Ministry of Education of the Republic of Belarus

Educational Institution
BELARUSIAN STATE UNIVERSITY OF INFORMATICS
AND RADIOELECTRONICS

UDC 004.93'12

Lu Gangfan

Voice User Authentication System

Abstract to master degree thesis

Specialty 1-98 80 01 Information security

_____

Scientific supervisor

Sergei Nikolaevich Petrov
Candidate of Engineering Sciences,
Assistant Professor

Minsk 2024

GENERAL DESCRIPTION OF WORK

Objective and tasks of the research

The **objective** of this dissertation is to develop a software module for user authentication by voice using machine learning.

To achieve the said objective, it is necessary to solve the following **tasks**:

– make an overview of voice services;

– to analyze threats to the information security of voice services and ways to protect against these threats;

– to study the main features of the speech signal and the basics of voice recognition of the speaker;

– to study the informative features of the voice signal used to verify the speaker;

– analyze machine learning methods to verify the user by voice and select the most promising ones;

– analyze datasets for machine learning and create training and test samples;

– explore machine learning tools for creating and training models, as well as for developing a software module;

– to conduct a comparative analysis of the effectiveness of machine learning methods and choose the best one;

– to develop a software module for user authentication by voice using the selected machine learning method.

The object of research: classification algorithms.

Subject of research: accuracy metrics (accuracy, precision, recall), classifier learning rate, classifier operation speed.

Connection with the priority areas of research and requests of the real sector of the economy

The topic of the thesis corresponds to paragraph 6 Ensuring the security of humans, society and the state (means of technical and cryptographic information protection, cryptology and cybersecurity) of the list of priority areas of scientific, scientific-technical and innovative activities for 2021–2025, approved by the Decree of the President of the Republic of Belarus dated 07.05. 2020 No. 156.

Personal contribution of Master's degree student

Contents of Dissertation work demonstrate personal contribution of the author. Main scientific and practical results were obtained personally by the author.

In works published in cooperation, the author focused on development of the algorithm of operation and architecture of the software module for user authentication by voice.

Dissertation advisor S.N. Petrov, doctor of philosophy, assistant professor, is a co-author of main publications. He formulated objectives and tasks of the research, decided on research methods, participated in work planning and results discussion, interpreted and summarized obtained results, completed scientific editing of Dissertation materials.

Practical approval and publication of Dissertation results

The main results of the dissertation were reported and discussed at: the II International Scientific Research Competition «Scientific Breakthrough – 2024» (Penza, March 20, 2024); the 60th Scientific conference of Graduate Students, Master Students and Students of BSUIR (Minsk, April 22–26, 2024).

According to the results of the research presented in the dissertation, 2 printed works were published.

Structure and scope of the thesis

The thesis consists of an introduction, General characteristics of the work, the main part of the three chapters, conclusion, references. The full volume of the thesis is 68 pages, including 51 illustrations, 7 tables, list of references of 42 titles, including 2 publications of the author.

INTRODUCTION

One of the directions of the field of voice biometrics is speaker verification (voice verification, speaker verification), which is the confirmation of a person's identity using a speech signal. In biometrics, the main scope of verification is access control systems. The process is carried out by comparing the presented sample with the models of speakers created earlier and then making a decision on the compliance of the presented sample with the speaker model.

Authentication is the verification of the authenticity of a user's identity by comparing a person's biometric data with data from a database. In fact, authentication is a component of verification. But if verification is more related to physical access control, then authentication refers to access to information systems.

Voice technology has found particular popularity in sectors such as banking, automotive and medical sectors. More and more banks are integrating voice authentication for an additional layer of security, especially for transactions that take place over the phone. It is expected that the growth of voice banking will ensure the growth of the market. It is expected that during the forecast period, the banking sector will register significant growth rates due to the wider adoption of voice recognition in mobile and web banking applications.

Machine learning, in particular classification algorithms, is increasingly being used to verify the speaker. Common classification algorithms include: logistic regression, naive Bayesian classifier, nearest neighbor, support vector machine, decision trees, random forest, neural networks.

The purpose of the research work is to develop a software module for user authentication by voice using machine learning.

To achieve this goal, it was necessary to solve a number of tasks. Namely, to choose the most effective classifier and determine the necessary and sufficient size of the training sample to achieve a high level of accuracy in recognizing the user by voice. Implement the software implementation of the module.

MAIN PART

Chapter 1: Introduction This chapter provides an overview of speaker recognition and its applications. Key challenges in speaker recognition are discussed. The chapter also outlines the aims and objectives of the research.

Chapter 2: Speech signal processing and feature extraction This chapter describes the methodology used in this research. The datasets used for model training and evaluation are introduced. Data preprocessing steps such as feature extraction are detailed. Various deep learning models investigated are specified, including their network architectures and training procedures. Performance evaluation metrics are defined.

Chapter 3: Experiments and Results This chapter presents results of experiments conducted to build and evaluate deep learning models for speaker recognition. Models are trained on selected datasets and validation results are reported. Performance is analyzed using defined metrics and comparisons are made between different models.

Chapter 4: Literature Review This chapter reviews previous work related to speaker recognition techniques. Traditional methods based on Gaussian mixture models and i-vectors are summarized. The chapter focuses on more recent approaches using deep learning models. Different network architectures that have been applied to speaker recognition are covered. The chapter identifies gaps in current research to be addressed.

CONCLUSION

The system utilizes the Mel Frequency Cepstrum Coefficient (MFCC) feature extraction technique to convert speech signals into feature vectors suitable for neural network processing. And this thesis compares the effectiveness of voiceprint recognition methods under different machine learning methods.

For neural network modeling, the system employs Long Short Term Memory (LSTM) network, which is a powerful RNN variant capable of learning long and short term dependencies in speech signals.The LSTM network is trained for speaker classification and optimized using cross-entropy loss function and Adam optimizer. To evaluate the performance of the system, a speech dataset containing different speakers was used. The system achieved high recognition accuracy on the test set, indicating that the system is able to recognize different speakers effectively. It was found through experiments that using recurrent neural networks and long and short term memory methods system significantly outperforms the traditional GMM-UBM as well as machine learning algorithms such as SVM, Random Forest algorithm and KNN.

In addition, the system was subjected to various reviews including confusion matrix, precision, recall and F1 score. In order to improve the performance of the model, this thesis optimizes the algorithm and the loss function, uses early termination and Dropout regularization to prevent model overfitting, and uses the Adam optimizer to train the model, which has fast convergence and good stability.The results of these reviews show that the system is robust and maintains high recognition accuracy even in the presence of noise and channel interference.

The work in this thesis can be further extended to explore other deep learning models and feature extraction techniques. In addition, speaker adaptation and speaker modeling techniques can be investigated to improve the performance of the system in real-world scenarios. With the development of deep learning techniques, we anticipate that significant progress will continue to be made in the field of voiceprint recognition.

The applicant's list of publications

1-A. Lu Gangfan Speaker recognition using neural networks // Materials of the 60th Scientific conference of Graduate Students, Master Students and Students of BSUIR (Minsk, April 22–26, 2024). – 2024. – P. (The materials is in the printing stage)

2-A. Petrov S. N., Lu Gangfan Chinese voice services market // Научный прорыв 2024: сборник статей II Международного научно-исследовательского конкурса, 20 марта 2024. – Пенза: МЦНС «Наука и Просвещение». – 2024. – С. 26-31.