Implementation of svm in clasiffaing sources of acoustic signals

Varbanov Vladimir
Institute of Metal Science, Equipment and Technology with Hydroaerodynamic Center at Bulgarian Academy of Sciences, 67 Shipchenski Prohod Street, 1574 Sofia, Bulgaria
vdvarbanov@abv.bg

Abstract: *The paper addresses the problem of acoustic signal classification and report the current state of the art in the field. More specifically the paper sheds light on the mathematical apparatus behind one specific method used for classification of sound sources – SVM (support vector machines). As well we propose use of histogram of oriented gradients (HOG) as features descriptor and explain how they work.*

Keywords: *SVM (Support vector machines), HOG (Histogram of oriented gradients) descriptor, machine learning*

1. Introduction

The problem of classification of audio signals is important task in development and design of multisensory devices. Often such devices operate in sleep mode in order to conserve energy, because their main sensory devices operates in physical domain that requires significant amount of power to operate in example radar, where the majority of power is used to emit electromagnetic waves, or camera / cameras where power consumption is high because of the processing hardware. In such cases it is often advisable all these high consumption subsystem device to be awakened by less consuming sensor in example acoustic or seismic[1,2]. Therefore a relatively simple and stable method is needed in order to correctly classify the signal and make decision to wake the primary sensor or not. Although there are numerous such methods and architecture in the domains of neural networks and machine learning, the algorithms we chose for the classification is support vector machine, combined with histogram of oriented gradients as feature descriptor. The main reason behind this decision is the short time needed for training as well as the ability of the classifier to be trained on relatively accessible hardware[6,7,8].

2. Decision rule for SVM

For illustrative reasons we will present two dimensional case with two classification features X_A and X_B, as well as two classes. and two classes – one positive and one negative.

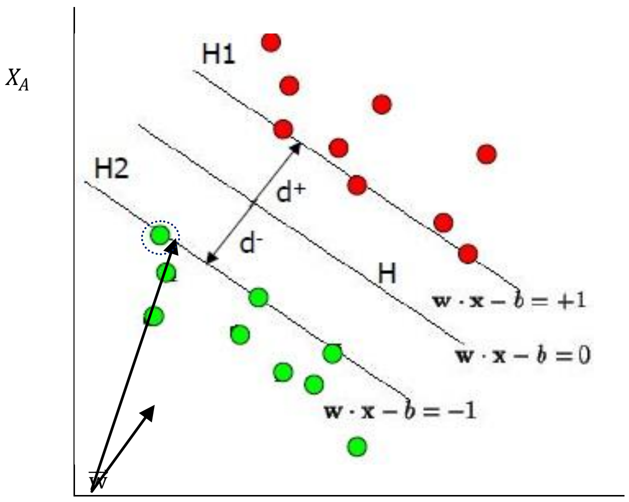

***Fig. 1** shows an separation in feature plane of objects belonging to two different classes.*

As it can be seen from Fig.1 every object can be put in one of the two classes with coordinates of the object presented as a vector with features X_A and X_B. Let's introduce vector $\vec{w}$, perpendicular to the separation line and unknown $\vec{u}$ (the object surrounded by dashed line). If we project the vector $\vec{u}$ onto the vector $\vec{w}$, we will find the distance, proportional to the direction of $\vec{w}$. There for if apply dot product on both vector we can compare the result with a constant b, which represents the distance from origin of the separation [4,5]line. Therefore whether the object belong to the red or green class can be decided by the equation:

$$\vec{w}\cdot\vec{u} \geq b, \tag{1}$$

which can be further developed into :

$$\vec{w}\cdot\vec{u} - b \geq 0 \tag{2}$$

Equation (2) is the decision rule. Next we can put the following additional constrains and write the following equations

$$\vec{w}\cdot\vec{x_r} - b \geq 1 \tag{3}$$

$$\vec{w}\cdot\vec{x_g} - b \geq -1 \tag{4}$$

where $\vec{x_r}$ and $\vec{x_g}$ are objects belonging to the red and green class. Equations (3) and (4) can be further derived into equation (5):

$$y_i(\vec{w}\cdot\vec{x_i} - b) \geq 1 \tag{5},$$

Which is more generalized form of the decision rule, where y_i obtains values -1 or $+1$, depending on whether the object x_i belongs to the red or green class.

3. Classification features

In order to be able to differentiate between two classes a suitable features must be selected for the objects belonging to the classes. In classification of acoustic signals such features are extracted from the spectrum of the signal. The problem with the spectral representation is that the carrier frequency might be shifted. This problem is usual solved by calculating the MFCC (mel frequency cepstral coefficients). MFCC manages to extract the form of the signal by applying the logarithm function on the frequency domain, followed by smoothing in order to find the envelope of the spectrum, then scaling it with Mel coefficients and followed by inverse Fourier transformation. Although this method was considered to be the most prominent in features extraction from audio signal it requires second Fourier on every sample which is not the most computationally efficient function. A novel method which we consider better is the use of HOG [3](histogram of oriented gradients). HOG is a feature extracting method which was develop for the needs of computer vision. In general; it works by finding contours in an image and creating histogram of this gradient's magnitude g and angle θ, thus providing a feature that is independent of the carrier frequency. HOG is implemented in the following steps:

1. creating series of Fourier samples on the signal temporal frequency spectrogram
2. processing the spectrogram like an image by

- applying gradient filter for the vertical component of the signal $-1 \quad 0 \quad 1$. This filter moves horizontally through the image and in current "pixel" of the image writes the difference between the next and the previous ""pixel".

- applying gradient filter for the horizontal component of the signal $\begin{matrix}-1\\0\\1\end{matrix}$.
  This filter moves vertically through the image and in current "pixel" of the image writes the difference between the next and the previous ""pixel".
- Finding the magnitude and the angles of the spectrogram using formulas (6) and (7)

$$g = \sqrt{g_x^2 + g_y^2} \quad (6)$$

$$\theta = arctg(g) \quad (7)$$

3. Save the values obtained with equations (6) and (7) into two matrices one with the values of the angles and another with the magnitude if the gradients
4. Split the spectrogram into blocks and then further split each block into four cells, as shown on Fig. 2

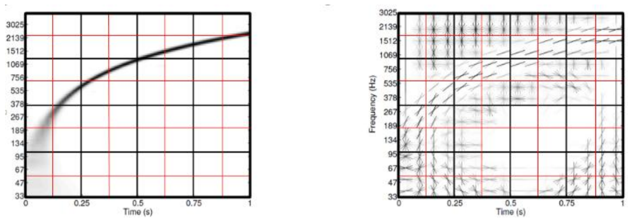

***Fig. 2** chirp generated in Matlab, the frame is split into blocks(black thick line) and each block is split into four cells*

5. Then for each cell create histogram of nine bins each representing the angle and its value representing the amplitude of the gradients

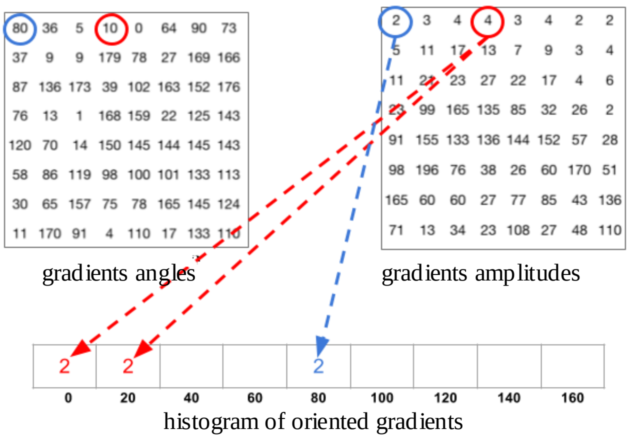

***Fig. 3** populating the histogram of oriented gradients with values*

6. Now that we have a histogram for each cell, all the four HOGs in a block are concatenated thus creating 32 dimensional feature vector for block and respectively 256 dimensional feature vector for a window moving window consisted of 8 blocks (4x2).

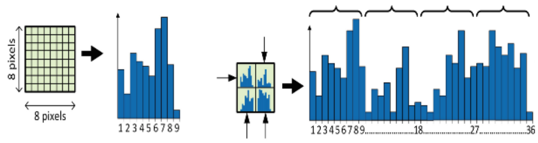

***Fig. 3** creating 32 dimensional feature vector from HOG descriptor*

7. Windows with dimensions 8 (4x2) blocks are taken from the signal and fed to the SVM classifier during the training stage. During the processing stage such window is moved over the spectrogram and decision rule based on equation (5) is applied for each window.

4. Overview

Classifier based on support vector machines provide stable results of correct detection in range between 75% and 82 %. Even though better results can be achieved using neural networks the latter require significantly more time to train, as well as expensive hardware (video cards). The probability of correct classification can be further increased by applying recursive decision filter based like Bayes filter or another filter based in Dempster-Shafer combinatorial rule in case the conditional probabilities are unknown or hard to obtain. Further research will include SVM classifier with LBP (local binary pattern) feature extraction. Unlike HOG, the LBP extracts texture from the spectrogram. Both features can be fused to further increasing the correct classification rate[5,6,7]. There are two methods of doing it early and late fusion. The former is done by combining both the LBP and HOG features, therefore using vector with higher dimensionality, while the latter is implemented by running the classifier twice with each of the feature vectors and the using some sort of logic for making decision. Still computational burden of such fusion need to be considered in order to design classifier that is balanced between speed and accuracy.

2014-2020.

5 References

1. M. Cowling and R. Sitte. Comparison of techniques for environmentalsound recognition. Pattern recognition letters, vol. 24, no. 15, pp. 2895–2907, 2003.

2. J. Dennis, H. D. Tran, and H. Li.Spectrogram image feature forsound event classification in mismatched conditions. IEEE Signal Process. Lett., Feb. 2011, pp. 130-133

3 N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol.1.IEEE, 2005, pp. 886–893.

4. M. Cowling and R. Sitte.Comparison of techniques for environmental sound recognition. Pattern recognition letters, vol. 24, no. 15, pp. 2895–2907, 2003.

5. D. Wang and G. Brown, Eds., Computational auditory scene analysis: Principles, algorithms and applications. Wiley-Interscience, 2006.

6. G. Yu and J. Slotine. Audio classification from time-frequency texture. in in Proceedings of IEEE International Conference in Acoustics, Speech and Signal Processing, 2009

7. Lichkov N. G, Dimitrov DL, Mihail Hristov, Stoyanov I .. Classification of chain technology as a source of propagation of surface seismic oscillations in different types of soils .. Proceedings of the International Scientific Conference on Security "CONFSEC 2020", 4, 1, NTS in Mechanical Engineering, 2020, ISSN: 2603-2945, 21-24

8. Lichkov N. G, Dimitrov DL, Mihail Hristov, Stoyanov I .. Classification of wheeled vehicles as a source of propagation of surface seismic oscillations in different types of soils .. Proceedings of the International Scientific Conference on Safety "CONFSEC 2020", 4, 1 , NTS in Mechanical Engineering, 2020, ISSN: 2603-2945, 17-20.