Voice pathology detection

Wahiba ABOU-ZBIBA, Sara SEKKATE and Abdellah ADIB

Master Informatique et Télécommunications

FACULTÉ DES SCIENCES ET TECHNIQUES DE MOHAMMEDIA
UNIVERSITÉ HASSAN II DE CASABLANCA

Morocco IndabaX

Abstract

Voice pathology detection requires the presence of a specialist doctor and time to treat each patient, but it is not always possible to have a doctor who can treat all patients at once and at one precise time. For residents of remote areas it is all expensive equipment that must be provided. Or even for people who may not be aware of having any voice pathology.
Our goal is to design a diagnostic aid system to detect whether the voice is pathological or healthy, so that the patient can be referred to a doctor or not without being moved from the start.
Our system is based on the classification, by SVM "Support Vector Machine", using MFCCs "Mel Frequency Cepstral Coefficients" extracted from the patient's voice.
The learning and testing of our system are done using the SVD database "Saarbruecken Voice Database".

Problematic and motivation

The detection of the voice pathology is a very important task, which must not be neglected before its progression. There are many precancerous pathologies that present a great danger, and other pathologies that are benign but have a negative effect on the voice. The voice becomes incomprehensible, which makes the communication with people very difficult. In order to avoid the progression of voice pathologies, it is necessary to follow a treatment in the first steps of the pathology. The problem that arises here is the early detection of these vocal pathologies, we can distinguish two types of methods of detection of pathologies; subjective and objective.
The subjective methods are those based on specialized medical techniques, but doctors are not always able to detect the pathology before reaching an advanced stage of evolution.
Objective methods are the methods based on voice signal processing.
In this work, we used a subjective method for the early detection of two pathologies 'dysphonia' and 'laryngitis', using machine learning techniques.

SVM

Support vector machines (SVMs) are a set of supervised learning techniques intended to solve problems of discrimination and regression. SVMs are a generalization of linear classifiers.

However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well.
The principle operation of the SVMs is to find the optimal hyperplane among the set of possible hyperplanes to correctly classify the data and maximize its distance between the nearest support vectors.

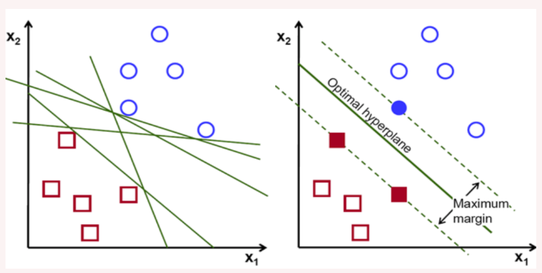

Figure1: Hyperplanes and the optimal hyperplane

In SVM, it is easy to have a linear hyper-plane between these two classes. But, another burning question which arises is, should we need to add this feature manually to have a hyper-plane. No, SVM has a technique called the kernel trick. These are functions which take low dimensional input space and transform it to a higher dimensional space i.e. it converts not separable problem to separable problem, these functions are called kernels. It is mostly useful in non-linear separation problem. Simply put, it does some extremely complex data transformations, then find out the process to separate the data based on the labels or outputs you've defined

Experiments

In this work we built a system that can detect two voice pathologies "dysphonia" and "laryngitis" and detect if the voice is healthy.
the database used in the learning and testing phases is SVD.
SVD is a free database available online.
It is a collection of voice recordings from more than 2000 persons, where a session is defined as a collection of:
– recordings of vowels /a/, /i/, /u/ produced at normal, high, low and lowhigh-low pitch.
– recording of sentence "Guten Morgen, wie geht es Ihnen?" ("Good morning, how are you?").

We focused on training the system to recognize the voice characteristics. One of the most popular feature extraction techniques is the Mel Frequency Cepstral Coefficients called MFCC as it is less complex in implementation and more effective. MFCC is designed using the knowledge of human auditory system. It is a standard method for feature extraction in speech recognition. Steps involved in MFCC are : Pre-emphasis, Framing, Windowing, Fast Fourier Transform (FFT), Mel filter bank, and Discrete Cosine Transform (DCT).
In our example we used a vector of characteristics composed of the mean of 12 MFCCs for each signal and its variance.

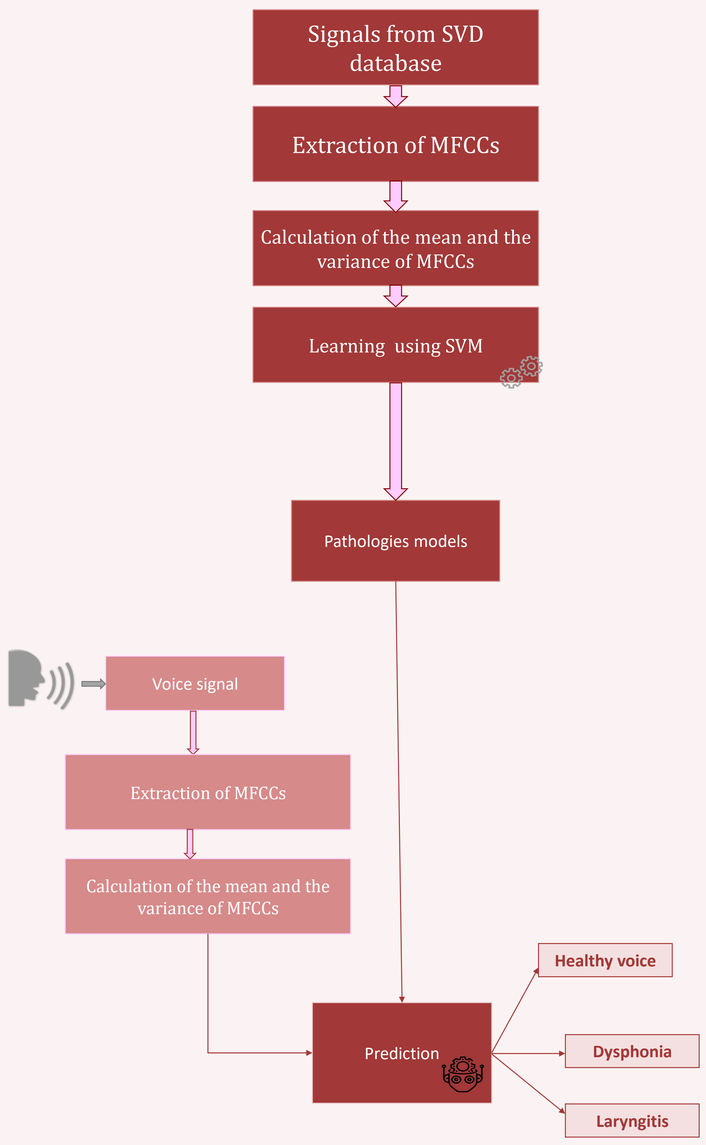

Figure 2: The system architecture

References

[1]- Voice Pathology Detection on the Saarbrucken Voice Database with Calibration and Fusion of Scores Using MultiFocal Toolkit- David Martınez, et al.
|[2]- An Approach to Extract Feature using MFCC -Parwinder Pal Singh, Pushpa Rani
[3]- Understanding Support Vector Machine algorithm from examples- www.analyticsvidhya.com