Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 138 (2018) 272–279

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN/HCist 2018

Classification of Control/Pathologic Subjects with Support Vector Machines

Felipe Teixeira[a], Joana Fernandes[a], Vitor Guedes[a,b], Arnaldo Junior[b], João Paulo Teixeira[a,c,*]

[a]*Instituto Politécnico de Bragança, Bragança 5300, Portugal*
[b]*Universidade Tecnológica Federal do Paraná, Câmpus Medianeira, Brasil*
[c]*Research Centre in Digitalization and Intelligent Robotics (CEDRI), Applied Management Research Unit (UNIAG), Instituto Politécnico de Bragança (IPB), Bragança 5300, Portugal*

Abstract

The diagnosis of pathologies using vocal acoustic analysis has the advantage of been noninvasive and inexpensive technique compared to traditional technique in use. In this work the SVM were experimentally tested to diagnose dysphonia, chronic laryngitis or vocal cords paralysis. Three groups of parameters were experimented. Jitter, shimmer and HNR, MFCCs extracted from a sustained vowels and MFCC extracted from a short sentence. The first group showed their importance in this type of diagnose and the second group showed low discriminative power. The SVM functions and methods were also experimented using the dataset with and without gender separation. The best accuracy was 71% using the jitter, shimmer and HNR parameters without gender separation.

Keywords: Vocal Acoustic Analysis, MFCCs, Jitter, Shimmer, HNR, SVM functions, SVM methods.

* Corresponding author. Tel.: +351 273 30 3129; fax: +351 273 30 3051.
E-mail address: joaopt@ipb.pt

10.1016/j.procs.2018.10.039

1. Introduction

Laryngeal pathologies can severely affect patients' quality of life. Such pathologies cause changes in speech quality, and these alterations are often not detected by the human ear. The use of common techniques for diagnose these pathologies causes discomfort to the patient, since they are invasive techniques and yet, they have the disadvantage of the cost associated with these techniques being expensive. However, the measurement of some parameters related to pathological speech, analyzed alone, are not very conclusive, but if they are adequately associated with an AI tool, the results are considerably better for voice pathologies diagnose [1] [2]. The artificial intelligence tools are an added value for this type of task, since, the analysis of a large number of data, with several variables, is not always possible for the human being.

Once the artificial intelligence system has been trained, it must be able to generalize, that is, by finding a situation that has never been seen previously, the system must be able to make a decision, based on similarities of parameters seen previously [2] [3] . Support Vector Machines (SVMs) are presented as binary classifiers, that is, they can only classify the data to be analyzed in one of two classes. Assuming it is possible to separate the data, it is up to the SVM to find a hyperplane that separates them.

The use of this tool is an adequate choice, since there are several linear classifiers, but the SVM hyperplane is created based on the concept of maximum margin, that is, the plane chosen between the data that maximizes the distance between the two options [4] [5].

The main objective of this work is to study and use SVMs in order to distinguish healthy speech from pathological speech. The pathologic speech includes dysphonia, chronic laryngitis and vocal cords paralysis.

In this study the relative jitter, relative shimmer, Harmonic to Noise Ratio (HNR) and Mel Frequency Cepstral Coefficients (MFCC) extracted from both vowels and continuous speech are used as acoustic analysis parameters.

The chapter 2 of this paper describes the pathologies in use, the used parameters and their mathematical expressions. The subjects used in this study are characterized according to their ages. Also the organization of the parameters is described.

In chapter 3 the results are present for the distinction between healthy subjects and subjects with pathology. Accuracy was calculated for three methods and several kernel functions.

The conclusions are presented in chapter 4.

2. Methodology

2.1. Pathologies

In this study three pathologies, dysphonia, chronic laryngitis and vocal cord paralysis were used. These diseases are the ones that most often cause disturbances in the human voice, being sometimes undetectable to the human ear.

Dysphonia is a disorder of the voice, often caused by abnormalities that affect the vibration of the vocal chords, this affects the ability to speak easily and clearly. Symptoms may include hoarseness, weak voice, changes in voice tone and may arise suddenly or gradually.

Chronic laryngitis consists of an inflammation that can result from inhalation of irritants or by the intensive use of voice. Symptoms include gradual loss of voice, hoarseness, and sore throat.

Paralysis of the vocal cords is the total interruption of the nervous impulse, being this total it happens in the two vocal folds, or being partial, occurs only in one of the folds. This disease can occur at any age and the problems associated with this pathology correspond to voice change, airway problems and swallowing problems.

2.2. Parameters

For this study it was necessary to extract a set of parameters from acoustic speech files. These parameters were relative jitter, relative shimmer [6] [7], HNR (Harmonic to Noise Ratio) and MFCCs (Mel Frequency Cepstral Coefficients) [4] [8] [9].

The jitter analysis for a speech signal is the mean absolute difference between consecutive periods, divided by the mean period and expressed as a percentage (Eq. 1).

$$jitter = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i-1}|}{\frac{1}{N}\sum_{i=1}^{N}T_i} \times 100 \quad (1)$$

Where T_i is the length of the glottal period i and N the total number of glottal periods.

The relative shimmer is defined as the mean absolute difference between magnitudes of consecutive periods, divided by the mean amplitude, expressed as a percentage (Eq. 2).

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A_{i+1} - A_i|}{\frac{1}{N}\sum_{i=1}^{N}A_i} \times 100 \quad (2)$$

Where A_i is the magnitude of the glottal period i and N the total number of glottal periods.

The HNR is a parameter in which the relationship between harmonic and noise components provides an indication of overall periodicity of the speech signal by quantifying the relationship between the periodic component (harmonic part) and aperiodic (noise) component. The overall HNR value of a signal varies because different vocal tract configurations imply different amplitudes for harmonics. HNR can be given by Eq. 3.

$$HNR = 10 \times log_{10}\frac{H}{1-H} \quad (3)$$

Where H is the normalized energy of the harmonic components and $1\text{-}H$ is the remaining energy of signal considered to be the non-periodic components.

The MFCC coefficients are obtained by the Fast Fourier Transform (FFT) of the signal, which are calculated through the spectrum of small windows of the speech signal. This spectrum is subjected to a bank of triangular filters equally spaced in the Mel frequency scale, in order to approximate the perception of the human ear to the frequencies of the sound.

At the output of the filters the discrete cosine transform is applied to calculate the Coefficients at the Mel frequency. Finally we calculate the energy and a factor that is designated by delta, in order to represent the dynamics of the signal from frame to frame [10].

2.3. Data

In this work we used the German Saarbrucken Voice Database (SVD) to extract the parameters associated with the various subjects used for the study. Parameters were extracted for 473 subjects, according to groups shown in table 1. The size and average age of each pathologic group are constrained by the subjects available in SVD.

Table 1 Dataset ages by each pathologic group and control group.

Groups	Sample size	Average age	Standard deviation of ages
Control	194	38,06	14,36
Dysphonia	69	47,38	16,37
Chronic Laryngitis	41	49,69	13,47
Vocal Cord paralysis	169	57,75	13,77
Total	473		

For every subject there are 9 speech sounds of sustained vowels. Three vowels, \a\, \i\ and \u\, spelled at low, normal, and high tones. For each speech file the parameters Jitter, Shimmer, HNR and MFCCs were extracted. Additionally, the SVD provides one speech file corresponding to the spelling of the sentence "Guten Morgen, wie geht en Ihnen?" ("Good morning, how are you?"). The MFCCs corresponding to continuous speech were also extracted from these sentences. In Table 2 we can see an excerpt for 3 subjects (one subject per line) from the organization of the matrix, relative to the parameter Jitter.

Table 2 Excerpt from the organization of the matrix regarding the parameter Jitter for 3 subjects

Jitter								
/a/			**/i/**			**/u/**		
high	**low**	**normal**	**high**	**low**	**normal**	**high**	**low**	**normal**
1,06	1,06	1,44	0,75	0,93	0,78	1,30	0,54	0,58
0,72	5,93	0,49	0,32	0,36	0,37	0,38	0,48	0,50
4,04	1,77	0,42	0,37	0,31	0,32	0,38	4,46	0,39

2.3.1 Organization of data

Initially all the sounds used for this study were downloaded by the 'Saarbrucken Voice Database'. Therefore the jitter, shimmer and HNR parameters were extracted using the software developed by Teixeira and Gonçalves [6] [7]. MFCCs coefficients, in a number of 13, were determined using the mfcc function of Matlab.

Three groups of parameters were considered as input features:

- Group I consists on the Jitter, Shimmer and HNR parameters extracted from the vowels /a/, /i/ and /u/ in the high, low and normal tones.
- Group II consists of the 13 MFCCs extracted also from the same vowels. The MFCC parameters for this group II were extracted in only one segment of speech of the sustained vowel. It was considered that MFCC parameters do not change significantly along the sustained vowel. Therefore, for every subject 13 MFCC parameters x 9 speech files were considered. Thus this matrix were converted into one row of 13x9=117 features to be used as input of the SVM for each subject.
- Group III consists in the extracted MFCCs parameters from the sentence: " Guten Morgen, wie geht es Ihnen? " ('Good morning, how are you?'). The MFCCs were extracted with a window length of 25 ms and using a displacement between segments of 10 ms. For the length of the sentence a total of 109 segments were used to extract the MFCCs. Therefore, for each subject was obtained a matrix (13x109) that were converted to a single vector with length 1417. The files with a length over 109 were stretched to 109 segments in order to have vector with same length for all subjects.

To sum up, the entrance of the SVM, for group I contains the parameters jitter, shimmer and HNR of the vowels in a total of 27 input nodes. For group II it contains the MFCC's of the vowels with length of 117, and for group III contains the MFCC's of the phrase with length of 1417.

The output of the SVM has only one binary node with 'zero' (0) for control or healthy subjects and 'one' (1) for subjects with pathology.

3. Results

The accuracy of the SVM's to classify between pathological and healthy subjects were experimented with various kernel functions and methods [11] [12]. The Kernel function is used to map the data in a given space and the method is used to find the hyperplane of separation between the mapped categories.

The 'QP' method consists of a smooth margin quadratic programming of 2 standards, the 'SMO' method consists of a minimal sequential optimization and the 'LS' method consists of the least squares method [12].

Accuracy was determined according to equation 4.

$$\text{Accuracy} = \frac{True\ positives + True\ negatives}{total\ of\ test\ subjects} \tag{4}$$

In this work, 75% of the subjects mentioned in table 1 were used to train the SVM and to perform the test the remaining 25%. Accuracy are presented only for the test set.

In the tables 3, 4, 5 and 6 diagnose was carried out by differentiating the gender of the subjects and was trained using 75% of female subjects and 75% of male subjects in a different SVMs. The accuracy shown was determined using the 25% used to perform the SVM test. For the further tables, * means that it was not possible to calculate the accuracy because the training was not possible for various reasons.

In Table 3 the group I parameters was used and it was concluded that the use of “LINEAR” and “RBF” functions give generally better accuracy for the 3 methods. The maximum accuracy was 70% when the "LINEAR" function and the "SMO" method are used.

In table 4, the procedure was similar to that of table 3. The data used were those of group II of parameters and some differences were obtained between some functions in the different methods, in this case the “POLYNOMIAL” and the “QUADRATIC” functions performed better. The maximum accuracy reached was 68%, using the functions "POLYNOMIAL" and " QUADRATIC" and both using the "QP" method.

In Table 5, group III of parameters is analyzed and "LINEAR" function achieved better performance. The maximum accuracy value of 69% is obtained when using the "LINEAR" function, using the "QP" or "SMO" method.

Table 3 Accuracy in test set for Group I parameters considering gender

	LINEAR	POLYNOMIAL	RBF	QUADRATIC	MLP
QP	69,0	59,5	68,0	59,5	*
SMO	69,8	57,8	68,0	65,5	63,8
LS	65,5	56,0	68,0	66,4	53,5

Table 4 Accuracy in test set for Group II parameters considering gender

	LINEAR	POLYNOMIAL	RBF	QUADRATIC	MLP
QP	56,0	68,0	58,6	68,1	*
SMO	*	66,4	58,6	66,4	60,3
LS	53,5	66,4	58,6	67,2	45,7

Table 5 Accuracy in test set for Group III parameters considering gender

	LINEAR	POLYNOMIAL	RBF	QUADRATIC	MLP
QP	69,0	58,6	58,6	58,6	*
SMO	69,0	58,6	58,6	62,9	63,8
LS	62,9	57,8	58,6	62,9	53,5

An attempt was made to obtain a greater precision than previously obtained, joining group I, group II and group III. Therefore the total number of input features is now 1561.

Table 6 Accuracy in test set for Groups I, II and III of parameters considering gender

	LINEAR	POLYNOMIAL	RBF	QUADRATIC	MLP
QP	69,8	58,6	58,6	58,6	*
SMO	69,8	58,6	58,6	60,3	69,8
LS	66,4	56,9	58,6	60,3	53,5

Analyzing table 6, it is noticed that the introduction of all the groups to be analyzed does not have great advantage. The best accuracy is equal as the case of group I.

In conclusion, the best accuracy considering gender is very similar for the 3 groups of parameters and is between 68 and 70%.

The study was also made without gender separation, this can be seen in table 7, where group I is used, in table 8 where group II is applied and in table 9 where group III is used, finally, in the table 10, the three groups are used as in table 6.

Table 7 Accuracy in test set for Group I parameters without gender separation

	LINEAR	POLYNOMIAL	RBF	QUADRATIC	MLP
QP	70,9	61,5	70,9	61,5	*
SMO	69,2	58,1	66,7	64,1	61,5
LS	61,5	61,5	70,9	67,5	54,7

Table 8 Accuracy in test set for Group II parameters without gender separation

	LINEAR	POLYNOMIAL	RBF	QUADRATIC	MLP
QP	58,1	63,2	59,0	63,3	*
SMO	*	58,1	59,0	60,7	59,8
LS	55,6	58,1	59,0	60,7	44,4

Table 9 Accuracy in test set for Group III parameters without gender separation

	LINEAR	POLYNOMIAL	RBF	QUADRATIC	MLP
QP	67,5	59,0	59,0	59,0	*
SMO	67,5	59,0	59,0	61,5	41,0
LS	62,4	58,1	59,0	61,5	49,6

Table 10 Accuracy in test set for Groups I, II and II without gender separation

	LINEAR	POLYNOMIAL	RBF	QUADRATIC	MLP
QP	70,1	59,0	59,0	59,0	*
SMO	70,1	59,0	59,0	61,5	41,9
LS	67,5	59,0	59,0	61,5	48,7

In table 7, for group I, again functions “LINEAR” and “RBF” give better accuracy at the level of 71 %. For table 8, group II, the accuracy become now lower than for the other groups of parameters, been the higher accuracy of only 63 %. For table 9, group II of parameters the accuracy also decrease to only 67,5%. Using the all parameters together in table 10 the accuracy achieved again the level of 70%.

Thus, this second group of experiments (without gender separation) show better results when using the group I of parameters. And the accuracy was slightly better than the accuracy achieved with gender separation.

With or without gender separation the better accuracy was achieved using “LINEAR” function with the “QP” or the “SMO” method.

4. Conclusions

This paper describes the experience of use SVM for diagnose between control subjects by subjects with one of the 3 pathologies (dysphonia, chronic laryngitis and vocal cord paralysis) using 3 groups of parameters extracted from the sustained vowels speech files or from a sentence. The analysis of the accuracy was carried out with and without gender consideration for different functions and methods of the SVM.

The function that indicates less favorable results is "MLP". The "QP" method is also the least favorable if used in conjunction with the "MLP" function. Therefore this function and method are not recommended to be used together.

It is understood that when analyzing group III the results are practically identical when adding all the data, which leads to conclude that a lot does not mean that a better learning is done by the SVM.

A maximum of 70,9% of accuracy in the test set was achieved using the jitter, shimmer and HNR parameters, "LINEAR" function with the "QP" method without gender separation. Anyhow very similar accuracy was achieved with other group of parameters.

In general, group II gives slightly lower results in relation to the remaining groups, whereas group I appears to contain better information. Therefore, the use of MFCCs for sustained speech is not recommended.

Concerning gender, the results of Teixeira and Fernandes [13] is reinforced, where it is affirmed that there is not a significant difference discriminating the gender of the subjects used. Analyzing the results obtained it is noticed that for both cases, whether or not gender is discriminated, the "LINEAR" function presents better results than the others.

Future work intends to measure the accuracy of diagnose of each individual pathology using SVM and other machine learning tools.

Acknowledgements

This work is supported by the Fundação para a Ciência e Tecnologia (FCT) under the project number UID/GES/4752/2016.

References

[1] J. Teixeira, J. Fernandes, F. Teixeira and P. Odete, (2018) "Acoustic Analysis of Chronic Laryngitis".

[2] J. Teixeira, N. Alves and P. Odete, (2017) "Vocal Acoustic Analysis - Classification of Dysphonic Voices with Artificial Neural Networks," pp. 19-26.

[3] R. May, G. Dandy and H. Maier, (2011) "Review of Input Variable Selection Methods for Artificial Neural Networks," In K. S. (Ed.), Methodological Advances and Biomedical Applications.

[4] H. T. Cordeiro, (2016) ""Reconhecimento de Patologias de Voz usando Técnicas de Processamento de Fala"," PhD thesis, FCT Universidade Nova de Lisboa..

[5] Andrew. Ng, (2012) “Support Vector Machines”, CS229 Lecture notes pt.1, Intell. Syst. their Appl. IEE, pp. 1-25.

[6] J. Teixeira and A. Gonçalves, "Accuracy of Jitter and Shimmer Measurements," *Procedia Technology, 16,,* pp. 1190-1199, 2014.

[7] J. Teixeira and A. Gonçalves, "Algorithm for jitter and shimmer measurement in pathologic voices," *Procedia Computer Science, 100,,* pp. 271-279..

[8] S. Davis. and P. Mermelstein, (1980) “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences" *IEEE Trans. Acoust.,* vol. 28.

[9] P. Boersma, (1993) “Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound,”," *IFA Proc.,,* vol. 17, pp. 97-110.

[10] X. Huang, A. Acero and H. Hon, "Spoken Language Processing: A guide to theory, algorithm, and system development.," *Prentice Hall.,* 2001.

[11] S. Haykin, "Neural Networks A Comprehensive Introduction.," *Prentice-Hall,* 1999.

[12] I. T. MathWorks, ""svmtrain,"," 2018.. [Online]. Available: https://www.mathworks.com/help/stats/svmtrain.html?s_tid=srchtitle..

[13] J. Teixeira and P. Odete, "Jitter, Shimmer and HNR classification within gender, tones and vowels in healthy voices.," *Procedia Technology,,* vol. 16, pp. 1228-1237., 2014.