Data Augmentation for Training Data by Using Signal Processing Techniques

G N Chaithanya

Department of Computer Science, Sri Venkateswara University, Tirupati

Abstract— *Speech-based virtual assistants, such as Amazon Alexa, Google assistant, and Apple Siri, typically convert users' audio signals to text data through automatic speech recognition (ASR) and feed the text to downstream dialog models for natural language understanding and response generation. The ASR output is error-prone; however, the downstream dialog models are often trained on error-free text data, making them sensitive to ASR errors during inference time. To bridge the gap and make dialog models more robust to ASR errors, we leverage an ASR error simulator to inject noise into the error-free text data, and subsequently train the dialog models with the augmented data. Data augmentation in data analysis are techniques used to increase the amount of data by adding slightly modified copies of already existing data or newly created synthetic data from existing data. It acts as a regularizes and helps reduce over fitting when training a machine learning model. Random value to generate the signals to show the result. Data augmentation is useful to improve performance and outcomes of machine learning models by forming new and different examples to train datasets. Here, we can use the four method to process the data by using signal processing techniques of the system. Data augmentation techniques enable machine learning models to be more robust by creating variations that the model may see in the real world.*

I. INTRODUCTION

Speech-based virtual assistants, such as Amazon Alexa, Google Assistant, and Apple Siri, have become increasingly powerful and popular in our everyday lives, offering a wide range of functionality including controlling smart home devices, booking movie tickets, and even chit-chatting. These speech-based virtual assistants typically contain the following components: an automatic speech recognition (ASR) module that converts audio signals from a user to a sequence of words, a natural language understanding (NLU) module that extracts semantic meaning from the user utterance, a dialog management (DM) module that controls the dialog flow and communicates with external applications if necessary, a natural language generation (NLG) module that converts the system response to natural language, and a text-to-speech (TTS) module that converts the text response to an audio Response. The errors made by the ASR module can propagate to the downstream dialog models in NLU and DM and degrade their performances. Data augmentation is useful to improve performance and outcomes of machine learning models by forming new and different examples to train datasets. Data augmentation techniques enable machine learning models to be more robust by creating variations that the model may see in the real world. Signal processing manipulates information content in signals to facilitate automatic speech recognition (ASR). It helps extract information from the speech signals and then translates it into recognizable words. Signal processing is an electrical engineering subfield that focuses on analyzing, modifying, and synthesizing signals such as sound, images, and scientific measurements. Signal processing is an electrical engineering subfield that focuses on analyzing, modifying, and synthesizing signals such as sound, images, and scientific measurements.

Speech recognition system is mainly based on three models, an acoustic model, a language model and a pronunciation lexicon. Besides, the performance of these models relies greatly on the amount of data used during training. During the last years, several researchers focused their works on improving two key elements in speech recognition: speech data, and acoustic model. The problem of automatic environmental sound classification has received increasing attention from the research community in recent years. Its applications range from context aware computing and surveillance to noise mitigation enabled by smart acoustic sensor networks. To date, a variety of signal processing and machine learning techniques have been applied to the problem, including matrix factorization, dictionary learning, wavelet filter banks and most recently deep neural networks. In particular, deep Convolutional neural networks (CNN) are, in principle, very well suited to the problem of environmental sound classification: first, they are capable of capturing energy modulation patterns across time and frequency when applied to spectrogram-like inputs, which has been shown to be an important trait for distinguishing between different, often noise-like, sounds such as engines and jackhammers. Second, by using Convolutional kernels (filters) with a small receptive field, the network should, in principle, be able to successfully learn and later identify spectro-temporal patterns that are representative of different sound classes even if part of the sound is masked (in time/frequency) by other sources (noise), which is where traditional audio features such as Mel Frequency Cepstral Coefficients (MFCC) fail. Yet the application of CNNs to environmental sound classification has been limited to date. For instance, the CNN proposed in obtained comparable results to those yielded by a dictionary learning approach (which can be considered an instance of "shallow" feature learning), but did not improve upon it. Deep neural networks, which have a high model capacity, are particularly dependent on the availability of large quantities of training data in order to learn a non-linear function from input to output that generalizes well and yields high classification accuracy on unseen data. A possible explanation for the limited exploration of CNNs and the difficulty to improve on simpler models is the relative scarcity of labeled data for environmental sound classification. While several new datasets have been released in recent years, they are still considerably smaller than the datasets available for research on, for example, image classification. An elegant solution to this problem is data augmentation, that is, the application of one or more deformations to a collection of annotated training samples which result in new, additional training data. A key concept of data augmentation is that the deformations applied to the labeled data do not change the semantic meaning of the labels. Taking an example from computer vision, a rotated, translated, mirrored or scaled image of a car would still be a coherent image of a car, and thus it is possible to apply these deformations to produce additional training data while maintaining the semantic validity of the label. By training the network on the additional deformed data, the hope is that the network becomes invariant to these deformations and generalizes better to unseen data. Semantics-preserving deformations have also been proposed for the audio domain, and have been shown to increase model accuracy for music classification tasks. However, in the case of environmental sound classification the application of data augmentation has been relatively limited, which used random combinations of time shifting, pitch shifting and time stretching for data augmentation) reporting that "simple augmentation techniques proved to be unsatisfactory for the UrbanSound8K dataset given the considerable increase in training time they generated and negligible impact on model accuracy".

Main Objective of Our Project

- The main aim our project to apply the data augmentation techniques by using signal processing.
- By using four types of method to augment the data by using Matlab.
- Finally, the data are augmented with the help of function file of the system.

1.1 Data Augmentation

Data augmentation in data analysis are techniques used to increase the amount of data by adding slightly modified copies of already existing data or newly created synthetic data from existing data. It acts as a regularizer and helps reduce overfitting when training a machine learning model. It is closely related to oversampling in data analysis. Data augmentation is a common strategy adopted to increase the quantity of training data. It is a key ingredient of the state of the art systems for image recognition and speech recognition. With the widespread adoption of neural networks in speech recognition systems which require a large speech database for training such a deep architecture, DA is very useful for small data sets. Indeed, it is possible to augment speech databases and to use the augmented database to achieve improved accuracy. Data Augmentation (DA) Technique is a process that enables us to artificially increase training data size by generating different versions of real datasets without actually collecting the data. The data needs to be changed to preserve the class categories for better performance in the classification task. The task of identifying what an audio represents is called audio classification. An audio classification model is trained to recognize various audio events. For example, you may train a model to recognize events representing three different events: clapping, finger snapping, and typing. It helps us to increase the size of the dataset and introduce variability in the dataset, without actually collecting new data. The neural network treats these images as distinct images anyway. Also Data Augmentation helps reduce over-fitting. Data augmentation helps to generate synthetic data from existing data set such that generalisation capability of model can be improved. We present Spec Augment, a simple data augmentation method for speech recognition. Spec Augment is applied directly to the feature inputs of a neural network (i.e., filter bank coefficients). The augmentation policy consists of warping the features, masking blocks of frequency channels, and masking blocks of time steps.

1.2 Signal Processing

Signal processing is an electrical engineering subfield that focuses on analyzing, modifying, and synthesizing signals such as sound, images, and scientific measurements. Signal processing techniques can be used to improve transmission, storage efficiency and subjective quality and to also emphasize or detect components of interest in a measured signal. Audio signal processing is a subfield of signal processing that is concerned with the electronic manipulation of audio signals. Audio signals are electronic representations of sound waves and longitudinal waves which travel through air, consisting of compressions and rarefactions. The energy contained in audio signals is typically measured in decibels. As audio signals may be represented in either digital or analog format, processing may occur in either domain. Analog processors operate directly on the electrical signal, while digital processors operate mathematically on its digital representation. Time-Frequency Signal Analysis and Processing (TFSAP) is a collection of theory, techniques and algorithms used for the analysis and processing of non-stationary signals, as found in a wide range of applications including telecommunications, radar, and biomedical engineering. Signal processor to interface the aircraft system bus and connected equipments through customer software under a real-time (RT) OS. The recorder and removable memory can be connected externally to the processor unit.

1.3 Filter of Signal

An audio filter is a frequency dependent circuit, working in the audio frequency range, 0 Hz to 20 kHz. Audio filters can amplify (boost), pass or attenuate (cut) some frequency ranges. Many types of filters exist for different audio applications including hi-fi stereo systems, musical synthesizers, effects units, sound reinforcement systems, instrument amplifiers and virtual reality systems.

Low-pass

Low-pass filters pass through frequencies below their cutoff frequencies, and progressively attenuates frequencies above the cutoff frequency. Low-pass filters are used in audio crossovers to remove high-frequency content from signals being sent to a low-frequency subwoofer system.

High-pass

A high-pass filter does the opposite, passing high frequencies above the cutoff frequency, and progressively attenuating frequencies below the cutoff frequency. A high-pass filter can be used in an audio crossover to remove low-frequency content from a signal being sent to a tweeter.

Band-pass

A band-pass filter passes frequencies between its two cutoff frequencies, while attenuating those outside the range. A band-reject filter attenuates frequencies between its two cutoff frequencies, while passing those outside the 'reject' range.

All-pass

An all-pass filter passes all frequencies, but affects the phase of any given sinusoidal component according to its frequency.

1.4 Method used in proposed method

Here, we can use the four type of data augmentation method of the system.

- Random Augmentation
- Random Independent Augmentation
- Sequential Augmentation
- Sequential Independent Augmentation

II. LITERATURE REVIEW

[1] AUTHOR NAME: Ghannay, S

TITLE: Acoustic Word Embeddings for ASR Error Detection.

DESCRIPTION: This paper focuses on error detection in Automatic Speech Recognition (ASR) outputs. A neural network architecture is proposed, which is well suited to handle continuous word representations, like word embeddings. In a previous study, the authors explored the use of linguistic word embeddings, and more particularly their combination. In this new study, the use of acoustic word embeddings is explored. Acoustic word embeddings offer the opportunity of an a priori acoustic representation of words that can be compared, in terms of similarity, to an embedded representation of the audio signal. First, we propose an approach to evaluate the intrinsic performances of acoustic word embeddings in comparison to orthographic representations in order to capture discriminative phonetic information. Since French language is targeted in experiments, a particular focus is made on homophone words. Then, the use of acoustic word embeddings is evaluated for ASR error detection. The proposed approach gets a classification error rate of 7.94% while the previous state-of-the-art Rebased approach gets a

CER of 8.56% on the outputs of the ASR system which won the ETAPE evaluation campaign on speech recognition of French broadcast news.

[2] AUTHOR NAME: Wei, J

TITLE: Easy data augmentation techniques for boosting performance on text classification tasks

DESCRIPTION: We present EDA: easy data augmentation techniques for boosting performance on text classification tasks. EDA consists of four simple but powerful operations: synonym replacement, random insertion, random swap, and random deletion. On five text classification tasks, we show that EDA improves performance for both Convolutional and recurrent neural networks. EDA demonstrates particularly strong results for smaller datasets; on average, across five datasets, training with EDA while using only 50% of the available training set achieved the same accuracy as normal training with all available data. We also performed extensive ablation studies and suggest parameters for practical use.

[3] AUTHOR NAME: Elnaz Lashgar

TITLE: Data Augmentation for Deep-Learning-Based Electroencephalography

DESCRIPTION: Data augmentation (DA) has recently been demonstrated to achieve considerable performance gains for deep learning (DL) increased accuracy and stability and reduced overfitting. Some electroencephalography (EEG) tasks suffer from low samples-to-features ratio, severely reducing DL effectiveness. DA with DL thus holds transformative promise for EEG processing, possibly like DL revolutionized computer vision, etc. We grouped DA techniques (noise addition, generative adversarial networks, sliding windows, sampling, Fourier transform, recombination of segmentation, and others) and EEG tasks (into seizure detection, sleep stages, motor imagery, mental workload, emotion recognition, motor tasks, and visual tasks). DA efficacy across techniques varied considerably. Noise addition and sliding windows provided the highest accuracy boost; mental workload most benefitted from DA. Sliding window, noise addition, and sampling methods most common for seizure detection, mental workload, and sleep stages, respectively

[4] AUTHOR NAME: Asuka Sakai

TITLE: Data augmentation methods for machine-learning-based classification of bio-signals

DESCRIPTION: Data augmentation methods for bio-signal classification are proposed. These methods improve recognition performance of human mental states showing intrinsic motivation from brain wave. Conventionally, data augmentation is used to image recognition research. Scaling, rotation, and distortion are applied to the original images to increase examples for machine learning. However, these augmentation methods are not effective for use with biological signals, as they involve spatial manipulation designed to represent the fluctuations of natural images. In the present study, we proposed four novel methods for data augmentation of biological signals. These methods are designed to represent variations inherent to bio-signals, especially for event-related signals. Electroencephalogram (EEG) data from participants engaged in an intrinsic motivation task were utilized to evaluate the feasibility of the proposed data augmentation methods. Our findings demonstrated that the proposed methods are particularly effective for improving prediction accuracy in small datasets.

[5] AUTHOR NAME: Fatemeh Fahimi

TITLE: Generative Adversarial Networks-Based Data Augmentation for Brain–Computer Interface

DESCRIPTION: The performance of a classifier in a brain-computer interface (BCI) system is highly dependent on the quality and quantity of training data. Typically, the training data are collected in a laboratory where the users perform tasks in a controlled environment. However, users' attention may be diverted in real-life BCI applications and this may decrease the performance of the classifier. To improve the robustness of the classifier, additional data can be acquired in such conditions, but it is not practical to record electroencephalogram (EEG) data over several long calibration sessions. A potentially time- and cost-efficient solution is artificial data generation. Hence, in this study, we proposed a framework based on the deep convolutional generative adversarial networks (DCGANs) for generating artificial EEG to augment the training set in order to improve the performance of a BCI classifier. To make a comparative investigation, we designed a motor task experiment with diverted and focused attention conditions. We used an end-to-end deep convolutional neural network for classification between movement intention and rest using the data from 14 subjects. The results from the leave-one subject-out (LOO) classification yielded baseline accuracies of 73.04% for diverted attention and 80.09% for focused attention without data augmentation. Using the proposed DCGANs-based framework for augmentation, the results yielded a significant improvement of 7.32% for diverted attention (p <; 0.01) and 5.45% for focused attention (p <; 0.01). In addition, we implemented the method on the data set IVa from BCI competition III to distinguish different motor imagery tasks. The proposed method increased the accuracy by 3.57% (p <; 0.02). This study shows that using GANs for EEG augmentation can significantly improve BCI performance, especially in real-life applications, whereby users' attention may be diverted.

III. PROBLEM DEFINITION

- In our Existing method, to generate the data to apply one augmentation techniques to show the result.
- One techniques are used to show the output in the form of signal format. Some drawback is come in the data.
- To Overcome this drawback, we propose the new method to show the result.

Drawbacks

- Less accuracy
- Inefficiency
- Low Performance

IV. DEVELOPMENT PROCESS

Input Design

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

Objectives

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus, the objective of input design is to create an input layout that is easy to follow.

4. For automatic detection of diabetic retinopathy in retinal images by using Machine Learning.

Output Design

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2. Select methods for presenting information.

3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

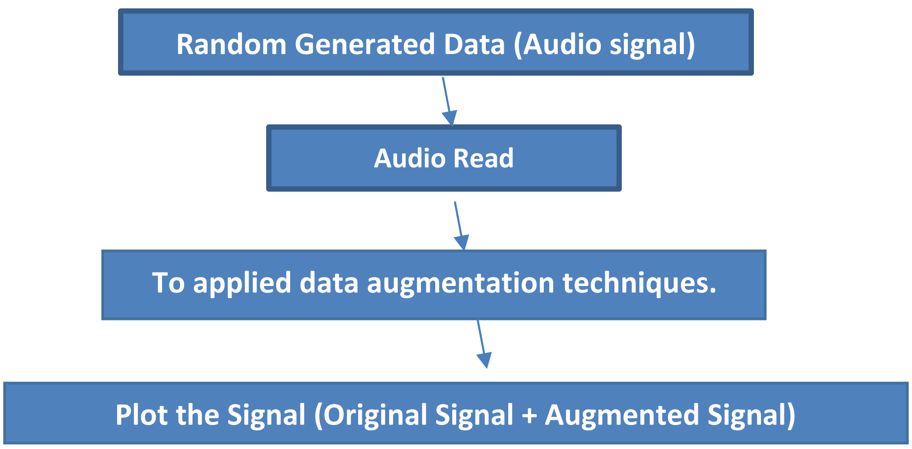
- ❖ Convey information about past activities, current status or projections of the
- ❖ Future.
- ❖ Signal important events, opportunities, problems, or warnings.
- ❖ Trigger an action.
- ❖ Confirm an action.

V. PROPOSED METHOD

- In our proposed method, to perform data augmentation method by using Signal Processing techniques with the help of Matlab Software.
- Here, they different types of data augmentation take place. They are Random augmentation, Sequential augmentation, Random Independent augmentation, Sequential Independent augmentation steps are used to show the results.
- So, here the data can be generated randomly or audio input can be taken to apply these different types of Augmentation techniques to show the Results.
- Finally, the output can be shown in the Signal of both Original signals as well as Augmented Signals of the system.

Advantages

- Better Performance.
- Result shows effectively high.
- Performance is high when compare to the other technique.
- Better Accuracy.



Modules Used

- Input Data
- Pre Processing

- Data Augmented

Input Data: An input data are taken in the audio format of the system. That data are processed in the matlab by using the function file to plot the signals of the system.

Pre-Processing: A Pre-Processing steps are used to reduce the noise in the signal to process the data

Data Augmented: Here, they are 4types of augmented data are used to augment the original by using the function file to process the output. In the four types of data augmentation, each and every steps to change the parameters of all the method of data augmentation. Finally to plot both the original and augmented signals of the system.

VI. APPLICATIONS

Processing methods and application areas include storage, data compression, music information retrieval, speech processing, localization, acoustic detection, transmission, noise cancellation, acoustic fingerprinting, sound recognition, synthesis, and enhancement (e.g. equalization, filtering, level compression, echo and reverb removal or addition, etc.).

1) Audio broadcasting

Audio signal processing is used when broadcasting audio signals in order to enhance their fidelity or optimize for bandwidth or latency. In this domain, the most important audio processing takes place just before the transmitter. The audio processor here must prevent or minimize over modulation, compensate for non-linear transmitters (a potential issue with medium wave and shortwave broadcasting), and adjust overall loudness to desired level.

2) Active noise control

Active noise control is a technique designed to reduce unwanted sound. By creating a signal that is identical to the unwanted noise but with the opposite polarity, the two signals cancel out due to destructive interference.

3) Audio synthesis

Audio synthesis is the electronic generation of audio signals. A musical instrument that accomplishes this is called a synthesizer. Synthesizers can either imitate sounds or generate new ones. Audio synthesis is also used to generate human speech using speech synthesis.

4) Audio effects

Audio effects alter the sound of a musical instrument or other audio source. Common effects include distortion, often used with electric guitar in electric blues and rock music; dynamic effects such as volume pedals and compressors, which affect loudness; filters such as wah-wah pedals and graphic equalizers, which modify frequency ranges; modulation effects, such as chorus, flangers and phasers; pitch effects such as pitch shifters; and time effects, such as reverb and delay, which create echoing sounds and emulate the sound of different spaces. Musicians, audio engineers and record producers use effects units during live performances or in the studio, typically with electric guitar, bass guitar, electronic keyboard or electric piano. While effects are most frequently used with electric or electronic instruments, they can be used with any audio source, such as acoustic instruments, drums, and vocals.

5) Computer audition

Computer audition (CA) or machine listening is general field of study of algorithms and systems for audio understanding by machine.[14][15] Since the notion of what it means for a machine to "hear" is very broad and somewhat vague, computer audition attempts to bring together several disciplines that originally dealt with specific problems or had a concrete application in mind. Inspired by models of human audition, CA deals with questions of representation, transduction, grouping, use of musical knowledge and general sound semantics for the purpose of performing intelligent operations on audio and music signals by the computer. Technically this requires a combination of methods from the fields of signal processing, auditory modelling, music perception and cognition, pattern recognition, and machine learning, as well as more traditional methods of artificial intelligence for musical knowledge representation.

VI. CONCLUSION

In this paper, we proposed a method for data augmentation in order to perform the data augmentation techniques has been Successfully implemented. Data augmentation is useful to improve performance and outcomes of machine learning models by forming new and different examples to train datasets. Data augmentation techniques enable machine learning models to be more robust by creating variations that the model may see in the real world.

REFERENCES

[1] Shorten, C. and T.M. Khoshgoftaar, A survey on image data augmentation for deep learning. Journal of Big Data, 2019. 6(1): p. 60.

[2] Kukačka, J., V. Golkov, and D. Cremers, Regularization for deep learning: A taxonomy. arXiv preprint arXiv:1710.10686, 2017.

[3] Deepa, V., Investigating the performance improvement by sampling techniques in EEG data. International Journal on Computer Science and Engineering, 2010. 2(6).

[4] Ragni, A., Knill, K.M., Rath, S.P., Gales, M.J.F.: Data augmentation for low resource languages. In: INTERSPEECH 2014, 15th Annual Conference of the International Speech Communication Association, Singapore, September 14–18, 2014, pp. 810–814 (2014)

[5] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25:1106–14.

[6] H.T. Duong, T.A. Nguyen-Thi, **A review: pre-processing techniques and data augmentation for sentiment analysis,** Computer. Soc. Netw., 8 (2021), p. 1, 10.1186/s40649-020-00080-x.