

РЕФЕРАТ

на магістерську дисертацію

виконану на тему: Спосіб нейромережевого розпізнавання ключових слів у аудіопотоці

студентом: Савченко Владислав Сергійович

Актуальність теми.

У сучасному світі інформаційні технології стрімко розвиваються, і зростаючий обсяг даних вимагає ефективних методів їх обробки та аналізу. Однією з найбільш динамічних областей є обробка аудіоінформації, зокрема, розпізнавання ключових слів у аудіопотоці. Ця технологія знаходить застосування в різноманітних сферах, від інтелектуальних систем допомоги до автоматизації бізнес-процесів і розваг. З розвитком глибокого навчання та нейромереж, розпізнавання ключових слів стало більш точним і ефективним. Нейромережеві моделі, особливо рекурентні нейронні мережі (RNN) та їх вдосконалені версії, такі як довготривалі короткочасні пам'яті (LSTM) та гейтовані рекурентні одиниці (GRU), дозволяють аналізувати послідовності даних та розпізнавати складні патерни у аудіопотоці.

Основною перевагою нейромережевих підходів є їх здатність адаптуватися до різноманітних умов і варіацій у мові, акцентах, інтонаціях та шумі фону.

Це робить їх надзвичайно корисними для застосування у реальних умовах, де аудіосигнали часто супроводжуються різними перешкодами. Крім того, сучасні методи глибокого навчання дозволяють зменшити потребу у великих обсягах вручну позначених даних, завдяки використанню методів навчання з підкріпленням та переднавчання на великих мовних корпусах. Розпізнавання ключових слів у аудіопотоці має широке коло застосувань. Наприклад, у сфері розумних помічників (такі як Siri, Google Assistant, Alexa), де користувачі взаємодіють з системою за допомогою голосових команд.

Уміння швидко і точно розпізнавати ключові слова дозволяє цим системам надавати більш персоналізовані та ефективні послуги.

У сфері безпеки та моніторингу нейромережеве розпізнавання ключових слів може використовуватися для виявлення потенційних загроз або для автоматичного транскрибування розмов у режимі реального часу. Це може бути корисним у контексті забезпечення громадської безпеки, моніторингу дзвінків у службах підтримки клієнтів або для збору розвідувальної інформації. Не менш важливим є застосування цієї технології у медицині, зокрема, для створення систем допомоги людям з обмеженими можливостями. Наприклад, системи розпізнавання мови можуть допомагати людям з порушеннями зору чи слуху отримувати інформацію та взаємодіяти зі світом.

Наукові дослідження у цій області також спрямовані на підвищення ефективності алгоритмів, зменшення затримок у обробці аудіосигналів та покращення точності розпізнавання. Це включає використання методів трансферного навчання, удосконалення архітектур нейромереж та інтеграцію різноманітних джерел даних для підвищення надійності системи.

З огляду на всі наведені аргументи, можна зробити висновок, що нейромережеве розпізнавання ключових слів у аудіопотоці є актуальною і перспективною темою, яка має величезний потенціал для розвитку та впровадження у різних галузях. Розвиток цієї технології сприятиме створенню більш інтелектуальних, адаптивних та корисних систем, які можуть значно покращити якість життя та ефективність роботи у різних сферах діяльності.

Метою дослідження є забезпечення ефективного розпізнавання ключових слів в аудіопотоці за допомогою комбінованого підходу з гнучкою нейромережевою моделлю.

Об'єктом дослідження є процес розпізнавання ключових слів в аудіопотоці.

Предметом дослідження є спосіб виявлення ключових слів в аудіопотоці.

Наукова новизна полягає в наступному:

1) Удосконалено спосіб розпізнавання ключових слів в аудіопотоці що за рахунок використання метод обробки вхідних даних для нейромережевої моделі аналізу голосового сигналу та за рахунок використання згорткової нейронної мережі адаптованої до умов задачі розпізнавання ключових слів забезпечує підвищення ефективності розпізнавання голосових сигналів в комп'ютерних системах загального призначення.

2) Удосконалено метод обробки вхідних даних для нейромережевої моделі аналізу голосового сигналу, що забезпечує підвищення швидкості навчання нейромережевої моделі.

Практична цінність :

Практична цінність даного дослідження полягає у створенні ефективного методу розпізнавання ключових слів в аудіопотоці, що може бути впроваджений у різні галузі. Завдяки новому підходу до обробки аудіоданих та покращенню нейронних мереж, результати дослідження можуть застосовуватися у розробці розумних помічників, систем безпеки, моніторингу та медичних пристроїв для людей з обмеженими можливостями. Підвищена точність і швидкість розпізнавання мовлення сприяють покращенню взаємодії користувачів з технологіями, забезпечуючи більш ефективну та надійну роботу систем. Це, в свою чергу, сприяє підвищенню якості життя та продуктивності у різних сферах діяльності.

Апробація роботи

1. Савченко, В.С., Терейковський, І.А. (2024). Методи нейромережевого розпізнавання ключових слів в аудіофайлі. Science — Technology — Innovation. – 2024.
2. Савченко, В.С., Терейковський, І.А. (2024). Методи нейромережевого розпізнавання ключових слів в аудіофайлі. «Перспективні напрямки наукових досліджень» – 2024.

Структура та обсяг роботи. Робота складається із вступу та чотирьох розділів. Загальний обсяг роботи: 99 аркушів основного тексту, 7 ілюстрацій, 2 таблиці, 2 додатки (лістинг програми, презентація). При підготовці використовувалася література з 55 різних джерел.

Ключові слова: розпізнавання ключових слів, нейронні мережі, розпізнавання голосу, аудіофайл, математичне моделювання, глибоке навчання, машинне навчання.

ABSTRACT

Relevance of the subject. In the modern world, information technology is rapidly evolving, and the growing volume of data requires efficient methods for processing and analysis. One of the most dynamic areas is the processing of audio information, particularly the recognition of keywords in audio streams. This technology finds applications in various fields, from intelligent assistance systems to the automation of business processes and entertainment.

With the development of deep learning and neural networks, keyword recognition has become more accurate and efficient. Neural network models, especially recurrent neural networks (RNNs) and their advanced versions such as long short-term memory (LSTM) and gated recurrent units (GRUs), enable the analysis of data sequences and the recognition of complex patterns in audio streams.

The main advantage of neural network approaches is their ability to adapt to various conditions and variations in language, accents, intonations, and background noise.

This makes them extremely useful for real-world applications where audio signals are often accompanied by different interferences. Moreover, modern deep learning methods reduce the need for large amounts of manually labeled data by utilizing reinforcement learning methods and pre-training on large language corpora. Keyword recognition in audio streams has a wide range of applications. For instance, in the field of smart assistants (such as Siri, Google Assistant, Alexa), where users interact with the system via voice commands. The ability to quickly and accurately recognize keywords allows these systems to provide more personalized and effective services. In the field of security and monitoring, neural network-based keyword recognition can be used to detect potential threats or for automatic transcription of conversations in real-time. This can be useful in the context of ensuring public safety, monitoring calls in customer support services, or for gathering intelligence information. Equally important is the application of this technology in medicine, particularly for creating assistance systems for people with disabilities. For example, speech recognition systems can help people with visual or hearing impairments

receive information and interact with the world. Scientific research in this area is also aimed at improving algorithm efficiency, reducing delays in audio signal processing, and enhancing recognition accuracy. This includes using transfer learning methods, improving neural network architectures, and integrating various data sources to increase system reliability.

Considering all the aforementioned arguments, it can be concluded that neural network-based keyword recognition in audio streams is a relevant and promising topic that has enormous potential for development and implementation in various fields. The development of this technology will contribute to the creation of more intelligent, adaptive, and useful systems that can significantly improve the quality of life and work efficiency in different areas of activity.

The aim of the research is to develop a method for recognizing keywords in audio streams using a combined approach with a flexible neural network model.

The object of the research is the process of keyword recognition in audio streams.

The subject of the research is the method of detecting keywords in audio streams.

The scientific novelty lies in the following:

A new approach to audio data processing and enhancing the efficiency of neural networks in speech recognition tasks has been implemented.

A new technique for converting audio signals into standard MFCC spectrograms has been developed and implemented, improving the quality of input data for further neural network processing. The use of a convolutional neural network consisting of seven layers has halved the recognition error.

The proposed methodology improves the model training speed, allowing for high speech recognition accuracy.