GPT2 model outperformed other language models on

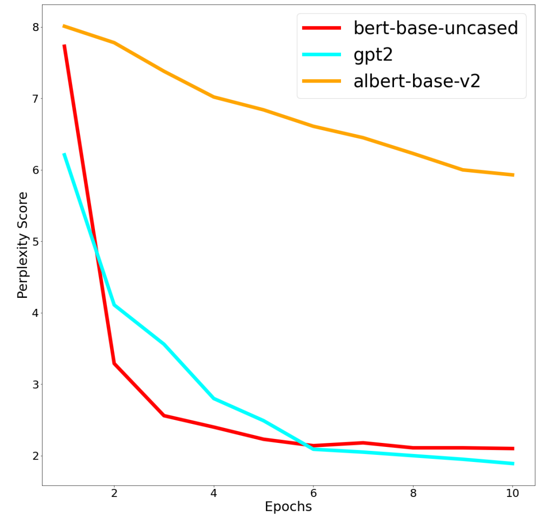

Figure 2: Perplexity graph for models distributed over 10 epochs

various benchmarks. We believe that through our proposed model, various companies would be able to generate qualitative insights from customer-generated feedback and understand the root cause of customer's issues.

References

Asghar, M. Z.; Khan, A.; Ahmad, S.; and Kundi, F. M. 2014. A review of feature extraction in sentiment analysis. *Journal of Basic and Applied Scientific Research* 4(3): 181–186.

Buche, A.; Chandak, D.; and Zadgaonkar, A. 2013. Opinion mining and analysis: a survey. *arXiv preprint arXiv:1307.3336* .

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* .

Lan, Z.; Chen, M.; Goodman, S.; Gimpel, K.; Sharma, P.; and Soricut, R. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942* .

Lavie, A.; and Agarwal, A. 2007. METEOR: An automatic metric for MT evaluation with high levels of correlation with human judgments. In *Proceedings of the second workshop on statistical machine translation*, 228–231.

Papineni, K.; Roukos, S.; Ward, T.; and Zhu, W.-J. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 311–318.

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; and Sutskever, I. 2019. Language models are unsupervised multitask learners. *OpenAI Blog* 1(8): 9.