The Novel Feature Selection Algorithm for Emotion Recognition

Ho-Duck Kim*, Young-Im Cho**, and Kwee-Bo Sim*

* School of Electrical and Electronics Engineering, Chung-Ang University
221, Heukseok-Dong, Dongjak-Gu, Seoul 156-756, Korea
e-mail : kbsim@cau.ac.kr

** Department of Computer Science, The University of Suwon
San2-2, Wauri, Bongdam-eup, Hwaseoung, Gyeonggi-do, 445-743, Korea
e-mail : ycho@suwon.ac.kr

Abstract

This paper presents an original feature selection method for Emotion Recognition which includes many original elements. Feature selection has some merit regarding pattern recognition performance. Thus, we developed a method called an '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. Our 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 (IFS)'. By performing the 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.

***Keywords*:** Reinforcement Learning, Feature selection, Emotion Recognition, SFS, GAFS, IFS

1. Introduction

Emotion recognition research has been typically attempted using four kinds of medium. They are speech, image, physiological signal, and gesture. 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. 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], 17 features were extracted from pitch, energy, speech rate and so on with sex also being classified. In addition, In [2], 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 [3]. In [1], a Forward Selection (FS) method was used. In [4], a Sequential Forward Selection algorithm was also used and the best feature subset was selected out of 39 candidate feature sets. 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 [5]. 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[7]. In this study, 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 2, it explains the emotion recognition method and Section 3 explains the proposed algorithm. The Section 4 shows a simulation and result of using the proposed algorithm. The Section 5 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.

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.

3. The Interactive Feature Selection 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 [5]. 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 [5]. 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 be 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.

$$R_t = \sum_{k=0}^{T} \gamma^k r_t + k + 1 \qquad (1)$$

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 eq. 1.

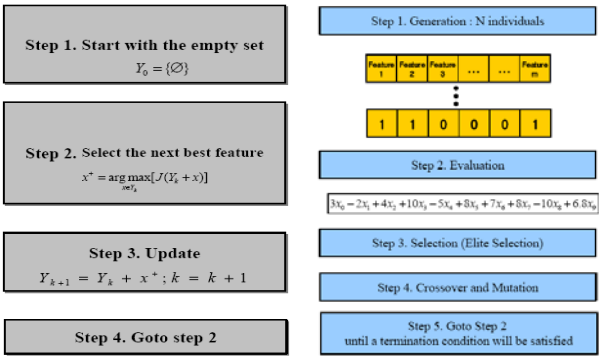
3.2 Sequential Forward Selection Algorithm

Sequential Forward Selection (SFS) is the simplest greedy search algorithm. In this paper, we will briefly explain this algorithm. Fig. 1(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

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 (Fig. 1(b)). So, we tried to search good feature set using the Simple Genetic Algorithm.



(a) SFS Algorithm (b) GAFS algorithm

Fig. 1. SFS and GAFS algorithm

3.4 Interactive Feature Selection Algorithm

The Interactive Feature Selection (IFS) 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.

Fig. 2 shows an IFS process. We assume that an emotion recognition system that includes this algorithm will be applied to a home robot or appliance. 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.

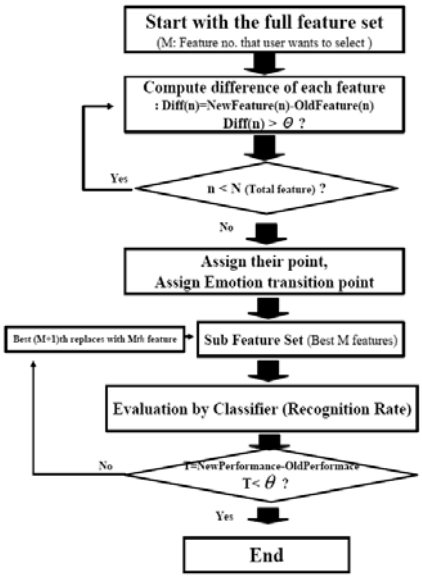


Fig. 2. IFS Algorithm

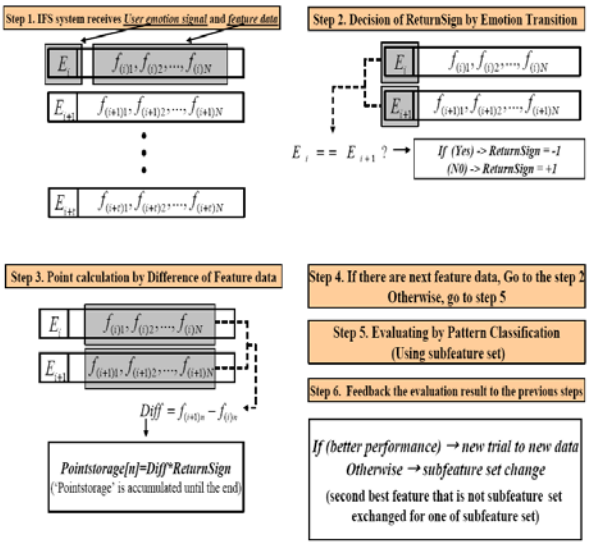


Fig. 3. IFS example

Fig. 3 shows an example of the IFS algorithm and is based on the fig. 2. First, this algorithm starts with a full feature set and when a new feature set and an emotion identifier is inputted, it assigns a +1 or -1 to the “return sign”(if an old emotion ID equals a new emotion ID then +1, Otherwise -1). Thereafter, the product of “return sign” and the difference of each feature is stored in an array “Point storage”. This iteration is repeated for one episode (user can arbitrarily define an episode). After the episode, the feature set that was selected first

is applied to an objective function(Pattern Classification System) and the evaluation result is stored. If the next evaluation result is worse than the previous, the worst feature of the selected feature set will be replaced with the best feature among those that were not selected (step 6 in fig. 3.

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.

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.

Fig. 4 is comparison graph of 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.

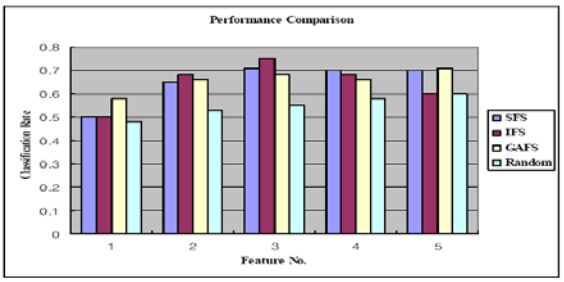


Fig. 4. Emotion Classification Rate Comparison

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.

Acknowledgment:

This research was supported by Seoul R&BD Program (No. : 106876) from Seoul city of Korea.

References

[1] 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.

[2] 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.

[3] P. Pudil and J. Novovicova, “Novel Methods for Subset Selection with Respect to Problem knowledge,” *IEEE Intelligent Systems, March,* pp. 66-74, 1998.

[4] 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.

[5] 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 Engineeringl, Vol.17, No.9,* pp. 1223-1232, 2005.

[6] G. H. John, R. Kohavi, and K.Pfleger, “Irrelevant Features and the Subset Selection Problem,” *Proceedings of Machine Learning,* pp. 121-129, 1998.

[7] R. S. Sutton and A. G. Barto, “Reinforcement Learning,” *An Introduction, A Bradford book, London,* 1998.

[8] 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.