THESIS REVIEW

For Presented in Partial Fulfillment for the Degree of Doctor of Science in Computing Systems and Software (degree code: 6D070400)

Thesis name: Automated emotional speech data mining for the speech emotion recognition.

Author Aisultan Shoiynbek

1. **The relevance of the research topic and its relationship with general scientific and national programs (requests for practice and development of science and technology)**

The study on the emotion recognition has particularly important realistic values in such aspects as enhancing the intelligence and humanity of computer, developing new human-machine environment, and improving speech recognition results. The first goal is to search the most useful features with analyzing the features related emotions. The second goal is to find a recognition model to make use of these features. The basic course of speech emotion recognition is introduced, which includes speech signal preprocess and speech feature extraction and speech emotion recognition. Theoretical definition, categorization of affective state and the modalities of emotion expression are presented. To achieve this study, an SER system, based on different classifiers and different methods for features extraction, is developed. There are major challenges to automatic ER including ambiguity about the definition of emotions, the lack of agreement on a manageable set of uncorrelated speech-based emotion relevant features, and the difficulty of collected emotion-related datasets under natural circumstances

2. Scientific results within the framework of the requirements for dissertations (clauses 1-3, clause 5 of the Rules for awarding academic degrees and passports of the relevant specialties of researchers)

The thesis is a qualified research study and is aimed at solving urgent theoretical and practical issues, consists of an introduction, five chapters, discussion a conclusion, a list of used literature. Unfortunately, without any appendixes. It would have been mandatory include an appendix with the developed software.

the introduction describes research relevance, practical significance, substantiates the scientific novelty, objectives of study, the following scientific statements are to be defended, and finally the organization of the thesis is shown in section.

In chapter one is focused on datasets regarding SER. The chapter one presents the challenges faced by SER as a pattern recognition task in this chapter. The contribution of the thesis is presented, followed by the publications related to this thesis in section.

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

In the third chapter, a ML model of deep NN 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 10 emotional-speech utterances, a classifier of speech and non-speech audio data, a classifier of speech emotions is proposed.

In the fifth chapter and conclusion, the main outcomes of the dissertation are defined based on results of presented study.

3. The degree of validity and reliability of each result (scientific position), conclusion and conclusions of the applicant, formulated in the dissertation

In this thesis, modern problems related to ER are described in detail. The scientific provisions formulated are obtained through the construction of machine learning methods, statistical methods for processing multidimensional data and have a rigorous justification. The reliability of the theoretical results obtained is confirmed by their consistency with the results of modeling using the tools proposed.

Thesis represents a complete study targeted formed on the interesting object and complex having practical value and containing new results, the accuracy of which is not in doubt.

4. The degree of novelty of each scientific result (position) and the conclusion of the applicant, formulated in the dissertation

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 AI 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.

5. Assessment of the internal unity of the results obtained

Thesis work has internal unity. In all sections, the proposed models for automated emotional speech data mining for the speech emotion recognition. a reference model is applied to achieve the objective within the objectives. In this work, the linkages and the intrinsic unity of results are clear.

6. The focus of the results obtained by the applicant on the solution of the relevant topical problem, theoretical and applied problem

The research results are aimed at solving the actual problem of real-time and there are available betta type of web service and mobile application for SER, unfortunately these developments have not reached the implementation certificate.

7. Confirmation of sufficient completeness of publications of the main provisions, results, conclusions, and conclusions of the thesis

The results of the work were reported in 10 publications; of them - 4 abstracts of reports at conferences, including international ones, including 1 foreign conferences (Great Britain), 2 articles in collections of scientific papers that meet the requirements of the Committee for Oversight and Certification in the Sphere of Education and Science of the Republic of Kazakhstan, 1 article in a foreign journal, 3 articles were published in foreign journals included in the international citation database SCOPUS: CiteScore:1.4, Quartile: Q3, percentile: 33.

8. Disadvantages of the content and design of the thesis

— general observations are differing in the number of pages in the thesis submitted, the candidate does not seem to be quite sure of the exact number of pages and thesis structure as well.

— it is exceedingly difficult to estimate exact number of datasets that is used for processing, due to the absence of the proposed listing e.g. appendix

— it would be convenient visually estimate the accuracy if everything appeared in the curves

— many drawings are screenshots from the literature review are low quality

— in section 5.4 feature extraction and DNN model neural network architecture is not clear for dataset

9. Compliance with the thesis requirements of paragraph 5 of the Rules of awarding of academic degrees

The obtained results are great practical value since with the recognition of speech emotions, it will become possible to understand human feelings and improve the quality of the services provided, receiving instant feedback. The practical value of the thesis lies in the possibility of qualitative improvement in service, in education, banking, insurance, public services, medicine, law enforcement agencies, and the military

In general, Submitted for Phd (Phd) degree in philosophy is at a high theoretical level, has scientific and practical value and deserves high rating. The Candidate is

acknowledged to be **WORTHY** of being awarded the academic Degree of Doctor of Science in Computing Systems and Software

Reviewer
Omirbekova Zh.Zh
PhD, assistant professor,
Automation and Control Department
Satbayev University

09.11.2020