

Optimal Feature Selection of Speech using Particle Swarm Optimization Integrated with mRMR for Determining Human Emotion State

S Rajarajeswari

Assistant Professor
Dept of Computer Science
MSRIT,Bangalore,India

Shree Devi B N

Dept of Computer Science
MSRIT,Bangalore,India

Sushma G

Dept of Computer Science
MSRIT,Bangalore,India

ABSTRACT

Speech is one of the most promising model through which various human emotions such as happiness, anger, sadness, normal state can be determined, apart from facial expressions. Researchers have proved that acoustic parameters of a speech signal such as energy, pitch, Mel frequency Cepstral Coefficient (MFCC) are vital in determining the emotion state of a person. There is an increasing need for a new Feature selection method, to increase the processing rate and recognition accuracy of the classifier, by selecting the discriminative features. This study investigates the use of PSO integrated with mRMR (Particle Swarm Optimization integrated with Minimal-Redundancy and Maximal-Relevance) technique to extract the optimal feature set of the speech vector, thus making the whole process efficient for the GMM.

General Terms

Speech Vector, Emotion Recognition, Feature Selection

Keywords

Emotion State Recognition, Feature Selection, GMM, Integrated PSO and mRMR, Speech characteristics, mRMR, MFCC, Pitch, PSO.

1. INTRODUCTION

The study of Human Emotion Recognition using Speech features is emerging as an extensively growing research topic in artificial intelligence due to its numerous applications such as in intelligent human-computer interaction, security systems, etc. Numerous approaches do exist for developing such a system. In this paper a proposal has been made to improve the efficiency of emotion state recognition by GMM, using an optimization algorithm such as PSO+mRMR, which selects the optimal feature set from a speech vector. The recognition process first requires pre processing of the voice signal followed by feature extraction. The extracted features are represented in vector format and subjected to recognition by GMM (Eslami and Sayadian, 2006).

A Gaussian Mixture Model (GMM) is a type of density model comprising a Gaussian component function (a

parametric probability density function) which is commonly used to represent model of the probability distributions of continuous features. Since vectors associated with a particular sound are clustered around the subspace of density range (Gaussian peak), the model can be used for estimating the parameters for the function and hence the emotion state.

Feature Selection plays a major role in reducing the dimensionality of extracted features, to produce an optimal subset, for increasing the speed and accuracy of the GMM classification. Various feature selection methods like Principle Component Analysis, LDA, exhaustive algorithms, forward selection and backward selection, filter approaches such as Fisher Score[1], Mutual Information[1] and AUC_S[1] have been proposed by many researchers with a maximum success rate for AUC_S based feature ranking method and GMM classifier. Since AUC for ROC curve has its own disadvantages which has been discussed in later section, the use of PSO+mRMR for optimal selection of speech characteristics has been proposed.

PSO (Kennedy and Elberhart, 1995) is a meta-heuristic optimization technique which seeks to explore the search space of a given problem to find the parameters required to maximize a particular objective, a concept inspired by the block movement of swarm of fishes and birds. mRMR is a filter based approach for measuring the relevance and redundancy of a feature. The measure of relevance of a feature for the class is determined by the mutual information between a feature and a class. The study investigates the integration of PSO and mRMR in feature selection. The following section enumerates the emotion recognition system without feature selection.

2. THE EMOTION RECOGNITION SYSTEM WITHOUT FEATURE SELECTION

This system requires analysis of speech inputs from various speakers, to recognize their different emotional states, namely happy, angry, sad and normal. Firstly, the speech inputs are pre-processed such as elimination of noise, followed by extraction of the speech characteristics such as energy, pitch, MFCC which are proved to be efficient for emotion state classification[2,3,9]. Each input is expressed as a frame,

given the average time for each sentence, to generate speech vector. The resultant vector is then subjected to further processing by GMM to ascertain one of the emotional states that the input indicates. Figure.1 shows the overview of the system with and without feature selection.

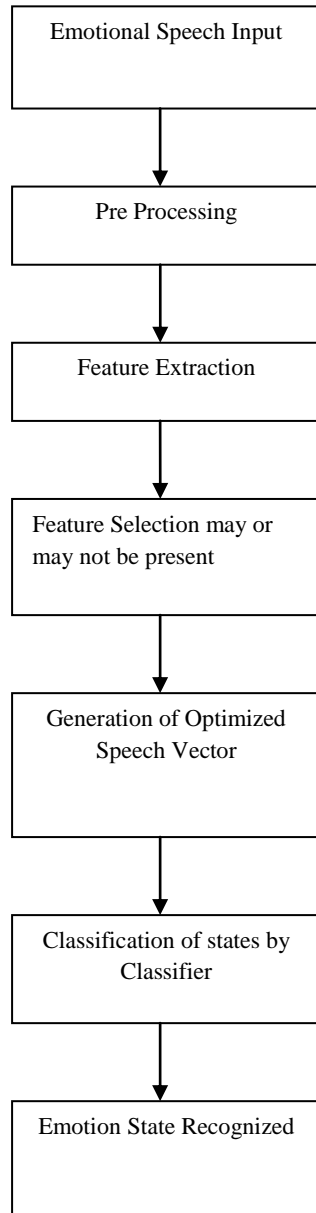


Figure 1: A Generalized Structure of the Emotion Recognition System

The GMM model is based on probabilistic modelling. The sub-populations are analysed with respect to the overall system and not with respect to their individual characteristics. The GMM works by analysing the density of particles. Systems are composed of various component functions. The functions are plotted and the density of the particles is obtained. The regions of largest density are indicators of the various speech characteristics [2,10]. Hence the speech vector must be subjected to Gaussian analysis so as to observe the

density of the components which are clear indicators of the emotional state.

The vector consists of 55 components[2] as listed below: MFCC coefficients, Energy(E); the first derivative of MFCC coefficients; the first derivative of energy(E'); the second derivative of MFCC coefficients; the second derivative of energy(E''); the first Formant frequency(F1); the second Formant frequency(F2); the third Formant frequency(F3); the first derivative of the first Formant frequency(F1'), the first derivative of the second Formant frequency(F2'), the first derivative of third Formant frequency(F3'); the logarithm of first Formant frequency(log F1), the logarithm of second Formant frequency(log F2), the logarithm of third Formant frequency(log F3), the first Formant frequency normalized(zF1); the second Formant frequency normalized(zF2); the third Formant frequency normalized(zF3); Pitch Frequency(F0), the logarithm of Pitch Frequency(log F0), the first derivative of Pitch Frequency(log F0') and the normalized Pitch(zF0). All the above components are subjected to Gaussian analysis and the different density ranges are studied and classified, with the help of trained dataset. One possible analysis would be as follows:

If the pitch and energy are very high then the state can be concluded as angry. If the pitch and energy are very low then the state can be concluded as sad, though the model is trained by the existing dataset. Analysing a 55 characteristic vector becomes very cumbersome for GMM model. For more efficient processing, feature selection should be applied for the selection of discriminative features as shown in Figure.1. PSO+mRMR technique has been proposed for the same.

3. SOME OF THE EXISTING FEATURE SELECTION METHODS AND THEIR DRAWBACKS

This section provides an overview of the existing techniques in feature selections.

3.1 Principle Component Analysis(PCA)

Principle Component Analysis (Karl Pearson, 1901) is a mathematical method that uses transformations to identify patterns in data. The components are arranged in a particular pattern with the component having highest variance occurring at the topmost level, followed by other components with high variance but totally uncorrelated with the previous components. Although PCA is a very old technique and a powerful tool, it has several disadvantages. Only orthogonal transformations are considered. Components with the highest variance are considered to be of topmost priority; only mean and variance are considered for all the components and is not very successful in eliminating higher dependencies.

3.2 Exhaustive Search

This is the most general problem solving strategy that is used to enumerate systematically through all the possible candidates to find whether any of these satisfy the problem's requirements. This is the easiest method to implement, however the cost is directly proportional to the problem size or the number of prospective candidates. For N predictor variables, the exhaustive search must traverse through $2^N - 1$ subsets. Hence it is the

best method to be used when the problem size is small but becomes intractable as the problem size increases. Therefore this method is seldom used for large systems with many candidates.

3.3 Forward and Backward Selection

The primary difference between forward search and exhaustive search is the tractability. Forward selection starts with an empty set and adds the predictor variables by selecting the one with the highest evaluation criterion. The problem with forward selection is that it is not always guaranteed to produce the subset with the highest evaluation criterion. Backward selection is vice versa of Forward Selection. As the selection proceeds, predictor variables are deleted one at a time keeping the predictor variables with the highest evaluation criterion in the subset. Backward selection is preferred to forward selection when there is some dependency among the variables. Still it has its problems similar to that of forward selection. Sometimes significant variables are dropped from the set. For both the methods in the worst case, at most $N((N+1)/2)$ subsets are evaluated.

3.4 Some of the Filter Based Approaches like MI, AUC_S

Mutual Information (MI) is a filter based approach which measures the dependence between two random variables. If the MI is zero, both variables are independent and contain no information about each other hence irrelevant to the target class. Higher MI values indicate more information about the target class and hence a higher relevance. Despite its advantages it has many drawbacks. It is complex and it requires a probability density function which is very difficult to obtain, and involves integration of those functions.

AUC_S is an AUC feature ranking method applied on SVM [1]. The area under the curve(AUC), is the performance measure determined from the ROC curve. Depending on their priorities for a given class such as anger, sad, etc, features are determined from the area beneath the ROC curve. Area Under Curve for each feature is calculated by determining the area under ROC curve including and excluding the feature to obtain priority measure of that feature, which is again a cumbersome process.

3.5 PSO

PSO is method based on social behaviour of birds which is used for eliminating less effective features to produce 49, 24 and 15 speech characteristic vector which can then be subjected to GMM classifier [2].

The subset can be further reduced when integrated with the mRMR technique thereby reducing the load on GMM, and making the whole system more efficient, as described in later section.

Due to all the disadvantages mentioned for the existing methods, a new method has been proposed for feature selection in the following section.

4. PROPOSED SYSTEM WITH PSO+mRMR

4.1 PSO

PSO is an optimization algorithm which maintains a population of particles which form the candidate solutions, by moving them in the search space with calculated velocities [11]. The following equations are used to perform the algorithm.

$$V_i(t+1) = \omega * V_i(t) + c_1 r_1 [P_i X_i(t+1)] + c_2 r_2 [Gbest X_i(t)] \quad (1)$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \quad (2)$$

Where $X_i(t)$ is the current position of the particle, $X_i(t+1)$ the updated position of the particle, $V_i(t)$ the current velocity of the particle, $V_i(t+1)$ the updated velocity of the particle, ω is the inertia coefficient which is a constant in the interval [0,1], c_1 the cognitive parameter of the particle, c_2 the social parameter of the particle, P_i best position of the i th particle, $Gbest$ the global best position of the swarm, and r_1 and r_2 are generated randomly in the interval [0,1].

The fitness identifies the best particle which communicates to its neighbouring particles through conduction, thus accelerating the particles towards its direction and slowly approaching the best solution [12]. Iterations are performed which applies the update equations given above until an exit criterion is met. The following computations are performed during every iteration:

1. Calculate the fitness function of each particle.
2. Update individual and global best fitness and positions.
3. Update velocity and position of each particle.

4.2 Feature Subset Selection Using PSO

For Feature Subset Selection using PSO, each particle in the swarm can be represented as a candidate feature subset. Accuracy of the learning algorithm is defined by the fitness function. Binary string representation of the position vector of a particle urges to modify the traditional PSO operations accordingly [4]. The modified operations for PSO can be given as follows [4]:

Let $x_{ia}(t)$ be the a th element of $X_i(t)$. Then $x_{ia}(t)$ sets its value to 0 and 1 iteratively as follows:

$$v_{ia}^{(t)} = v_{ia}^{(t-1)} + c_1 (p_{ia} - x_{ia}^{(t-1)}) + c_2 (p_{ga} - x_{ia}^{(t-1)}) \quad (3)$$

$$s(v_{ia}^{(t)}) = \frac{1}{1 + \exp(-v_{ia}^{(t)})} \quad (4)$$

$$\text{if } \rho_{ia} < s(v_{ia}^{(t)}), \text{ then } x_{ia}^{(t)} = 1, \text{ else } x_{ia}^{(t)} = 0 \quad (5)$$

, where $v_{ia}^{(t)}$, p_{ia} , p_{ga} are the a th elements of $V_i^{(t)}$, P_i , P_g ; $s(v_{ia}^{(t)})$ is the sigmoid function to normalize the value of $v_{ia}^{(t)}$ into [0.0,1.0], and ρ_{ia} is a vector which takes values

between 0.0 to 1.0 randomly. The velocity parameter $v_{ia}^{(t)}$ affects the choice of selection. If $v_{ia}^{(t)}$ is high ($s(v_{ia})$ is close to 1.0), the individual $x_{ia}^{(t)}$ tends to choose 1, and if its low it tends to choose 0. c_1 and c_2 can be set to 2.[4].

4.3 Minimum Redundancy Maximum Relevance (mRMR)

mRMR is a filter based feature selection method, which depends on mutual information of the features and is proved to have shown good results for various other fields[5], as well as in hierarchical approaches[6]. The mutual information quantifies the mutual dependencies between two random variables. If X and Y are two discrete random variables then their mutual information is defined as [4]:

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left[\frac{p(x, y)}{p(x)p(y)} \right] \quad (6)$$

, where $p(x, y)$ is the joint probability distribution function of X and Y , and marginal probability distribution functions of X and Y are $p(x)$ and $p(y)$ [4]. Hence the mutual information quantifies the information shared by X and Y by measuring the reduction in uncertainty of one variable when the other variable is concentrated upon.

In mRMR technique the relevance of a feature for the class is estimated from the mutual information between a feature and class. If there exist data D for a speech signal that consist of N features y_1, y_2, \dots, y_N and a class c such as angry, sad, etc, then the relevance of feature y_n ($1 \leq n \leq N$) is defined by the mutual information between feature y_n and the class c [4]:

$$relevance(y_n) = I(y_n; c) \quad (7)$$

The above eq(7) reflects the dependency of feature y_n on class c and the largest dependency on the class is by the feature y_i with the largest relevance. The mRMR technique redundancy of each feature is also considered by the mutual information of them. [4] Let y_n be a feature in a feature subset S then its redundancy is defined by following equation:

$$redundancy(y_n) = \frac{1}{|S|-1} \sum_{y_i \in S, y_i \neq y_n} I(y_i; y_n) \quad (8)$$

The above eq(8) shows the redundancy between 2 features. For example, the class discriminative power would not change much if one of feature were removed, when two features are highly dependent on each other and thus have high redundancy. Mutually exclusive features can be obtained by finding features with low redundancy.

Combining the above two notions a technique called as Minimal-Redundancy and Maximal-Relevance (mRMR) can be obtained. A highly dependent feature for a class is the one with lower redundancy and high relevance and hence mutually exclusive to other features. The mRMR measure of feature y_n is defined as [4]:

$$mRMR(y_n) = relevance(y_n) - redundancy(y_n) \quad (9)$$

First the mRMR values of each feature is calculated and represented in $D = y_1, y_2, \dots, y_N$ as m_1, m_2, \dots, m_N . Then the average (m) and the standard deviation (s) of the values

are calculated. Also the data D is divided into three sets which satisfy

$$D = D1 \cup D2 \cup D3 \quad (10)$$

, to separate the features into three different categories as $D1$ comprises of features with high mRMR, lower mRMR values in $D3$, and the rest in $D2$ by choosing the threshold as half of standard deviation [5].

4.4 PSO+mRMR

Combining PSO and mRMR reduces the feature subset subjected to GMM and eliminates the drawbacks of other techniques mentioned before. It reduces the complexity of mRMR and aims at selecting features with high significance and eliminates features with low significance. It is obtained by just altering eq(5) [4]. In the combined algorithm, the relevance of a feature is checked to see in a subset $D3$ and $D1$ and the values are set to 1 or 0 correspondingly.

PSO+mRMR technique have been used for feature selection in various applications such as Audiology, Dermatology, Image Segmentation and many others. Hence the technique can be applied for the feature selection of speech with GMM to determine Human Emotion.

5. BENEFITS OF PROPOSED SYSTEM

Particle Swarm Optimization is based on swarm intelligence hence can be applied to a variety of applications.

1. PSO requires no knowledge of the fitness function for optimization, unlike other optimization techniques which expect gradient information of the function.
2. Since the optimal solution of the system is followed by all the other particles, it seems to be attractive due to its ease of computation unlike PCA (Principle Component Analysis) which requires higher extraction effort and is exhaustive.

Minimum Redundancy and Maximum Relevance technique helps in filtering of features and extracting the ones with high significance and hence can be integrated with PSO to give a simpler process and better results thereby reducing the complexity of mRMR alone. mRMR is a filter based approach and hence classifier independent unlike wrapper methods such as Forward and Backward Selection. Therefore it gives almost same result when used with SVM or GMM.

AUC_S has the following disadvantages:

1. If ROC curves cross then AUC can give misleading results.
2. It is fundamentally incoherent in terms of misclassification costs.

Many Hybrid approaches do exist, but makes the system overly complicated whereas PSO and mRMR can be implemented with ease. Hence a simple and integrated technique can be applied for the feature selection of speech.

PSO+mRMR for feature subset selection gives good results for various fields[4] and since mRMR is classifier

independent, PSO+mRMR can be used for Speech Feature Subset Selection which is subjected to GMM to resolve all problems. Selecting only the optimal features reduces the work load on GMM model for classifying the states. The recognition process is simplified for the GMM, by selecting just the sufficient characteristics. Adding on, it also eliminates the complexity of mRMR alone by integration with PSO.

Hence the proposed system is extremely efficient.

6. CONCLUSION

This study shows the use of PSO+mRMR algorithm for optimizing Feature subset, for the recognition of the emotional states of various speakers by GMM. The technique selects the optimal subset thereby reducing the load on GMM for recognition and also eliminates the complexity incurred by mRMR alone, by integrating with PSO.

PSO has variety of applications like Image processing, Biometrics, Neural networks, Movies, Market Segmentation, Multi objective optimization, etc. mRMR is used widely for feature selection. Hence both of them make a good combination for feature selection as proposed in this paper.

Further research can be conducted on other variants of PSO to extend the field searching space and on new classifiers. A modified Gaussian model such as booster Gaussian models or the use of EM algorithm for clustering of the emotion states along with GMM is of great interest for further research.

7. ACKNOWLEDGMENTS

Our thanks to our professor S.Rajarajeswari for her guidance and encouragement without which this would not have been possible.

8. REFERENCES

- [1] Mohammad Khadem Safdarkhani, Shokooh Peymani Mojaver, Saeid Atieghchechi, Alireza Molanoori, Masume Sadat Riahi, 2012. Emotion Recognition of Speech Using ANN and GMM, Australian Journal of Basic and Applied Sciences, 6(9):45-57, 2012 ISSN 1991-8178.
- [2] Nasim Chagha Ghasemi, Khosro Rezaee, 2012. "The Reduction of Speech Characteristic Vector Using PSO Algorithm and the Evaluation of the Effectiveness of Different Speech Characteristic in Recognition of Persian Language State", Journal of Academic and Applied Studies, Vol. 2(3) March 2012, pp. 32-42.
- [3] Yasunari Yoshitomi, Sung-Il Kim, Takako Kawano and Tetsuro Kitazoe. Audio-visual feature selection and reduction for emotion classification, unpublished.
- [4] Chulmin Yun, Byonghwa Oh, Jihoon Yang and Jongho Nang, 2011, Feature Subset Selection Based on Bio-Inspired Algorithms, JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 27, 1667-1686 (2011).
- [5] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max dependency, max-relevance, and min-redundancy", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, 2005, pp. 1226-1238.
- [6] Panagiotis Giannoulis, Gerasimos Potamianos. A Hierarchical Approach with Feature Selection for Emotion Recognition from Speech, unpublished.
- [7] Bjorn Schuller, Stephan Reiter, Ronald Muller, Marc Al-Hames, Manfred Lang, Gerhard Rigoll. "Speaker Independent Speech Emotion Recognition By Ensemble Classification", 0-7803-9332-5/05/20.00 © 2005 IEEE.
- [8] Ashish B. Ingale, Dr. D. S. Chaudhari. "Speech Emotion Recognition Using Hidden Markov Model and Support Vector Machine", International Journal of Advanced Engineering Research and Studies E-ISSN 2249-8974.
- [9] Iker Luengo, Eva Navas, Inmaculada Hernáez, Jon Sánchez. "Automatic Emotion Recognition using Prosodic Parameters". INTERSPEECH 2005.
- [10] Daniel Neiberg, Kjell Elenius and Kornel Laskowski. "Emotion Recognition in Spontaneous Speech Using GMMs". unpublished.
- [11] Ioan Cristian Trelea. "The particle swarm optimization algorithm: convergence analysis and parameter selection". Elsevier, 12th September 2002.
- [12] F Van Den Bergh. An Analysis of Particle Swarm Optimizers. PhD thesis, Faculty of Natural and Agricultural Science, University of Pretoria, Vol. 200, November 2001.