Nitte Meenakshi Institute of Technology,

Department of Computer Science and
Engineering

18CSE751 Introduction to Machine Learning

Learning Activity Proposal

Baby Cry Sound Classification

Gaurav Raj Shah, Abhishek Kushwaha, Khush Dassani

Abstract

Neonatal infants communicate with us through cries. The infant cry signals have distinct patterns depending on the purpose of the cries. Preprocessing, feature extraction, and feature selection need expert attention and take much effort in audio signals in recent days. In deep learning techniques, it automatically extracts and selects the most important features. For this, it requires an enormous amount of data for effective classification. This work mainly discriminates the neonatal cries into pain, hunger, and sleepiness. The neonatal cry auditory signals are transformed into a spectrogram image by utilizing the short-time Fourier transform (STFT) technique. The deep convolutional neural network (DCNN) technique takes the spectrogram images for input. The features are obtained from the convolutional neural network and are passed to are passed toother machine learning technique classifies neonatal cries.

Introduction

Babies convey their needs through cries. Experienced baby care persons and parents can understand the reason for the baby's cries. Some young working parents struggled to interpret the baby's cries. The baby's cries imply their emotions, physical needs, and pathological problems from internal or external stimulation. Humans can listen to the audio signal in the frequency range from 50 to 15,000 Hz for music, 20 to 20,000 Hz for sounds, and 100 to 4,500 Hz for speech. Within this range, humans can discriminate the audio. Babies do not have control over their vocal tract so that it is more sensitive than adults. Baby cries contain information, and their crying pattern varies based on their physical and emotional state. The researchers found that there is a pattern for each kind of cry. Infant cry classification can be considered pattern recognition or speech recognition. An abnormal cry of the infant can indicate a genetic or pathological problem. Childcare experts can differentiate it. The baby cry-based recognition approach will help us know the infant's feelings from their cries. Techniques such as signal preprocessing, feature extraction, feature selection, and classification are the steps involved in baby cry classification.

Data Set

We will use the dataset from the Donate-a-cry campaign. The dataset is obtained from the following link –

https://github.com/gveres/donateacry-corpus

The dataset is present wav file and will contain 3000 entries upon conversion.

Machine Learning Methods

- **KNN:** K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most like the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. KNN algorithm at the training

phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much like the new data.

- **SVM:** SVM executes classification by differentiating the data points with a larger margin using hyperplane as a decision boundary. SVM classifier includes hyperplane, margin hyperplane, kernels, and soft margin. The hyperplane is the line that differentiates the discrete data points. Margin is the distance between data samples and the hyperplane. The margin hyperplane divides the dissimilar data with maximum distance from one another. The data samples which are near the hyperplane are named as support vectors.

- **CNN:** Convolutional neural networks are distinguished from other neural networks by their superior performance with image, speech, or audio signal inputs. They have three main types of layers, which are: Convolutional layer, Pooling layer and fully connected (FC) layer. The convolutional layer is the first layer of a convolutional network. While convolutional layers can be followed by additional convolutional layers or pooling layers, the fully connected layer is the final layer. With each layer, the CNN increases in its complexity, identifying greater portions of the data. Earlier layers focus on simple features.

Assessment:

Accuracy scores of all the three models will be used for assessment.

Presentation and Visualization

Accuracy scores of the three models will be displayed. Moreover, an audio file can be used to describe the use of the models.

Roles

Abhishek Kushwaha – KNN implementation.

Khush Dassani – SVM implementation.

Gaurav Raj Shah – CNN implementation.

Schedule

The schedule is a table of dates and tasks that you plan to complete.

Date	Tasks to be Completed
17/01/21	Tasks completed by chosen date
18/01/22	Tasks to be completed by the final report/ presentation date

Bibliography

- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8222524/
- https://www.ibm.com/cloud/learn/convolutional-neural-networks/
- https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47
- https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761