

Affects of Named Entities on Amazon Sentiment Analysis

Abstract

1 Introduction

Sentiment Analysis looks into extracting subjective information from a source. Some words has more weight in sentiment analysis compared to other words. With reviews from Amazon, name entities are mentioned in the review text.

There has been plenty of models on predicting sentiment analysis of Amazon reviews. The top 3 models; accurancies unsupervised data augmentation for consistency training [5], deep pyramid convolutional neural networks for text categorization [1], and disconnected recurrent neural networks for text categorization [4]; all has been tested with Amazon product reviews from users. Although those models accurancies are between 60%-70% percent, does simplifying the name enitities into one word affect Amazon sentiment analysis in positive or negative accurancies. The possiblility of removing the context of named entity affects the training and tests of sentiment analysis.

2 Related Work

Specific words has neutral, positive, or negative polarity when it comes to sentiment analysis. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis research determined a new stragety of figuring out polarity of words in senitment analysis. The research did not look into the part of speech of these words, but showed how some words can have no impact to sentiment analysis.

Research into subjectivity compared with objectivity of a word looked into parts of speech affects of sentiment analysis from Bing Liu [2]. It has been noted from Bing Liu sentiment analysis on subjectivity that adjectives, adverbs, nouns, and verbs can obtain sentiment detail. Furthermore,

an object in object extraction from sentiment is a noun, proper noun, and sometimes a verb. The paper didn't look into named entities give the impacts of training with the context of the named entities.

3 Task

The task is to see if named entities contains necessary information needed for sentiment analysis in Amazon product reviews models. The difference between each similar model is one data is trained on Amazon product review while the other model is trained on same Amazon product review that replaced named entities with a token. The goal is to compare the two models to see if removing named entities has any impacts on sentiment analysis.

4 Method

Amazon product review data comes with a rating and review that can be used for training. The product must parse out named entities with a replacement token to represent there is a named entity, however, removes the context of the named entity. An example of removing named entities 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 named entity, so 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 entities in an Amazon product review. LUKE-500K pre-trained data had an F1 score of 94.3 on named entity recognition from

CoNLL-2003 dataset [6]. This gives the capability for machine to detect the named entity in a passage. Knowing where the entity names are in the passage, those words can be replaced with one token, so the context of the named entity is not known. To conclude, LUKE will parse out the named entity before sending the review text to the model.

5 Experiments

The Amazon product review data came from University of California, San Diego Amazon product review dataset [3]. The data consists of the rating and the review text that came with the rating for training and testing. The rating from 1-5 will be simplified to 1-2 being negative, 3 being neutral, 4-5 being positive.

5.1 Parsing Named Entities

From each category of Amazon product review, there is two copies of the same product review. One copy has the original Amazon product review text. The second copy has the review text parsed out named entities with a token (NNP). There is 10,000 Amazon product reviews for each category that was used for the models.

5.2 Training the Models

The will be used for detecting sentiment analysis. Each model requiring training will be done with 60% training, 20% validation, and 20% testing. The models will be tested on accuracy between the three different Amazon product reviews, one being the unparsed review, the other two having proper nouns replaced with ambiguity word. The only change with the models is the text has proper nouns parsed out. This will show how the model compares with or without the context of the proper noun.

5.3 Results

6 Conclusion

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] Jianmo Ni, Jiacheng Li, and Julian McAuley. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. 2019.
- [4] Baoxin Wang. Disconnected recurrent neural networks for text categorization. 2018.
- [5] Qizhe Xie, Zihang Dai, Eduard Hovy, Minh-Thang Luong, and Quoc V. Le. Unsupervised data augmentation for consistency training. 2020.
- [6] Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, and Yuji Matsumoto. LUKE: Deep contextualized entity representations with entity-aware self-attention. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 2020.