Project Report

Meme Sentiment Analysis

Prepared by: Raghav Prasad (2017A7PS0297G)
Rahul Rajeev Karajgikar (2017A7PS0050G)
Krishna Datta (2017A7PS0007G)

Under the guidance of: Prof. Tirtharaj Dash

In partial fulfilment of the requirements for the course
BITS F312: Neural Networks and Fuzzy Logic
25th November 2019

Table of Contents

INTRODUCTION

Multimodal sentiment analysis is a new dimension of the traditional text-based sentiment analysis, which goes beyond the analysis of texts, and includes other modalities such as audio and visual data. It can be bimodal, like this project, which includes Visual data and Textual data. With the extensive amount of social media data available online in different forms such as videos and images, the conventional text-based sentiment analysis has evolved into more complex models of multimodal sentiment analysis, which can be applied in the development of virtual assistants, analysis of YouTube movie reviews, analysis of news videos, and emotion recognition such as depression monitoring, among others.

Similar to the traditional sentiment analysis, one of the most basic task in multimodal sentiment analysis is sentiment classification, which classifies different sentiments into categories such as positive, negative, or neutral. The complexity of analyzing text, audio, and visual features to perform such a task requires the application of different fusion techniques, such as feature-level, decision-level, and hybrid fusion. The performance of these fusion techniques and the classification algorithms applied, are influenced by the type of textual and visual features employed in the analysis.

DATA PREPROCESSING

In the data that we had, there were certain images that could not be read as they were corrupted, and there were some columns in the text data with data missing. Due to these discrepancies, we only picked Image and Text pairs that contained valid data that we could use. This gave us 6506 such pairs.

Then we processed the data according to a convention called Decision Level (Late) Fusion, where the different aspects of the data, Visual and Textual here, are processed separately and then combined to give the multimodal model.

For the Visual Analysis, we used a pretrained model called Xception, based on Google's Inception, along with two dense layers.

For the Textual Analysis, we used GloVe embeddings with a simple LSTM based RNN.

An additional remark may be made regarding the data and it's distribution. Two extra columns (we labelled them *funny* and *extra_info*) were part of the dataset. The extra columns may have been of some use in the 3 classification tasks. In order to determine their usefulness, we found out pairwise correlations for all variables, as shown in the table below:

	Funny	Extra_Info	Offensive	Motivational	Positive
Funny	1.000000	0.138989	0.130904	0.051580	0.203712
Extra_Info	0.138989	1.000000	0.410248	0.153923	0.030827
Offensive	0.130904	0.410248	1.000000	0.278053	-0.025606
Motivational	0.051580	0.153923	0.278053	1.000000	0.089615
Positive	0.203712	0.030827	-0.025606	0.089615	1.000000

We see that there is no appreciable correlation between any pair of variables. This led us to discard *funny* and *extra_info*. However, the more noteworthy observation from this correlation table is the absence of strong positive correlations that one would have expected; such as between *Motivational* and *Positive*. It seems very intuitive that something that is motivational would also be, more often than not, positive. However, this is not the case and it leads us to believe that the dataset is flawed.

VISUAL ANALYSIS

As mentioned, we used the Xception pre-trained model. Xception is a CNN based model which is an improvement on Inception V3. The usage of Xception is from the Keras library.

We used Xception without the top layer to be able to fit our images of dimensions (150, 150, 3), or we would've had to give the input as (299, 299, 3).

After this, we added a Flatten() layer, and two Dense layers, with shape (8) and (2).

This was then concatenated with the Textual Analysis part of the model.

TEXTUAL ANALYSIS

We used GloVe embeddings, as taught to us in previous labs, and connected that to a Spatial Dropout layer to prevent overfitting.

This was connected to an LSTM layer, and then to a Dense layer of dimensions (2).

This was then concatenated with the Visual Analysis model.

FINAL MODEL STRUCTURE

Once these two parts were concatenated, we use two more Dense layers, one with dimensions (8), and the other varies based on the task.

Task A - (2)

Task B - (5)

Task C - (4)

RESULTS

This particular model design gave us the following validation accuracies:

Task A - 0.66052

Task B - 0.8

Task C - 0.75038

CONCLUSION

The models trained were not able to learn the given task well. This was probably because of class imbalance and incorrectly labelled data leading to contradictory labels throughout the entire dataset. Another cause could be the text present as part of the images in the training dataset.

Possible solution: In order to tackle the problem of the incorrectly labeled data, leading to contradictions, we can consider using soft labels for the classes in the proportion that they have been incorrectly labelled. However, this would require a priori knowledge of an estimate of the proportion of incorrectly labelled classes.

REFERENCES

1. *Multimodal Sentiment Analysis To Explore the Structure of Emotions,* https://arxiv.org/abs/1805.10205

2. *Multimodal sentiment analysis*, https://en.wikipedia.org/wiki/Multimodal_sentiment_analysis

3. *Transfer learning and Image classification using Keras on Kaggle kernels*, https://towardsdatascience.com/transfer-learning-and-image-classification-using-keras-on-kaggle-kernels-c76d3b030649

4. *Xception: Deep Learning with Depthwise Separable Convolutions*, https://arxiv.org/pdf/1610.02357.pdf

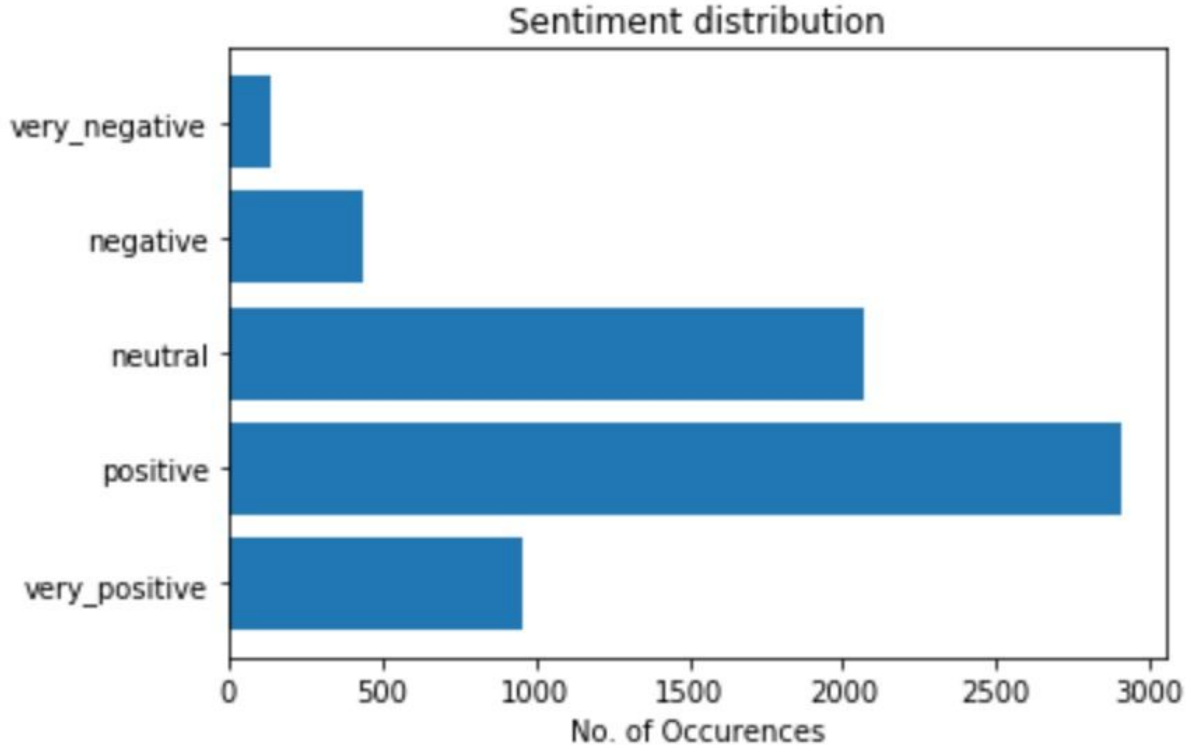
5. *Keras: Multiple Inputs and Mixed Data*, https://www.pyimagesearch.com/2019/02/04/keras-multiple-inputs-and-mixed-data/

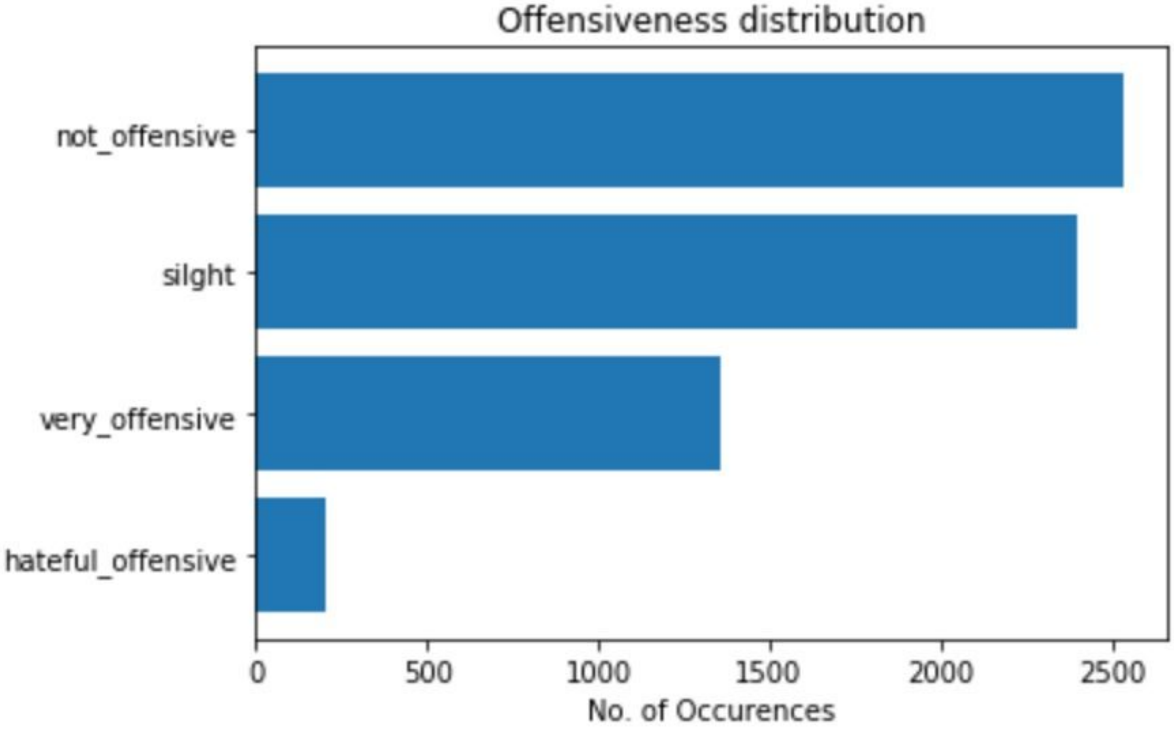
6. *Learn to Combine Modalities in Multimodal Deep Learning,* https://arxiv.org/pdf/1805.11730.pdf

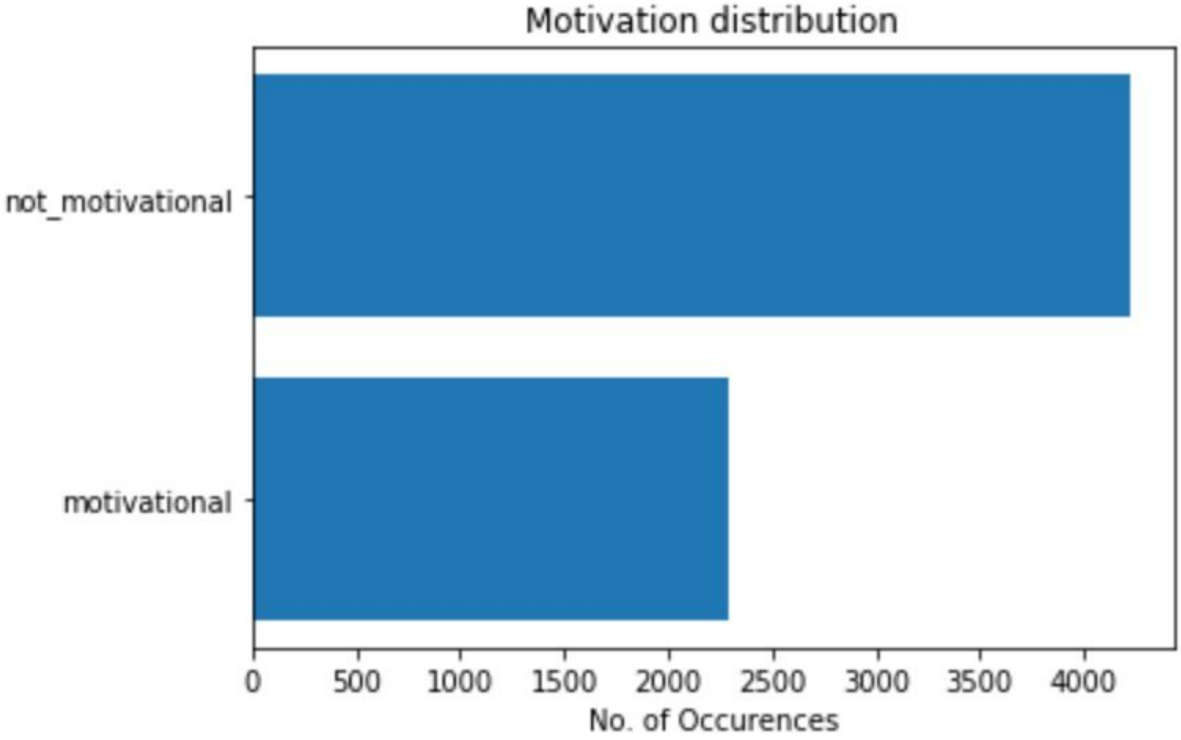
7. *Classification on Soft Labels is Robust Against Label Noise,*

   https://www.christianthiel.com/publications/Classification_on_Noised_Labels.pdf

APPENDIX


Sentiment distribution
very_negative
negative
neutral
positive
very_positive
0
500
1000
1500
2000
2500
3000
No. of Occurences


Offensiveness distribution
not_offensive
silght
very_offensive
hateful_offensive
0
500
1000
1500
2000
2500
No. of Occurences


Motivation distribution
not_motivational
motivational
0
500
1000
1500
2000
2500
3000
3500
4000
No. of Occurences

```
Model: "model_1"
_________________________________________________________________________________________________
Layer (type)                 Output Shape              Param #     Connected to
=================================================================================================
xception_input (InputLayer)  (None, 100, 100, 3)  0
_________________________________________________________________________________________________
embedding_1_input (InputLayer)  (None, 105)           0
_________________________________________________________________________________________________
xception (Model)             (None, 3, 3, 2048)   20861480    xception_input[0][0]
_________________________________________________________________________________________________
embedding_1 (Embedding)      (None, 105, 300)     3259200     embedding_1_input[0][0]
_________________________________________________________________________________________________
flatten_1 (Flatten)          (None, 18432)        0           xception[1][0]
_________________________________________________________________________________________________
spatial_dropout1d_1 (SpatialDro (None, 105, 300)     0           embedding_1[0][0]
_________________________________________________________________________________________________
dense_1 (Dense)              (None, 8)            147464      flatten_1[0][0]
_________________________________________________________________________________________________
lstm_1 (LSTM)                (None, 64)           93440       spatial_dropout1d_1[0][0]
_________________________________________________________________________________________________
dense_2 (Dense)              (None, 5)            45          dense_1[0][0]
_________________________________________________________________________________________________
dense_3 (Dense)              (None, 5)            325         lstm_1[0][0]
_________________________________________________________________________________________________
concatenate_1 (Concatenate)  (None, 10)           0           dense_2[0][0]
                                                             dense_3[0][0]
_________________________________________________________________________________________________
dense_4 (Dense)              (None, 8)            88          concatenate_1[0][0]
_________________________________________________________________________________________________
dense_5 (Dense)              (None, 5)            45          dense_4[0][0]
=================================================================================================
Total params: 24,362,087
Trainable params: 21,048,359
Non-trainable params: 3,313,728
_________________________________________________________________________________________________
```

Positive Model - Summary

```
Model: "model_1"
_________________________________________________________________________________________________
Layer (type)                 Output Shape              Param #     Connected to
=================================================================================================
xception_input (InputLayer)  (None, 100, 100, 3)  0
_________________________________________________________________________________________________
embedding_1_input (InputLayer)  (None, 105)          0
_________________________________________________________________________________________________
xception (Model)             (None, 3, 3, 2048)   20861480    xception_input[0][0]
_________________________________________________________________________________________________
embedding_1 (Embedding)      (None, 105, 300)     3259200     embedding_1_input[0][0]
_________________________________________________________________________________________________
flatten_1 (Flatten)          (None, 18432)         0           xception[1][0]
_________________________________________________________________________________________________
spatial_dropout1d_1 (SpatialDro (None, 105, 300)     0           embedding_1[0][0]
_________________________________________________________________________________________________
dense_1 (Dense)              (None, 8)             147464      flatten_1[0][0]
_________________________________________________________________________________________________
lstm_1 (LSTM)                (None, 64)            93440       spatial_dropout1d_1[0][0]
_________________________________________________________________________________________________
dense_2 (Dense)              (None, 4)             36          dense_1[0][0]
_________________________________________________________________________________________________
dense_3 (Dense)              (None, 4)             260         lstm_1[0][0]
_________________________________________________________________________________________________
concatenate_1 (Concatenate)  (None, 8)             0           dense_2[0][0]
                                                                 dense_3[0][0]
_________________________________________________________________________________________________
dense_4 (Dense)              (None, 8)             72          concatenate_1[0][0]
_________________________________________________________________________________________________
dense_5 (Dense)              (None, 4)             36          dense_4[0][0]
=================================================================================================
Total params: 24,361,988
Trainable params: 21,048,260
Non-trainable params: 3,313,728
_________________________________________________________________________________________________
```

Offensive Model - Summary

```
Model: "model_1"
_________________________________________________________________________________________________
Layer (type)                 Output Shape              Param #     Connected to
=================================================================================================
xception_input (InputLayer)  (None, 100, 100, 3)  0
_________________________________________________________________________________________________
embedding_1_input (InputLayer)  (None, 105)          0
_________________________________________________________________________________________________
xception (Model)             (None, 3, 3, 2048)    20861480    xception_input[0][0]
_________________________________________________________________________________________________
embedding_1 (Embedding)      (None, 105, 300)      3259200     embedding_1_input[0][0]
_________________________________________________________________________________________________
flatten_1 (Flatten)          (None, 18432)         0           xception[1][0]
_________________________________________________________________________________________________
spatial_dropout1d_1 (SpatialDro (None, 105, 300)      0           embedding_1[0][0]
_________________________________________________________________________________________________
dense_1 (Dense)              (None, 8)             147464      flatten_1[0][0]
_________________________________________________________________________________________________
lstm_1 (LSTM)                (None, 64)            93440       spatial_dropout1d_1[0][0]
_________________________________________________________________________________________________
dense_2 (Dense)              (None, 2)             18          dense_1[0][0]
_________________________________________________________________________________________________
dense_3 (Dense)              (None, 2)             130         lstm_1[0][0]
_________________________________________________________________________________________________
concatenate_1 (Concatenate)  (None, 4)             0           dense_2[0][0]
                                                               dense_3[0][0]
_________________________________________________________________________________________________
dense_4 (Dense)              (None, 8)             40          concatenate_1[0][0]
_________________________________________________________________________________________________
dense_5 (Dense)              (None, 2)             18          dense_4[0][0]
=================================================================================================
Total params: 24,361,790
Trainable params: 21,048,062
Non-trainable params: 3,313,728
_________________________________________________________________________________________________
```

Motivational Model - Summary