Evaluation Report on Doctoral Thesis

Author: Aisultan Shoiynbek

Title: Automated emotional speech data mining for the speech emotion recognition

University: Suleyman Demirel University

Novelty of the topic:

The author presents a new methodology for the analysis of emotional content in oral expression focused on the Russian and Kazakh languages. An additional contribution is the automated labelling of speech content based on the analysis of facial expressions. This is of great importance in particular for languages where there has been less applied effort to the construction of the necessary datasets for use in machine learning.

Methodology:

The manuscript does an extensive analysis on emotional expression in available datasets for various languages, in particular on the fidelity of these datasets. As most of them are acquired using actors to express emotion, in spite of their professionalism and ability, there is always a tendency to end with stereotyped expressions that do not cover the span of their representation in natural circumstances.
The role of emotions in daily human lives and in particular in enriching spoken content as a way to convey information in communication between people is analysed and compared to how difficult it will be to achieve the same performance level with computers. An interesting survey is also done about how voiced emotions go beyond languages and speech content as a form of universal communication. This study goes up to the point of analysing the emotional expressiveness in singing performances and the accuracy in their perception by emotion type.

The author does an analysis of typically used characteristics for characterisation of voice contents and on tools used to extract them. Finally a deep neural network architecture is presented and the contribution of each of these characteristics in the classification process is analysed.
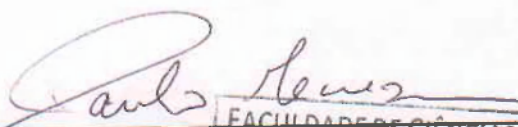
An architecture for automated voice acquisition and emotional labelling is presented based on the joint face-based emotion recognition. This is undoubtedly interesting as producing the necessary datasets for training machine learning algorithms require an enormous effort, typically only available to large corporations that in turn apply them for target markets' populations. The methodology is clearly explained and its application is validated through the training of a DNN whose performance results are presented.

General evaluation:
The work presented in this manuscript is a real contribution to the field on voiced emotion recognition which is of great interest in the field of human-machine interaction, where it may be used to identify user experience issues frequently manifested via frustration or other negative expressions.

My report on this thesis is favourable with a recommendation to explore the possibility to improve the results and review the text.

Coimbra, 20th July, 2020

(Dr. Paulo Menezes)

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA
Departamento de Engenharia
Electrotécnica e de Computadores