Development of Interactive Feature Selection Algorithm(IFS) for Emotion Recognition

Hyun-Chang Yang, Ho-Duck Kim, Chang-Hyun Park, and Kwee-Bo Sim*

School of Electrical and Electronic Engineering, Chung-Ang University
221, Heukseok-Dong, Dongjak-Gu, Seoul, 156-756, Korea

Abstract

This paper presents an original feature selection method for Emotion Recognition which includes many original elements. Feature selection has some merits regarding pattern recognition performance. Thus, we developed a method called the 'Interactive Feature Selection' and the results (selected features) of the IFS were applied to an emotion recognition system (ERS), which was also implemented in this research. The innovative feature selection method was based on a Reinforcement Learning Algorithm and since it required responses from human users, it was denoted an 'Interactive Feature Selection'. By performing an IFS, we were able to obtain three top features and apply them to the ERS. Comparing those results from a random selection and Sequential Forward Selection (SFS) and Genetic Algorithm Feature Selection (GAFS), we verified that the top three features were better than the randomly selected feature set.

Key words : Reinforcement Learning, Feature selection, IFS, Emotion, Recognition, SFS, GAFS

1. Introduction

The emotion recognition research has been typically attempted using four kinds of medium. They are speech, image, physiological signal, and gesture. In addition, our IEEE survey papers published from 1990 to 2005 show that papers using the speech medium have been published more often than others have. The reason for this result is probably due to feature set extraction from speech and image being easier than physiological signal or gesture and the possibility of classification is higher. In particular, EEG, ECG, and SC sensors are used to obtain a physiological signal but the signal from those sensors may be obstructed by electrical signals from fluorescent lamps or electric home appliances. This problem is the one obstacle in emotion recognition using a physiological signal. For an image, this means facial expression recognition and the main problem in this case is usually lighting conditions, which often change, or personal accessories like glasses which affect recognition performance. A problem of gesture recognition is similar to that of image recognition and the bigger problem is that it may not include much information regarding emotion. Apart from above the problems which these three media present, speech signal can send much more information regarding emotion. For example, talking over the telephone, one can recognize emotions and this shows the validity of speech

signal for emotion recognition. Even a cheap microphone can be used sufficiently as a sensor for collecting speech signals and noise will not affect the extraction of the feature set unless it is too loud to be classified as a coming from the source of the signal. These reasons are why most researchers have focused on speech signal and why we have selected this medium for our paper. The commonly used feature set for emotion recognition from speech consists of pitch, energy, formant, and speech rate. Some researchers select all four of the feature sets, others select only one, and the features are generally extracted statistically from the four feature sets. In [1], 87 features were extracted from pitch, energy, and formant and they were classified into five emotions. In [2], 17 features were extracted from pitch, energy, speech rate and so on with sex also being classified. In addition, In [3], 11, 40, and 13 features were extracted. The fact that feature set selection is not fixed suggests that features may or may not be relevant to emotion recognition. This problem will plague researchers in this field until exceptional results are obtained. For this case, there is a GA based selection method, Floating search method and so on which can somewhat reduce difficulties for researchers [4]. Especially, the Sequential Forward Selection and Sequential Backward Selection methods of a Floating search method have been frequently used. In [2], a Forward Selection (FS) method was used and in [1], the 10 best features were selected out of 87 features by using a Sequential Floating Forward Selection algorithm (The extended version of SFS). In [5], a Sequential Forward Selection algorithm was also used and the best feature subset was selected out of 39 candidate feature sets and In [6], a good feature set was found using genetic programming for the music genre classification related

Manuscript received Oct. 9, 2006; revised Dec. 14, 2006.

* Corresponding author

This research was supported by the Chung-Ang University Research Grants in 2006

problem. These feature selection methods provided a good solution for “The curse of dimensionality” and contributed to the performance of pattern recognitions. In addition, feature selection methods included supervised and unsupervised cases. Generally, a supervised case is employed more often than an unsupervised case. This is due to unsupervised feature selection methods having a high probability of incorrect results for corresponding patterns regarding perceived speech [7]. Although, there are many cases that cannot obtain an explicit supervised value, the unsupervised method has advantages. We propose a method using reinforcement learning taking advantage of both the supervised and unsupervised method, which can alleviate the shortcomings of both methods. Researches of the reinforcement learning have been proceeded using many methods, i.e. Dynamic programming, Monte Carlo method, TD method, Q learning etc. proposed by Sutton and Barto. Since there is such a variety of methods and the main elements such as “state”, “action” and “reward” may be freely defined and implemented by a developer, this method is thought to be a very important one for machine learning techniques[9]. In this study, rather than using a specific reinforcement learning method, we propose a method which selects feature sets by calculating rewards received when an action is performed in a state. In particular, this method does not only calculate the frequency of emotion transit but also the sum of the rewards for the evaluation of a feature selection. Therefore, this method has the advantage that the more frequently it contacts a user, the better its performance becomes. The outline of the paper is as follows, In Section II, it explains the emotion recognition method and Section III explains the proposed algorithm. The Section IV shows a simulation and result of using the proposed algorithm. Section V concludes and shows future works.

2. Emotion Recognition Method

This paper addresses emotion recognition by extracting features of speech. The emotion recognition with speech is largely divided into cases using acoustic information and language or discourse information. The former is a method that uses some feature sets such as pitch, formant, speech rate, timbre, etc. and the latter uses the meaning of a word. That is, whether the word is positive or negative to whether it represents a happy or sad state. The process of emotion recognition consists of collecting emotional speech, the acoustic analysis, implementing DB, feature set extraction and such features are trained and classified with emotions using a pattern classification method.

2.1 Database and Preparation

Emotional speeches were collected from 10 male graduate students. Their ages ranged from 24 to 31 years old and they were asked to say with 10 short sentences emotionally. The choice of the 10 sentences (scripts) was decided upon from the result of another survey or experiment. In the first stage, 30 sentences had been prepared and all the 30 sentences were asked to say by the subjects. After the recording, the speeches were listened to by other people and they were asked the question “What emotion do you feel when listening to the given recording?” The emotions conveyed in the 10 sentences that were read and the answers given by the subjects in this experiment were in agreement 90% of the time. In addition, the length of the prepared sentences was limited from 6 to 10 syllables. The recording format was 11Khz, 16bit, mono and the subjects were asked to keep a distance of 10 cm between themselves and the microphone. Since the distance of the microphone affected loudness or intensity, maintaining the required distance was very important. Recorded files were preprocessed and stored in DB(MS-ACCESS). In the preprocessing stage, there were several processes to signals such as FFT (extracting spectrum), Pitch extraction (by an autocorrelation method), IR(Increasing Rate) of pitch, CR(Crossing Rate), VR(Variance), and statistical values etc[10]. Fig. 1 shows emotion recognition system we implemented, which has the function of recording, preprocessing as uttered above and pattern recognition.

Fig. 1. Emotion recognition system

Table 1. Parameters of ANN

Parameter	Values
Input Units	3~5
Hidden Units	11
Output Units	2
Learning Rate	0.003
Tolerance	0.25
Sigmoid Function	$1 / (1 + e^{-3x})$

2.2 Pattern classification method

We used an artificial neural network for pattern classification, which commonly performs well and is robust to a signal with noise. It has been the most popular method to use in the pattern recognition field. This method commonly uses a Back Propagation Algorithm for tuning network parameters. In this study, we fixed the setting to ANN as follows. The number of Input Units and Hidden Units and Output Units and Learning rate and Tolerance and Sigmoid function are 3~5 and 11 and 2 and 0.003 and 0.25 and $1/(1 + e^{-3x})$ respectively as shown in Table 1.

3. The Interactive Feature Algorithm

Typically, researchers in the emotion recognition use various feature sets. Some researchers looked into the relation between acoustic analysis and emotion and used the feature sets based on that relation. However, because this method is subjective, it may easily lead to local minima. For this reason, recent studies consider a feature selection method for finding small superior features (4~10) out of as many as 30 to 90 features. Most researchers do not use all features because they cannot confirm whether they are valid or not and noises with every features may deteriorate. Therefore, feature selection methods are popular in the pattern classification field[7]. According to John et al.[8], there are two main approaches to FS: filtering and wrapping. In the former, a feature subset is selected independently of the learning method that will use it. In the latter, a feature subset is selected using an evaluation function based on the same learning algorithm that will consider this subset of features. Although wrapper approaches have been shown to perform better, they can be rather time-consuming and it is sometimes infeasible to use them[7]. For this reason, the proposed algorithm tries to combine the characteristics of both the wrapper and filtering.

3.1 Reinforcement Learning Algorithm.

Reinforcement learning consists of an agent and environment and is a learning method that leads the agent to perform a target action for a user. The process of learning is as follows, given an environment, an agent firstly performs an action and the agent receives a reward for the action from the environment. At that time, each time step is denoted as t , an environment state which the agent may include is denoted as $s_t \in S$ (S is a set of possible environments) and an action is denoted as $a_t \in A(s_t)$ ($A(s_t)$ is a set of possible actions in a state). A reward for an action is denoted as r_t and when an episode has been completed, the r_t is expressed as the following equation.

$$R_t = \sum_{k=0}^T \gamma^k r_{t+k} + k + 1 \quad (1)$$

Where γ is a discount coefficient in the above equation and does not make the sum of rewards an infinity in the case of being defined as $t = \infty$. In addition, if the discount coefficient is zero, it means that only the current reward value is admitted. That is, we can give the weight to a future value differently according to the discount coefficient. Finally, reinforcement learning is a method that determines a policy to maximize the equation 1.

3.2 Sequential Forward Selection (SFS) Algorithm.

Sequential Forward Selection is the simplest greedy search algorithm. In this paper, we will briefly explain this algorithm. The Fig. 2(a) shows the algorithm. Starting from the empty set, sequentially add the feature x^+ that results in the highest objective function $J(Y_k + x^+)$ when combined with the feature Y_k that has already been selected.

3.3 Genetic Algorithm Feature Selection Algorithm (GAFS)

The Genetic Algorithm is popular method for finding an optimized solution. This algorithm has also good performance to the problems like nonlinear problems, which are hard to be solved by using the classic optimization techniques. The problem we are treating is also a nonlinear problem and thus, we think this problem may be solved by Genetic Algorithm. So, we tried to search good feature set using the Simple Genetic Algorithm. The Fig. 2(b) is the algorithm of the GAFS.

Fig. 2. SFS and GAFS algorithm

The Fig. 3(a) and 3(b) are the performance graph of Genetic algorithm feature selection. Because this method is included in the part of wrapping, an evaluation function is needed. So, the function which is used for evaluation is as follows,

$$3x_0 - 2x_1 + 4x_2 + 10x_3 - 5x_4 + 8x_5 + 7x_6 + 8x_7 + 10x_8 + 6.8x_9 \quad (2)$$

In the Fig. 3(a), we can see the best fitness converged on the about 8 generations and the Fig. 3(b) shows the average fitness is being increased. This simulation verifies that the GAFS is working well.

(a) GAFS Best fitness

(b) GAFS Average fitness

Fig. 3. Performance graph of GAFS (Crossover rate: 0.5, Mutation rate: 0.02, Chromosome length: 10, Population No.: 100)

3.4 Interactive Feature Selection Algorithm (IFS)

The Interactive Feature Selection algorithm we are proposing is an algorithm based on reinforcement learning. Specially, popular algorithms such as SFS, SBS and so on, are deductive algorithms but our proposed algorithm is inductive. Also, these feature selection algorithms are based on the rationale of correlation and information-theoretic measures. Correlation is based on the rationale that good feature subsets contain features highly correlated with a class, yet are uncorrelated with each other. The IFS is also based on the correlation concept. Moreover, the feature selection algorithms consist of a search strategy and an evaluation by objective function but the conventional methods are incompetent in the search strategy part. Therefore, an IFS focuses on both the search strategy and evaluation by objective function. Fig. 4(a) shows an IFS process. We assume that an emotion recognition system that includes this algorithm will be applied to a home robot or appliance. If such a situation is considered, users may be comfortable inputting emotional speech and a user's emotional state at that time (as a supervisor value). Due to this characteristic, this algorithm is a user adaptive system that can efficiently solve a problem and the

more a user is in contact with this system, the better it will perform.

(a) IFS Algorithm

(b) IFS example

Fig. 4. IFS algorithm and an example

4. Experimental Results

We applied 11 original features to the IFS simulator; Pitch features (max, min, mean, median, crossing rate, increasing rate), Loudness; Intensity (max, min, mean), Sect. No and Speech rate. This program was made only for an IFS and the results from this program were applied to the Emotion Recognition System using ANN. That is, the feature set applied to ERS was the features previously outputted by IFS and then the emotion recognition experiment was performed. Classification was attempted using four methods. Results are shown in the Fig. 5(a) and 5(b). The Fig. 5(a) shows the fitness movement on every generation.

From the Fig. 5(a), we can see that the IFS system searched for better results and improved gradually. In the algorithm, because the searching work was performed again when the new evaluation result was worse than the previous one, there was some range in the steady state (The range of 3~5 and 6~8 in the Time axis). In addition, from Fig. 5(b), this graph compares four methods (IFS, SFS, GAFS and random selection) with the changed feature no. As expected, random selection performed poorly but IFS and SFS and GAFS similarly performed better. In the IFS and SFS and GAFS case, the results show a subtle distinction but the IFS with features 1, 2 and 3 was better. However, with features 4 and 5, SFS showed better results.

(a) Fitness movement for 3 methods

(b) Emotion Classification Rate Comparison

Fig. 5. Experiment Result

Table 2. Experiment Results

Feature No.	Features
1	Pitch mean
2	Pitch mean, Speech rate
3	Pitch mean, Speech rate, Loudness
4	Pitch mean, Speech rate, Loudness, Sect. No.
5	Pitch mean, Speech rate, Loudness, Sect. No., Intensity mean

5. Conclusions

This paper presents a solution to feature selection when there is an emotion recognition problem. The solution called the IFS performed as well as an SFS. In particular, it is reinforcement based learning and supplements the role of search strategy in the feature selection process. Using the IFS simulator, we found some of the best features and used them in the emotion recognition experiment and results were compared to those of SFS and Random selection. Performance was slightly better than SFS. However, IFS has some disadvantages. If the amount of training data is too small, selection results may be not good. SFS does not require much training data. It is also sufficient that training data be only one set. If an objective function is clear, SFS will be adequate. However, in the case of emotion recognition, SFS may not perform as well as it had. In this case, the correlation-based method like the IFS will be better.

References

- [1] D. Ververidis and C. Kotropoulos, "Emotional speech classification using Gaussian mixture models," *Proceedings of ISCAS*, vol. 3, pp. 2871-2874, 2005.
- [2] C. M. Lee and S. S. Narayanan "Toward detecting emotions in spoken dialogs," *IEEE Transactions on Speech and Audio Processing*, vol.13, pp. 293-303, 2005.
- [3] J. Wagner, J. H. Kim, and E. Andre, "From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification," *Proceedings of ICME*, pp. 940-943, 2005.
- [4] P. Pudil and J. Novovicova, "Novel Methods for Subset Selection with Respect to Problem knowledge," *IEEE Intelligent Systems*, March, pp. 66-74, 1998.
- [5] Y. L. Lin and W. Gang, "Speech Emotion Recognition based on HMM and SVM," *Proceedings of Machine Learning and Cybernetics*, Vol.8, pp. 4898-4901, 2005.
- [6] F. Morchen, A. Ultsch, M. Thies, and I. Lohken, "Modeling

Timbre Distance With Temporal Statistics From Polyphonic Music," *IEEE transaction on Audio, Speech and Language Processing*, Vol.14, Issue 1, pp. 81-90, 2006.

- [7] E. F. Combarro, E. Montanes, I. Diaz, J. Ranilla, and R.Mones, "Introducing a Family of Linear Measures for Feature Selection in Text Categorization," *IEEE transactions on Knowledge and Data Engineering*, Vol.17, No.9, pp. 1223-1232, 2005.
- [8] G. H. John, R. Kohavi, and K.Pfleger, "Irrelevant Features and the Subset Selection Problem," *Proceedings of Machine Learning*, pp. 121-129, 1998.
- [9] R. S. Sutton and A. G. Barto, "Reinforcement Learning," *An Introduction, A Bradford book, London*, 1998.
- [10] C. H. Park and K. B. Sim "The Implementation of the Emotion Recognition from Speech and Facial Expression System," *Lecture Notes in Computer Science(LNCS)*, vol. 3611, pp. 85-88, 2005.

Hyung-Chang Yang received his B.S. and M.S. degrees from the Department of Industrial Engineering of Soongsil University, Seoul Korea, in 2002. He is currently pursuing the Ph.D. degree at the Chung-Ang University. His research fields

are Intelligent Robot, Home Network, Smart Home, Ubiquitous Sensor Network, Soft Computing, etc.

Ho-Duck Kim received the B.S. degrees in the School of Electrical and Electronics Engineering from Chung-Ang University, Korea, in 2005. He is currently pursuing the M.S. degree at the Chung-Ang University. His research interests include Evolvable

H/W, Processor Design, Embedded Linux System, Genetic Algorithm, DARS, etc.

Chang-Hyun Park received his B.S. and M.S. and Ph.D. degrees in the School of Electrical and Electronics Engineering from Chung-Ang University in 2001, 2003, and 2006 respectively. In 2006, He joined the ETRI(Electronics and Telecommunications

Research Institute) of Korea, where he is Senior Research in the Radio Technology Research Group of Radio Broadcasting Research Division. His research interests include Emotion Recognition, Pattern Recognition, Machine Learning, Evolutionary Computation, Intelligent Robot etc.

Kwee-Bo Sim received his B.S. and M.S. degrees in the Department of Electronic Engineering from Chung-Ang University, Korea, in 1984 and 1986 respectively, and Ph.D. degree in the Department of Electrical Engineering from The University of Tokyo,

Japan, in 1990. Since 1991, he has been a faculty member of the School of Electrical and Electronics Engineering at Chung-Ang University, where he is currently a Professor. His research interests are in artificial life, emotion recognition, ubiquitous intelligent robot, intelligent System, computational intelligence, intelligent home and home network, ubiquitous computing and Sense Network, adaptation and machine learning algorithms, neural network, fuzzy system, evolutionary computation, multi-agent and distributed autonomous robotic system, artificial immune system, evolvable hardware and embedded system etc. He is a member of IEEE, SICE, RSJ, KITE, KIEE, KFIS, and ICASE Fellow. He is currently President of the KFIS.

Phone : +82-2-820-5319

Fax : +82-2-817-0553

E-mail : kbsim@cau.ac.kr