JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Review Paper on SPEECH TO TEXT USING MACHINE LEARNING

[1] **Prof. Rahul Raut,**

[2] **Pranav Baisane,** [3] **Pratik Tamboli,**

[4] **Siddhavi Raul**

[1]Assistant Professor, Department of Information Technology, Sandip Institute of Technology and Research Center, Nashik, India,
[2]Department of Information Technology, Sandip Institute of Technology and Research Center, Nashik, India,
[3]Department of Information Technology, Sandip Institute of Technology and Research Center, Nashik, India,
[4]Department of Information Technology, Sandip Institute of Technology and Research Center, Nashik, India

ABSTRACT:

Ever wish you could just speak your thoughts into a document instead of writing or typing them? Doing so may be easier than you think. Speech-to-text technology has been around for decades in one form or another. It was made popular by technology companies such as IBM, the Department of Defense, and medical offices. Over the last few years it has become increasingly more popular with the general public and much more accurate at deciphering what we are saying. This has positively impacted everyone from commuters wanting to dial a number without looking at the keypad to individuals with disabilities needing to send an SMS. A number of voice recognition systemsare available on the market. The most powerful can recognize thousands of words. However, they generally require an extended training session during which the computer system becomes accustomed to a particular voice and accent. Such systems are said to be speaker dependent. A speaker dependent system is developed to operate for a single speaker. These systems are usually easier to develop, cheaper to buy and more accurate, than but not as flexible as speaker adaptive or speaker independent systems. Speaker dependent software works by learning the unique characteristics of a single person's voice, in a way similar to voice recognition. New users must first "train" the software by speaking to it, so the computer can analyze how the person talks.

This often means users have to read a few pages of text to the computer before they can use the speech recognition software.

Keywords**:** Machine Learning, sk-learn,pandas, , data analysis, pyaudio, pyaudio, portaudio.

I. INTRODUCTION

Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format. Rudimentary speech recognition software has a limited vocabulary of words and phrases, and it may only identify these if they are spoken very clearly. This project presents techniques for speech-to-text and speech-to-speech automatic summarization based on speech unit extraction and concatenation. For the former case, a two stage summarization method consisting of important sentence extraction and word-based sentence compaction is investigated. Sentence and word units which maximize the weighted sum of linguistic likelihood, amount of information, confidence measure, and grammatical likelihood of concatenated units are extracted from the speech recognition results and concatenated for producing summaries. For the latter case, sentences, words, and between-filler units are investigated as units to be extracted from original speech. These methods are applied to the summarization of unrestricted-domain spontaneous presentations and evaluated by objective and subjective measures.

II LITERATURE REVIEW

1.TITLE:VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT

VOICE RECOGNITION SYSTEM:SPEECH-TO-TEXT is a software that lets the user control computer functions and dictates text by voice. The system consists of two components , first component is for processing acoustic signal which is captured by a microphone and second component is to interpret the processed signal, then mapping of the signal to words. Model for each letter will be built using Hidden Markov Model(HMM). Feature extraction will be done using Mel Frequency Cepstral Coefficients(MFCC). Feature training of the dataset will be done using vector quantization and Feature testing of the dataset will be done using viterbi algorithm. Home automation will be completely based on voice recognition system.

2. TITLE: Evaluating Speech-to-Text Automatic Transcription of Digitized Historical Oral Sources

Conducting “manual” transcriptions andanalyses is unsustainable for most his-torical oral archives because they requirea remarkable amount of funds and time.The FONTI 4.0 project aims at exploringthe suitability of automatic transcriptionand information extraction technologies for making historical oral sources avail-able. In this work, we conducted an exper-iment to test the performance of two com-mercial speech-to-text services (GoogleCloud Speech-to-text and Amazon Transcribe) on digitized oral sources. We cre-ated an eight-hour corpus made of man-ually transcribed and annotated historicalspeech recordings in TEI format. The re-sults clearly show how audio quality anddisturbing elements (e.g., overlaps, for-eign words, etc.) impact on the automatictranscription, showing what needs to beimproved for implementing an unsuper-vised transcription chain

3.TITLE:SPEECH RECOGNITION USING LONG SHORT TERM MEMORY RNN

The sound waves are fed into the PC and are required to convert into content. Since sound it waves are analog signal, they are to be tested through Nyquist hypothesis. This inspected signal is directed towards the neural system, however earlier handling of flag is improved outcome and exact expectations of verbally expressed words. Pre-handling is gathering of an extensive tested flag into 20-milliseconds little pieces. Pre-handled tested information which is in computerized position is presently encouraged to the Recurrent Neural Network (RNN). Types utilized in STT Engine are talked about in more areas.Speech Recognition is otherwise called Computer Speech recognition which implies influencing the PC to comprehend what we talk. As a rule program, a PC c listen to us, anunderstands and convert it into texts. Therefore speech recognition is additionally referred as Speech to Text change technology .It comprises of amplifier for people to talk, acknowledgment of speech programming and a PC to do the work. The essential acknowledgment of speech framework is appeared.

4. TITLE: Application of the Speech Recognition Technology in Language Education

As technology advances and innovative communication products emerge, the interaction between people, globally, changes. Therefore, technology plays a major role in influencing both language and culture. The purpose of this paper is to answer the question of how the speech recognition technology available on the smartphones and tablets can support language learning. The very latest and advanced features of the smartphones such as the speech recognition capabilities were utilised in experimenting with language learning. The languages included French, German, Italian, Japanese and Mandarin. In all cases, the established learning concepts such as learning by guidance were considered. This paper has demonstrated how the latest technologies such as speech recognition can be utilised to produce effective educational materials for immersive language education. It is concluded that Technology can enable a learner to simulate the learning by guidance approach in the absence of the tutor or the opportunity of being in the actual environment.

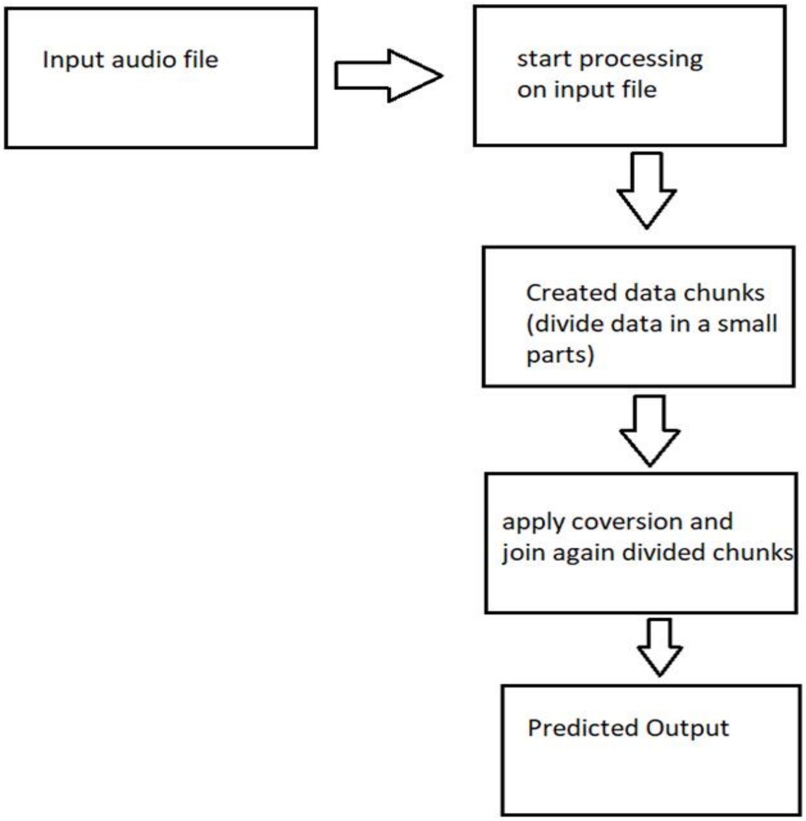
5. Title: Pushing the Limits of Semi-Supervised Learning for Automatic Speech Recognition

We employ a combination of recent developments in semi-supervised learning for automatic speech recognition to obtain state-of-the-art results on LibriSpeech utilizing the unlabeled audio of the Libri-Light dataset. More precisely, we carry out noisy student training with SpecAugment using giant Conformer models pre-trained using wav2vec 2.0 pre-training. By doing so, we are able to achieve word-error-rates (WERs) 1.4%/2.6% on the LibriSpeech test/test-other sets against the current state-of-the-art WERs 1.7%/3.3%.

6. Title: Scribosermo: Fast Speech-to-Text models for German and other Languages

Recent Speech-to-Text models often require a large amount of hardware resources and are mostly trained in English. This paper presents Speech-to-Text models for German, as well as for Spanish and French with special features: (a) They are small and run in real-time on microcontrollers like a RaspberryPi. (b) Using a pretrained English model, they can be trained on consumer-grade hardware with a relatively small dataset. (c) The models are competitive with other solutions and outperform them in German. In this respect, the models combine advantages of other approaches, which only include a subset of the presented features. Furthermore, the paper provides a new library for handling datasets, which is focused on easy extension with additional datasets and shows an optimized way for transfer-learning new languages using a pretrained model from another language with a similar alphabet.

III. PROPOSED SYSTEM


Input audio file
start processing on input file
Created data chunks (divide data in a small parts)
apply coversion and join again divided chunks
Predicted Output

The FAU Aibo Sentiment Corpus is more sentimental and contains speech of children. All three databases are German, however, the methods described later in this chapter work languageindependently. At the beginning of the work reported in this thesis, databases with spontaneous sentimental speech were sparsely available. Nowadays, the situation has improved, though it is still far from abundant. Besides the databases already mentioned an overview of the most important corpora has been compiled by Ververidis and Kotropoulos. While classification on the Berlin database is a relatively easy task, it is not very realistic as it contains acted speech obtained under ideal conditions, a scenario one would scarcely find in an application. Furthermore, it is limited in size. The Aibo and the SmartKom database are much harder tasks, because sentiments are not as prototypical and clear, but they are larger and close to realistic conditions. Thus, by evaluating these three databases, that are described in detail in the following, a wide variety of sentiments is covered and results can be expected to be general.

i. Data Modeling

Data analysis is a core practice of ultramodern businesses. Choosing the right data analytics tool is grueling, as no tool fits every need. To help you determine which data analysis tool stylish fits your association, let’s examine the important factors for choosing between them and also look at some of the most popular options on the request moment.

There are a many effects to take care of before assessing the available tools. You should first understand the types of data your enterprise wants to assay, and, by extension, your data integration conditions. In addition, before you can begin analysing data, you ’ll need to elect data sources and the tables and columns within them, and replicate them to a data storehouse to produce a single source of verity for analytics. You ’ll want to assess data security and data governance as well. However, for illustration, there should be access control and authorization systems to cover sensitive information, If data is participated between departments.

ii. Feature Execution

In machine literacy, pattern recognition, and image processing, point birth starts from an original set of measured data and builds deduced values (features) intended to be instructional and non-redundant, easing the posterior literacy and conception way, and in some cases leading to better mortal interpretations. Point birth is related to dimensionality reduction.

When the input data to an algorithm is too large to be reused and it's suspected to be spare (e.g. the same dimension in both bases and measures, or the repetitiveness of images presented as pixels), also it can be converted into a reduced set of features (also named a point vector). Determining a subset of the original features is called point selection. The named features are anticipated to contain the applicable information from the input data, so that the asked task can be performed by using this reduced representation rather of the complete original data.

CONCLUSION

The system and music classification both are difficult and problematic tasks individually. In this paper, we have studied the algorithms or techniques which helps to classify both the facial expression and the music too. Research works are going on to collaborate both the works. Soon there will be a music system which will play music according to the emotions of the person.We have explored the automatic classification of audio signals into an hierarchy of musical genres. More perform full, basic feature sets for representing the texture, rhythmic content and pitch content are proposed. The performance and relative importance of the proposed features is investigated by training statistical pattern recognition classifiers using real-world audio collections. Both whole file and real-time frame-based classification schemes are described. The use of proposed feature sets, classification of 63% for ten musical genres is achieved. This result is comparable to results reported for human musical genre classification.

REFERENCES

[1] P. iCortez, iA. iCerdeira, iF. iAlmeida, iT. iMatos iand iJ. iReis. iModelling iwine ipreferences iby idata imining fromiphysicochemical iproperties. In iDecision iSupport iSystems, iElsevier, i47(4):547-553, i2009

[2] K.Thakkar,J.Shah,R.Prabhakar,A.Narayan, A.Joshi, "AHP and MACHINE LEARNING TECHNIQUES for Wine Recommendations,"*International Journal of Computer Science and Information Technologies*,7(5),pp. 2349-2352 ,2016

[3] Reddy, Y. S., & Govindarajulu, P. (2017). An Efficient User Centric Clustering Approach for Product Recommendation Based on Majority Voting: A Case Study on Wine Data Set. *IJCSNS,* 17(10), 103.

[4] P.Appalasamy and A. Mustapha,(2012) "Classification based Data Mining Approach for Quality Control in Wine Production", Volume 12 (6), Journal of Applied Sciences.

[5] Shen Yin,(2013) "Research Article: Quality Evaluation Based on Multivariate Statistical Methods", Article ID 639652, 10 pages.

[6] Paulo Cortez1, Juliana Teixeira1, Ant´onio Cerdeira2."Using Data Mining for Wine Quality Assessment".

[7] Cortez, P., Cerdeira, A., Almeida, F., Matos, T., Reis, J., 2009. Modeling wine preferences by data mining from physicochemical properties. Decis. Support Syst. 47, 547–553. https://doi.org/10.1016/j.dss.2009.05.01