MULTIMEDIA DATA ANALISYS FOR EMOTION RECOGNITION

Author: Fernando García Novo
Thesis Advisor: Carmen García Mateo
Departamento de Teoría do Sinal e Comunicación

MOTIVATION OF THE WORK

The motivation that has led to the choice of this area of knowledge for the Doctoral Thesis is quadruple:

•The Major Depressive Disorder (MDD), is a mental disorder what affects approximately 3% of the population. Fortunately, medical studies show that the depression is curable, and early detection of depression is very important to be successful with the treatment. Traditional approaches of depression analysis are prevailingly dependents on the verbal reports of patients, and the mental status examination such as SANS, HRSD, BDI-II, PHQ-8, etc. Besides, they commonly require extensive human expertise an are time consuming, therefore, very expensive. So, if we want to carry out mass detection campaigns for the detection of the we have to focus on Automatic Depression Detection (ADD).

• The MDD is a pathological emotion characterized by a pervasive and persistent low mood, and for what has been said previously, the focus of our research.

• It is a field of research much less developed than automatic speech recognition, ADD has not been investigated until 2009.

•The rapid development in machine learning, especially with regard to the Deep Neural Networks. The possibility of applying these new techniques in the field of automatic depression classification will open many lines of research with promising results.

•Increasingly, there are better databases available to study this problem.

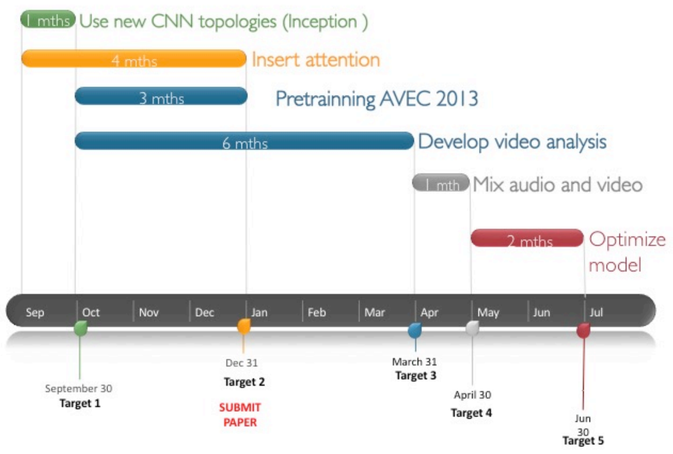
THESIS OBJECTIVES

Develop a methodology that allows to detect the depression of the people through multimedia data.

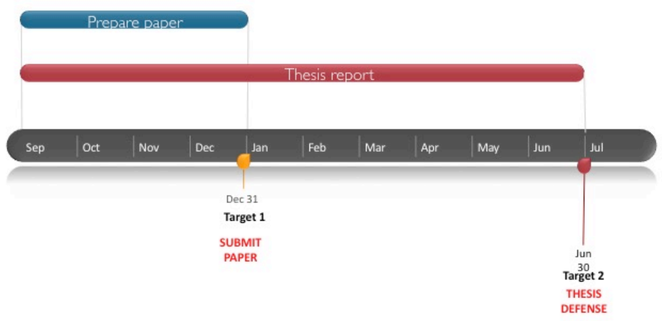
ACHIVED GOALS THIS YEAR

- Change DDBB from AVEC 2013 to AVEC 2016.
- Delete long silences and segment in preprocessing phase.
- Introduce GRU and LSTM cells in the architecture. GRU selected
- Introduce techniques to resolve the imbalanced problem in the DDBB.
- Study different techniques to detect the depression in utterance level. Develop new method.
- Study different preprocessing techniques (MFCC and STFT magnitude), and they impact in the algorithm performances.
- Preleminary results using Inception nets.
- Study different initialization techniques.

NEXT YEAR PLANNING:



RESEARCH PLANNING:



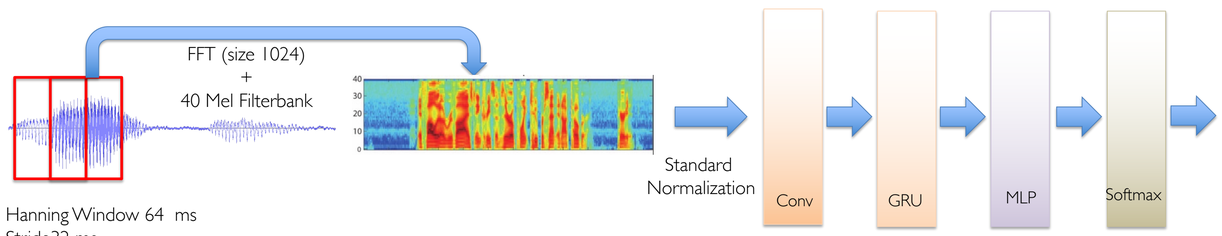
FUTURE

- Finish the inception study to use in ADD problem.
- Insert attention techniques to improve the GRU performances to detect relationships across the time in a window.
- Introduce pretrainning to improve the optimization algorithms convergence.
- Study ADD problem using video analysis to improve the performances
- Study the best way to mix the voice and video analysis.

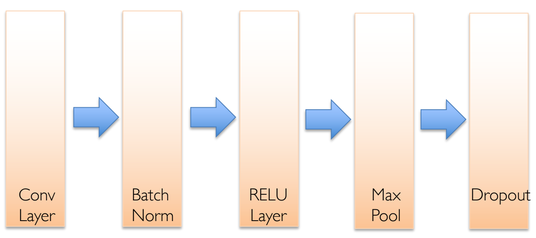
REFERENCES

[1] Xingchen Ma, Hongyu Yang, Qiang Chen, Di Huang, and Yunhong Wang. 2016. DepAudioNet: An Efficient Deep Model for Audio based Depression Classification. In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge (AVEC '16).
[2] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich;"Going Deeper with Convolutions", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1-9
[3] Valstar,Michel and Gratch, Jonathan and Schuller and others. AVEC 2016: Depression, Mood and Emotion Recognition Workshop and Challenge. In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge (AVEC '16). DOI:10.1145/2988257.2988258

PROPOSED ARQUITECTURE BASED IN DEEP NEURONAL NETWORKS



Conv structure:



PARAMETER SETTINGS:

DDBB:
•AVEC 2016
• Using Random Sampling [1] to resolve Uneven sample distribution.

Preprocessed:
• Time Segment = 4 seg.
• Low/High Freq = 140/6854 Hz.
• Freq. Sample= 16 Khz.

Conv:
• kernels, dimensions 40x3
•Max pool kernels 1x3 with stride 1x3

GRU:
•128 cells.

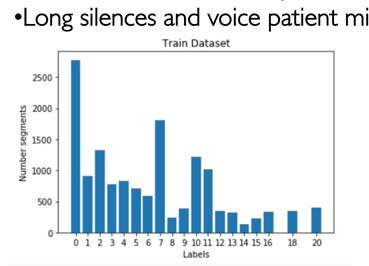
MLP:
• RELU activation
• layer 128 neurons

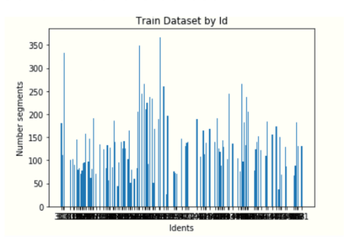
Softmax:
• 2 outputs (depressed or not)

Other techniques:
• Early Stop
• Adam optimization
• Dropout: 0.5
•Clipping

DAIC-WOZ

- Multimedia Data Base AVEC 2016 and 2017 (Voice, Video, Text).
- Use PHQ-8 questionarie to detect depression.
- 128 patients, only 37 depressed.
- The time recorded is very different for each patient.
- Long silences and voice patient mixed with others voices





PRELIMINARY RESULTS (Values of non-depression in brackets)

	F1	Precision	Recall	MAE	RMSE
Result (Sum prob)	0.545(0.792)	0.4(0.95)	0.857(0.678)	0.286	0.534
Result (Mode)	0.50(0.667)	0.33(1.00)	(1.00)0.50	0.40	0.632
DepAudionet [1]	0.52(0.70)	0.35(1.00)	(1.00)0.54	-	-
Base [3]	0.462(0.682)	0.316(.938)	0.857(0.54)	-	-

SUM-PROB

- Usually in utterance level [1] and [3] majority vote method over the whole segments from the same spaker is used to depression prediction.
- Sum prob: sum the log over the whole segments, because the softmax ouput is the probability to be depreseed or not in each segment.

fgnovo@gmail.com