ANNOTATION

of Aisultan Shoiynbek

PhD thesis on the specialty 6D070400 – «Computing Systems and Software» on the topic «Automated emotional speech data mining for the speech emotion recognition»

Relevance. Emotions take a significant place in interpersonal human interactions and relationships. Emotion affects our life, because it is a human inside reaction on surrounding and occurring circumstance. It helps us to make intelligent decisions, feel the mood of other person better, convey feelings as feedback for understanding reaction, and adapt to the reality of human behavior. All these facts motivated scientists to study and measure the level of human emotional states. Considering today's realities, during the development of robotics and artificial intelligence, one of the popular areas for research is the recognition of emotions by a machine with the further ability to reproduce emotion. The primary objective of Emotion Recognition (ER) is to aid human to machine interaction. Therefore the emotion recognition is significant in AI products since it will make human-computer interface (HCI) more friendly and improve the user experience.

There are three main steps to achieve successful results in SER, namely, (1) collect emotional speech data or choose already existing database, (2) preprocess the data and extracting useful features, (3) designing robust classification models based on machine learning algorithms.

Many scientists consider that the main issue is to define and explore features with more emotional information and try to discover that area. Other researchers concentrate on improving classifiers of machine learning algorithms; however, in most researches it was used the emotional speech collected under controlled conditions Almost all existing results have been received from databases, which was recorded by actors imitating and artificially producing emotions in front of a microphone. Existing emotion corpuses have next limitations: (1) emotions are limited by quantity and weak in their variety of speaking for machine learning and get impressive results, (2) emotions are not natural, and it means that machine learns to recognize fake emotions.

The relevance of the topic is based on the need to create effective approaches for collecting and labeling natural emotional statements for the recognition of speech emotions with high accuracy.

Aim and objectives of research. The goal of the thesis is developing automatic data collection and labeling methods and algorithms, for speech emotion recognition task with a high level of accuracy, not less than 80%.

Objectives of the study. In accordance with the aim, the following objectives are identified to be solved in this work:

- to study and analyze the exist emotional corpuses

- to study and analyze the current data mining methods

- to study theoretical base of human speech and voice and how to transform to machine language

- to analyze and justify the choice of robust feature transformation method for speech emotion recognition task.

- to analyze and justify the choice of classificatory for speech emotion recognition task.

- to develop the DNN model to separate speech and nonspeech audio content.

- to develop the methods and algorithm which extract audio emotional content from video.

- to develop the automated labeling methods and algorithm.

Object of research. The research focuses on data mining automated methods mining for the speech emotion recognition

Research methods. The objectives assigned were solved by carrying out theoretical and empirical research. As part of the research we used conceptual positions of artificial intelligence, classical machine learning theories, studies of leading foreign and domestic scientists in the field of emotion recognition and speech recognition, probability theory, mathematical statistics, numerical analysis.

Scientific novelty of the work. The novelty of the dissertation is to design an automated method for collecting and labeling speech emotional data. The results obtained in this dissertation will significantly advance the field of artificial intelligence in recognizing speech emotions. Using the method of collecting emotional data, scientists will be able to collect emotional datasets in all languages of the world.

The following scientific statements are to be defended:

Methods and Algorithms for emotional data mining

Methods and algorithms for automated search and download of video containing emotional utterances.

Methods and algorithms for extracting emotional utterances from videos.

Methods and algorithms for the classification and labeling of emotional utterances.

An automated system for collecting and labeling emotional utterances for the task of recognizing speech emotions.

Designed deep neural network model for the classification of human speech from various sounds (noises)

Practical significance of the research results. The practical value of the thesis lies in the possibility of qualitative improvement in service of call-centers, in education, banking, insurance, public services, and medicine.

The practical significance of the study is determined by the possibility of applying its results and recommendations in the: development recognizing true or false emotion systems, drawing up an emotional portrait of the offender by law enforcement agencies, identification of depressive and suicidal tendencies in order to prevent child suicide.

Approbation of the work. The main results of the study were reported and discussed at international conferences: « IV International Scientific and Practical Conference "Informatics and Applied Mathematics" dedicated to the 70th anniversary of professors Biyarov T.N., Waldemar Vuytsik and the 60th anniversary of professor Amirgaliev E.N. » (Kazakhstan, Almaty, 2019); «III International Scientific Conference "Informatics and Applied Mathematics" dedicated to the 80th anniversary of Professor RG Biyashev and the 70th anniversary of Professor M.B. Aidarkhanov»(Kazakhstan, Almaty, 2018); « Trends of Modern Science » (UK, Sheffield, 2019).

Publications. In the framework of this thesis, 10 research papers were prepared and published on this topic, including:

- two articles are published in publishing houses that meet the requirements of the higher Attestation Commission of the Ministry of Education of Science of the Republic of Kazakhstan;
- fourth articles are published in the proceedings of international conferences;
- one article in a foreign publication
- three in international peer-reviewed scientific journal.

Structure and scope of the dissertation. Thesis consists of introduction, four chapters, conclusion, and references list. It is presented on 110 pages of typewritten text, contains 51 figures, 31 tables, a list of references that includes 115 titles.

The first chapter provides an overview of existing research.

In the second chapter, the nature of voice and speech emotions is studied in detail.

In the third chapter, a machine learning model of deep neural networks is proposed, as well as, through a comparative analysis, the most effective feature for transforming vocal emotions into a machine form is determined.

In the fourth chapter, a method for the automatic collection of marked emotional-speech utterances, a classifier of speech and non-speech audio data, a classifier of speech emotions is proposed.

In conclusion, the main outcomes of the dissertation were defined based on results of presented study