ISSN No: 2350-1278 (Print)
Peer Reviewed & Refereed Journal
National Research Journal

MACHINE LEARNING FOR VOICE RECOGNITION

Karamjeet Kaur

Assistant Professor, PG Department of Computer Science, Sri Guru Teg Bahadur Khalsa College, Sri Anandpur Sahib

ABSTRACT

Although verbal communication is crucial for human interaction, there are still certain obstacles when interacting verbally with robots. Currently, researchers are trying to devise a different method of talking with a machine that is more akin to speaking with a human, for which speech and voice systems have already been specified. Speech recognition places a strong emphasis on the speaker's independence, while voice recognition uses the speaker's voice tone—which can be impacted by their physical characteristics—to identify who the speaker is. Because of this, in order to provide the data to a system, one must first verify these distinctive tonal qualities. Once this recognition is successfully accomplished, the system can additionally benefit from speech recognition by allowing the speaker's features to be adjusted and knowing how the speaker makes their voice or sound.

KEYWORDS: Machine Learning, Communication, Voice Recognition, Speech Recognition, Security and Biometric Authentication

1. INTRODUCTION TO MACHINE LEARNING

This machine learning course gives you a thorough overview of the field's principles and allows you to experiment with a variety of methods, such as reinforcement learning, supervised learning, and unsupervised learning. The goal of machine learning, a branch of artificial intelligence (AI), is to create systems that perform better based on the data they process. Artificial intelligence is the term used to describe systems or machines that mimic human intelligence. Although the terms machine learning and artificial intelligence are commonly used interchangeably, they do not always refer to the same thing when spoken together.

A important variation is that, all AI is not machine learning while all machine learning is AI

1.1 WHAT IS MACHINE LEARNING?

The area of education known as "machine learning" offers the possibility of learning without requiring precise programming. We can therefore conclude that machine learning is the most exciting technological advancement ever made, and the word itself provides evidence that it mimics human learning processes in a very similar way to computer behavior. Nowadays, it is employed in far more fields or sectors than one could imagine.

1.2 FEATURES OF MACHINE LEARNING

1. The best feature of machine learning is its data-driven technology. Large volumes of data are produced daily by the organizations. Thus, organizations can make better judgments by identifying significant links in the data.
2. The machine has the ability to autonomously learn from its past.
3. It is capable of identifying various patterns in the data from the provided dataset.
4. Recognize that large companies place a high value on branding and that using machine learning (ML) will make it simpler to target the right audience.
5. We can all agree that it is comparable to data mining because it works with a lot of data.

ISSN No: 2350-1278 (Print)
Peer Reviewed & Refereed Journal
National Research Journal

2. INTRODUCTION: WHAT IS VOICE RECOGNITION (SPEAKER RECOGNITION)?

What does speech recognition, or speaker recognition, mean?the ability of a machine or program to understand and carry out spoken orders or to receive and interpret dictation.Artificial intelligence (AI) has grown in prominence, and examples of this include voice recognition systems like Apple's Siri and Amazon's Alexa. With speech recognition technology, we can quickly communicate with customers by just speaking to it, confirming requests made hands-free, prompting, and performing other straightforward tasks. ASR-automated speech recognition software makes it simple for voice recognition to recognize and distinguish between voices. For more accurate speech to text discussions, some automatic speech recognition algorithms require users to educate the system to recognize their voices. Voice recognition systems evaluate a speaker's intonation, frequency, and speech pattern. Although voice recognition and speech recognition are sometimes used interchangeably, they are not the same in reality, necessitating an analytical comparison. While voice recognition locates the speaker, speech recognition assesses that as well.

2.1 HOW DOES VOICE RECOGNITION WORK?

Voice recognition softwares needs analog audio to be changed into digital signals on computer. It is known as analog-to-digital altering. A computer must have a digital database of works or consonant and a rapid process for differentiate this data to signals. When any program runs the speech design are stored on the hard drive and then loaded into the memory. An analogy tests these stored designs opposite the output of the A/D converter. This action is known as pattern or design recognition.

3. SPEECH RECOGNITION AI: WHAT IS IT AND HOW DOES IT WORK?

AI speech recognition is a type of technology that facilitates the understanding of human speech data by computers and apps. For many years, this particular trait has always been there, but in more recent times, it has become more precise and sophisticated. Artificial intelligence is used in speech recognition to comprehend any human speaker's words or language and then translate that information into text. Although this technology is still in its infancy, it is rapidly improving in accuracy.

Speech recognition aids in the comprehension and interpretation of spoken data by apps, software, and other organizations for business purposes. Artificial intelligence is used in speech recognition models to comprehend or evaluate your language and voice. It ascertains what we say by examining the words, after which it outputs the words on the screen as text data or model content with decoding veracity.

Speech recognition applications heavily rely on artificial intelligence (AI). The AI mimics human behavior by absorbing information from its surroundings and equipment. Speech recognition software and computers are used to understand what people are saying, which enables them to process data fast and accurately. Both Siri and Alexa serve as models for voice recognition in speech recognition, enabling users to interact with computers through the use of content or transcription data. Recent developments have made speech recognition technology far more accurate and widely utilized than it was in the past. It is also utilized in a variety of industries, including entertainment, healthcare, education, and customer service.

Despite that, there any many challenges that we need to overcome, for example handling of drawls and language and the trouble of understanding in noisy environment. Regardless of these challenges, speech recognition is a thrilling area of AI with the great ability of future growth.

3.1 HOW DOES SPEECH RECOGNITION AI WORK?

Voice recognition as well as speech recognition is a complicated process that includes audio accuracy over many steps and language or information solutions, along with:

ISSN No: 2350-1278 (Print)
Peer Reviewed & Refereed Journal
National Research Journal

i) Understanding the words, models as well as content in the user's audio or speech. Many steps are needs for his model to be train for business accuracy to determine each word in your vocabulary or audio cloud.

ii) Modifying those languages or audios into the text. This step includes transform recognized audios into letters or numbers such that other parts of the AI software solutions system can process those models decides what was said. Next, AI determines the meaning of the words which are frequently used together and most often spoken together. This concept is known as predictive modeling.

iii) Looking into commands from the rest of audio or speech content is also known as elucidating.

4. SPEECH RECOGNITION AI AND NATURAL LANGUAGE PROCESSING

Artificial intelligence's other crucial part is natural language processing (NLP) which requires analyzing data related to natural language and transform it into the machine graspable format. AI as well as speech recognition plays a crucial role in NLPs in improving the efficiency and perfection in human language recognition.

To improve the customer experience as well as enhancing the business applications many of the business now involves speech recognition or speech-to-text software .companies can reprint calls, meeting etc. by using natural language processing (NLP) as well as speech recognition AI together. To provide the perfect customer experience many of the companies like Google, Amazon and apple are manipulating AI- based voice or speech recognition applications.

5. USE CASES OF SPEECH RECOGNITION AI

For many business benefits speech recognition AI is being used in different industries and applications. AI is helping people in different areas like call centers and voice-activated audio content assistants as well as ATMs to interact people with technology and software with better data record accuracy than ever before.

5.1 CALL CENTERS

The most admired used of speech AI is speech recognition in call centers. With the help of this technology you will be able to listen what customer is saying and then you can use that information via cloud models to counter suitable.

For the audio and voice biometrics speech recognition technology can also be used, which means we can use the voice design as evidence of recognize or permit for access solutions or services without depend on passwords or other trick or method like eye scans or fingerprints. This can help us in business problems like forgetting password or came to terms of security codes in favor of anything more secure like your voice.

5.2 BANKING

Financial and banking institutions are helping customers with their business problems by using speech AI applications. For example, you can ask any bank about your details like bank balance or the current interest rate on your saving account. This will help us in saving time because if customer service representatives have to answer the question then they would have to search and then look at the cloud data, which will take much time and if we use AI then it will work faster than the representative and provides the better customer services.

ISSN No: 2350-1278 (Print)
Peer Reviewed & Refereed Journal
National Research Journal

5.3 TELECOMMUNICATIONS

AI is a speech-enabled technology that is obtaining friction in the telecommunications industries. Speech recognition is a great technology model with enables or helps the calls to be examine and managed more efficiently. To deliver the better customer service it helps the agents to distinct on their highest-value task.

Customers can now interact with businesses in real-time 24/7 via voice transcription solutions or a text messaging application, which makes them feel more connected with the company and improves their overall experience.

5.4 HEALTHCARE

In many different areas like transcription solutions Speech AI is used which is a learning technology. One of the most important is healthcare, as it helps the hospital staff like doctors and nurses to take care their patients very well. Voice-activated devices that uses learning model allows patients to communicate nurses, doctors as well as other healthcare professionals without typing on keyboard or using their hands. Via cloud data doctors can use speech recognition to help patients to understand their feelings and why they feel that way. It is much easier for them than reading from the pamphlet or the brochure- and it is more appealing. Speeches AI can also note down the patient histories and also help with the medical interpret**.**

5.5 MEDIA AND MARKETING

Tools such as dictation software use speech recognition and AI to help users type or write more in much less time. Roughly speaking, copywriters and content writers can transcribe as much as 3000-4000 words in as less as half an hour on an average.

Accuracy, though, is a factor. These tools don't guarantee 100% foolproof transcription. Still, there are many advantages of using this as it helps in marketing people as well as media in formulating their first drafts.

6. WHAT ARE THE CHALLENGES IN WORKING WITH SPEECH RECOGNITION AI?

When using voice AI, there are a variety of obstacles to overcome. Cloud computing and technology are both relatively young and are evolving quickly. Because of this, it is difficult to estimate with precision how long it will take any business to develop a speech-activated product. Another issue with voice AI is that evaluating your data requires the appropriate tools. Because so many individuals demand access to the cloud and technology, it may occasionally be challenging to locate the ideal solution for your needs because it requires a lot of time and work.

You have to utilize proper syntax and terminology when developing your algorithms on the cloud. It can get challenging at times since it has to comprehend how people and machines communicate. Because of this, voice recognition technology needs to improve, and occasionally it can be challenging for computers to comprehend all you say. You must train speech recognition software on your voice so that it can comprehend what you say. This may take longer since you have to pay close attention to how your voice and other sounds differ from those of other individuals. The other concern is that medical records are subject to certain privacy restrictions.

Because these kinds of rules vary from state to state, you should consult your authority before using voice AI technology. It's crucial that you train your employees on the technology and demonstrate its functionality if you choose to employ speech artificial intelligence.

ISSN No: 2350-1278 (Print)
Peer Reviewed & Refereed Journal
National Research Journal

6.1 BIOMETRIC AUTHENTICATION

For improving the security, one requires to manage the authenticate someone to make sure that a person or human is who he or she claims to be. Voice recognition's authentication is generally split into the need to:

i) First verify the person with comparison with the previous stored data for that person if he or she is who they claim to be.

ii) Identification aims to identify a person out of a larger crowd, by seeing whether that person's profile equals that of anyone in the database or not. Obviously, one can only verify or identify a person if the relevant information has been previously obtained. When thinking about current authentication methods, We need to be careful of the fact that the passwords can easily be hacked or guess but can easily be changed, with the biometric solutions it may be harder to mimic, but can never be changed as they are connected with the person's physical body. So once they have been compromised there is a serious challenge for that person and/or the verification/identification system.

6.2 SPEAKER VERIFICATION

So the biometric authentication's main focus is to use a person's unique features. As a technology it has gained popularity throughout the world as it can help reduce fraud cases because it uses physical characteristics to identify oneself. Biometric comes from the Greek, "bio" which means: "life" and "metric" which means: "to measure", and when combined they describe how to: "measure one's life". Biometric characteristics of individuals are unique and can be used to authenticate one's identity for different systems. There are two types of biometric technology properties which can be either behavioral or physical. Physical properties involves: fingerprints, DNA, facial recognition and retina/iris scan, while the behavioral properties involves signatures, handwriting and voice.

6.3 SPEAKER IDENTIFICATION

Biometric voice recognition systems mainly focus on identifying these unique characteristics and then store those within a database. This needs one to realize which of the parameters are important to uniquely realize one's voice and examine that these parameters are physical but cannot required to be measured easily; one need to understand how to check these features form the voice signals. To gain the information it is important one to work backwards, as there might have some features that one would choose to exclude from the profile to halt the accuracy going down because one has e.g.cold. there are many challenges like imitation, as in this situation the standard physical features are "adopted on purpose" to provide a different output, and it is yet to be seen how this could be spot accordingly, that is why system should be capable of handling these challenges. For speech recognition tool voice profile data could also be useful if one can automatically detect who speaks and adjust the speech recognition profile for the person or people to improve the overall accuracy.

6.4 SECURITY IN VOICE RECOGNITION

A voice recognition system is designed to identify an administrator voice. The administrator voice can be verified by using the MATLAB software for coding the voice recognition. For further analysis and processing the key is to change the speech waveform to a type of parametric representation.

Mel-frequency Cestrum Coefficients (MFCC) are just one of several options available for parametrically describing the speech signal for the voice recognition system.Using the MFCC techniques, the computer will compare the recorded input voice signals with the signal stored in the

ISSN No: 2350-1278 (Print)
Peer Reviewed & Refereed Journal
National Research Journal

database. The single word recognition is the foundation of the voice-based biometric technology. In order to train and store, an administrator must change the password just once during the training session. If there is a match, the user can enter the password again to complete the recognition test. The output of a MATLAB simulation can be used to determine if the user is accepted or rejected. After the system has been tested, it will identify the correct user's voice and reject the other user's. Finally, we can state that the accuracy of the entire system recognizes the user's voice successfully. It is the medium range of a security level system.

7. CONCLUSION

The ability to recognize the factors influencing voice and speech is necessary to enhance the human-machine interface. Gaining an understanding of these factors and their significance will aid in enhancing artificial systems' ability to process human speech and voice more naturally. Being able to distinguish between the factors that pertain to voice and/or speech is also crucial. Knowing these speech recognition criteria is especially crucial if you want to use voice as a biometric identification method to increase security. Nonetheless, it is anticipated that one will be able to enhance speech recognition systems in general by developing and comprehending these voice-related factors.

REFERENCES

Asci, F., Costantini, G., Di Leo, P., Zampogna, A., Ruoppolo, G., Berardelli, A., ...&Suppa, A. (2020). Machine-learning analysis of voice samples recorded through smartphones: the combined effect of ageing and gender. Sensors, 20(18), 5022.Barbosa,

Bourlard, H. A., & Morgan, N. (2012). Connectionist speech recognition: a hybrid approach (Vol. 247). Springer Science & Business Media.

Deng, L., & Li, X. (2013). Machine learning paradigms for speech recognition: An overview. IEEE Transactions on Audio, Speech, and Language Processing, 21(5), 1060-1089.

F. G., & Silva, W. L. S. (2015, August). Automatic voice recognition system based on multiple Support Vector Machines and mel-frequency cepstral coefficients. In 2015 11th International Conference on Natural Computation (ICNC) (pp. 665-670). IEEE.

Hegde, S., Shetty, S., Rai, S., &Dodderi, T. (2019). A survey on machine learning approaches for automatic detection of voice disorders. Journal of Voice, 33(6), 947 - e11.

Kinkiri, S., Melis, W. J., &Keates, S. (2017). Machine learning for voice recognition. Padmanabhan, J., & Johnson Premkumar, M. J. (2015). Machine learning in automatic speech recognition: A survey. IETE Technical Review, 32(4), 240-251.

Prasad, V. (2015). Voice recognition system: speech-to-text. Journal of Applied and Fundamental Sciences, 1(2), 191.

Shin, J. W., Chang, J. H., & Kim, N. S. (2010). Voice activity detection based on statistical models and machine learning approaches. Computer Speech & Language, 24(3), 515-530.

Yu, D., & Deng, L. (2016). Automatic speech recognition (Vol. 1). Berlin: Springer.