Affects of Proper Nouns on Amazon Sentiment Analysis

Abstract

Abstraction needs to be written after writing everything else

1 Introduction

Sentiment Analysis looks into extracting subjective information from a source. It has been noted from Bing Liu sentiment analysis on subjectivity that adjectives, adverbs, nouns, and verbs can obtain sentiment detail [2]. Furthermore, an object in object extraction from sentiment is a noun, proper noun, and sometimes a verb. Even though nouns are a way of obtaining subjective information, further investigation has not been looked into for proper nouns.

There has been plenty of models on predicting sentiment analysis of Amazon reviews. The top 3 models; accurancies unsupervised data augmentation for consistency training [4], deep pyramid convolutional neural networks for text categorization [1], and disconnected recurrent neural networks for text categorization [3]; all has been tested with Amazon product reviews from users. Although those models accurancies are between 60%-70% percent, does simplifing the proper noun into one word affect Amazon sentiment analysis in positive or negative accurancies.

2 Related Work

3 Task

The task is to see if proper nouns contains necessary information needed for sentiment analysis in Amazon product reviews.

4 Method

Amazon product review data comes with a rating and review that can be used for training. The product must parse out proper nouns with a replacement token to represent their is a proper noun, however, removes the context of the proper noun. An example of removing the proper noun context can be shown.

```
"It's just like having $100. Except I
couldn't get Walmart to accept it.
Apparently you can only use it on the
interweb at Amazon."

Detect Amazon and Walmart as proper noun
replaced with <NNP>

"It's just like having $100. Except I
couldn't get <NNP> to accept it.
Apparently you can only use it on
the interweb at <NNP>."
```

4.1 LUKE

LUKE is an entity extraction model that can be used for extracting proper nouns in an Amazon product review.

4.2 NLTK Speech Tagging

NLTK library comes with speech tagging library that can be used for tagging the proper nouns in a text.

4.3 Unsupervised Data Augmentation

Unsupervised data augmentation is a way to train the model. Can make comparsion on data with proper noun context and without.

4.4 Deep Pyramid Convolutional Neural Networks for Text Categorization

Deep Pyramid Convolutional Neural Networks for Text Categorization is another method for training a model for sentiment analysis. Can make comparsion on data with proper noun context and without.

4.5 FastText

FastText can be used in a bag of tricks for efficient text classifcation. Can make comparsion on data with proper noun context and without.

5 Experiments

6 Conclusion

References

[1] Rie Johnson and Tong Zhang. Deep pyramid convolutional neural networks for text categorization. 2017.

[2] Bing Liu. Sentiment analysis and subjectivity.

[3] Baoxin Wang. Disconnected recurrent neural networks for text categorization. 2018.

[4] Qizhe Xie, Zihang Dai, Eduard Hovy, Minh-Thang Luong, and Quoc V. Le. Unsupervised data augmentation for consistency training. 2020.