Develop a Sentiment Analyser

Student Name: Saleh Mohamed

Student Number: 30001162

Abstract

Sentiment Analysis is also commonly referred to as **Opinion Mining** and revolves around obtaining the opinions which people express and conducting an analysis of those opinions in order to see what emotions they carry. For example, whether they are positive, negative or neutral. Moreover, with the growth of social media platforms, sentiment analysis is becoming increasingly important in modern society as a means of understanding how people feel towards various topics and what emotions they are conveying. [1] Birjali M, Kasri M, Beni-Hssane A et al., 2021

The purpose of this dissertation was to develop a sentiment analyser which could analyse the opinions expressed by people and understand what emotion they are conveying. The analyser obtains data from the **IMDb** website revolving around movie reviews in order to see what people think about certain movies. It also makes use of a **Recurrent Neural Network**.

Introduction

Sentiment Analysis, when applied to various social networking sites can be used as an efficient and unique way of galvanising peoples views and opinions. However, despite its success at opinion mining, sentiment analysis has also been held back by issues faced by natural language processing (NLP).

Deep learning models have been viewed as the solution to the issues of NLP. Recurrent Neural Networks (RNN) are a type of neural networks whereby a directed cycle is formed through their connections between neurons.

This leads to the production of loops for feedback that are contained in the RNN model. One unique type of RNN is long short-term memory (LSTM) which utilises memory that is long as a source of input for activation functions within a hidden layer [2]. Dang et al., 2020

Moreover, as well as RNNs, Convolutional Neural Networks (CNN) are another frequently used type of network when it comes to the processing of textual data. Whilst RNNs are useful when it comes to modelling which is sequential, CNNs are efficient in understanding patterns that are local. [3]. Basiri et al., 2021

However, Recurrent Neural Network models make use of being able to obtain time effect as well as propagating data regarding sentiment labels. Furthermore, due to RNNs ability of undertaking the modelling of data that is sequential, they are now often used in studies. The two main types of Recurrent Neural Networks that are widely used are LSTM and GRU. [4]. Aydln & Gungor, 2020

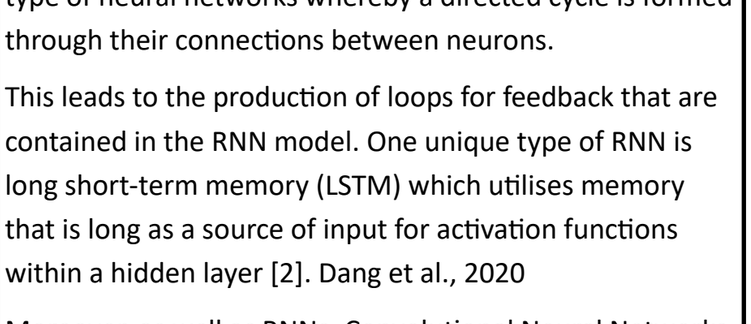


Figure 1—Example of how the system works

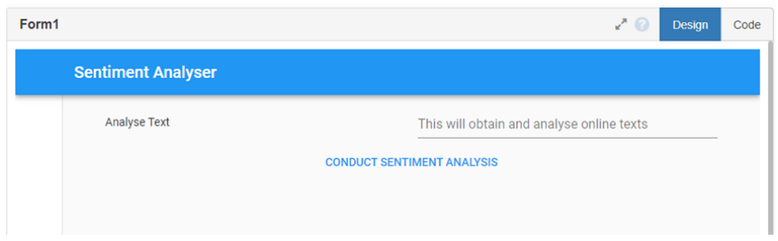


Figure 2— An image of the user interface of the system

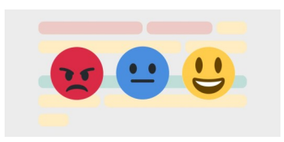
Methods

The dissertation contained a variety of forms of research and analysis into sentiment analysis including the different forms of sentiment analysis and models in order to understand how to design and produce a sentiment analyser which could analyse digital online texts.

Different approaches to developing a sentiment analyser were investigated and considered including Twitter-based sentiment analysis but ultimately, PyTorch was chosen as the preferred method of producing the sentiment analyser. This was because PyTorch provided useful features in regards to producing a sentiment analyser and did not require the use of a Twitter account.

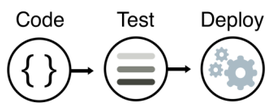
When designing the system to be produced, the main source code to enable the system to work and function as expected was produced in Google Colab after examining different means of programming the system and this code was expanded and improved over time. With more and more features being given to it.

After testing the first draft of the system (source code), whilst the code was detailed and contained important functionality for the sentiment analyser, it at first was not working correctly.



From conducting an in—depth analysis of the source code, it was made evident that the main issues within the system were the result of the poor structure of some of the code as well as certain library features not being installed properly on the given device which was running it.

Thus, with the second draft of the code, it was refined and modified to ensure that the code was of a higher quality than the previous draft and that it contained the necessary elements for the system to work optimally.



The structure of the code was placed into different classes and functions, with each one having an important role in the system as a whole. In addition, the device running the code was checked to ensure that it had the essential library features installed on it.

The way in which the system undertook its tasks can be seen from looking at **Figure 1.**

Results

The final code was **tested** robustly and attempts were made to ensure that it was able to function correctly and complete its tasks properly. The end result was a success, after thoroughly checking the code, it was made clear that it was able to complete its objectives effectively. The system was able to successfully obtain data from the IMDb dataset and output the obtained data to indicate the sentiment that it contained.

After the code had been completed, a **user interface** was then designed and produced in order for the user to be able to interact with it as can be seen in **Figure 2.** The purpose of the interface was to provide the user with an **easy to use** way of accessing the system and conducting sentiment analysis. As a result, when operating the user interface, the user would be able to **utilise the features and functionalities** contained within the source code.

However, despite the fact that the source code itself appeared to work as expected, the user interface encountered issues when reading from the source code. This seemed to be due to the way in which the code was **designed** and thus this seemed to confuse the user interface. As a result, this was one area that could have been improved.

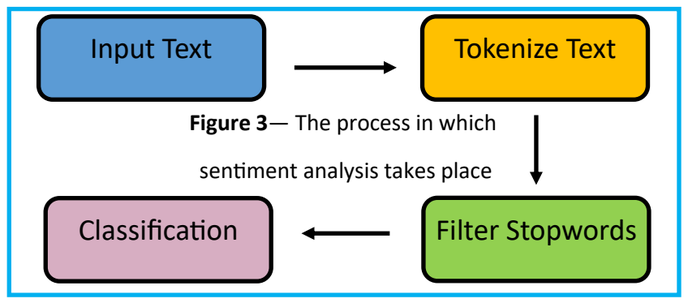


Figure 3— The process in which sentiment analysis takes place

Discussion

Whilst the system developed was designed in immense detail and contained significant functionality as well as provided with a professional look to it, it could have been improved in a number of areas as it could have been given more features, for example, the ability to enable the user to enter input data into it.

The code itself was sufficient for its purpose and was laid out in a neat manner but may have benefitted from being structured slightly differently and having a wider range of coding styles.

Moreover, the system would also benefit from including a login section for the user to be able to use their details to log into the system and conduct sentiment analysis.

These improvements could be considered in potential future iterations of the given system in order to improve on its authenticity.

Conclusion

The final version of the system was able to successfully conduct sentiment analysis from analysing data from the IMDb dataset. It conducts an in-depth analysis of the given data that it has obtained in order to understand the view that it is conveying.

The code was designed in significant detail whilst the interface itself was provided with a professional design with the intention of it being straightforward for the user to use. Thus, the system as a whole was an overall success.

References

[1] Birjali M, Kasri M, Beni-Hssane A et al., 2021 https://www.sciencedirect.com/science/article/pii/S095070512100397X?casa_token=BUkvXYncXdAAAAAA:yBkkRThjv_qDcGudrCTRFCadNWpy9zhyJsqjV-D0bJCjEJSLzRpwZs2bfpZj00W8cOOcWT1tK6M

[2] Dang et al., 2020 https://www.mdpi.com/2079-9292/9/3/483/htm

[3] Basiri et al., 2021 https://www.sciencedirect.com/science/article/pii/S0167739X20309195?casa_token=OyZY2kt3OS8AAAAA:6uuLVe_HeVAfScwxzueprqBNzCcYls20U1eNMP_i5jf4p6Dd43-CGAtuqcmwTuvJ5pkBCOc6RqI

[4] Aydln & Gungor, 2020 https://ieeexplore.ieee.org/abstract/document/9078752