Use Case Prediction using Deep Learning

Tinashe Wamambo[1], Cristina Luca[1], Arooj Fatima[1], and Mahdi Maktab-Dar-Oghaz[1]

Anglia Ruskin University, Cambridge Campus, East Rd, Cambridge CB1 1PT
tinashe.wamambo@pgr.aru.ac.uk
{cristina.luca,arooj.fatima,mahdi.maktabdar}@aru.ac.uk

Abstract. Research into utilising text classification to analyse product reviews from e-commerce websites has increased tremendously in recent years. Machine Learning and Deep Learning classifiers have been utilised to organise, categorise and classify product reviews, enabling the identification of polarity and sentiment within product reviews. In this paper, we propose a methodology to classify product reviews using machine learning and deep learning with the intention to identify and predict the activity (use case) in which the consumer used the product they have reviewed.

Keywords: Natural Language Processing, Text Classification, Machine Learning, Deep Learning

1 Introduction

In the modern world the internet is the most valuable resource for learning, getting ideas, buying and selling products and services. E-commerce retail websites such as Amazon, Ebay, etc experience a large armount of internet traffic as consumers purchase products for their websites. Everyday millions of reviews are generated by consumers as they provide feedback about products and services, and their experience using them. The increased popularity of e-commerce websites and the explosion of product reviews in record high numbers has seen increased research into sentiment analysis and text classification.

Sentiment analysis (also known as opinion mining) is the process of analysing text documents to extract and understand the sentiments expressed in the text. A combination of natural language processing (NLP) with a machine learning capability (also known as text classification) is utilised to determine the polarity of a text document .i.e. identifying whether the opinion expressed in a text document is positive, negative or neutral. Text classification is also utilised to determine a text document's sentiment orientation, i.e. identifying whether the opinion expressed in a text document is subjective or objective. Product reviews are packed full of subjective opinions because consumers provide feedback on their experience using a product or service.

Millions of product reviews have been generated by consumers who have purchased a product and have had experience using and benefiting from that product. New consumers looking to purchase the same or similar products utilise these product reviews as part of their decision making process to make sure they make informed purchasing decisions. However, given the large amount of reviews generated for a given product, the average consumer will most likely not read all of the reviews to gain a holistic view of the product and the sentiments other consumers who have already bought and used/experienced the product have towards that product. Currently the quickest and easiest decision making element that is tied to product reviews that consumers utilise is the star rating, the higher the star rating the more likely a consumer will purchase that product. However, a star rating is ambiguous and prone to grade inflation, e.g. having a 4.8 star rating does not mean the product is exceptional, the difference between a product with a 4.5 star rating and one with a 4.8 star rating could be massive which makes it very difficult for consumers to differentiate between OK products and very good ones.

Another issue with star ratings is the fact that they are shallow, they do not truly provide summarised information about the sentiments expressed by consumers or why consumers expressed such sentiments in the product reviews. Star ratings also leave a lot of room for assumptions to be made about a product's suitability for certain tasks .e.g. a pair of shoes with a high star rating may be suitable for running but not for hiking. New consumers looking to purchase that pair of shoes would not be provided with this information in a quick and easy manner.

This research aims to identify the activities other consumers used a product for alongside the polarity and sentiment they expressed in the product reviews using sentiment analysis, text classification and machine learning. The intention is to provide new consumers with valuable summarised information that they can use to decide if a product is suitable for the activity they intend to perform. For this research, the activity a consumer used a product for will be known as a 'use case'.

For this research, a use case is an action or activity performed using the product as described in a product review. An example of this is "I bought these boots for walking my dogs around the local park, they have been fit for purpose thus far". The use case in this example is "walking".

The rest of the paper is organised as follows: Section 2 describes the related work, Section 3 presents the proposed approach to detect and predict a use case, Section 4 discusses the experiments performed and the results, and Section 5 finally draws the Conclusions.

2 Related Work

This research is motivated by advancements in machine learning techniques, sentiment analysis and text classification. In many reviews, users express their opinions towards a product's features and their user experience. So, aspect/feature based sentiment analysis is a suitable direction to pursue. Action words/terms (verbs) are of particular importance for this research, so Parts of Speech (POS) tagging will be considered for feature extraction. Many machine learning approaches have been implemented over the years, Support Vector Machines (SVM) and Naive Bayes have significant popularity at effectively performing text classification with high accuracy and dealing with large datasets.

2.1 Parts of Speech

Part of Speech (POS) tagging has been used for feature extraction. Devi, *et. al* [1] performed sentence level classification to detect words tagged as nouns because aspects/features are usually described by nouns or noun phrases. Alfrjani, *et. al* [2] applied POS tagging to determine if tokens in reviews were nouns, verbs, adjectives, adverbs, etc with the intention of extracting the POS tags as token features. Hemmatian and Sohrabi [7] utilised frequent based nouns as well as order and similarity based filtering to improve feature extraction using POS tagging.

Devi, *et. al* [1] used POS tagging to extract product features from product reviews, however Alfrjani, *et. al* [2] used POS tagging to categorise words in a review as part of an integration process between a semantic domain ontology and natural language processing. Likewise, Hemmatian and Sohrabi [7] used POS tagging to extract features described as nouns or noun phrases. However, their approach included word frequency, whilst Devi, *et. al* [1] identified grammatical dependencies between words in sentences.

Feature extraction strategies used by Devi, *et. al* [1] and Hemmatian and Sohrabi [7] have been considered for this research because product features are expected to be described as nouns or noun phrases, so POS tagging is vital.

2.2 Deep Learning

Deep learning is a branch of machine learning that aims to enable machines to learn and evolve, similar to the way humans learn from their memories and experiences throughout their lifetime. Instead of using predefined equations, deep learning sets up “basic parameters about the data and trains the computer to learn on its own by recognizing patterns” [14] using neural networks. A neural network consists of “multiple hidden layers that can learn increasingly abstract representations of the input data” [5] using weights that are adjusted during training [3] to produce better predictions.

Parvathi and Jyothis [11] proposed a text classification strategy that involved using a Convolutional Neural Network (CNN) to identify relevant text to determine the category a document belonged to. The proposed strategy used “layers of neurons and a bag of words approach” [11] to analyse text documents. In their conclusions, Parvathi and Jyothis [11] reported accurate and positive results. However, they noted that deep learning models have a limitation of learning through observation which means they only contain knowledge provided in the training data instead of learning in a generalised manner.

Parwez, *et. al* [12] highlighted that traditional machine learning models suffered from a limitation of “relying on the bag of words representation of documents to generate features in which word order and context are ignored” [12] which could cause data sparsity problems. They proposed a neural network architecture that involved Convolutional Neural Networks (CNN) to classify short text documents, e.g. tweets. The CNN models used a combination of generic and domain specific word embeddings to predict class labels, whilst considering the contextual information within text documents. Results showed that CNN models outperformed traditional machine learning models in terms of validation accuracy and optimal feature generation which was used to analyse unlabelled text documents. Parwez, *et. al* [12] concluded that the proposed approach could be used to perform social media surveillance focused on predicting disease outbreaks.

Kolekar and Khanuja [8] performed a comparison between machine learning algorithms and neural networks to find out the better approach to classifying the polarity of tweets. They applied Term Frequency and Inverse Document Frequency (TF-IDF) word embedding technique to the tweets and fed the features to Support Vector Machines (SVM), Naive Bayes and Convolutional Neural Network (CNN) developed using Keras and Tensorflow. Results showed that “using deep learning approach has given better result compare to traditional machine learning technique like SVM and NB” [8].

Subramani, *et. al* [15] implemented a neural network based topic modelling architecture to anaylse text documents and cluster highly similar documents together. They coined this approach as the Neural Topic Modelling approach. Their architecture used Latent Dirichlet Allocation (LDA), Keras and Tensorflow. According to the researchers, the approach provides a scalable and unsupervised learning framework that accurately discovers topics for a text corpus by considering the “semantic meanings of the words ensuring the usefulness of the topics” [15]. Results from testing with short and large text documents showed positive results, leading Subramani, *et. al* [15] to conclude that their proposed topic modelling approach had real world applications, i.e. movie recommendations and news clustering.

To identify similar documents based on the semantic meaning of their text, Mo and Ma [10] built DocNet, a clustering system that combined word embedding vectors, a deep neural network and euclidean distance. The expectation was for small document to have “small distances among their vectors while distinct document have large distances” [10]. Results showed this to be true with DocNet performing better than traditional clustering techniques, i.e. TF-IDF and K-means clustering. However, Mo and Ma [10] stated that DocNet's performance heavily depended on the similarity of “data distribution between data fed to DocNet and data to classificaton or clustering” [10] which means the DocNet system will most likely perform poorly on new datasets.

At the Google I/O conference in 2019, Sara Robinson [13] a developer advocate at Google presented a text classification model that predicted the topic of a Stack Overflow question. As part of pre-processing the train/test dataset, words that specified the topic of the Stack Overflow question within the text were replaced with the word avocado. This was done to prevent the machine learning model from using signal words to perform classification, but instead generalise to find patterns within a dataset because some Stack Overflow questions may not specify the topic. A bag of words approach was used to encode words into matrices, applying a multi-hot encoding technique to convert the Stack Overflow questions into vocabulary size arrays. The training labels were also converted into a multi-hot array because the model was going to have the ability to identify and predict on multiple labels. A deep learning neural network was developed, it took the bag of words matrices as input data, feeding the data into hidden layers. The output layer of the neural network used sigmoid to compute the model's output. Sigmoid returns a value between zero and one for each label which corresponded to the probability of the label being associated with the Stack Overflow question.

To develop the deep learning text classification neural network, Sara Robinson [13] used Pandas, Scikit-learn and Keras to pre-process the data, an 80/20 train/test split was applied to the dataset and the neural network model architecture was built using TensorFlow. The model had 95% accuracy.

The techniques applied in various research in sentiment analysis and text classification have been primarily focused on improving the classification of text to extract features, polarity, opinions and emotions expressed towards products by consumers. This has been valuable for understanding the sentiment expressed by consumers, however, it has not been able to identify why consumers hold those sentiments towards products or the activities consumers have used the products for which is a major limitation. This research aims to resolve this limitation by understanding why consumers express particular opinions towards products, through identifying the activities consumers have used the products for.

3 Proposed Approach

The proposed approach is a deep learning neural network that has been developed to perform text classification on product reviews to detect and predict a use case. This approach has similarities to the approach presented by Sara Robinson [13] at the 2019 Google I/O conference. However, key signal words that specify the use case within the product review text have not been replaced with the word avocado. Multi-hot encoding has been applied to the product reviews and their labels producing a dataset of matrices. The product reviews (extracted from Amazon) used for this research have only one label and, therefore, the model will not classified on multiple labels as [13], where the Stack Overflow questions being classified had multiple labels.

A 90/10 train/test split has been applied to the dataset for this iteration, whereas Sara Robinson [13] applied an 80/20 train/test split to her dataset. In similar fashion to Sara Robinson [13], TensorFlow has been utilised to build, train and test a deep learning neural network model.

This approach focuses on the use cases listed below:

- Run
- Walk
- Hike
- Climb
- Swim
- Unknown

30,000 product reviews have been extracted from Amazon to train and test the proposed model. They consist of 5000 reviews for each of the use cases on focus to make sure the train/test dataset is balanced. This provides a greater probability for the neural network model to have balanced classes.

Product reviews have been pre-processed and analysed, spelling mistakes have been corrected and stop words have been removed using NLTK's natural language processing capbilities. TextBlob, a Python library that sits on top of NLTK has been utilised to extract features that have been used to create a multi-hot encoded bag of words matrices. A MultiLabelBinarizer class which is part of the Scitkit-learn library has been used to multi-hot encode the labels.

Parts of Speech (POS) tagging has been utilised to identify and extract nouns, noun phrases and verbs as features which is different to the approach proposed by Devi, *et. al* [1] who only extracted nouns. The verbs that have been extracted as features describe the actions that have been performed as denoted in the review text and the activity (use case) in which the consumer used the product is expected to be described by the verbs in some way.

NLTK is referred to as "The Conqueror" in EliteDataScience [4]. It is a "leading platform for building Python programs to work with human language" [16] that provides an easy to use interface to a suite containing a variety of text processing libraries. Many breakthroughs have been made in the field of analysing and processing text using NLTK as it is "responsible for conquering many text analysis problems" [4].

TextBlob, referred to as "The Prince" in [4], is a text processing library that "sits on the mighty shoulders of NLTK" [4]. It provides a "simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification", [9] and it also has a "gentle learning curve while boasting a surprising amount of functionality" [4]. TextBlob also allows the use of NLTK tools along side its own tools, enabling access to the NLTK tool kit and all of its benefits.

The deep learning neural network has been created using a Sequential class which is part of the Keras library. Dense layers which are also part of the Keras library have been added to the neural network as three layers that are used to classify a data matrix in chunks spread across various hidden layers. The multi hot encoded bag of words matrices and the multi hot encoded labels have been provided as input to the neural network. The neural network trains using the training data matrices over five epochs, this means the neural network repeatedly goes through the entire training dataset five times.

4 Experiments and Results

4.1 Datasets Description

Amazon product reviews have been used as training/testing data for this research. This is because Amazon has a large amount of free text product reviews it holds due to Amazon's vast product range and user base. A fantastic extensive dataset [1] containing millions of Amazon product reviews has been used. Approximately 5 million reviews from the clothing and shoes categories have been retrieved.

4.2 Metrics

Table 1. Neural Network Metrics

Accuracy	Precision	Recall
90%	96%	44%

Table 1 shows the accuracy, precision and recall of the neural network model. As shown in the table, the model has high accuracy and high precision, but unfortunately it has low recall. This means 9 out of 10 positive predictions have a high probability of being correct (precision), however only 4 out of 10 positive predictions have a probability of actually being correct (recall). As a result, the model may not generalise well and has a high probability of producing a significant amount of incorrect predictions. According to Google Developers [6], "improving precision typically reduces recall and vice versa" [6] because of the tension that is present between precision and recall. The good news is that the neural network is not biased towards positive predictions because the prerequisites required for the model to behave in such a manner are a low precision and very high recall.

Table 2. Neural Network Performance

Accurately Predicted Use Cases	Inaccurately Predicted Use Cases
61%	39%

Table 2 shows the performance of the neural network classifier when it is exposed to 1200 brand new product reviews extracted from the extensive dataset described in section 4.1.

Even though the neural network classifier has a very high classification accuracy as shown in table 1, it did not accurately predict an extremely large amount

[1] https://nijianmo.github.io/amazon/index.html

of use cases as expected. The classifier accurately predicts the use case for the majority of the 1200 product reviews. This is positive reflection of the neural network and its performance given it managed to classify completely new product reviews and accurately predict the use cases for a relatively large amount of the product reviews, even though the neural network recorded a low recall. The most likely cause for the neural network failing to accurately predict the use case for a larger amount of product reviews is that the classifier does not generalise well enough.

5 Conclusions

This paper proposes an approach that develops a deep learning neural network to classify product reviews and predict the activity (use case) in which the consumer used the product. Natural language processing techniques, text classification and machine learning have been applied to develop the neural network. As shown by the metrics, the neural network has high accuracy and high precision, however low recall showed that results have a high probability of consisting of false positives. This was evidenced by exposing the neural network to completely new product reviews it had never consumed. The neural network accurately predicted the use case for the majority of the product reviews, however approximately 40% of the product reviews were incorrectly classified which is a significant number of reviews for a neural network to incorrectly classify.

The aim of this research is to classify product reviews and accurately predict the activity (use case) a consumer used the product for as described in the review text. Results and metrics from testing the neural network show that text classification recall needs to be improved to limit the prevalence of false predictions and make sure the accurate predictions are reliably produced as output. As further work is undertaken within this research, this will be the focus.

References

1. Devi, D.V.N., Kumar, C.K., Prasad, S., 2016. A Feature Based Approach for Sentiment Analysis by Using Support Vector Machine. Published In: 2016 IEEE 6th International Conference on Advanced Computing (IACC). doi:10.1109/IACC.2016.11
2. Alfrjani, R., Osman, T., Cosma, G., 2016. A New Approach to Ontology-Based Semantic Modelling for Opinion Mining. Published In: 2016 UKSim-AMSS 18th International Conference on Computer Modelling and Simulation (UKSim). doi: 10.1109/UKSim.2016.15
3. Allibhai, E., 2018. Building A Deep Learning Model using Keras. https://towardsdatascience.com/building-a-deep-learning-model-using-keras-1548ca149d37
4. EliteDataScience, 2017. 5 Heroice Python NLP Libraries. https://elitedatascience.com/python-nlp-libraries
5. EliteDataScience, 2018. Keras Tutorial: The Ultimate Beginner's Guide to Deep Learning in Python. https://elitedatascience.com/keras-tutorial-deep-learning-in-python

6. Google Developers, 2020. Machine Learning Crash Course. https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall
7. Hemmatian, F., Sohrabi, M.K., 2017. A survey on classification techniques for opinion mining and sentiment analysis. Published In: Artificial Intelligence Review (2017). https://link.springer.com/article/10.1007/s10462-017-9599-6
8. Kolekar, S.S., Khanuja, H.K., 2018. Tweet Classification with Convolutional Neural Network. Published In: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). doi:10.1109/ICCUBEA.2018.8697397
9. Loria, S., 2018. TextBlob: Simplified Text Processing.. https://textblob.readthedocs.io/en/dev/
10. Mo, Z., Ma, J., 2018. DocNet: A document embedding approach based on neural networks. Published In: 2018 24th International Conference on Automation and Computing (ICAC). doi:10.23919/IConAC.2018.8749095
11. Parvathi, P., Jyothis, T.S., 2018. Identifying Relevant Text from Text Document Using Deep Learning. Published In: 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET). doi:10.1109/ICCSDET.2018.8821192
12. Parwez, MD., A., Abulaish, M., Jahiruddin., 2019. Multi-Label Classification of Microblogging Texts Using Convolution Neural Network. Published In: IEEE Access (Volume: 7). doi:10.1109/ACCESS.2019.2919494
13. Robinson, S. 2019. Live Coding A Machine Learning Model from Scratch (Google I/O'19), Google Cloud Platform. https://www.youtube.com/watch?v=_RPHiqF2bSs
14. SAS Institute, 2020. Deep Learning What it is and why it matters. https://www.sas.com/en_us/insights/analytics/deep-learning.html
15. Subramani, S., Sridhar, V., and Shetty, K., 2018. A Novel Approach of Neural Topic Modelling for Document Clustering. Published In: 2018 IEEE Symposium Series on Computational Intelligence (SSCI). doi:10.1109/SSCI.2018.8628912
16. The NLTK Project, 2017. Natural Language Toolkit. https://www.nltk.org/