Sound Classification using machine learning

Advance Data MiningAnurag Abhay Singh  
Msc. Data Analytics

x18104053

National College of Ireland

**Abstract：**

Deep learning is becoming more and more popular as a way to avoid laborious pre-processing of data. Various pre-processing can not only improve the prediction of prediction but also constitute a fascinating topic. The sound classification problem is such a situation. Sound classification is an evolving research area with a variety of applications for large-scale, content-based multimedia indexing and retrieval. In this paper, we concentrate on the cross-industry standard process for data mining (CRISP-DM), CRISP-DM is an open, freely usable data mining process. Business understanding, data understanding, data preparation, modeling, model evaluation, and model release are six standard stages of CRISP-DM. CRISP-DM has no specific tool limitations and no specific domain limitations. Compared to other existing data mining methodologies, the CRISP-DM methodology is more advantageous. Besides, random forest classifier and k-nearest neighbor (K-NN) are applied to model. By taking various sound samples through the data processing stage, the spectrograms are obtained, and the sound is classified according to the frequencies in the spectrograms.

Key words: CRISP-DM, sound classification, random forest, K-NN.

1. **Introduction**

The sound consists of random noise and regular audio signals. A regular signal is an analog signal that is connected to a periodic change. The audio frequency that human hearing can feel is 20hz~20khz, and the sound signal below 20hz or higher than 20khz belongs to multimedia information audio. Sound waves are a timely extension of the signal, and at each point in time, the current signal depends on the past. In the time domain, the continuous amplitude of the signal age is composed. In the frequency domain, the signals are represented by different frequencies. The conversion from time domain to frequency domain can use Fourier transform, according to the time and whether the signal itself is continuous or discrete, there are different types of Fourier transform, in the "real world", using discrete Fourier transform, because when dealing with digitized signals, it actually means virtual [1]. Accurate classification of sound signals has been applied in the fields of audio monitoring, hearing aids, hospitals and so on. Urban sound classification faces many difficulties. For example, the sound source generation mechanism is different, there is a possibility that there is noise in the source (such as the sound of air conditioner or engine), the urban auditory scene can represent almost unlimited configuration, lack of voice, etc [2]. To the advanced structure and so on. So classifying city sounds is a challenging task.

Based on the above, the research question of this report is:

How to accurately classify different sources of sound through deep mechanical learning?

By reading the relevant literature and combining the research content, the random forest and K-NN methods are used in this report to model. A random forest is a classifier that contains multiple decision trees, and the output category is determined by the mode of the category of the individual tree output. The advantage of random forests is that they can handle a large number of input variables. When determining the category, the importance of the parameters can be evaluated. When constructing the forest, the step can be used to estimate the error after the generalization even if the data is lost. K-NN is to calculate the distance between the input feature and the feature points in the training set, and then select the smallest k values ​​among the distances and predict the classification of the input instance according to the category of the data points corresponding to the k values. Distance metrics, classification decision rules, and k-values ​​are three basic features. K-NN does not require training, especially suitable for multi-classification problems.

The motivation of this paper is to process the data from different sources, and then standardized data especially for audio or image finally obtain the spectrogram to classify the sound. Spectral separation, spectral distribution, harmonicity, onset and offset, correlation amplitude and frequency variation, spatial separation, and time separation are all auditory features that play a key role in auditory grouping [3].

The contribution of this paper is to propose a new hybrid model to accurately classify sounds. The remaining of the paper is structured as follows: Section 2 the literature review about the sound classification methods and Section 3 is the background of the dataset; Section 4 is the model; Section 5 presents the results and Section 6 is the conclusion and future work.

1. **Literature review:**

Sound is generated by different frequency signal with related to time. Feature extraction showed important classes out of which 12 significant classes were chosen by Pearson correlation co-efficient. 7 optimized classes was designed as some of 12 classes failed to show any instances after classification technique. Higher accuracy was achieved in classifying sounds of 7 class model by ANN, SVM after PCA [4]. Spherical k-means was used for feature engineering on noise complaints dataset which was classified by Random forest (RF) Algorithm in [2]. The results evaluated by 10-fold cross validation showed greater accuracy in classification of stationary noise and non-stationary noise. The identification of background noise was done by classification model in [5]by extracting pitch, zero crossing rate (ZCR), Mel Cepstral co-efficient (MFCC), spectral flux (SF), spectral centroid (SC) and short time energy later tested by various classification algorithm like SVM, RF, DNN out which RF achieved higher accuracy in predicting background noise in less time followed by SVM and DNN. The dependency on feature extraction was contradicted in [6]where heart sounds classification was done from intensity map using CNN and Soft max regression (SMR). The intensity map was generated on 48x48 square matrix by using features of heart beats. The obtained results clearly indicated that CNN outperformed SMR by automating feature extractions process by adjusting its hyper parameters to accurately classify between abnormal and normal heart beats from heart sounds. Performance of CNN is revived and higher accuracy is achieved if it is trained at initial stage with chromatic spectrogram image generated from sound data. 47 class of insect sounds was classified up to 97% in less time [7]. Evaluating results of [6] a combinational method was proposed for classification of animal sounds in [8] using CNN for extraction of features then feature transformation by LDA and then classifying it by kernel SVM.. CNN extracted sound features of insect anuran and birds and trained dataset. The classification accuracy was in range of 70 % to 97% by SVM after LDA. A test was conducted in classifying lung sounds generated by sonar and stethoscope to fitness [9]. The extracted features of sound by Essentia library were later used to build classification model by boosted regression trees which served 85% accuracy in classifying patient’s class. On similar dataset vector auto regressive (VAR) was performed to build a spatial-temporal features model impacting sound which was later trained on SVM and Gaussian mixture model(GMM) for classification, out of which GMM yielded 85% accuracy in predicting label of class on basis of evaluating MAP and probability score [10]. Abnormalities in lung sound was classified using a Recurrent neural network (RNN), classifying 89% of class compared to GBM which showed only 53% accuracy [11].

Extracted significant classes of sounds of local area and global level were aggregated to form dictionary type model for classification. Local features were extracted using SVR and for global features Bayes law was used. Kernel density estimator (KDE) classified the sound from different source at higher accuracy. The results clearly show that dictionary of sound is directly proportional to classification accuracy [12].

Convolutional neural networks (CNN) have been successful in the processing of speech and music signals, especially in the classification of environmental and urban sound sources. By extending CNN and introducing environmental sound classification (ESC) problems, [13] have achieved better results than CNN and the largest collections and other Advanced method. A real-time hierarchical classification method that can distinguish environmental sound signals is developed by [14] which consists of three levels. Calculating the effective features according to needs at different levels of the hierarchy is the main feature of the method. Compared with the traditional part of the class methods, [14] has higher classification rate and computational efficiency. The severity and wheezing of asthma patients were classified using spectral integration (SI) characteristics. [15] using the overall ensemble (ENS), support vector machine (SVM) and k-nearest neighbor (KNN) methods to classify asthma severity levels in the group, the results show that the best positive predictive value (PPV) for mild, moderate and severe samples is 95 %(ENS), 88% (ENS) and 90% (SVM). Using the Lanczos kernel and deep neural network (DNN) classification to reduce the spectrogram image features (SIF) function, the results show that although MFCC, MPEG7 and Gabor get better continuity under noise, the available SIF technology provides results to some extent [16]. According to the polysomnography data, the zero-crossing rate and the formant of the sound signal classify the breathing and snoring segments, [17]first calculate the number of zero crossings, the logarithm of the signal energy and the first formant and then pass the Fischer linear discriminant method The three-bit feature is transformed into a one-dimensional space, and finally the Bayesian threshold is applied to the transformed features to classify the sound segments as snoring or breathing.

Summary of literature review:

Based on the literature review, it is easy to find that features such as ZCR, MFCC, SF, and SC are effective for sound classification in algorithms such as SVM, RF, ANN and DNN. Besides, The Bayesian threshold is applied to the transform feature, KDE, and GMM also has a high accuracy for sound classification.

# References:

|  |  |
| --- | --- |
| [1] | S. Keydana, "Audio classification with Keras: Looking closer at the non-deep learning parts," 6 February 2019. [Online]. Available: https://blogs.rstudio.com/tensorflow/posts/2019-02-07-audio-background/. |
| [2] | J. Salamon and J. P. Bello, "UNSUPERVISED FEATURE LEARNING FOR URBAN SOUND CLASSIFICATION," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, QLD, Australia, 2015. |
| [3] | S. Allegro, M. Büchler and S. Launer, "Automatic Sound Classification Inspired by Auditory Scene Analysis," 2001. [Online]. Available: academia.edu. |
| [4] | V. Bountourakis, L. Vrysis and G. Papanikolaou, "Machine Learning Algorithms for Environmental Sound Recognition: Towards Soundscape Semantics," 2015. |
| [5] | P. Dhakal, P. Damacharla, A. Y. Javaid and V. Devabhaktuni, "Detection and Identification of Background Sounds to Improvise Voice Interface in Critical Environments," in *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, Louisville, 2018. |
| [6] | J. X. Low and K. W. Choo, "Classification of Heart Sounds Using Softmax Regression and Convolutional Neural Network," in *Proceedings of the 2018 International Conference on Communication Engineering and Technology*, Singapore, 2018. |
| [7] | X. Dong, N. Yan and Y. Wei, "Insect Sound Recognition Based on Convolutional Neural Network," in *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, 2018. |
| [8] | K. Ko, S. Park and H. Ko, "Convolutional Feature Vectors and Support Vector Machine for Animal Sound Classification," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018. |
| [9] | G. Chambres, P. Hanna and M. Desainte-Catherine, "Automatic Detection of Patient with Respiratory Diseases Using Lung Sound Analysis," in *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*, La Rochelle, France, 2018. |
| [10] | I. Sen, M. Saraclar and Y. P. Kahya, "A Comparison of SVM and GMM-Based Classifier Configurations for Diagnostic Classification of Pulmonary Sounds," *IEEE Transactions on Biomedical Engineering ,* vol. 7, no. 62, pp. 1768 - 1776, 2015. |
| [11] | V. Mladenov, P. Koprinkova-Hristova, G. Palm, A. E. P. Villa, B. Appollini and N. Kasabov, "Artificial Neural Networks and Machine Learning – ICANN 2013,," in *23rd International Conference on Artificial Neural Networks Sofia, Bulgaria, September 10-13, 2013.* , Bulgaria, 2013. |
| [12] | J. Ye, T. Kobayashi and M. Murakawa, "Urban sound event classification based on local and global features aggregation,”," *Applied Acoustics,* vol. 117, pp. 246-256, 2017. |
| [13] | Y. Chen, Q. Guo, X. Liang, J. Wang and Y. Qian, "Environmental sound classification with dilated convolutions," *Applied Acoustics,* vol. 148, pp. 123-132, 2019. |
| [14] | F. Saki and N. Kehtarnavaz, "Real-time hierarchical classification of sound signals for hearing improvement devices," *Applied Acoustics,* vol. 132, pp. 26-32, 2018. |
| [15] | F. G. N. K. S. K. L. Palaniappan, "Characterization and classification of asthmatic wheeze sounds according to T severity level using spectral integrated features," *Computers in Biology and Medicine,* vol. 104, pp. 52-61, 2019. |
| [16] | I. Ozer, Z. Ozer and O. Findik, "Lanczos kernel based spectrogram image features for sound classification," *Procedia Computer Science ,* vol. 111, pp. 137-144, 2017. |
| [17] | A. Yadollahi and Z. Moussavi, "Automatic breath and snore sounds classification from tracheal and ambient sounds recordings," *Medical Engineering & Physics,* vol. 32, pp. 985-990, 2010. |
|  |  |