УДК 004.85

REVIEW AND COMPARISON OF MACHINE LEARNING METHODS FOR RECOGNIZING HYDROACOUSTIC SIGNALS

DOI 10.36994 / 2788-5518-2022-01-03-18

Andriy Verlan, *Dr.habil.,Prof. National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" Kyiv, Ukraine. verlander@gmail.com*
Artur Oleksii. *National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" ,Kyiv, Ukraine. arturoleksii@gmail.com*

Abstract: *Article presents machine learning algorithms which can be used as a classifier of hydroacoustic signals. Firstly, the preparation and methods of data presentation are considered, with the inquiring of the necessary characteristics. Next, we consider machine learning models which are used to classify hydroacoustic signals. When describing the algorithms, theoretical data and the accuracy of the classification of algorithms are given. The paper compares the areas of application of machine learning algorithms, its advantages and disadvantages. Based on the described data, recommendations on the choice of approaches to data preparation and the use of machine learning algorithms for signal identification are formulated*.
Key words*: identification of hydroacoustic signals, machine learning, signal processing, decision tree, support vector machine, neural network.*

ОГЛЯД ТА ПОРІВНЯННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ РОЗПІЗНАВАННІ ГІДРОАКУСТИЧНИХ СИГНАЛІВ

Верлань А.І.,*д.т.н., проф. Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна. verlander@gmail.com*
Олексій А.О. *Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна. arturoleksii@gmail.com*

Анотація: *У статті представлені алгоритми машинного навчання, які можна застосуванти яя класифікатор гідроакустичних сигналів. Спочатку, розглядався процес вилучення характеристик сигналу, серед яких були представлені методи: розріджене розкладання, генеративний процес створення сигналу та перетворення Фур'є. Розріджене розкладання не часто застосовується, оскільки, майже всі наявні алгоритми розрідженого розкладання мають значну обчислювальну складність, що серйозно впливає на практичне застосування цих алгоритмів і обмежує розріджене представлення за переповненим словником. Генеративний процес хоч і може згенерувати звуки з бажаними показниками, створення сигналу є досить затратним та не завжди відображає звуки в реальному середовищі. Ефективним та перевіреним способом є перетворення Фур'є, що широко застосовується для обробки сигналів в багатьох областях. Уніерсальність цього методу була показана у порівнянні обробки гідроакустичного сигналу та сигналу зображення, де цей підхід мав місце у обох прикладах. Були розглянуті наступні підходи машинного навчання: метод опорнх векторів, дерево рийняття рішень та нейронна мережа. Якщо є велика вибірка даних і потрібна висока точність класифікації, тоді варто обрати нейронну мережу. Значна кількість типів моделей та алгоритмів нейронних мереж дозволяє мати широкий спектр можливостей вирішення поставлених задач. Якщо вибірка мала, потрібна висока точність і швидкість, то добрим вибором буде метод опорних векторів. Метод опорних векторів був створений для задач бінарної класифікації. Застосування алгоритму для вирішення завдань багатокласової класифікації потребує додаткових модифікацій. При застосуванні методу опорних векторів, слід остерігатися великих обсягів даних, при яких можливе перенавчання. Якщо потрібно не витрачати багато зусиль на підготовку даних, потрібен простий та непараметричний алгоритм у цьому разі варто обрати дерево прийняття рішень. Дерево прийняття рішень застосовується для вирішення задач бінарної та мультикласової класифікації..*
Ключові слова*: ідентифікація гідроакустичних сигналів, машинне навчання, обробка сигналу, дерево прийняття рішень, метод опорних векторів, нейромережа*

Introduction

The study of acoustic signals has significant scientific and applied potential, the results of which can be applied in the recognition of human speech, identification of objects as sources of hydroacoustic radiation, and many other areas.

When identifying hydroacoustic signals, a number of problems arise, which include the receipt of damaged data, reverberation and big data. In the presence of big data, a significant amount of human effort is required to quickly determine acoustic features. It should be borne in mind that the input data may contain features that cannot be interpreted by human perception. Even with machines running, there can be problems interpreting corrupted data or having effective and accurate ways to identify signals. These problems are solved by methods of artificial intelligence, particularly machine learning.

Machine learning in acoustics is an area that is evolving rapidly and has many ways to solve these problems. The use of machine learning technology has made significant progress in the analysis of hydroacoustic signals. Let us consider the main approaches used in the extraction of useful data from signals and in the recognition of hydroacoustic signals.

Data preparation

Before classification, hydroacoustic signals must be pre-processed, since the quality of the classification will depend on its quality. Signal processing includes data preparation, as well as the further usage of algorithms to eliminate characteristics that may be useful in subsequent stages. Pre-processing includes noise reduction, estimation of the degree of randomness, removal of short-term local features, and preliminary filtering.

Modern algorithms use preprocessing to represent the signal in a different space from which the desired characteristics can be extracted. The signal spectrum reflects the distribution of signal energy over frequencies. To represent the spectrum of the signal, it is necessary to apply the Fast Fourier Transform. In spectral decomposition, the following features can be selected: spectral density measure [1], mel-frequency cepstral coefficients and gammaton-frequency cepstral coefficients. All these decompositions have a finer spectral resolution at lower frequencies, close to biological hearing and may be sufficient for the recognition of hydroacoustic signals [2]. In addition to the spectral-time domain, there are algorithms which extract characteristics from other representations of acoustic data, but they are used in more specific cases, like the DEMON algorithm, which is used to identify ships by analyzing the sonar signal of the ship's propellers [3].

The sparse decomposition provides the basis for optimally decomposing the waveform into a set of features from which the original signal can be roughly reconstructed. This can be used to optimize source recognition algorithms and, in particular, in the form of decomposition of non-negative matrices, provides the studied set of functions for signal recognition [4].

Another approach to the selection of acoustic characteristics for classification is to consider the processes by which signals with the desired characteristics are created. In some cases, the physical processes that create sound are well characterized and can be modeled using physical models [5].

Fully physical simulations allows to create datasets of the required size, which will simplify the training of classification algorithms. Such physical synthesis models enable classifiers to be trained that can go beyond recognizing broad-sounding classes and be able to recognize small physical features. Human perception can easily detect such traits, although the principle by which this occurs is unknown [6]. Detailed and flexible judgments can be made using a generative model that explicitly identifies relevant causal factors, which is, physical parameters. Such generative models can be used to identify objects and surfaces from images, movement of the vocal tract during speech, simple sounds from simulated scenes. However, physically synthesizing a signal is computationally complex and slow. It is not yet clear how to apply such approaches to real-world scenes.

Raw samples form a one-dimensional time series signal that is fundamentally different from two-dimensional images. Hydroacoustic signals can be transformed, for example, into a two-dimensional representation of frequency and time and so utilized for processing, but the time and frequency axes are not uniform, like the horizontal and vertical axes in the picture. Images are snapshots of a target, which can be raster or vector models, and are often analyzed as a whole or in specific areas with little ordering restrictions. However, acoustic signals should be studied in chronological order. Before submitting an image for classification, it can be processed.

The next step is the segmentation of the image. With segmentation, you can separate the unimportant aspects of the image. At the segmentation stage, the incoming image is divided into subgroups consisting of many pixels. There are the following types of segmentation: semantic, essential and panoptic. Various filters can also be applied to filter the image.

For hydroacoustic signals, it is possible to remove noise or to isolate the desired signal using the Fourier transform [7].

For hydroacoustic signals, it is possible to create a spectrogram. To create a spectrogram, the amplitude of each frequency is calculated over successive time frames using a Fourier transform. To create a visual representation of the spectrogram, these intensities can be represented using colors and frequency frames. In some cases, the preprocessing step may also extract only those aspects of the signal which are relevant to the task and discard other information. For example, the signal can be converted to a mel-frequency cepstrum to more accurately represent the way data is processed in the human auditory system. This is useful for processing that simulates human perception, such as speech processing.

You can remove the following types of features for an image:

1. Statistical features: features derived from a statistical distribution to increase speed and reduce data complexity. There are various statistical characteristics such as zoning, intersection, and distance.

2. Features of the global transformation and expansion of the series: these functions are immutable to global deformations such as translation and rotation. The continuous signal contains enough information for linear classification. Can contain the following data: Fourier transform, Gabor transform, wavelets

3. Geometric and topological features: These functions can reflect global and local properties and have high tolerances for distortion and style variation. These topological features can encode certain knowledge about the contour of an object or require certain knowledge about what type of components the object consists of.

Support vector machine

Support vector machine is a classification method that belongs to the group of limit methods and defines classes using the boundaries of ranges. Support vectors are objects of the set which belong to these limits. The classification is considered successful if the range between the boundaries is empty.

Support vector machine is a group of algorithms used for regression analysis and classification problems. It belongs to supervised learning type algorithms. A characteristic feature of the support vector machine is a constant decrease of the empirical error of the classifier and an increase in the range. Also, this method is called the maximum clearance classifier method. The essence of the method is to transfer input vectors to a range with a different dimension and search for the resolution of a hyperplane with a maximum interval in range. There is a construction of two parallel hyperplanes on opposite sides of the hyperplane, which is located between the studied classes. A visualization of the support vector machine operation can be seen in Fig. 1.

As the resolution of the hyperplane, the plane will be selected, which makes the maximum distance to the parallel hyperplanes. For the algorithm, the true assumption will be that the greater the difference or distance between the parallel planes, the less will be the average error of the classifier [8].

Support vector machine is not robust to noise in input data and when data is standardized. If the training sample contains noise emissions, this method is used irrationally. Among the advantages, given the correctness of the input data, the method is one of the best algorithms for identifying audio signals. This is the fastest way to find decisive functions. It solves quadratic programming problems in releasing areas that always have a single solution. The method reveals the dividing plane of maximum width, which further allows for a more accurate classification. Support vector machine was applied in [9] and showed an accuracy of 89.5%.

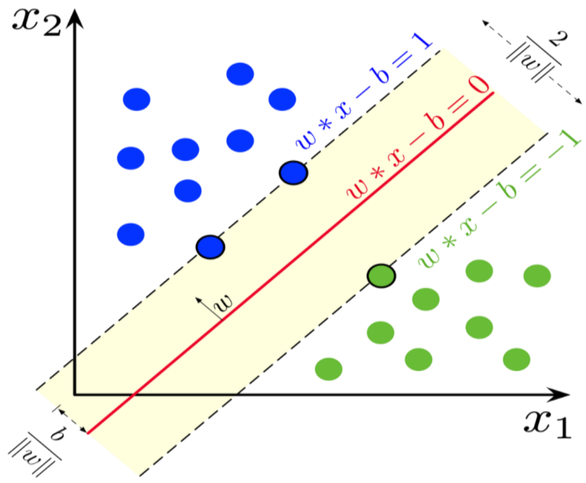

Fig. 1. Support Vector Machine Visualization

Decision tree

A decision tree (also called a classification tree or regression tree) is a decision support tool used in statistics and data analysis for predictive models. It helps with classification and regression problems.

The training set must have a target value for the examples because decision trees are supervised learning models. By the type of the variable, two types of trees are distinguished:

classification tree - when the target variable is discrete;

regression tree - when the target variable is continuous;

Decision tree is a method of representing decision rules in a certain hierarchy, which includes elements of two types - nodes and leaves. The nodes include rules and check the examples against the selected attribute of the training set.

A simple case: the samples get to a node, they are tested and split into two subsets:

the first - those that satisfy the established rule;

the second - those that do not meet the established rule.

Then the rule is applied to each subset again and the procedure is repeated. It continues until the condition for stopping the algorithm is reached. The last node, if not checked and split, becomes a leaf. An example of a decision tree is shown in Fig. 2.

The leaf defines the solution for each input sample. For the classification tree, this is the class associated with the node, and for the regression tree - the modal interval of the target variable corresponding to the leaf. A sample is included into the leaf if it matches all the rules on the way to it. There is only one path to each leaf. Thus, a sample can fall into only one leaf, ensures the uniqueness of the solution.

This method is simple and straightforward, does not require the preparation of input data, and makes it possible to assess the reliability of the model using tests. When used for identification of acoustic signals, the method shows rather high indicators of completeness and accuracy of the result. However, the disadvantages of the algorithm include a complicated way of describing certain concepts. Another disadvantage is the need to retrain the model with a significant increase in the size of the tree. The decision tree method was applied in [10] and showed a classification rate of 99%.

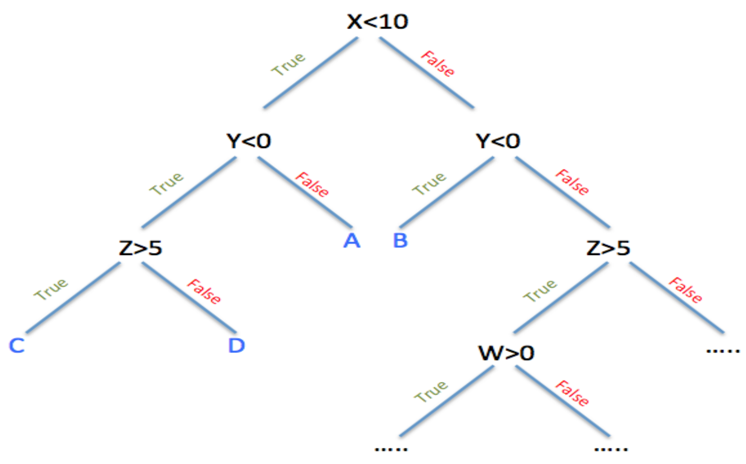

Fig. 2. Decision tree visualization

Neural network

A neural network is a computing system based on a principle similar to the brain of an animal. Neural networks are used for a number of tasks such as speech recognition and image recognition. The classical network uses a three-layer feedforward structure to solve classification problems. An example of a three-layer neural network architecture is shown in Fig. 3:

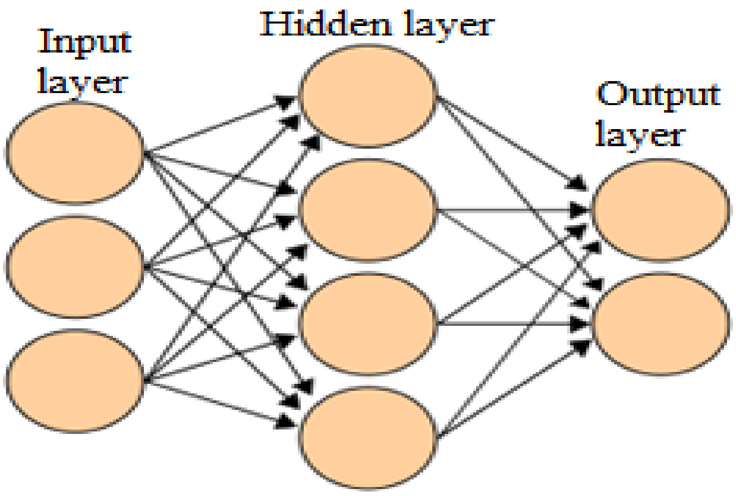

Fig. 3. Three-layer neural network architecture

Information enters the neural network through the input layer, which is the primary outer layer. The final level through which information passes is the output level. The input and output layers may or may not have additional layers in between. Layers, if any, are present between the input and output layers, called hidden layers. A neural network is considered a "deep" neural network if it has several hidden layers. As a rule, each neuron of any layer is interconnected with all neurons of neighboring layers. Each layer of the neural network breaks down the input data into a simple form for interpreting and classifying the content. The large number of layers means that there may be more paths for information to move through the network, potentially allowing for quite complicated tasks. In deep learning networks, each level of nodes is trained on a separate set of functions based on the results of the previous level. The further work goes in the neural network, the more difficult the functions can recognize the nodes, since they generalize and recombine the functions of the previous layer.

One of the main algorithms that is used to train a neural network is the error backpropagation algorithm, which is an iterative gradient method designed to minimize the mean square error between the expected output and the actual output for a particular input in the neural network. However, the backpropagation algorithm has two drawbacks. The first is that the learning process is very slow. The second is the presence of local minima. The reasons for this are a complex nonlinear cost function in the weight space and the lack of optimality in the choice of the learning rate, which is associated with different weights. There are many modifications to the backpropagation algorithm. Neural networks were applied in [11] and showed an accuracy ranging from 95% to 98%.

Neural network has been widely used in the field of pattern recognition because of its highly parallel distributed processing, association, self-learning and self-organization capabilities and strong nonlinear mapping capabilities. Based on the time-frequency analysis of the convolutional neural network, in the feature extraction application, a high-dimensional time-frequency map is selected, which contains a large number of signal features [12]. There are differences in the time-frequency diagrams, and image recognition technology can be used for pattern recognition and classification of underwater acoustic communication signals.

In order to use a deeper and wider network while reducing overfitting, reducing the amount of parameters, and improving the utilization of computing resources, a network structure that combines sparsity and intensive computing can be used, and on this basis, GoogLeNet[13] Inception v1 version, through the short-time Fourier transform (STFT) or wavelet transform of the underwater acoustic signal, after extracting the time-frequency diagram, combined with the convolutional neural network, the GoogLeNet is migrated and learned [14], reducing the amount of convolution operation , So that the deep neural learning network used for image recognition, after optimization and streamlining, can be used in the classification and recognition of underwater acoustic signals.

Conclusions

When considering the topic of data preparation, the following approaches were considered: sparse decomposition, generative process of signal creation and Fourier transform. An efficient and proven method is the Fourier transform. Fourier transform is widely used for signal processing in many fields. The versatility of this method was shown by comparing sonar and image signal processing, where this approach was used in both examples.

Machine learning algorithms were considered and compared, among which the following approaches were described: support vector machine, decision tree, neural network. Neural networks can be used for both regression and classification. A significant number of types of neural network models and algorithms allows to have a wide range of capabilities for solving specific problems. In comparison with others machine learning algorithms, neural network gives plausible results while working with large amounts of data. Support vector machine has high accuracy and speed when working with a small sample, but it should be understood that the support vector machine requires a lot of resources, significant and complicated work with parameters. Decision tree is easy and non-parametric algorithm, which do not require much effort for data preparation. It can be used for binary and multiclass classification problems. The main advantage of a decision tree is its ease of perception and implementation.

References

1. Johnston J. Transform coding of audio signals using perce tual noise criteria // Sel. Areas Commun. IEEE. 1988. Vol. 6, № 2. p. 314-323.

2. Theunissen F. E., Elie J. E. Neural processing of natural sounds. Nature Reviews Neuroscience. vol. 15, pp. 355–366, 2014

3. Kil W. C., Alexander S., Alexander S., Michael B., DEMON Acoustic Ship Signature Measurements in an Urban Harbor. Advances in Acoustics and Vibration, vol. 2011, Article ID 952798 , 2011.

4. Cho Y.-C., Choi S., Nonnegative features of spectro-temporal sounds for classification. Pattern Recog. Lett. vol. 26, pp. 1327–1336, 2005

5. Bonneel N., Drettakis G., Tsingos N. , Viaud-Delmon I., James D., Fast modal sounds with scalable frequency-domain synthesis, ACM Trans. Graph. vol 27, 2008.

6. G. Lemaitre and L. M. Heller, “Auditory perception of material is fragile while action is strikingly robust,” J. Acoust. Soc. Am. vol. 131 №2, pp. 1337–1348, 2012.

7. Purwins H., Li B., Virtanen T., Schlüter J., Chang S., Sainath T., Deep Learning for Audio Signal Processing. In Journal of Selected Topics of Signal Processing, Vol. 13, № 2, May 2019, pp. 206–219.

8. T. Hastie, R. Tibshirani, and J. H. Friedman, The elements of statistical learning: data mining, inference, and prediction, Second. New York, NY, USA: Springer, 2009

9. Robotham H., Castillo J., Bosch P., Perez-Kallens J. "A comparison of multi-class support vector machine and classification tree methods for hydroacoustic classification of fish-schools in Chile", Fisheries Res., vol. 111, № 3, pp. 170-176, Nov. 2011.

10. Lavner Y., Ruinskiy D.: A decision-tree-based algorithm for speech/music classification and segmentation. EURASIP Journal on Audio, Speech and Music Processing 2009.

11. Chin-Hsing C., Jiann-Der L., and Ming-Chi L., Classification of underwater signals using wavelet transforms and neural networks. Mathematical and Computer Modelling. vol. 27 pp. 47-60, 1998.

12. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

13. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going Deeper with Convolutions. In Proceedings of the CVPR, Boston, MA, USA, 7–12 June 2015; pp. 1–9.

14. Wang N., He M., Wang H., Zhou L., Shang Xi. A fast dimensionality reduction convolution model for underwater target recognition. Journal of Harbin Engineering University: vol. 40, 2019. pp. 1327-1333.

15. Wan S. Recognition Research of Underwater Sound Signal[D]. Xiamen University, 2017.