

Higher school of communication

Project Report

**Sentiment Analysis**

Realized by : **JEBALI Rani**

**OUERGHIE Belhassen**

**KHLIFI Houssem**

Supervised by : **Mme Ben lTAIFA Asma**

**Mme ABDELWEHED Nawres**

1

Table of contents

Introduction ……………………………………………………………………………………3

1.1.Problematic ………………………………………………………………………..3

1.2.Solution ………………………………………………………………………………3

Creating the model …………………………………………………………………………4

2.1.Data pre-processing ……………………………………………………………4

2.2 Loading & Preparing GloVe for word representation …………..5

2.3. Building the Model …………………………………………………………….5

2.4 Training and validation ………………………………………………………..6

Conclusion……………………………………………………………………………………….6

2

**1.Introduction:**

Sentiment analysis is the use of natural language processing (NLP), machine learning, and other data analysis techniques to analyze and derive objective quantitative results from raw text. It helps in detecting positive or negative sentiment in text. It’s often used by businesses to detect sentiment in social data, gauge brand reputation, and understand customers.

1.1.Problematic :

Since customers express their thoughts and feelings more openly than ever before, sentiment analysis is becoming an essential tool to monitor and understand that sentiment. Automatically analyzing customer feedback, such as opinions in survey responses and social media conversations, allows brands to learn what makes customers happy or frustrated, so that they can tailor products and services to meet their customers’ needs.

1.2.Solution :

To perform this task, artificial intelligence and more precisely neural networks were used to establish a model capable of predicting the feeling or the opinion of the customer from his comment and this by assigning a satisfaction score ranging from 1 to 5.

3

**2.Creating the model:**

In this part we’ll discuss the different stages we passed through to build

our model.

2.1.Data pre-processing:

Data preprocessing can refer to manipulation or dropping of data before it is used in order to ensure or enhance performance. Data preprocessing aims at making the raw data at hand more amenable to neural networks. This includes vectorization, normalization, handling missing values, and feature extraction.

* handling the missing values using the “dropna” function
* creating the data dataframe and target dataframe
* encoding the labels using “to\_categorical” function
* turning the text in our data\_set into lists of integer indices
* splitting our data\_set into training\_set and test\_set
* Padding our Data so that they all have the same length, turn them into an integer tensor of shape (samples, word\_indices), and then use as the first layer in your network a layer capable of handling such integer tensors(Embedding layer in our case

4

2.2 Loading & Preparing GloVe for word representation

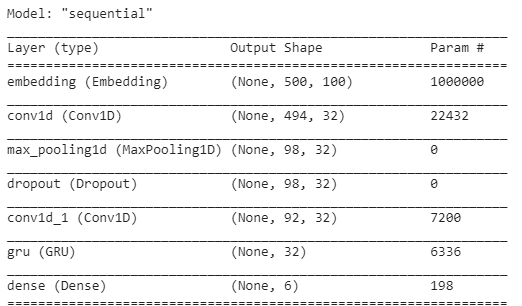
There are various precomputed databases of word embeddings that we can download and use in a Keras Embedding layer. Word2vec is one of them. Another popular one is called Global Vectors for Word Representation This embedding technique is based on factorizing a matrix of word co-occurrence statistics. Its developers have made available precomputed embeddings for millions of English tokens, obtained from Wikipedia data and Common Crawl data.

This is how we used GloVe embeddings in our Keras model:

* Parsing the GloVe word-embeddings file
* Preparing the GloVe word-embeddings matrix

2.3. Building the Model:

We chose this architecture:



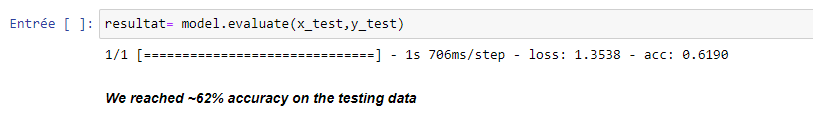
After our model is ready. We froze the Embedding layer to avoid loosing the learned information

5

2.4 Training and validation

Training a neural network is very simple. First, we need to do a forward pass in which we feed the model with the data and calculate the loss function. Next, the backpropagation is preformed in which the gradient of loss function is calculated and then update the weights of the trainable parameters. This step is repeated for a specific number of iterations (epochs). Each epoch will

produce a model that we can save and use in a production environment, but how do we know which one to choose?



**Conclusion**

In summary, most of the time was spent in training the model, specifically in experimentation with different values and fine-tuning the model. We achieved a decent accuracy.

Finally, we personally appreciated working on this project. It was a very enriching technical experience as far as it was a good practice to work on new technologies we never used before. It allowed us to set in practice some of the knowledge acquired during the two academic years in SUP’COM. To sum up, it was an excellent opportunity to develop our observation, analysis, design, and programming skills with a good introductive experience in the machine-learning field.

6