TURDALYULY MUSSA

END-TO-END CONTINUOUS SPEECH RECOGNITION USING DEEP RECURRENT NEURAL NETWORK MODELS

ABSTRACT

of philosophy doctor (PhD) thesis in 6D070400 – Computer engineering and Software by Turdalyuly Mussa

Relevance of the research topic. The development of technology and science associated with the evolution of the interaction of man and machine. Currently, speech interfaces are gaining popularity in human-machine interactions. This is the most natural means of human relations. Automatic speech recognition is an important component of human speech.

One of the most difficult problems in the field of automatic speech recognition is speech recognition in pronunciation. Occurring with the obvious participation of the speakers, but not carried out in advance prepared. The complexity of these tasks stems from the following conversations and speakers, the presence of accent and character in the language of speech, from a variety of quantitative formal words. The presence of hezitation increases the complexity of the task. Parasites, non-lexical extraneous input sounds, “parasitic words”, breaking sentences, exchanging words, repeating, sticking, illegal, indefinite sentences. In talking about pronouncing speech. Therefore, the problem of the relevance of its recognition increases.

The system of recognition of continuous speech is one of the most popular and relevant problems. For example, information media or analysis of large archives of speech at various meetings. But there are a number of features impairing the quality of the speech recognition system, the speaker’s own speech environment. They are: limited frequency band in the range of 0-4000 Hz, the presence of additive and non-specific channel deviations, as well as loss of information in the process of encoding a speech signal. These features further complicate the task of continuous speech recognition.

In English speech recognition studies, the Switchboard-1 corpus in English (300 hours), the Fisher corpus (2000 hours), etc. are used. Most English researchers have paid great attention to the selection of HLIB5 Evai 2000 test results made in the Linguistic Data Consortium (Linguistic Data Consortium, LDC). These bases were used by researchers from IBM Brian Kingsbury, George E. Dahl, from Microsoft Li Deng, Dong Yu, Frank Seide, from Google Andrew Senior, Tara Sainath, etc. In today's time, each country attracts its scientists to the process of recognizing their language. Today, advanced speech recognition systems for the English language make it possible to reduce the level of recognition to 15%.

Recognition of the Russian spoken language and the recognition of uniform samples are devoted to the work of researchers at the St. Petersburg Institute of

Informatics and Automation of the Russian Academy of Sciences Andrei Ronzhin, Alexei Karpov and other scientists from Russia.

Today in our country, scientists from the L.N. Gumilyov Eurasian National University are engaged in speech recognition of the Kazakh language. A. Sharipbay, G. Bekmanov, scientists of the Al-Farabi Kazakh National University U. Tukeev, D. Rakhimov, also from the Institute of Information and Computational Technologies of Ye. Amirgaliev, R. Musabayev.

A. Sharipbay, together with the students, proposed a mathematical theory, morphological and syntactic order, synthesis and analysis of words, a speech recognition algorithm and an application based on the regularities of the formal phonetics of the Kazakh language.

Research studies of the Kazakh language have shown that fluent speech cannot yet be achieved. This situation indicates that the system of recognition of continuous speech in the Kazakh language, corresponding to the level of speech recognition in the Kazakh language, has not yet been created. There are a number of reasons why it is not sufficient to effectively recognize spoken fluent speech in Kazakh speech. First, the inaccessibility of the necessary corpus for assessing the quality of the recognition system of colloquial spoken speech in the Kazakh language; secondly, the Kazakh language, as an agglutinative language, has a much larger number of word forms than analytical languages. If the Kazakh language is spoken in several tens of thousands of words, then the Kazakh language will need a dictionary containing hundreds of thousands of words. Thirdly, the spoken language in the Kazakh language can be called phonetic features, such as articulatory weak form, assimilation phenomenon (adaptation of sounds), reduction (reduction of sound duration). For the Kazakh language, it is necessary to create a recognition system that strongly influences the acoustic variation of the distinctive features of speech arising from these phonetic features.

Based on the above, it can be concluded that the need for algorithms, methods and software to ensure the accuracy of continuous speech recognition in Kazakh speech is very important.

The purpose of the thesis. Development of methods, algorithms and software to ensure the accuracy of recognition of spoken fluent speech in the Kazakh language and its rapid introduction into the system, which is sufficiently used for practical tasks.

The tasks of the research, realizing the purpose of the dissertation work:

1. Analysis of modern speech recognition methods in pronunciation.
2. Development of an acoustic corpus and a text corpus of continuous speech of the Kazakh language.
3. Development of a language model, a dictionary of transcription and an acoustic model, which is part of the system of recognition of continuous speech in pronunciation in the Kazakh language.

4. Assessment of the quality of work of the developed system of recognition of continuous speech in pronunciation in the Kazakh language, as well as a comparison with foreign systems.

5. Preparation of methods, algorithms and software for recognition of continuous speech in pronunciation in the Kazakh language.

The object of study. Automatic speech recognition systems

The subject of study. Methods, algorithms and software for recognition of continuous speech in pronunciation in the Kazakh language

Research methods. It is widely used in applied scientific research: preparing goals and objectives, analyzing the state of research and existing literature, developing algorithmic and software solutions, evaluating the effectiveness of the solutions developed, testing and analyzing the results. In the experimental part of the study, it was carried out only with the help of natural language material, here the test sorting converges with the given using the pronunciation or the composition of the speakers.

Numerical methods of signal processing, probabilistic and mathematical statistical theories, machine learning, applied linguistics, and software development methods were used as research methods.

Novelty of the received results:

– An acoustic corpus and a text corpus of continuous speech of the Kazakh language were developed.

– A speech signal compression algorithm has been developed to increase the quality of acoustic models when recognizing spoken speech.

– Using the developed acoustic and text corpora for the first time, a language model of the continuous speech of the Kazakh language was developed based on deep recurrent neural network models.

– Software developed for the continuous speech recognition system in pronunciation in the Kazakh language, which allows the use of acoustic and language models created using the provided algorithms in the dissertation.

The theoretical and practical importance of the research. The theoretical significance of this work is to improve existing and new algorithms for acoustic models based on deep neural networks for speech recognition problems, as well as in experimental research and development of a new type of receiving symbols that were more advanced than those used previously. The practical significance of the dissertation research lies in the following results: the use of algorithms and software developed during the creation of a speech recognition system and speech in the Kazakh language; demonstration of the use of quality recognition and quick response in practical tasks, such as automatic conversion to text archives, thematic clustering of records.

The main findings of the defense:

The use of deep neural network models in the tasks of continuous recognition of continuous speech of the Kazakh language increases the quality of recognition of automatic speech recognition systems.

Personal contribution of the researcher. The researcher personally solved the problems of dissertation work. Methods and speech recognition algorithms for pronunciation are developed. An experimental evaluation of the developed methods and algorithms has been carried out. Developed software included in the system of continuous speech recognition in pronunciation in the Kazakh language.

Relationship of the dissertation topic with the plans of research programs. The dissertation research was carried out within the framework of the grant financing project "Development of technologies for multilingual automatic speech recognition using deep neural networks". (2018-2020, state registration number: 0118RK00139) at the Institute of Information and Computational Technologies of the Scientific Committee of the Ministry of Education and Science of the Republic of Kazakhstan.

The scope and structure of the work. The thesis consists of introduction, 4 chapters and conclusion. The total amount of the thesis is 95 pages, 41 figures, 6 tables. References consists of 111 items.

In the **introduction**, the relevance of the work was determined and the problems associated with the topic were shown. The idea of the work, the purpose and objectives of the research, the scientific novelty and practical value of the research, the research methods are shown.

The first section describes the classification of automatic speech recognition systems, their creation problems and other models used.

The second section describes the work to create a speech and text corpus in the Kazakh language. First, the stages of creating a speech and text corpus were identified. The process of collecting textual information in the Kazakh language is described. Speakers were announced dubbing speech for work on the creation of an acoustic body and were carried out their identification work. The stages of creating a speech and text corpus were identified. The tools used to collect speech information were identified. Work has been done to create a dictionary consisting of the words of the created cases, and as a result, a dictionary has been created to develop systems for the automatic recognition of continuous speech in the Kazakh language.

In **the third section**, with the help of the Kazakh language speech corpus, work was carried out on the construction of the acoustic and language models of this language. Developed and based speech compression algorithm. The tools for creating acoustic and language models were chosen, with the help of this tool, the Kazakh language speech corpus was adapted to the system training device. The learning process of the system is described using the Hidden Markov model, Gaussian mixtures and deep neural networks. After the learning process of the system, to determine the quality of the acoustic and language models of speech of the Kazakh language, practical work was carried out with monophonic, triphonic,

deep neural networks and deep recurrent networks, and the results of speech quality were obtained.

In the fourth section, with the help of the deep recurrent neural network models obtained from the practical work carried out in previous chapters, a software package was developed which transparently recognizes the Kazakh language using acoustic and language models. The software package consists of a software application running on Windows and Linux operating systems, a Web application running on a client-server architecture, and a mobile application running on Android and iOS systems. Disclosed functions working in the application of the software complex.

In **conclusion**, the main results and conclusions of the thesis are presented and their relationship with future work is indicated.

Approbation of the work. The validity and reliability of the study correspond to the well-founded responsibilities of the task, the analysis of the criteria and the state of research in this area, the large number of experiments conducted and their successful implementation in practice. The results of the thesis were discussed and reported at the following scientific and methodological conferences:

1. II International Scientific and Practical Conference "Information and telecommunication technologies: education, science, practice" (Almaty, December 3-4, 2015).

2. XLII International Scientific and Practical Conference "Innovative Technologies in Transport: Education, Science, Practice" in the framework of the Message of the President of the Republic of Kazakhstan N.A. Nazarbayeva "New development opportunities in the fourth industrial revolution." (Almaty, April 18, 2018).

3. XIV international Asian school-seminar "Problems of optimization of complex systems" (Cholpon-Ata, Kyrgyzstan, 2018).

4. III International Scientific and Practical Conference "Informatics and Applied Mathematics", dedicated to the 80th anniversary of Professor Biyashev RG and the 70th anniversary of Professor Aidarkhanov MB (Almaty, September 26-29, 2018).

5. 3rd International Conference Applied Mathematics, Computational Science and Systems Engineering (Rome, Italy, 2018).

6. 11th Asian Conference on Intelligent Information and Database Systems (Yogyakarta, Indonesia, 6-12 April 2019.

Scientific publications:

1. Yasser Mohseni Behbahani, Bagher BabaAli, and Mussa Turdalyuly. Persian sentences to phoneme sequences based on recurrent neural networks // Open Computer Science. - 2016. - № 6. - p. 219-225. (Scopus)

2. Bagher BabaAli, Waldemar Wojcik, Oken 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). - P. 90-93. (Clarivate Analytics)

3. K. Kalimoldayev, O. Mamyrbayev, M. Turdalyuly, K.E. Nurlan, A.E. Ibraimkulov. Automatic speech recognition by Neural Networks // KazTSTU Herald. - 2016. -№ 5 (117). - p. 435-438.

4. O. Mamyrbayev, N.O. Mekebaev, M. Turdylyly. Using MFCC to ASR // KazNTRU Herald. - 2018. -№ 2 (126). - p. 389-392.

5. O. Mamyrbayev, N.O. Mekebayev, M. Turdylyly. Genetics algorithm of ASR for gender identification // Bulletin of the Almaty University of Energy and Communications. - 2018. - special edition. - pp. 120-129.

6. O. Mamyrbayev, M. Turdalyuly, N.O. Mekebayev. The recognition system of continuous Kazakh speech based on deep neural networks // Bulletin of the Almaty University of Energy and Communications. - 2018. - special edition. - pp. 130-135.

7. O. Mamyrbayev, M. Turdalyuly, N.O. Mekebayev. End-to-end Kazakh speech recognition system // KBTU Herald. - 2018. - №3 (46). - B. 129-133.

8. O. Mamyrbayev, M. Turdalyuly, N.O. Mekebayev, I. Akhmetov. Dictor identification system by MFCC // KazNTRU Herald, № 2 (132), 2019

9. O. Mamyrbayev, M. Turdalyuly, N.O. Mekebayev, K. Alimkhan, G.S. Nabiyeva, B. Mamyrbayev. Phonetically representative text for creating systems for automatic recognition of Kazakh speech // Science and World. - 2018. - № 6 (58). - T. 2 - p. 49-52.

10. O. Mamyrbayev, M. Kalimoldaev, M. Turdalyuly, B. BabaAli. Methods for the construction of multimodal speech recognition // Proceedings of the II International Scientific and Practical Conference "Information and telecommunication technologies: education, science, practice". - Almaty, 2015. - V. 1. - p. 217-221.

11. O. Mamyrbayev, N.O. Mekebayev, M. Turdylyly. Phonetically representative text for creating and researching systems for automatic recognition of Kazakh language // Proceedings of the XLII Proceedings of the “Innovative technologies in transport: education, science, practice "in the framework of the Message of the President of the Republic of Kazakhstan N.A. Nazarbayeva "New development opportunities in the fourth industrial revolution”. - Almaty, 2018. - T. 2. - p. 81-87.

12. O. Mamyrbayev, M. Turdalyuly, N.O. Mekebayev. Developmen of acoustic and language copuses of the Kazakh language // Proceedings of the XIV International Asian School-seminar "Problems of optimization of complex systems." - Almaty, 2018. - T. 2. - p. 344-347.

13. O. Mamyrbayev, N.O. Mekebayev, M. Turdalyuly. Algorithms and Architectures of Speech Recognition Systems // Proceedings of the 3rd International Scientific Conference "Informatics and Applied Mathematics" dedicated to the 80th

anniversary of Professor RG Biyashev. and the 70th anniversary of Professor Aidarkhanov MB - Almaty, 2018. - T. 2. - p. 108-121.

14. 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

15. Orken Mamyrbayev, Mussa Turdalyuly, Nurbapa Mekebayev, Alimukhan Keylan, Aizat Kydyrbekova and Turdalykyzy Tolganai. Automatic Recognition of Kazakh Speech Using Neural Networks // 11th Conference on Intelligent Information Systems and Database Systems. - Yogyakarta, Indonesia, 2019

16. O. Mamyrbayev, A.S. Kydyrbekova, M. Turdalyuly, N.O. Mekebayev. Review of the methods for identifying and authenticating users by voice // Proceedings of the scientific conference of the Institute of Informatics and Information Technologies of the Ministry of Education and Science of the Republic of Kazakhstan “Innovative IT and Smart Technologies”, dedicated to the 70th anniversary of Professor Utepbergenov I. Almaty, 2019.

17. Author's certificate of "System of automatic creation vocabulary for ASR" dated January 22, 2019, No. 1425.