GR-113

NLP Sentiment Analysis On Amazon Reviews

Swetha Pailla, Sushma Aladhalli Shivakumar and Sireesha Hasti

Advisor: Dr. Mahmut Karakaya

Abstract

A lot of times, companies want to understand the public opinion on their products and figure out what's responsible for the growth and backfall. For a better understanding, they perform sentiment analysis on their product reviews. Sentiment analysis has been on the rise because of the availability of new analysis techniques in deep learning, and there is an incomprehensibly large amount of data being generated everywhere. In this project, we propose a deep neural network architecture (RNN-LSTM) with NLP techniques to evaluate the performance of proposed model in terms of accuracy, precision, sensitivity, specificity, F1-score, confusion matrix and classification report.

Introduction

Sentiment Analysis is a natural language processing (NLP) technique used to determine whether data is positive, negative or neutral. Our objective is to help the organizations to understand their emotions about their products and act accordingly. In this project, we propose a deep neural network architecture Long Short Term Memory (LSTM) using NLP technique to evaluate the sentiment of Amazon reviews.

Research Question(s)

The research questions lead towards creating a framework that can analyze the polar emotions over Amazon reviews. We have three questions we wish to answer:
1. Why LSTM model to analyze sentiment of Amazon reviews?
2. How does feature extraction useful in the process?
3. Why NLP techniques are used?
4. How does the model compare with a traditional approach to interpret the customer review process?
5. What is the impact of this project to society or organizations?

Materials and Methods

- Used *Amazon Reviews dataset* from Kaggle having 4 million customer reviews of all products.
- Dataset has classes _label_ 1 and _label_ 2
- _label_ 1 (Bad Reviews) represents 1-2-star reviews
- _label_ 2 (Good Reviews) represents 4-5-star reviews
- 3-star reviews with neutral sentiment are not included in the dataset.
- Used NLP (Natural Language Processing) for data preprocessing techniques
- LSTM (Long Short-Term Memory) is an RNN (Recurrent Neural Network) architecture which analyses data over multiple time steps.
- Embedding Layer → LSTM Cell → Single Unit (Sigmoid)

Results

<table>
<thead>
<tr>
<th>Classification Report</th>
<th>N = 40000</th>
<th>Predicted [NO]</th>
<th>Predicted [YES]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label 1</td>
<td>0.93</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>Label 2</td>
<td>0.94</td>
<td>0.92</td>
<td>0.93</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.93</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>Macro avg</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>Weighted avg</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Performance Metrics</th>
<th>Accuracy</th>
<th>Precision</th>
<th>F1-Score</th>
<th>Recall/Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.930</td>
<td>0.935</td>
<td>0.925</td>
<td>0.924</td>
<td>0.915</td>
</tr>
</tbody>
</table>

Conclusions

In this project, the proposed model overcome the limitations of traditional/manual approaches to interpret the customer review process and achieved best accuracy in predicting bad (93%) and good reviews (94%) on test dataset using NLP and LSTM techniques. This model is a powerful marketing tool which helps the organizations and product managers to understand the customer's emotions about their products and to discover the reason about customers leaving some negative or positive feedback.

Acknowledgments

We would like to express our special thanks of gratitude to our professor Dr. Mahmut Karakaya, who gave us the golden opportunity to do this wonderful project of sentiment analysis on "Amazon Reviews", who also gave us constant feedback to improve the research questions and results lead us to achieve the successful results. Secondly, we would like to thank class CS1267 for contributing to evolve our thought process and achieve the best results.

Contact Information

https://www.linkedin.com/in/sireesha-hasti-3b62277a
https://www.linkedin.com/in/swetha-spailla@students.kennesaw.edu
https://www.linkedin.com/in/sushma-aladhalli
https://www.linkedin.com/in/mahmut-karakaya

References