Proceedings of the International MultiConference of Engineers and Computer Scientists 2011 Vol I,
IMECS 2011, March 16 - 18, 2011, Hong Kong

Iterative Forward Selection Method Based on Cross-validation Approach and Its Application to Infant Cry Classification

Yoshifumi Okada, Kentarou Fukuta, and Tomomasa Nagashima

Abstract—**Feature selection is an important step in building an effective and efficient classifier. In this paper, we propose an iterative forward selection method (IFSM) inspired by the cross-validation approach. IFSM ranks all variables by a score called the discriminant power score (DPS). DPS represents the discriminant ability of each variable, which is estimated using different training datasets through cross-validation. In this paper, IFSM is applied to an infant cry dataset. The following results were derived from the experiment: 1) the variable with a higher DPS is effective for building a better classification model, and 2) IFSM shows a drastic improvement in classification accuracy compared to non-iterative IFSM. We believe that IFSM is a promising method in the field of pattern recognition.**

Index Terms—**feature selection, forward selection method, cross-validation, classification, baby cry**

I. Introduction

Feature selection is an important step for building an effective and efficient classifier. The forward selection method (FSM) is widely used as a powerful approach [1]. In this method, variables (also called attributes or features) are progressively added one by one to the classification model in descending order of discriminant power. In our previous study, we proposed an F-value-based FSM (FSM in the following text) and showed that effective variables for classification can be extracted using a gene expression dataset [2] and an infant cry dataset [3]. However, a critical issue of FSM was that classification accuracy significantly varied depending on the number of variables used for the model.

The aim of this study is to extract variables that have a robust and high discriminant power. Here we propose iterative FSM (IFSM in the following text) that is inspired by cross-validation. IFSM conducts FSM for different training datasets generated by cross-validations. A major point of IFSM is that all variables can be ranked by a score called the discriminant power score (DPS). A variable with a higher DPS is expected to be more effective for class discrimination. In this study, we verify the performance of IFSM using a cry dataset for infants with a genetic disease in their throat called ankyloglossia with deviation of the epiglottis and larynx ADEL [4, 5].

This paper is organized as follows. Section II briefly describes the FSM algorithm and then explains the procedure of IFSM. Section III describes the evaluation experiments using the infant cry data. Section IV shows the results and discussion of the performance. Finally, Section V summarizes our conclusions and suggests future work.

Manuscript received January 28, 2011. This work was supported in part by A-STEP No.AS221Z01841F from JSTA.

Y. Okada is with College of Information and Systems, Muroran Institute of Technology, 27-1, Mizumoto-cho, Muroran 050-8585, Japan (corresponding author to provide phone: +81-143-5498; fax: +81-143-5498; e-mail: okada@csse.muroran.-it.ac.jp).

F. Kentaro is with the Satellite Venture Business Laboratory, Muroran Institute of Technology, 27-1, Mizumoto-cho, Muroran 050-8585, Japan (e-mail: fukuta@mail.svbl.muroran-it.ac.jp).

T. Nagashima is with College of Information and Systems, Muroran Institute of Technology, 27-1, Mizumoto-cho, Muroran 050-8585, Japan (e-mail: nagasima@csse.muroran.-it.ac.jp).

II. Method

A. *Forward Selection Method (FSM)*

Suppose that we have a two-class dataset as shown in Fig.1. Each sample is characterized by a set of variables (also called attributes or features). The aim here is to select discriminative variables between the two classes. FSM extracts the differentially represented variables using a simple statistic called F-value, as defined by equation (1).

$$F = \frac{(n_1 + n_2 - p)n_1 n_2 (D_{p+1}^2 - D_p^2)}{(n_1 + n_2 - 2)(n_1 + n_2) + n_1 n_2 D_p^2}, \quad (1)$$

where n_1 and n_2 are the number of samples in class 1 and class 2, respectively, p is the number of variables before adding a new variable, and D_p^2 is the squared Mahalanobis distance [6] between the mean vectors of the two classes. Using the F-value defined by equation (1), FSM incrementally extracts variables by the following FSM:

Step 1: The first variable v_i ($i = 1$) is determined by ranking the F-value for all variables by substituting $p = 0$.

Step 2: For the i (≥ 2)-th variable, we choose an i-th variable from the remaining variables and add it to the set of $i - 1$ variables. Then, the F-value is calculated for the total set of i variables by substituting $p = i - 1$.

Step 3: Step 2 is repeated for all variables in the remaining set, and the i-th variable is determined by selecting the variable with the largest F-value.

Step 4: Steps 2 and 3 are repeated for $i \leq n_1 + n_2 - 2$ until the ranking of variables is accomplished.

Finally, we obtain the following ranking vector:

$$[v_1, v_2, \dots, v_{n_1+n_2-2}],$$

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

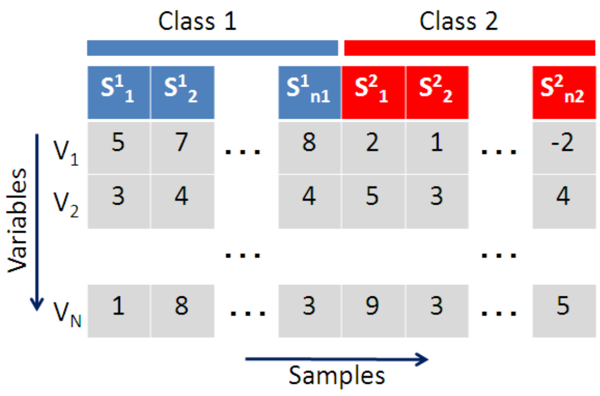

Fig. 1. Data format used in this study.

Fig. 2. Procedure of IFSM.

where the variables are sequenced in the order of the one selected in Step 3.

B. *Iterative FSM based on cross-validation (IFSM)*

IFSM is an iterative FSM inspired by cross-validation. In this study, we utilize the leave-one-out cross-validation method (LOOCV) [7] for cross-validation.

Figure 2 shows the procedure of IFSM. First, one test sample is taken from the dataset. Then, the remaining samples are used as a training dataset. Subsequently, we apply FSM to the training dataset and obtain a ranking vector of variables. In addition, each variable is scored according to its rank order. These steps are repeated for every test sample, and accordingly, the scores of the variables are also updated every time. The resulting score is called a discriminant power score (DPS). DPS takes a value 0–1. Algorithm 1 is a pseudo code for the DPS calculation, and RM (i, j) is the index number of the variable ranked jth by FSM for the training dataset i. A variable with a higher DPS is expected to be more effective for class discrimination. Thus, we build a classifier using a set of variables with high DPSs under a certain threshold.

Algorithm 1 Calculation of DPS

```
Input: ranking matrix RM, number of variables N, total number of samples M (=n1+n2)
for i=1 to N do
    DPS (i) = 0
end for
for i=1 to N do
    for j=1 to M do
        DPS (RM(i,j)) = DPS(RM(i,j)) + (N+1) − i
    end for
end for
for i=1 to N do
    DPS (i) = DPS (i) / (N * M)
end for
```

The above process is performed for 49 samples. Finally, we obtain a labeled dataset in which the column and row are the variables and samples, respectively, and each element corresponds to a spectral power.

C. *Evaluation*

We evaluate the classification accuracy using variables obtained by IFSM. The classifier used here is linear discriminant analysis using Mahalanobis distance [6]. Classification accuracy is estimated by LOOCV [7].

First, we extract one sample as a test sample from the dataset. Next, IFSM is applied to the remaining samples and the variables are ranked in descending order of DPS. Starting from the variable that has the maximum DPS, we build a classification model and conduct a classification test for the test sample. By cumulatively adding a variable in descending order of DPS, we repeat the classification test for the test sample using the classification model constructed each time.

The above process is repeated for all test samples. For each cumulative size, we calculate the percentage of correctly classified test samples, namely classification accuracy.

III. EXPERIMENTS

A. *Baby Cry Samples*

We focus on the cries belonging to a class of painful cries caused by ADEL [4, 5], and aim to discriminate between the ADEL cries and others. In this study, we use a set of baby cry samples provided by Wang *et al.* [3], which consists of 26 samples from ADEL babies and 23 samples from normal babies. A sample indicates waveform data per breath, which is manually clipped from the time-series data.

B. *Preprocessing*

For each sample, we perform the following process. First, L frames are created from a sample. Each frame has a length of 51.2 ms with an overlap of 50% for each other and is weighted by a Hamming window. Next, these L frames are transformed into spectral envelopes by linear predictive coding (LPC) [8]. LPC is widely used for generating a smooth curve for the graph of FFT magnitudes, and can highlight the spectral peaks of the formant frequencies. In this study, the LPC order (the number of LPC coefficients used) is fixed to 16 following the report by Hariharan *et al.* [9], and the number of variables (frequency points in the power spectrum) is fixed to 80 based on preliminary experiments.

IV. RESULTS AND DISCUSSION

Figure 3 is the DPS graph for all variables (frequencies) in which the horizontal and vertical axes indicate the frequency and DPS value, respectively. This DPS graph is created by applying IFSM to the entire sample (49 samples). We see that the DPSs of the top 5 variables are significantly higher than those of the other variables. Thus, these variables are expected to be effective for discriminating between ADEL cries and normal cries. Figure 4 shows the classification accuracies with a cumulative increase of the number of

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

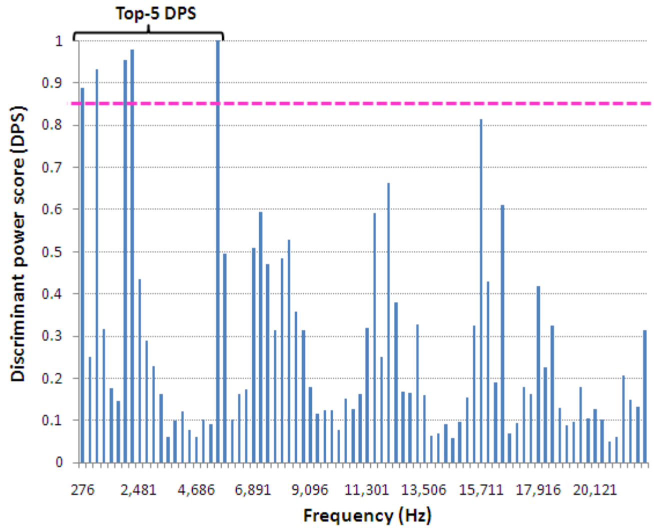

Fig. 3. DPS graph generated using the entire sample (49 samples).

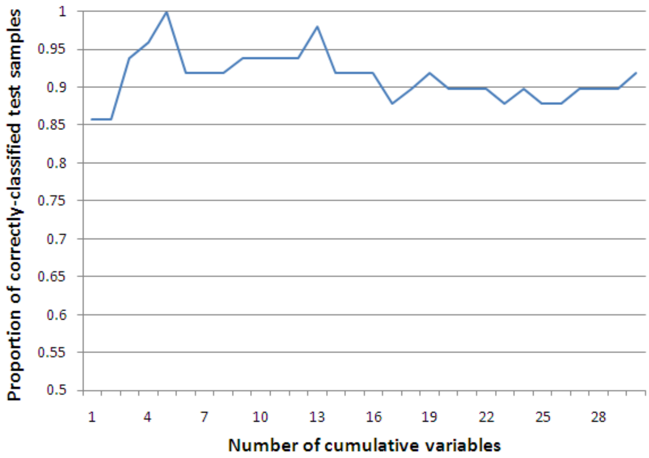

Fig. 4. Classification accuracy by IFSM.

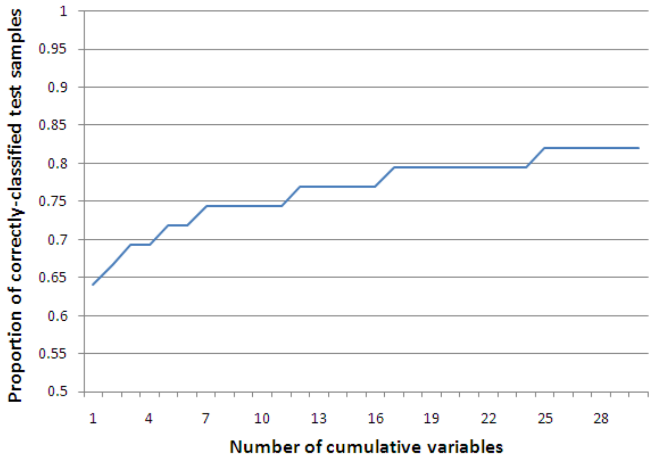

Fig. 5. Classification accuracy by FSM.

variables (up to the 30th) in descending order of DPS. As expected, the classification accuracy exhibits a rapid increase up to the 5th variable and accomplishes 100%. An unstable and downward trend is observed from the 6th variable and above; this is caused by the fact that the discrimination power of the variable weakens with the increase of ranks. Nevertheless, it is noteworthy that it keeps accuracies from being larger than 87.5%.

Figure 5 shows the classification accuracies obtained by FSM, which is our previous approach. Note that the horizontal axis is the cumulative number of variables ranked by a single FSM. We see that the classification accuracy of FSM is considerably inferior compared to that of IFSM. FSM trains only a single dataset. In contrast, IFSM learns different training datasets by cross-validation, and preferentially extracts variables showing a stronger discriminant power. We consider that such a strategy of IFSM works effectively for finding discriminative and robust variables.

Mukai *et al*. reported that in ADEL, the vocal tract is narrowed in the vicinity of the glottis [5]. This means that an ADEL baby finds it more difficult to produce a low-frequency voice than a normal baby. In fact, higher DPSs are distributed in relatively low frequencies. In this way, DPS is potentially capable of being related to qualitative features between different classes. A more detailed discussion of DPS interpretation will be published in a separate paper.

V. CONCLUSIONS

We proposed a new feature selection method called IFSM that is an improved algorithm of the F-value-based FSM. IFSM iterates FSM for different training datasets generated by cross-validation, and computes a discriminant power score (DPS) for each variable. In this paper, we applied IFSM to an infant cry dataset and evaluated the classification accuracy by LOOCV. As a result, it was shown that IFSM significantly enhanced the classification accuracy by FSM, and that the variables with higher DPSs were effective for building better classification models.

In this study, linear discriminant analysis using Mahalanobis distance was employed as a classifier. In future work, we will test various combinations of IFSM and other classifiers such as a support vector machine. In addition, we will apply IFSM to other various datasets as well as audio datasets.

ACKNOWLEDGMENT

The authors would like to express their thanks to Dr. Mukai for his advice and cooperation. This work was supported in part by A-STEP No.AS221Z01841F from JSTA.

REFERENCES

[1] Y. Nagata and M. Munechika, “Introduction to Multivariable Analysis Methods,” Saiensu-sha Co., Ltd. Publishers, 2007 (in Japanese).
[2] H. Mitsubayashi, S. Aso, T. Nagashima, and Y. Okada, “Accurate and robust gene selection for disease classification using a simple statistic” *Bioinformation*, vol. 3, no. 2, pp.68–71, 2008.
[3] X. Wang, T. Nagashima, K. Fukuta, Y. Okada, M. Sawai, H. Tanaka and T. Uozumi, “Statistical method for classifying cries of baby based on pattern recognition of power spectrum,” *International Journal of Bioimetrics*, vol. 2, no. 2, pp.113–123, 2010.
[4] S. Mukai and K Asaoka, “Congenital ankyloglossia with deviation of the epiglottis and larynx,” *Annals of Otolaryxgol*, vol. 100, pp.1–20, 1991.
[5] S. Mukai and M. Nitta, “Correction of the Glosso-Larynx and Resultant Positional Changes of the Hyoid Bone and Cranium,” *Acta Oto-laryngologica*, vol. 122, no. 6, pp.644–650, 2002.
[6] P. C. Mahalanobis, “On the generalised distance in statistics,” *Proceedings of the National Institute of Sciences of India*, vol. 2, no. 1, pp.49–55, 1936.
[7] P. A. Lachenbruch and M. Goldstein, “Discriminant Analysis,” *BIOMETRICS*, vol. 35, pp.69–85, 1979.

[8] F. Itakura, “Minimum Prediction Residual Applied to Speech Recognition,” *IEEE Trans. on ASSP*, vol. 23, no. 1, pp.67–72, 1975.
[9] M. Hariharan, L.S. Chee, S. Yaacob, “Analysis of Infant Cry Through Weighted Linear Prediction Cepstral Coefficients and Probabilistic Neural Network,” *J. Med. Syst.* 2010.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)