

Study on Optimization of Pathological Voice Features

Xiaojun Zhang^a, Yating Shao, Di Wu, Baoyin Sun, Zhi Tao^b

College of Physics, Optoelectronics and Energy, Soochow University, Suzhou, China

^azhangxj@suda.edu.cn, ^btaoz@suda.edu.cn

Keywords: classification, optimal selection, pathological voice, recognition

Abstract. This paper proposes a method to optimize Pathological Voice Features based on classification and experimental design. Firstly, all voice parameters of features are classified, then pick the right amount of each class of representative and typical parameters based on the principle of mathematical statistics. Finally, recognition experiments are performed with this combination. C4.5 decision tree algorithm is used to recognize vocal cord or non-vocal cord damage voice, results show that recognition rate of optimized features is 5% higher than non-optimized, which is up to 89%.

Introduction

Automatic identification of pathological voice has become an important research field. However with the increase of the amount of parameters, the complexity and memory consumption will be significantly increased in model training or recognition [1-3]. Several proper features which have the best discrimination capability should be selected [4-6]. Currently there are two main methods about selection of parameters, one is dividing the characteristic parameters into several categories based on the knowledge of phonetics and auditory aspects, then select a class from the various types of characteristic parameters by comparing experiment. The other method is to calculate the number of dimensions of feature vectors by doing experiment on each type of the feature parameters, including figuring out the number of dynamic characteristics. Both two methods obtained a single type of characteristic parameters or simply add several other parameters, they can't find a combination of different types of characteristic parameters systematically.

To solve this problem, this paper presents a method based on classification and experimental design to optimize the parameters of features. This method complete the optimization of the acoustic parameters by researching the correlation among acoustic parameters of pathological voice. Then optimized parameters is selected which contain more information and mostly reflect the pathological features of voice. Results show that recognition rate with optimized features is 5% higher than non-optimized, which is up to 89%.

Classification of voice features

The merits and the relation of the characteristics need to be evaluated, because there will be a close correlation between the same type of parameters. If these parameters are used to characterize the voice features of vocal cord diseases, there must be lots of redundancy. So it is necessary to study the correlation of these features. Based on this problem, all the parameters is divided into three categories. Three combinations of different types of parameters will each adopt the highest rate of property optimization algorithm to screen the best parameters. Finally aggregate all selected parameters. The flow chart of experiment is showed in figure 1.

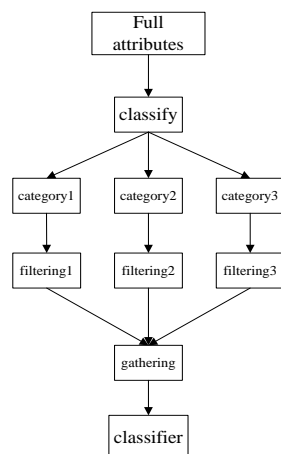


Fig. 1: flow chart of experiment

To reduce the redundancy between parameters, the correlation of these features is studied. It concludes that each parameter have different effects on different types of voice disease. Therefore it is necessary to study the merits of the various features to describe different voice diseases. In this study, all the features are divided into three categories.

Features about fundamental frequency reflect the stability degree of the frequency of the voice signal or pitch period. STD is a measure of the overall degree of stability of the fundamental frequency, it is closely related to the degree of relaxation of vocal cords. When sound becomes rough, STD significantly increased. Pitch frequency variation coefficient (VFo) reflects the deviation degree relative to the mean Fo. Value change of Fo causes the increase of VFo. The first feature set shows in Table 1.

Table 1 Category1: The first set of frequency-dependent features

The first set of frequency-dependent features		
pitch frequency(F_0)	highest pitch frequency (Fhi)	lowest pitch frequency (Flo)
pitch period(T_0)	standard deviation of frequency (STD)	pitch frequency variation (VFo)

Jitter refers to the change rate of acoustic wave frequency during the adjacent period. It represents the degree of change during the relative periods. The relative average perturbation (RAP) and pitch period perturbation quotient (PPQ) are parameters reflecting the change degree of Jitter. RAP is equal to the third-order perturbation, PPQ is equal to the fifth-order perturbation. Sf is the smoothing factor of smoothed pitch period perturbation quotient (SPPQ).The second feature set shows in Table 2.

Table2 Category2: The second feature set associated with the frequency jitter

The second feature set associated with the frequency jitter			
frequency of Fo jitter (Fftr)	frequency of amplitude jitter (Fatr)	absolute frequency jitter (Jita)	frequency of jitter percent (Jitt)
relative average perturbation (RAP)	period perturbation quotient (PPQ)	Smoothed PPQ (SPPQ)	Fo-Tremor intensity index (FTRI)

Amplitude characteristics describe the degree of stability of the voice amplitude. The percentage of amplitude jitter (Shim) gives an evaluation in percent of the variability of the peak-to-peak amplitude within the analysed voice sample. Peak amplitude variation (vAm) represents the relative standard deviation of the period-to-period amplitude. In this study, a small number of energy-

related characteristics are attributed to this category. Noise to harmonic ratio (NHR) is an average ratio of energy of the in-harmonic component in the range 1500-4500 Hz to the harmonic component energy in the range 70-4500 Hz. SPI is an average ratio of the lower-frequency to the higher-frequency harmonic energy. The third feature set shows in Table3.

Table3 Category3: The third feature set associated with the amplitude jitter

The third feature set associated with the amplitude jitter		
shimmer in dB (ShdB)	shimmer percent (Shim)	amplitude perturbation quotient (APQ)
Smoothed APQ (sAPQ)	peak amplitude variation (vAm)	noise to harmonic ratio (NHR)
voice turbulence index (VTI)	the soft pronunciation index (SPI)	the amplitude tremor intensity index(ATRI)

Searching algorithm of feature selection

Attribute selection problem is a very important method to preprocess data in the field of mining data. Attribute selection process is mainly composed of two parts. One is the generation process of properties subset, the other is the evaluation of the properties. Currently study on generation of attribute subset focused on two aspects those are complete style and heuristic style. Attribute selection methods commonly use search method forward, backward delete method, exhaustive search method, relief algorithm and packaging method and so on[7-9]. This paper uses Cfs algorithm evaluation methods and gain ratio evaluation.

Stepwise forward selection procedure starts from an empty set of features. In each subsequent iteration, we select the best features of the current feature set (Evaluation function J is max when this feature combines with the already selected features together), then add it to the collection.

Cfs is a filtering algorithm which sorts feature subsets based on evaluation function. Deviation of the evaluation function can determine the relevance of features. It evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Cfs feature subsets evaluation function as shown:

$$J = \frac{k r_{cf}}{\sqrt{k + k(k-1)r_{ff}}} \quad (1)$$

Where J is the evaluation function, r_{cf} is the average correlation between the characteristics and classification category, r_{ff} represents the mean value among interrelated characteristics.

Ranker is a searching method for subset of the attributes, ranking attributes by their individual evaluations and using in conjunction with attribute evaluators (Relief, GainRatio, Entropy etc). GainRatio evaluates the worth of an attribute by measuring the gain ratio with respect to the class.

$$\text{GainRatio}(S, A) = \frac{\text{Gain}(S, A)}{\text{SplitInfo}(S, A)} \quad (2)$$

Where $\text{SplitInfo}(S, A)$ is on behalf of the breadth and uniformity of property A splitting sample set S.

$$\text{SplitInfo}(S, A) = - \sum_{i=1}^c \frac{|S_i|}{|S|} \times \log_2 \left(\frac{|S_i|}{|S|} \right) \quad (3)$$

Follow the above steps, select the properties most wanted with the most suitable property evaluation algorithm respectively, and meet demand of good integrity of data, less redundancy of data, less correlation among properties, etc. Selection results show in Table 4.

Table4 Selection results

Category1		Category2		Category3	
Algorithm chosen	Attribute selection	Algorithm chosen	Attribute selection	Algorithm chosen	Attribute selection
	Fhi		RAP		Vam
	Flo		PPQ		NHR
BF+Cfs	STD	Ra+Gain	Jitt	BF+Cfs	VTI
			Jita		
			SPPQ		

Experimental results and data analysis

Currently common diseases causing voice changes can be divided into vocal and non-vocal cord damage. Vocal cord damage includes vocal nodules, vocal fold edema, vocal fold polyps and others. In this study, the main job is to identify vocal or non-vocal cord damage. 100 groups of hyperthyroidism data was chosen from KAY Database and 100 groups of vocal cord damage data from the First Affiliated Hospital of Soochow University.

The experiment roll up the three feature sets, classify it with J48 classifiers and use 10-fold cross-validation method [10]. Finally the average classification error rate is calculated. The Kappa statistic and mean absolute error rate are chosen as the evaluation index.

Kappa analysis is a multivariate statistical method to evaluate the classification accuracy. The larger Kappa value is, the more accurate the classification is. As shown in Figure 2, through experimental comparison, the classification with the individual category of parameters is less accurate than with the full-featured parameters, and the Kappa statistic is maximum using optimized features. As shown in Figure 3, similar conclusions can be seen from the mean absolute error rate curve.

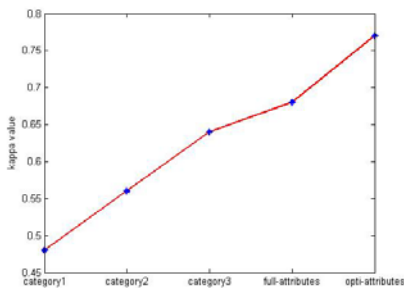


Fig. 2: Kappa value comparison

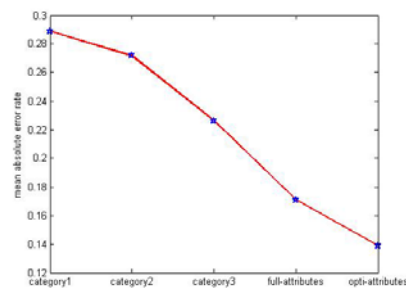


Fig. 3: Mean absolute error rate comparison

Table 5 shows the recognition rate under different types of parameters. When selecting all features (dimension is 23), it reached 84% correct recognition rate. After optimization, there are 11 optimized characteristic parameters left which obtaining 89% recognition rate. Compared with traditional methods, this method not only achieves the purpose of dimensionality reduction, but also improves the recognition rate at the same time.

Table5 Recognition comparison

Combination regimen	Dimension	Recognition rate (%)	Time consuming (ms)
Full attributes	23	84	80
Optimized attributes	11	89	10

Conclusions

Pathological voice recognition is an important part of voice signal analysis field, and the impact of disease on vocal voice is more common and significant. Extracting the right features is key to

successfully recognize vocal disease and non-vocal voice disorders. However, there is redundancy among similar characteristics. So characteristic optimization and selection is also important. In this paper, the feature classification, parameters selection and classification algorithms are combined, consisting of a new feature vector to conduct identification experiment. Experimental results demonstrate the effectiveness of the theory.

Acknowledgements

The research work was supported by National Natural Science Foundation of China under Grant No. 61271359.

References

- [1] Emmanouilidis C., Hunter A., MacIntyre J., A Multiobjective Evolutionary Setting for Feature Selection and a Commonality-Based Crossover Operator[C], CEC'2000 , 2000:309-316.
- [2] Lee J., Jeong S., Hahn, M., Pathological Voice Detection Using Efficient Combination of Heterogeneous Features[J]. IEICE transactions on information and systems, 2008, (2): 367-370.
- [3] Kreiman J., Gerratt BR., Perception of aperiodicity in pathological voice[J]. J Acoust Soc Am, 2005, 117(4):2201-2211.
- [4] Ververidis D., Kotropoulos C., Pitas I., Automatic emotional speech classification, ICASSP' 2004, 2004: 593-596.
- [5] McGilloway S., Cowie R., Douglas-Cowie E., et al., Approaching automatic recognition of emotion from voice: a rough benchmark[J], Proc Isca Workshop Speech & Emotion, 2000:207-212.
- [6] Blom ED., Pauloski BR., Hamaker RC., Functional outcome after surgery for prevention of pharyngospasms in tracheoesophageal speakers. part 1:speech characteristics[J]. Laryngoscope, 1995,105(10):1093-1103.
- [7] Derksen S., Keselman H.J., Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables[J]. British Journal of Mathematical & Statistical Psychology, 1992, 45(2):265–282.
- [8] Khan J.A., Aelst S.V., Zamar R.H., Building a robust linear model with forward selection and stepwise procedures[J]. Computational Statistics & Data Analysis, 2007, 52(1):239–248.
- [9] Fang W.B., Wang Q.Q., Zhi H., Feature Selection Method Based on Adaptive Relief Algorithm[C]. ICCEE'2010, 2010:21-24.
- [10] Kohavi R., A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection[J]. International Joint Conference on Artificial Intelligence, 1995:1137-1143.