Hierarchical Emotion Classification using Genetic Algorithms

Ba-Vui Le
Computer Engineering Department, Kuyng Hee University
01 Seocheon-dong, Yongin-si, Giheung-gu
Gyeonggi-do, South Korea
lebavui@oslab.khu.ac.kr

Jae Hun Bang
Computer Engineering Department, Kuyng Hee University
01 Seocheon-dong, Yongin-si, Giheung-gu
Gyeonggi-do, South Korea
jhb@oslab.khu.ac.kr

Sungyoung Lee
Computer Engineering Department, Kuyng Hee University
01 Seocheon-dong, Yongin-si, Giheung-gu
Gyeonggi-do, South Korea
sylee@oslab.khu.ac.kr

ABSTRACT

Emotion classification from speech signal is an interesting subject of machine learning applications that can provide the emotional or psychological states from speakers. This implicit information is helpful for machine to understand human behavior in more comprehensive way. Many feature extraction and classification methods have being proposed to find the most accurate and efficient method, but this is still an open question for researchers. In this paper, we propose a novel method to select features and classify emotions in hierarchical way using genetic algorithm and support vector machine classifiers in order to find the most accurate binary classification tree. We show the efficiency and robustness of our method by applying and analyzing on Berlin dataset of emotional speech and the experiment results show that our method achieves high accuracy and efficiency.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; I.5.4 [**Pattern Recognition**]: Applications

General Terms

Algorithms

Keywords

emotion classification, emotional speech, support vector machine, genetic algorithm, hierarchical classification

1. INTRODUCTION

There is no exact definition of emotions from literature. However, we can intuitively imagine that emotions are human interactions with changes of environment [9]. Emotions can be represented by different ways as people want to express their psychological status such as emotional speech, facial expression, body gestures, or physiological signals of body. Research of emotion is not new, many works about emotion have been done so far. However, application of emotion awareness in human computer interaction is a new approach. By understanding emotional states of users, intelligent machine can interact in a more natural and interesting way [4]. The interaction between machine and users can be changed naturally depending on psychological states of users.

Emotional speech is one of different ways people use to express their psychological states [12]. In face to face conversation, emotional speech, facial expression and body gestures convey mostly implicit information of speakers to listeners. However, in some situations such as communication with blind people or call center system, speech is the only way that can be used to convey human messages. In the general framework of a supervised machine learning technique, there are two major parts of feature extraction and classification. When feature extraction step provides the most relevant features for a particular task of recognition, classification step uses these features to formulate models and to recognize unknown input. In emotional speech recognition, a lot of algorithms for feature extraction and classification have being proposed, but there is no consistent conclusion on what the most relevant features and classifiers for emotion classification from speech signal are [11].

In recent decade, emotion recognition interests many researchers in finding new methods as well as applying state-of-the-art speech recognition methods of feature extraction. As a result, many features are proposed to classify emotional speech such as pitch-related, formants, energy-related, duration, voice quality, linear predictor cepstral coefficients (LPCC), mel-frequency cepstral coefficients (MFCC), etc [5] [1]. Similarly, various classification methods are utilized, most popular classifiers are artificial neural network (ANN), hidden markov model (HMM), Gaussian mixture model (GMM), support vector machine (SVM). Different feature selection techniques are also applied to select the most relevant features so that the accuracy level can be improved, common methods are principal component analysis (PCA), linear discriminant analysis (LDA), sequential floating feature selection (SFFS), mutual information based, etc. The selection of appropriate methods strongly depends on

SoICT'13 December 05 - 06 2013, Danang, Viet Nam

http://dx.doi.org/10.1145/2542050.2542075

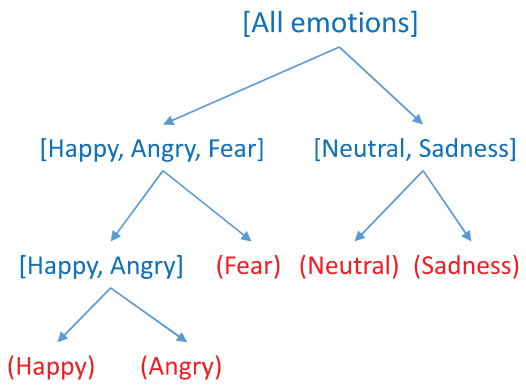

Figure 1: An example of binary classification tree with dataset of 5 emotions.

different kind of application, input dataset, as well as different emotions.

Several emotion research works tried to separate the original complex multiple emotion classification problem by applying hierarchical approaches with combination of different classifiers [14, 8, 7]. Instead of solving the complex problem of multiple emotions classification, hierarchical approach tries to solve multiple simpler binary classifications of two classes. Different feature extraction, feature selection, classifier algorithms can be applied to a binary classification. Most of existing hierarchical approaches manually predefined the structure of a classification tree based on the preliminary knowledge of input dataset and emotions.

In this paper we propose a method that is able to flexibly generate a binary classification tree using genetic algorithm and support vector classifiers for a given dataset. The method can work well with different datasets by generating different structures of the classification tree.

This paper is structured as follows: Section II, we described our method in details. It is subdivided in 3 subsections: first, we briefly introduce support vector classifier which is used as the key classifier in our method. The second subsection presents the details about the training stage used to generate binary classification tree. The third subsection describes about classification stage. In section III, we present experiment results of the proposed method and the discussion. Finally, Section IV concludes the paper.

2. PROPOSED METHOD

In this section, we described in details our method of feature selection and classification that utilize genetic algorithm (GA) and support vector machine (SVM) classifier to find the binary classification tree in order to achieve the most accurate classification results.

One example of a binary classification tree is shown in Figure 1, where the root of this tree is the original group of all emotions, each node is a smaller group of two or more emotions, and each leaf is an individual emotion class. Each node of classification tree is considered as one classification stage where classifying of the input group of emotions into two further groups, is performed. If classified group contains only one emotion, it will be assigned as a leaf of tree, otherwise it will be used as the input group for the next classification stage. In order to classify an unknown input sample, we follow the classification tree until it reaches to a leaf that is considered as the output label of a given sample. We can apply different feature selection for a particular classifier at each node to increase the accuracy level of this classification stage in order to improve the accuracy of binary classification.

Depending on different datasets and different emotion labels, the binary classification tree can have different structures, they will be constructed in order to get the most accurate and efficient classification result.

Therefore, the main problem is how to generate the binary classification tree automatically without any preliminary knowledge about input dataset or user interaction. And then this method of generating classification can be applied with different systems and different datasets.

In this paper, we propose to use GA and SVM classifier to generate the binary classification tree for given dataset. GA will be used to search the optimal solution of separation between groups and SVM will be used for the searching criteria. As a well-aware algorithm, two major requirements of GA is how to represent its chromosomes or solutions for the input problem by an array of number that can be a bit string or real numbers. The second requirement is how to measure the fitness of a particular solution in comparison with others so that GA can assign what the good and the bad solutions for next generations are.

2.1 Support Vector Machine Classifier

In recent researches about speech emotion classification, researchers proposed many classification algorithms, such ANN, GMM, HMM, SVM, etc. Each of algorithm has particular uniqueness and limitation depends on different characteristics of features and input dataset. SVM [2] is a powerful classifier that maps the feature vector to a higher dimensional vector space, and then establishes a maximum interval hyper plane in this space. SVM will find two parallel hyper planes on both sides of the hyper plane and establish a suitable direction hyper plane to make the maximum distance between the two hyper planes. A kernel function is used to map the original input feature vector to a higher dimensional space and then obtain an optimal classification. Since SVM is a simple and efficient computation classifier especially for two classes problem, and under the conditions of limited training data, it can have a very good classification performance compared to other classifiers. Thus we applied SVM to classify the speech emotion in this paper.

2.2 Genetic Algorithms for Searching Optimal Solution

GA is an optimization algorithm that mimics the natural evolution to find the most optimal solution for a particular problem by evaluating the population of solutions in parallel through each generation [13]. New population of solutions generated by crossover and mutation tends to find and keep the better solution for next generation. Basically GA is an iteration algorithm that iterates the generation and evaluation of solutions until the predefined stop condition is satisfied. The algorithm is stopped when it reaches to the number of created generations or there is no significant change in the fitness value. In GAs, there are two requirements need to be considered in order to apply into a particular problem. The first requirement is how to represent the output of problem

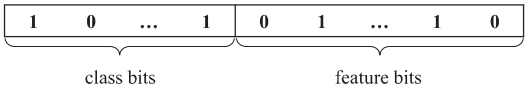

Figure 2: Chromosome structure consists of class bits and feature bits.

Table 1: Confusion matrix of binary classification

	Left node	Right node
Left node	**tp** (true positive)	**fn** (false negative)
Right node	**fp** (false positive)	**tn** (true negative)

by chromosome and the other is how to evaluate candidate solutions.

2.2.1 *Chromosome Representation*

GA is exploited in order to achieve two missions at each node of binary classification tree: how to separate the input classes into two groups and find the feature subset so that these two groups are the most separable ones. In other words, when we apply SVM classifier on these selected features to classify two candidate groups, it will returns the best accurate result in comparison with all available cases. Based on this point, each chromosome represented by a bit string will have two parts, one part represents the separation between classes and the remaining part represents the selected features.

The first mission is how to split input emotions into two groups that have maximum separation or tend to archive the best classification accuracy. We assign these groups as left branch and right branch with bit 0 and bit 1 respectively. So that emotion has class bit 0 will belong to the left branch and emotion has class bit 1 will belong to the right branch.

The second mission of GA at each node is how to select features so that the classification between two separated groups has the largest accuracy level. We simply use feature bits of a chromosome as a mask so that the feature with bit 1 will be selected and the feature with bit 0 will not be selected for the classification.

2.2.2 *Fitness Function*

Other important requirement of GA is how to measure the efficiency of solutions or chromosomes so that the useful chromosome will be kept to the next generation. In our method, we utilize SVM classifier to measure the efficiency of classes separation and feature selection.

In general, the efficiency of a classification is measured by the confusion matrix as described in Table 1 showing the correct and incorrect classification between two nodes. We use average accuracy as the measurement for individual classifier as follows:

$$average\,accuracy = \frac{1}{2}\left(\frac{tp}{tp+fn} + \frac{tn}{tn+fp}\right) \quad (1)$$

where tp is true positive and tn is true negative that describe the correct classification, while fp is false positive and fn is false negative that describe the incorrect classification. Given a solution of class bits and feature bits, we generate the classification problem to classify between the left group and the right group with selected features. To increase the reliability of training stage, we apply k-fold cross-validation and take the average accuracy after the validation as the fitness function of GA.

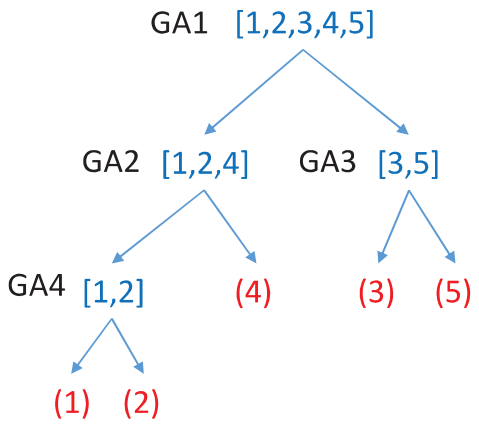

Figure 3: Breadth-firth search strategy is applied to generate binary classification tree

$$f = \frac{1}{k}\sum average\,accuracy \quad (2)$$

The ultimate objective of GA is finding the optimal solution that has the maximum fitness function value in comparison with other solutions.

2.3 GA Queue

Once a particular GA is processed completely, input classes are separated into the left group and the right group. If number of emotion in one group is equal or greater than two, this group will be processed further to generate two other groups. The similar sequence of GA problem will be applied repetitively for every nodes that have number of classes equal or greater than two until all emotion classes are assigned to leaves of the classification tree. In the example described in Figure 1, to construct this tree, we have to generate 4 classifiers:

(Happy, Angry, Fear) vs. (Neutral, Sadness)
(Happy, Angry) vs. (Fear)
(Happy) vs. (Angry)
(Neutral) vs. (Sadness)

Once a GA process is constructed, the next problem is how to manage generated classifiers and perform the respective GA automatically. We apply a simple method that is similar breadth-first search technique. From the root of tree, each tier of tree is considered in order to generate next tier if a node has 2 or more emotion classes. We use a queue to store generated GA problems, so that first GA put into the queue will be processed first.

Figure 3 and Figure 4 describe one example of proposed method taking input from 5 emotion classes and generate the binary classification tree. A GA queue is initialized by putting the GA1 with input of all classes in the queue, after GA1 process completes, it generates 2 groups of (1, 2, 4) and (3, 5). Because both groups have more than 1 class so that two GA2 and GA3 are generated and put into the

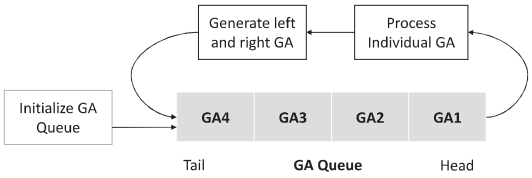

Figure 4: GA queue is used to store generate GA problems

queue. Because GA2 is added first so it is performed before GA3. Once GA2 is performed, it generates GA4 and a leaf that contains emotion label number 4. And then GA3 and GA4 are performed until the queue is empty so that all classes are assigned to leaves completely and the binary tree is constructed successfully.

Algorithm 1 Algorithm generates binary classification tree

Require: Initialize GA *queue*, put the first GA into *queue*
while *queue* is not empty **do**
  take one GA from *queue*
  perform GA process until satisfy its condition stop
  assign input classes to left and right group using *classbit* respectively
  if number classes in left group > 1 **then**
    generate new GA
    add new GA to *queue*
  end if
  if number classes in right group > 1 **then**
    generate new GA
    add new GA to *queue*
  end if
end while
generate binary classification tree
return *classification model*

Algorithm 2 Algorithm perform individual GA problem

Require: Initialize GA parameters
while *true* **do**
  assign input classes to 2 groups with respective *class bit*
  select features that has *feature bit* 1
  train SVM classifier to classify 2 groups with selected features
  estimate the fitness function
  if fitness value satisfies stop condition **then**
    exit
  end if
  reproduce new population by performing crossover and mutation
end while
return *optimal solution*

Algorithm 1 and 2 are used to construct binary classification where algorithm 1 continuously generates and performs GA processes until all input classes are assigned to leaves and algorithm 2 performs standard a GA process with the evaluation of fitness function is estimated by training and testing with the SVM classifier for a candidate solution.

2.4 Classifying stage

Once the optimal binary classification tree is generated for a particular input dataset, the classifying of an unknown input sample is performed easily following the structure of classification tree. The output emotion label is generated by performing multiple binary SVM classifiers from the root to a leaf of the constructed classification tree.

3. EXPERIMENT AND RESULTS

3.1 Emotion Dataset

In the experiments, we evaluate the efficiency of proposed method in term of accuracy using the Berlin Dataset of Emotional Speech (EMO-DB) [3]. This dataset is one of the most popular datasets used for emotion recognition because of public availability, and usually used for comparison between existing works. Ten actors including of 5 males and 5 females were asked to speak 10 everyday sentences (five short and five long) in German, they can be categorized in all of seven acted emotions. The dataset was evaluated by a subjective perception test with 20 listeners to produce 535 sentences in total. The numbers of speech files for the seven emotion categories are: anger (127), boredom (81), disgust (46), fear (69), joy (71), neutral (79) and sadness (62). We divide the dataset in two parts randomly, one part is used for training to generate classification model and remaining part is used for evaluation. Number of sample for each part is described as in Table 2.

Table 2: Berlin Dataset of Emotional Speech (EMO-DB)

Emotion	Training	Testing	Total
Neutral (**N**)	63	16	79
Anger (**A**)	102	25	127
Fear (**F**)	55	14	69
Happiness (**H**)	57	14	71
Sadness (**S**)	50	12	62
Disgust (**D**)	36	10	46
Boredom (**B**)	65	16	81
Total	428	107	535

3.2 Feature Extraction

We use the acoustic feature set largely based on the findings by [10], these features are extracted using the OpenSmile toolbox [6]. The feature set includes 16 low level descriptors consisting of prosodic, spectral envelope, and voice quality features. These low level descriptors are zero crossing rate, root mean square energy, pitch, harmonics-to-noise ratio, and 12 mel-frequency cepstral coefficients and their deltas. And then 12 statistical functions were computed for every low level descriptor per utterance: mean, standard deviation, kurtosis, skewness, minimum, maximum, relative position, range, two linear regression coefficients, and their respective mean square error. This results in a collection of 384 acoustic features.

3.3 Experiment result

After training stage, the binary classification tree is generated automatically as in Figure 5 where each node is corresponded to a classifier that classifies between the left branch

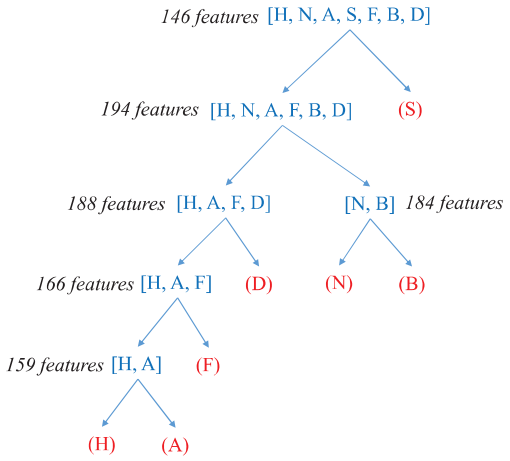

Figure 5: Generated binary classification tree with dataset of 7 emotions.

and the right branch. Each leaf is corresponded to an emotion class. The number of selected features with corresponding classifier are presented in the figure also.

The classification tree consists of 6 classifiers. From the top of this tree, the less confusion cases are taken first. Two classifiers are generated to classify Sadness, Neutral and Boredom. After that, Disgust and Fear are classified, and at the bottom Happy and Angry, two most confused emotions, are classified. These separations between emotion classes are reasonable in term of activation, potency, and valance values. At the first stage, Sadness is classified with remaining emotions since it has lest activation and valance. After that, Happiness, Angry, Fear, and Disgust are classified with Neutral and Boredom due to different activation level. And then Happiness and Angry are classified with Disgust and Fear because of potency level.

Table 3 presents evaluation results when using testing data for each classifier.

We can see that the most misclassification cases happen with group has only one emotion class such as Sadness with remaining emotions, or Disgust with Happiness, Angry, and Fear because of unbalance in the number of samples between two group. This is one of further research challenges needed to improve in future works.

Since the binary classification tree is generated automatically, it may have different structure when applying with different separation of training and testing dataset.

4. CONCLUSIONS

In this paper, we proposed a method to automatically generate hierarchical binary classification tree for multiple emotions classification from speech signal. We applied genetic algorithm in combination with support vector classifier to find the optimal separation between two groups of emotions at each node of classification tree and select the most relevant features so that the classification result achieves the best accuracy. Our method can work well with various datasets with different emotions because of the adaptation and robustness of algorithms. Once the binary classification tree is generated completely, it can be applied in different applications and environments with high efficient and accuracy level.

Table 3: Classification result at each node of binary decision tree

(H, N, A, F, B, D) vs. (S)

	(H, N, A, F, B, D)	(S)
(H, N, A, F, B, D)	**96.84%**	3.16%
(S)	25%	**75%**

(H, A, F, D) vs. (N, B)

	(H, A, F, D)	(N, B)
(H, A, F, D)	**96.83%**	3.17%
(N, B)	9.38%	**90.63%**

(H, A, F) vs. (D)

	(H, A, F)	(D)
(H, A, F)	**100%**	0%
(D)	30%	**70%**

(N) vs. (B)

	(N)	(B)
(N)	**100%**	0%
(B)	6.25%	**93.75%**

(H, A) vs. (F)

	(H, A)	(F)
(H, A)	**97.44%**	2.56%
(F)	21.43%	**78.57%**

(H) vs. (A)

	(H)	(A)
(H)	**85.71%**	14.29%
(A)	16%	**84%**

5. ACKNOWLEDGMENTS

This research was supported by the MSIP(Ministry of Science, ICT & Future Planning), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency)" (NIPA-2013-(H0301-13-2001))

References

[1] A. Batliner, S. Steidl, B. Schuller, D. Seppi, T. Vogt, J. Wagner, L. Devillers, L. Vidrascu, V. Aharonson, L. Kessous, and N. Amir. Whodunnit - searching for the most important feature types signalling emotion-related user states in speech. *Computer Speech and Language*, 25(1):4–28, 2011.

[2] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.

[3] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, and B. Weiss. A database of german emotional speech. In *in Proceedings of Interspeech, Lissabon*, pages 1517–1520, 2005.

[4] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. Taylor. Emotion recognition in human-computer interaction. *Signal Processing Magazine, IEEE*, 18(1):32–80, 2001.

[5] M. El Ayadi, M. S. Kamel, and F. Karray. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recogn.*, 44(3):572–587, Mar. 2011.

[6] F. Eyben, M. Wöllmer, and B. Schuller. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the international conference on Multimedia*, MM '10, pages 1459–1462, New York, NY, USA, 2010. ACM.

[7] C.-C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan. Emotion recognition using a hierarchical binary decision tree approach. *Speech Commun.*, 53(9-10):1162–1171, Nov. 2011.

[8] Q.-R. Mao and Y.-Z. Zhan. A novel hierarchical speech emotion recognition method based on improved ddagsvm. *Comput. Sci. Inf. Syst.*, 7(1):211–222, 2010.

[9] K. Scherer. What are emotions? and how can they be measured? *Social Science Information*, Jan. 2005.

[10] B. Schuller, A. Batliner, D. Seppi, S. Steidl, T. Vogt, J. Wagner, L. Devillers, L. Vidrascu, N. Amir, L. Kessous, and V. Aharonson. The relevance of feature type for the automatic classification of emotional user states: Low level descriptors and functionals. In *INTERSPEECH*, pages 2253–2256. ISCA.

[11] B. Schuller, A. Batliner, S. Steidl, and D. Seppi. Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Communication*, 53(9-10):1062–1087, 2011.

[12] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. A. MÃijller, and S. Narayanan. Paralinguistics in speech and language - state-of-the-art and the challenge. *Computer Speech & Language*, 27(1):4–39, 2013.

[13] D. Whitley. A genetic algorithm tutorial. *Statistics and Computing*, 4:65–85, 1993.

[14] Z. Xiao, E. Dellandrea, W. Dou, and L. Chen. Automatic hierarchical classification of emotional speech. In *Multimedia Workshops, 2007. ISMW '07. Ninth IEEE International Symposium on*, pages 291–296, 2007.