MEKEBAYEV NURBAPA OTANOVICH

«ALGORITHMS DEVELOPMENT AND RESEARCH FOR IDENTIFICATION AND FEATURES PROCESSING IN SPEECH RECOGNITION PROBLEMS USING MACHINE LEARNING»

ABSTRACT

of the dissertation thesis submitted for the PhD degree in the specialty

«6D060200 – Computer Science»

Research rationale. Since information technology is now rapidly developing and widely used in information exchange and communication, it plays an important role in the development of speech recognition.

In everyday life, language is a natural expression of human nature. It is well known that for many years scientists and engineers have been studying the question of verbal communication between the user and the machine as science and technology developed.

Today, many research projects are impossible without information and communication technologies. They will help to increase economic competitiveness, improve the efficiency of public administration and local self-government. The introduction of information technologies in various industries is rapidly increasing interest in speech recognition technology.

Currently, the most urgent task of solving speech recognition problems is to identify the speaker in automatic speech recognition.

Improving existing speech recognition systems greatly facilitates human-computer interaction. The use of speech recognition systems is also important in the work of law enforcement agencies. The relevance of research on this problem is explained by the low efficiency of speech signals processed in combination with noise in modern systems, as well as the dependence of the results on the detector, and the low speed of the systems.

Modern speech recognition systems are mainly based on hidden Markov models (HMM), which determine the probability that one phoneme in speech can be combined with another phoneme. Thus, the dispersion of the observed signal is achieved by modeling the possible division of characters by a Gaussian distribution. This method was proposed by Lawrence Robimer in 1989 and served as the basis for modeling long speech. Due to the rapid development of the network, Deep Belief Networks are an alternative to SMM and provide high accuracy of the recognition process.

Since the publication of L.Rainer's work, melancholic coefficients (MFCC Mel Frequency Cepstral Coefficients) have been used to describe the speech signal in automatic speech recognition systems that were founded by Paul Mermelstein. In addition, the MFCC marker has recently been replaced with changes in the flow dynamics, which helps to create more reliable systems.

Currently, there are many foreign companies (Agnitio, Nuance, Voice Security Systems) and Russian companies (Speech technologies, Speech Technologies Center) that develop voice biometrics. Most developed systems have a 1-3% error rate, but these applications have a number of drawbacks.

Foreign universities, in particular Carnegie Mellon University (USA), University of Illinois (USA), University of Oregon (USA), University of Eastern Finland headed by researcher Tomi Kinunen, have made significant progress in speech identification. Shelepova V.Yu., researcher at the State University of Informatics and Artificial Intelligence in Donetsk, Ukraine, A.A. Karpov, researcher at St. Petersburg State University of the Russian Academy of Sciences, and Agranovskyi A.B., Lednov D.A. ., Balakirev N.E., Malkov M.A. and domestic scientists from the Eurasian National University. L.N. Gumilev A. Sharipbay and Kazakh National University named after Al-Farabi U.A. Tukeyev, scientist from Nazarbayev University, Esenbayev Yu.A., scientists of the Institute of Information and Computational Technologies of the National Academy of Sciences of the Republic of Kazakhstan Amirgaliyev Y.N., Mussabayev R.R., Mamyrbayev O.Zh.

Machine learning has become a key part of the speech recognition process as one of the areas of artificial intelligence. Creation of speech interfaces for quick control of information and communication technologies, problems of various speech recognition are one of the urgent problems that have not yet been completely solved.

A complex system of artificial intelligence is developing in the country. There are a number of issues that are still waiting to be resolved in this process. One of them is automatic speech recognition, the definition of which speech features have not yet been studied, which determines the relevance of the topic.

The purpose of the thesis: Development and research of optimal speech recognition algorithms and models using machine learning in speech recognition.

Purpose of research. To achieve the purpose, it is necessary to solve the following tasks:

- analysis of modern speech signal recognition systems and methods and speaker identification in the field of speech recognition;
- design of an acoustic enclosure for the audio signals and the identification data of the speaker in the speech recognition process;
- comparison of modern classification models and speaker identification algorithms;
- development of a model and algorithm for speaker identification and gender characteristics based on neural networks;

Object of research. Speech recognition and identification system.

Field of research: Acoustic speech data and speaker identification methods and algorithms in the process of automatic speech recognition.

Research methods: Information theory, signal theory, neural network theory and technology, speech recognition methods and technologies, video recognition theory and technology, software design and construction technologies.

Scientific novelty of the work:

- an acoustic case has been created for the speaker's speech sounds and speaker identification data;

- developed classification algorithms for speaker identification using machine learning;
- developed models and algorithms for determining gender characteristics and speaker identification based on neural networks;
- using the model and algorithm obtained in the course of the study, a software application was created to characterize speech characteristics and identify the speaker.

Theoretical and practical significance of the work. The theoretical significance of the study is to improve neural network models and algorithms for identifying gender differences in speech recognition. At the same time, the theoretical significance of the work is characterized by the development and experimental study of new methods of identifying extended and specific features of speech.

The practical significance of the dissertation research is the use of modern neural network models to identify speech signals; built-in acoustic corpus allows for research in the field of speech recognition.

The main findings of the defense. An acoustic corpus of the Kazakh language has been developed, consisting of more than 36 hour-long speeches intended for recognition and identification of speech signs. Speech features of speakers registered in the database are automatically filtered out and speech features are detected.

The general structure and advanced architecture of neural networks in speech recognition are considered, and the algorithms and models used to identify gender characteristics have been developed.

Personal contribution of the researcher. The applicant has solved the tasks. An algorithm and model of gender identity and speaker identification have been developed. An experimental assessment of the effectiveness of the developed algorithms has been carried out. A model of identification and gender characteristics and speaker identification has been created.

Connection of the thesis topic with research plans. The dissertation research was carried out within the framework of the grant funding project "Development of multilingual automatic speech recognition technology using deep neural networks" at the Institute of Information and Computational Technologies of The Committee of science of the Ministry of Education and Science. (2018-2020, state registration number: 0118RK00139).

The volume and structure of the work. The dissertation consists of an introduction, three sections, a conclusion, and a list of references. The total volume of the thesis: 108 pages of written text, including 35 figures, 10 tables. The reference list of 110 links contains two titles and two appendices.

The introduction defines the relevance of the work and identifies problems related to the topic under study. The idea of the work, the purpose and objectives of the research, scientific novelty and practical significance of the work, and research methods are presented.

The first section analyzes the features of methods and symptoms of speech signal preprocessing, characteristics of speech signals, methods and models of speech recognition and recognition.

The second section is devoted to the use of machine learning algorithms and models in speech recognition problems and neural networks in machine learning. An acoustic corpus for speech recognition and recognition was created. Classification algorithms for speaker recognition were analyzed, and SVC showed the best results.

In the third part, algorithms and models are developed for identifying and processing characters using machine learning in speech recognition tasks. Gender specificity was revealed using the MCC in the preprocessing of the speech signal.A comparison of the MLP and CNN neural network architecture with respect to gender specificity and sound recognition in dynamics was found to demonstrate good CNN performance.

In conclusion, the main conclusions and conclusions of the dissertation are presented, as well as the connection with the future work.

Approbation of the work. Reliability and validity of the research results are confirmed by responsibility for the task, an assessment of the criteria and research in this area, a large number of experiments conducted and their successful implementation in practice. The results of the dissertation were reported and discussed at the following scientific and methodological conferences:

1. «Төртінші өнеркәсіптік революция жағдайындағы дамудың жаңа мүмкіндіктері» атты ҚР Президенті Н. Назарбаевтың Жолдауын іске асыру шеңберінде «Көліктегі инновациялық технологиялар: білім, ғылым, тәжірибе» атты XLII Халықаралық ғылыми-практикалық конференциясында (Алматы, 18 сәуір, 2018);
2. Профессор Р.Г. Бияшевтың 80 жылдығына және профессор М.Б. Айдархановтың 70 жылдығына арналған «Информатика және қолданбалы математика» атты III Халықаралық ғылыми-практикалық конференциясында (Алматы, 26-29 қыркүйек 2018);
3. O.Zh.Mamyrbayev, N.O.Mekebayev, M.Turdalyuly. Speaker identification using MFFC / / IV international scientific and practical conference "Informatics and applied mathematics" dedicated to the 70th anniversary of professors T.N. Biyarov, Valdemar Vuitsik and the 60th anniversary of professor Y.N.Amirgaliyev. – Almaty, 2019. – V. 2. – pp. 384-392.
4. 3rd International Conference Applied Mathematics, Computational Science and Systems Engineering (Rome, Italy, 2018);
5. 11th Asian Conference on Intelligent Information and Database Systems (Yogyakarta, Indonesia, April 6-12 2019).

20 articles were published on the topic of the dissertation and 2 author's certificates were obtained:

1. О.Ж. Мамырбаев, Н.О. Мекебаев, М. Тұрдалыұлы. Сөйлеулерді тану үрдісінде MFCC алгоритмін қолдану // ҚазҰТЗУ хабаршысы. – 2018. -№ 2 (126). – Б. 389-392.
2. О.Ж. Мамырбаев, Н.О. Мекебаев, М. Тұрдалыұлы. Генетикалық алгоритм көмегімен сөйлеуді автоматты танудағы гендерлік сәйкестендіру // Алматы энергетика және байланыс университетінің хабаршысы. – 2018. – специальный выпуск. – Б. 120-129.

3. O.Zh.Mamyrbayev, N.O.Mekebayev, M.Turdalyuly. Recognition system of continuous Kazakh speech based on deep neural networks// Bulletin of the Almaty University of Energy and Communications. – 2018. – special issue. –pp. 130-135.

4. О.Ж. Мамырбаев, М. Тұрдалыұлы, Н.О. Мекебаев. Кіріккен қазақ сөйлеін тану жүйесі // ҚБТУ хабаршысы. – 2018. - № 3 (46). – Б. 129-133.

5. О.Ж. Мамырбаев, М. Тұрдалыұлы, Н.О. Мекебаев, Ахметов И. MFCC негізіндегі дикторды анықтау жүйесі // ҚазҰТЗУ хабаршысы. – 2019. – № 2 (132). – Б. 155-160.

6. O.Zh.Mamyrbayev, M.Turdalyuly, T.Turdalykhyzy, A.S.Shayakhmetova Automatic recognition of Kazakh speech using DNN // KBTU Bulletin. – 2019. – No.2 (49). – pp. 134-142.

7. O.Zh.Mamyrbayev, A.S. Kydyrbayeva, A.T. Akhmediyarova, M.Turdalyuly, N.O.Mekebayev. Systematic review and analysis of voice identification features // KBTU Bulletin. – 2019. – No.2 (49). –pp. 120-133.

8. Қалимолдаев М.Н., Мамырбаев О.Ж., Н.О. Мекебаев., Тұрдалыұлы М. Машиналық оқытуды қолдануда дауыстың гендерлік жіктелінуі // ҚазҰТЗУ хабаршысы – 2019. – № 6 (136). – Б.229-233.

9. Bagher BabaAli, Waldemar Wojcik, Orken Mamyrbayev, Mussa Turdalyuly, Nurbapa Mekebayev. Speech Recognizer-Based Non-Uniform Spectral Compression for Robust MFCC Feature Extraction // Przeglad Elektrotechniczny. ISSN: 0033-2097 – 2018. – No. 6 (94). – pp. 90-93. (Scopus) (Clarivate Analytics)

10. Orken Mamyrbayev, Mussa Turdalyuly, Nurbapa Mekebayev. Automatic Recognition of Kazakh Speech Using Deep Neural Networks // Lecture Notes in Computer Science. 11432 LNAI, pp. 465-474. (Scopus)

11. Mamyrbayev O, Toleu A, Tolegen G, Mekebayev N. Neural Architectures for Gender Detection and Speaker Identification // Journal Cogent Engineering. ISSN: 2331-1916. – 2020. Volume 7, - Issue 1. (Scopus)

12. Orken Mamyrbayev, Turdalyuly, Nurzhamal Oshanova,Tolga Ihsan Medeni, Aigerim Yessentay. Voice Identification Using Classification Algorithms // We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists. June 25, 2019**.** London.

13. Automatic Recognition of Kazakh Speech Using Deep Neural Networks. 11th Asian Conference on Intelligent Information and Database Systems, ACIIDS 2019 (Scopus).

14. О.Ж. Мамырбаев, Н.О. Мекебаев, М. Тұрдалыұлы. Қазақ сөйлеуін тануда іргелі және қолданбалы зерттеуге арналған фонетикалық мәтін // «Төртінші өнеркәсіптік революция жағдайындағы дамудың жаңа мүмкіндіктері» атты ҚР Президенті Н. Назарбаевтың Жолдауын іске асыру шеңберінде «Көліктегі инновациялық технологиялар: білім, ғылым, тәжірибе» атты XLII Халықаралық ғылыми-практикалық конференцияның материалдары. – Алматы, 2018. – Т. 2. – Б. 81-87.

15. О.Ж. Мамырбаев, М. Тұрдалыұлы, Н.О. Мекебаев. Қазақ тілі сөйлеуінің акустикалық және тілдік модельдерін құру // Материалы XIV

Международной Азиатской школы-семинара «Проблемы оптимизации сложных систем». – Алматы, 2018. – Т. 2. – Б. 344-347.

16. O.Zh.Mamyrbayev, N.O.Mekebayev, M.Turdalyuly. Algorithms and architectures of speech recognition systems // Materials of the III International Scientific Conference "Informatics and Applied Mathematics" dedicated to the 80th anniversary of professor R. Biyashev and the 70th anniversary of professor Aidarkhanov M.B. – Almaty, 2018. – V. 2. – pp. 108-121.

17. O.Zh.Mamyrbayev, N.O.Mekebayev, M.Turdalyuly. Speaker identification using MFFC// IV international scientific and practical conference "Informatics and applied mathematics" dedicated to the 70th anniversary of professors T.N. Biyarov, Valdemar Vuitsik and the 60th anniversary of professor Y.N.Amirgaliyev.– Almaty, 2019. – V. 2. – pp. 384-392.

18. Orken Mamyrbayev, Mussa Turdalyuly, Nurbapa Mekebayev, Kuralay Mukhsina, Alimukhan Keylan, Bagher BabaAli, Gulnaz Nabieva, Aigerim Duisenbayeva and Bekturgan Akhmetov. Continuous Speech Recognition of Kazakh Language // AMCSE 2018 - International Conference on Applied Mathematics, Computational Science and Systems Engineering. - Rome, Italy, 2018, v24 - 2019

19. O.Zh.Mamyrbayev, A.S. Kydyrbekova, N.O.Mekebayev, M.Turdalyuly. Review of methods of identifying and authenticating users by voice // Materials of the scientific conference of the IICT CS MES RK "Innovative IT and Smart Technologies" dedicated to the 70th anniversary of professor Utepbergenov I.T., Almaty, 2019. – pp. 315-321.

20. O.Zh.Mamyrbayev, M.Turdalyuly, N.O.Mekebayev, К. Alimkhan, G.S. Nabiyeva, B.Zh Mamyrbayev. A phonetically representative text for creating systems for automatic recognition of Kazakh speech // Science and World. – 2018. – No. 6 (58). – V. 2 – pp. 49-52.

2 copyright certificates were obtained:

1. Computer software "Automatic vocabulary creation system for ASR" Author's certificate dated January 22, 2019 No. 1425.

2. Software for computers "MultiSpeech Multi-Spacing Multiprocessing", author's certificate dated January 9, 2020. No. 7844.