Speech emotion recognition system using RNN deep learning model

Swati Tiwari
Assistant Professor, Kalinga University ,

Shilpi Choubey
Assistant Professor , Kalinga University

Abstract

Advanced technology is leading towards automation of everything like taking commands from the user to providing different feed-forward to the machines so that every machine can be operated with the voice command f the user. The technology is developed to catch the voice command and then convert it with an analog or digital form to make it understandable by the algorithms and then they fulfill the wish of the user. However, the RNN deep learning model is one of the most popular and important inventions of technology which cannot only recognize the voice command but also it can determine the emotion of the person to whom the voice belongs. The current study is going to be developed to understand the inner engineering of the RNN Deep learning model and its style of working with bet results.

Keywords: *Deep learning model, speech recognition, Algorithm, Speech emotion determination*

DOI Number: 10.14704/nq.2022.20.13.NQ88319 Neuro Quantology 2022; 20(13):2564-2568

Introduction

Language is the primary way of transferring human emotion to another person. Human interaction is very effective by using the same language between two people talking to each other. The language and the inner development of the language help the other person to understand the exact meaning of the speech provided by the person. Similarly in the case of machines that can understand the human language or the human voice by recording the speech of humans to their system. In this case, the devices are used to understand the details of the speech by converting the speech signal into an analog or digital waveform. However, for the past few years, the major topic of the researchers is about emotion recognition from speech signals. The current study is also going to find out the speech emotion recognition system using the RNN deep learning model.

Aims

The study aims to find out the methods of speech emotion recognition by the lichen learning algorithm with the help of the RNN deep learning model. The existing challenges regarding the deep learning middle to recognize speech emotion by different algorithms are also going to be adequately discussed.

Objectives:

1. To find out the internal structure of the RNN deep learning model.

2. To find out the effectiveness of Speech emotion recognition using the RNN deep learning model

3. To determine the current Challenges faced by the RNN deep learning model

Overview of the RNN deep learning model

A deep learning model of the Recurrent Neural Network is very effective for modeling the sequential data before the set of the model of attention. Every set of data has a particular parameter. The specific parameters for every elementary have the element of sequence which has the deep feed-forward model (Anvarjon and Kwon 2020). A recurrent neural Network has the ability of similar weight for each element related to the sequence. The decrease of the parameter allows the model to make the sequence of varying length in general. RNN has the outcome to become structured data rather than sequential data, like geographical data or graphical data due

to the various design of the data captured by the model.

Advantages	Disadvantages
The ability to self-learning	The Blockchain nature and uncertain prediction rates
Ability to work with insufficient data	Long Training processes and limited data efficiency
The ability to parallel processing	Very expensive economically and computationally

Table1: various advantages and disadvantages of recurrent neural Networks

(Source: Yoon et al.2018, p.77)

The RNN deep learning model is comparatively old like other machine learning or deep learning method available in the market and has been implemented by different companies in terms of completing different jobs. In the 1980s the first deep learning model was developed and still, its effective news is making the developers launch a more updated version for the old ones comparatively (Huang et al.2019). The previous models before the RNN neural network model were not able to handle sequential data, considered only the current inputs and they were also not able to memorize the previous units. Thus the RNN model is developed by attaching these entire facilities o that the implementation of RNN would make the users work without facing the mentioned challenges. After the development of the RNN model, the current time is also using the RNN model very effectively, so that the entire job of speech emotion recognition is becoming extremely easy.

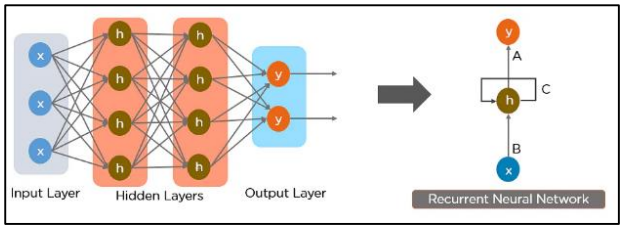

Figure 1: working method of the recurrent neural Network

(Source: Anvarjon and Kwon 2020, p.67)

Various applications of the RNN model

As earlier mentioned, the best form and working method of the RNN deep learning method, hence it gets used by several types of sections in different countries. The usage of the RNN model is to…

- It is very effective to predict various problems related to different aspects of any type of data or in any type of form.
- It can translate the form of the data to any type of form whatever is needed by the user (Parry et al.2019). Like in speech emotion recognition is continually active to translate audio data into analog or digital form.
- Speech recognition is one of the most important specialties of the RRN deep learning model.
- It has many more capabilities to perform with perfection like generating image descriptions, text summarization, Face detection, OCR application as image recognition, and other application also.

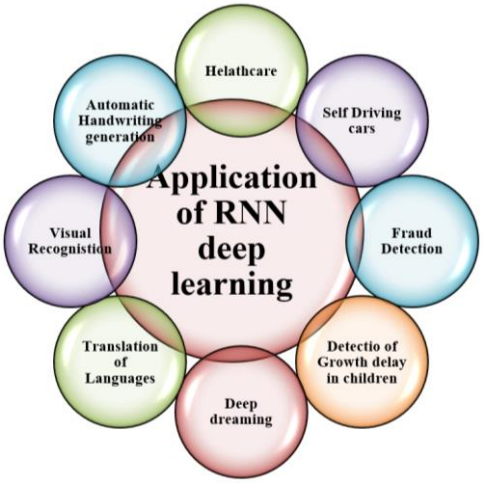

Figure 2: Different use of the recurrent neural Network

(Source: Parry et al.2019,p.198)

Speech emotion recognition using the RNN deep learning model

The recurrent neural network is frequently used in speech emotion recognition or the natural language processing system, even though it can perform any type of speech recognition-related prediction. Recurrent neural Networks are used

to recognize the sequential characteristics of the speech or other voice data, then it predicts the possible sequence that gets repeated with the same pattern or with the same gap between the previous or the next sequence, thus the prediction becomes successful in most of the times and the RNN model is taking over the entire market of speech emotion recognition technology (Khalil et al.2019). The development of the determination of the speech emotion behind the speech has been recorded is done with the help of different algorithms, where the majority of the different algorithms are entitled to perform different jobs. The majority of the jobs are to first convert the voice nodes to the analog or digital format then find out the sequence among the voice nodes so that among the different sequences the majority of the most common sequences can get found by the algorithm then the computation process to the sequences helps to predict the exact next step of the voice node, thus the emotional view related to the sequences help the algorithm to find out the node of the voice (Yoon et al.2018). The commutation method of the process is the key factor in speech emotion recognition.

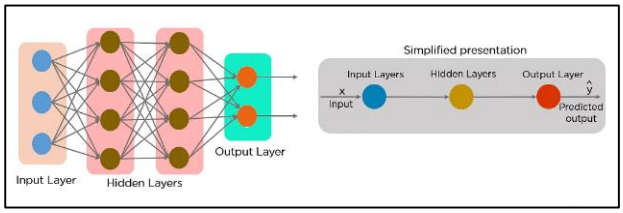

Figure 3: Speech emotion recognition method of the recurrent neural Network

(Source: Khalil et al.2019, p.93)

Challenges faced by the RNN deep learning model

Despite having a huge impact on different important works through the addition of different algorithms by the recurrent Neural Network, it has some problems too. The developer of the RNN is still trying to recover the system from the current issue related to the working method of the RNN. The major problem of the RNN is the vanishing Gradient. The Gradient is used to carry information that is used in the RNN, in case the RNN is too small then the parameter upgradation becomes insignificant (Issa et al.2020). Thus, the long data sequence becomes very difficult.

Another major aspect is the entire RNN method is an extremely complex technology and other several numbers of operations have been done by the algorithm, hence entire processing to get the result from the model is time costly and thus the entire process becomes terribly slow. Other minor problems are like the exploding of the server or too long time waiting while working with long sequences etc. the training of the RNN is an exceedingly difficult process.

The RNN model works with different sequences and thus previous sequence and the next sequence are not to be mandatory same, even though the sequence could be of different types, hence, the majority of the training of the model to find out the determination of the sequences is not an easy task to perform, the major problem occurs when the sequences are continuously developing with getting more complex day by day.

Even the RNN model is developed with the majority of the sequences and parameters of the data, hence when it comes the time of parallelizing the training in the RNN or a simple stacking up of the training it becomes very difficult to do due to the connectivity of the nodes (Atmaja and Akagi 2019). The nodes in the RNN are connected because the computation on the present node is done based on the previous node thus parallelizing is difficult for RNN.

Data analysis and findings

The current study has been developed with most of the data from various references, books, and published materials with the method of the secondary data collection method. The entire data has been sorted out affect been collected altogether, the sorting method of all the data was done what the inclusion ad exclusion criteria methods thus the data is eligible to get selected to be used in the study. Most of the data, which is outdated, not readable or understandable, or published in another language than English gets rejected. Hence the correct and exact relevant data is used to implement the study. After gathering the data from different sources, the entire data is analyzed by the qualitative data analysis method. The qualitative data analysis method is amazingly effective to find out the best quality results from any type of data or the

expected outcome (Ho et al.2020). The majority of the data after analysis shows the various possibility of the RNN model being used in various models.

	MLP	RNN	CNN
Data	Tabular Data	Sequence Data	Image Data
Recurrent connections	NO	Yes	No
Parameter sharing	NO	Yes	Yes
Spatial relationship	NO	No	Yes
Vanishing and Exploding Gradient	Yes	Yes	Yes

Table 2: comparison with other deep learning models of the recurrent neural Network

(Source: Yoon et al.2018, p.87)

Findings:

After gathering the data related to speech emotion recognition the RNN deep learning model and the qualitative data analysis method helped to find out the exact details about certain points related to the various applications of the RNN deep learning model and the current challenges of the model along with its different ability to recognize the speech emotion form various audio pieces (Latif et al.2018). Even the study has also developed different types of graphs or other diagrams showing the various algorithms being used in the study showing the exact details from the gathered data. The various algorithms and their necessary upgrades to remove the current challenges are also mentioned with visual representation. Hence the study is completed with all types of necessary relevant data.

Type	variants	Network structure	Applications
CNN	Le Net Alex Net Egg Net	Input layer Convolution layer Pooling Layer Full Connected layer	Image processing, speech signal, Natural Language processing
RNN	LSTM	Input layer Hidden Layer Output Layer	Time series analysis Emotion analysis Natural Language Processing
GAN	DUGAN	Discrimination Layer Generation Model	Image generation Video generation

Table 3: different learning networks and their usage with network structure

(Source: Latif et al.2018, p.77)

Conclusion

As conclusion, it is established the conclusion that the study is the RNN deep-learning model is extremely beneficial to understand speech emotion which is recorded by the machine learning method by the machine. As earlier mentioned most machines are used to convert speech in the form of analog or digital format, hence speech learning emotion detection is amazingly effective using the RNN Model. The entire gas was developed with the overview of machine learning or the current challenges of the RNN machine learning with the speech emotion detection system. Even the statistical data via the diagrams or other visual representations are also developed with very helpful to understand the details of the machine learning method and its relationship with the speech emotion recognition method.

Future scopes

The current study is developed about all the entails related to the RNN deep learning method to complete the job of determining the emotional detection from the speech signal of a human being. The study has developed all the necessary details to make understand the related details of the speech recognition method. Hence the detailed study might be greatly beneficial for most of the researchers working on the same topic (Zhao et al.2019). Even the speech emotion recognition from the limited speech of the system is the current topic for the researchers to understand about the special ability of the algorithm, that circumstance the current study with all the necessary data might help the researchers with a huge amount of data. Thus, in near future, the study might be studied very effectively by most of the researchers working on the same topic.

References

Anvarjon, T. and Kwon, S., 2020. Deep-net: A lightweight CNN-based speech emotion recognition system using deep frequency features. *Sensors*, *20*(18), p.5212.

Atmaja, B.T. and Akagi, M., 2019, July. Speech emotion recognition based on speech segment using LSTM with attention model. In *2019 IEEE International Conference on Signals and Systems (ICSigSys)* (pp. 40-44). IEEE.

Ho, N.H., Yang, H.J., Kim, S.H. and Lee, G., 2020. The multimodal approach of speech emotion recognition using multi-level multi-head fusion attention-based recurrent neural network. *IEEE Access*, *8*, pp.61672-61686.

Huang, K.Y., Wu, C.H., Hong, Q.B., Su, M.H. and Chen, Y.H., 2019, May. Speech emotion recognition using deep neural networks considering verbal and nonverbal speech sounds. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (pp. 5866-5870). IEEE.

Issa, D., Demirci, M.F. and Yazici, A., 2020. Speech emotion recognition with deep convolutional neural networks. *Biomedical Signal Processing and Control*, *59*, p.101894.

Khalil, R.A., Jones, E., Babar, M.I., Jan, T., Zafar, M.H. and Alhussain, T., 2019. Speech emotion recognition using deep learning techniques: A review. *IEEE Access*, *7*, pp.117327-117345.

Latif, S., Rana, R. and Qadir, J., 2018. Adversarial machine learning and speech emotion recognition: Utilizing generative adversarial networks for robustness. *arXiv preprint arXiv:1811.11402*.

Parry, J., Palaz, D., Clarke, G., Lecomte, P., Mead, R., Berger, M. and Hofer, G., 2019, September. Analysis of Deep Learning Architectures for Cross-Corpus Speech Emotion Recognition. In *Interspeech* (pp. 1656-1660).

Yoon, S., Byun, S. and Jung, K., 2018, December. Multimodal speech emotion recognition using audio and text. In *2018 IEEE Spoken Language Technology Workshop (SLT)* (pp. 112-118). IEEE.

Zhao, J., Mao, X., and Chen, L., 2019. Speech emotion recognition using deep 1D & 2D CNN LSTM networks. *Biomedical signal processing and control*, *47*, pp.312-323.