International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 3 (2011) pp. 159-164

Audio Classification Based on Closed Itemset Mining Algorithm

Yoshifumi Okada[1], Takahiro Tada[2], Kentaro Fukuta[3] and Tomomasa Nagashima[1]

[1] College of Information and Systems, Muroran Institute of Technology,
27-1 Mizumoto-cho, Muroran 050-8585, Japan
okada@csse.muroran-it.ac.jp, nagasima@csse.muroran-it.ac.jp

[2] Department of Information and Electronic Engineering, Muroran Institute of Technology,
27-1 Mizumoto-cho, Muroran 050-8585, Japan
tada@cbrl.csse.muroran-it.ac.jp

[3] Satellite Venture Business Laboratory, Muroran Institute of Technology,
27-1 Mizumoto-cho, Muroran 050-8585, Japan
fukuta@mail.svbl.muroran-it.ac.jp

***Abstract*: Automatic audio classification is a major topic in the fields of pattern recognition and data mining. This paper describes a new rule-based classification method (classification rule extraction for audio data, cREAD) for multiclass audio data. Typically, rule-based classification requires much computation cost to find rules from large datasets because of combinatorial search problems. To achieve efficient and fast extraction of classification rules, we take advantage of a closed itemset mining algorithm that can exhaustively extract non-redundant and condensed patterns from a transaction database in a reasonable time. A notable feature of this method is that the search space of the classification rules can be dramatically reduced by searching for only closed itemsets that are constrained by "class label item." In this paper, we demonstrate that our method is superior to other salient methods for accurately classifying a real audio dataset.**

***Keywords*:** classification, audio, data mining, closed itemset, pruning, baby cry.

I. Introduction

Automatic audio classification is a major topic in the fields of pattern recognition and data mining. It has been used in various applications, such as musical instrument sound identification [1], [2], music retrieval [3], and personal recognition [4-6]. The most popular approach in audio classification is a statistical machine learning technique such as a support vector machine (SVM) [7]. This approach constructs a discriminant mathematical model to discriminate among different classes by using feature quantities (*e.g.*, power spectrum data) derived from audio sample data. Although a statistical machine learning approach exhibits high classification accuracy in many cases, it is not easy to interpret the resulting models, because the learning process is "black box."

On the other hand, a rule-based approach identifies distinctive patterns (classification rules) among different classes. The advantage of a rule-based classifier is the explicitness and comprehensibility of the resulting model in addition to relatively high classification accuracies [8]. Thus, we can not only infer important features to characterize each class, but also use them as useful knowledge. However, a rule-based approach has some drawbacks. Typically, the extraction of classification rules requires heavy computation because of the combinatorial searching of large-scale data. Thus, the number of rules generated is extremely large, so a further process is necessary to select only the discriminative rules between classes. Therefore, existing rule-based classification methods have focused on finding rules as possible local solutions by using stochastic or heuristic approaches [9], [10].

In recent years, much attention has been paid to discovering co-occurrence patterns called closed itemsets [11], [12] from transaction databases (see Section II) in the field of data mining. Frequent pattern mining, such as the Apriori algorithm [13], searches for all frequent patterns (even if these are included in other patterns), whereas closed itemset mining can extract maximal and condensed patterns by excluding redundant patterns having inclusive relations.

Here, we propose a new rule-based classification method for audio data based on an efficient closed itemset mining algorithm called linear-time closed itemset miner (LCM) [12], [14]. In our method (classification rule extraction for audio data, cREAD), audio power spectra of multiple classes are transformed into a transaction database that includes a "class label item" for each transaction. The classification rules are extracted using the following two processes: 1) an exhaustive search of the closed itemsets having the class label item by using the LCM algorithm and 2) the greedy rule selection approach. A notable feature of this method is that it drastically reduces search space by omitting the search for unnecessary closed itemsets that have no class label item. In this paper, we show the results of comparing the performance of two salient classifiers and the effect of pruning operation by using a real audio sample dataset.

This paper is organized as follows. Section II describes briefly closed itemset and LCM algorithm and then provides the basic idea of our pruning approach. Section III describes the computational procedure of cREAD. Section IV explains

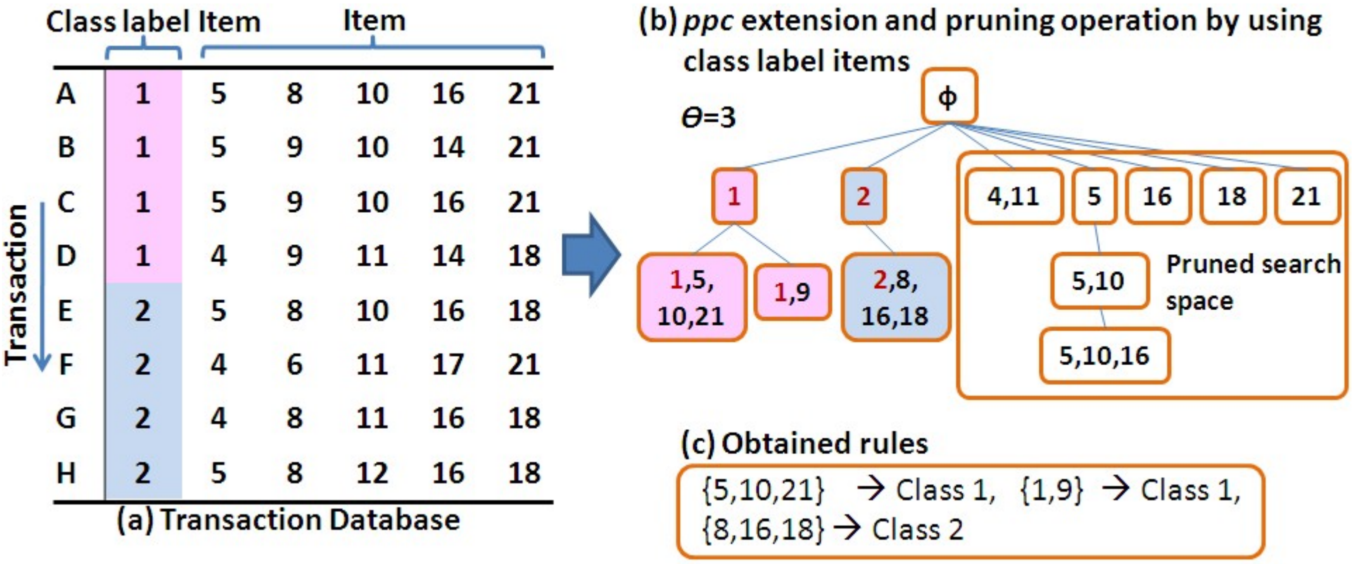

Figure 1. Transaction database and basic idea of our method

the datasets used in this study and the experimental method to evaluate the performance of cREAD. Section V discusses the results and performance. Finally, Section VI summarizes our conclusions and suggests future work.

II. Extension of Closed Itemset Mining Algorithm to Rule-Based Classifier

A. *Closed Itemset*

A transaction database is a set of transactions with a set of items, as shown in Figure 1(a). Each transaction with an item I is called an occurrence of I, and a set of occurrences of I is termed an occurrence set. For a given constant $\Theta \geq 0$, a frequent itemset is a set of items that is included in at least Θ transactions, where Θ is called a minimum support. A closed itemset is defined as the maximal itemset (with respect to inclusion) among the set of itemsets having the same frequency and appearing in the same transaction [11], [12]. In Figure 1(a), {1, 5, 10} is a frequent itemset under $\Theta = 3$, but it is not a closed itemset because {1, 5, 10, 21}, which is a maximal itemset (*i.e.*, a closed itemset), appears in the same transactions as {1, 5, 10}. Thus, the closed itemset represents non-redundant and condensed patterns in contrast to frequent itemsets.

B. *LCM Algorithm*

As might be expected, the mining of closed itemsets by a naive full search requires considerable computation time because of its combinatorial search requirements. LCM proposed by Uno *et al.* is the fastest algorithm that can enumerate frequent closed itemsets in linear time depending on the database size [12], [14]. This algorithm uses a prefix-preserving closure extension (*ppc* extension), which is an extension of a closed itemset to another closed itemset. Because this extension generates a new frequent closed itemset from the previously obtained itemset without duplication by using the depth-first search technique, unnecessary non-closed frequent itemsets can be completely pruned.

C. *Basic Idea of Our Method*

In this paper, LCM is extended to a rule-based classifier for multiple classes. Assume that we have a multiclass transaction database in which each transaction has a class label. Figure 1(a) shows a two-class transaction database, *i.e.*, every transaction has a class label item of either “1” or “2.” The aim here is to exhaustively extract only closed itemsets having a class label item by using LCM. That is, closed itemsets occurring in multiple classes are ignored; hence, we can drastically reduce the computational time as well as the search space required for enumerating candidates of classification rules. Figure 1(b) shows an illustrative example of the pruning operation based on class label items. As shown in this figure, only the closed itemsets having a class label item are extracted and the subsequent search is stopped, because the other closed itemsets are unusable as rules. The extracted closed itemsets are represented with the rules as shown in Figure 1(c), in which each class label item is denoted as the consequent of the rule.

III. Method

For simplicity, we describe the method of cREAD by using an example of a two-class audio sample. Figure 2 illustrates the procedure of cREAD.

A. *Preprocessing*

Consider a Fourier transform power spectra classified into two classes ($n1$ samples in Class 1 and $n2$ samples in Class 2) and having p frequency points for each audio sample. First, all powers in each frequency f_{ij} are normalized to have a mean of 0 and a variance of 1 over the two classes, where $i = 1, 2,\ldots, n1 + n2,$ and $j = 1, 2,\ldots, p$. Next, these normalized powers (nf_{ij}) are quantized to q levels by uniformly dividing the range of the maximum and minimum values. In this way, each power

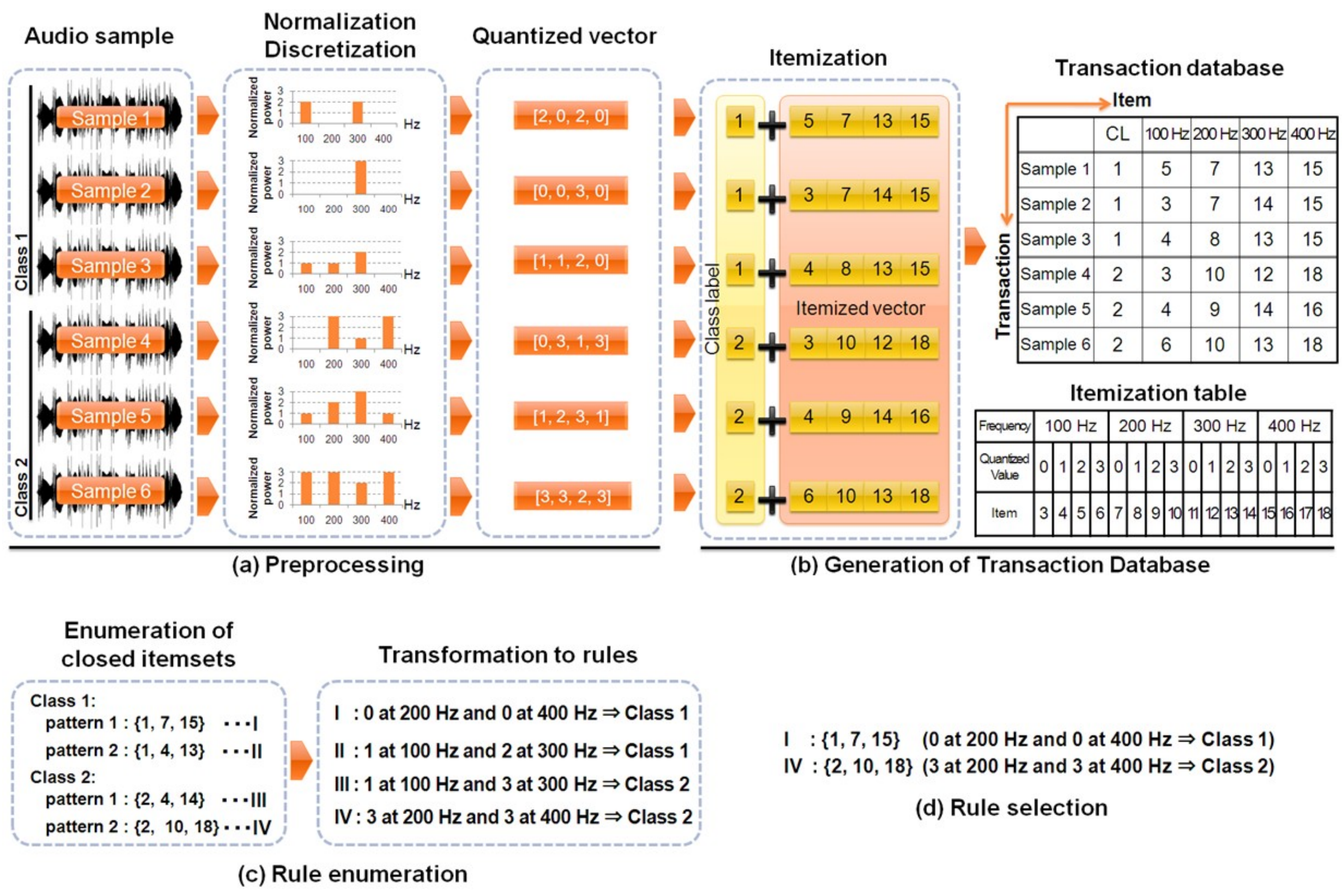

Figure 2. Procedure of cREAD

spectrum is transformed into a p-dimensional vector having a quantized value in each element.

B. Generation of Transaction Database

The transaction database is created by referring to an itemization table, as shown in Figure 2(b). In the itemization table, each item corresponds to a unique quantized value obtained for each dimension (*i.e.*, each frequency point) of the quantized vector. Note that items 1 and 2 in Figure 2(b) are used as the class label items. According to the itemization table, we transform each quantized vector into an itemized vector having a class label item. Finally, these itemized vectors are summarized into the transaction database in which each itemized vector corresponds to a transaction.

C. Rule Enumeration

The aim here is to extract rules such as “A and B→ class1,” whose antecedent and consequent include itemsets and a class label item, respectively. Figure 2(c) shows the closed itemsets (and their rule representations) extracted from the transaction database in Figure 2(b). Note that we can easily transform each item in the antecedent of the rule into a quantized value (or a normalized power) corresponding to a frequency point by referring to the itemization table. The original LCM program does not guarantee preferential enumeration of the closed itemsets having a class label item. In contrast, we sort items in ascending order at every *ppc* extension and preferentially search for the closed itemsets having a class label item. The enumeration of the closed itemsets is stopped when class label items cannot be included in the closed itemsets.

D. Rule Selection

We select final discriminative classification rules from the enumerated rules by the following greedy approach:

1) Select the most frequent rule (*i.e.*, the most frequent closed itemset) in the transaction database. If there are multiple rules with the same occurrence frequency, then we perform the following steps with respect to those rules:

1-1) Select a rule, and calculate the inner product between the quantized values of the rule’s antecedent and those (in the same frequency points as the rule) of each vector of other classes.

1-2) Calculate the mean value of the inner products obtained in Step 1-1.

1-3) Repeat Steps 1-1 and 1-2 for all rules in Step 1.

1-4) Select a rule in which the absolute mean value of the inner products is minimal.

2) Exit if the union of the occurrence sets of the rules selected in Step 1 covers all transactions, otherwise repeat Step 1.

E. Classification

Here, we explain the classification method for unknown (or test) samples by using the final rules obtained from the above procedure. First, each unknown sample is transformed into a quantized vector having p dimensions in the same manner as that in preprocessing. Second, the quantized values of an unknown sample are compared with those of the classification

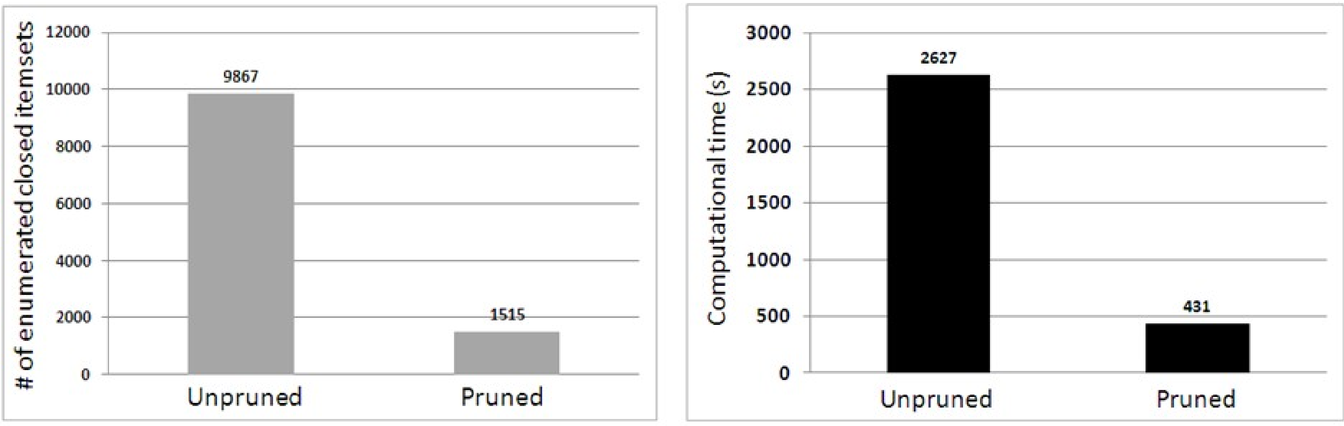

Figure 3. Effect of our pruning operation

rules obtained from each class by the *k*-nearest neighbor approach. In this process, we calculate the Euclidian distances between the quantized values of the classification rules and those of each unknown sample. Finally, we extract the top-*k* classification rules in ascending order of their Euclidian distances, and assign each unknown sample into a class by the majority vote of its neighbors.

IV. Experiments

A. Dataset

We use the baby-cry dataset generated by Wang *et al*. [15] to evaluate the usefulness of our method. This dataset consists of two classes: a class of cries belonging to the genetic disease ankyloglossia with deviation of the epiglottis and larynx (ADEL) [16] and a class of normal cries. The ADEL class and normal class include 17 and 22 audio samples (waveform data), respectively. These samples are transformed into the frequency domain by the Fourier analysis in the frequencies 0-22050 Hz and in a resolution of 43Hz (513 frequency points).

B. Evaluation of Classification Accuracy

The evaluation is conducted using leave-one-out cross-validation (LOOCV) method [17]. In LOOCV, first, we extract one sample (*i.e.*, a quantized vector having *p* dimensions) as a test sample from the dataset and generate classification rules using the remaining samples. Second, a test sample is assigned to a class by our classification method. We repeat these processes for all samples and calculate the accuracy of correctly classified samples. In the evaluations, the number of quantization bins and the minimum support in the closed itemset mining are set to 21 and 4, respectively.

The classification accuracy is compared to those of the two salient classifiers, C4.5 [18] and SVM [7]. C4.5 is a statistical classifier based on the decision tree, and SVM is one of the latest and most successful kernel-based machine-learning methods. We downloaded free software available in [19] and [20], and [21], and used them with their default parameters.

Table 1. Average classification accuracy (%)

Method	Classification accuracy (%)
cREAD	69.2
C4.5	38.5
L-SVM	66.7
R-SVM	61.5

V. Results and Discussion

Figure 3 shows the number of enumerated closed itemsets (left) and the computational time (right) before and after the pruning operation. These graphs are obtained by using all samples of the ADEL dataset. From these figures, we can infer that the pruning operation results in a drastic decrease in the number of enumerated closed itemsets and computational time.

Table 1 shows the classification accuracies of cREAD, C4.5, and SVM based on the average accuracies obtained by the LOOCV test. For SVM, we used two kernel functions: the linear kernel and the radial basis function kernel, which are shown by L-SVM and R-SVM in Table 1, respectively. As a result, cREAD exhibits the best accuracy (69.2%) in the LOOCV test. Notably, this score is far superior to that of C4.5, which is one of the rule-based classifiers.

As observed in most bio-signal data, the dataset used in this study has “underdetermined problem” in which the number of samples is extremely small compared to that of the attributes (frequencies). We expect that cREAD will become a promising method for bio-signals having such high dimensional attributes as well as audio spectrum data. However, accuracy of this method is still insufficient for practical purposes. In this paper, the final classification rules are selected by using a simple greedy approach. For further accuracy improvement, we will incorporate a rule-optimization process for selecting more sophisticated rules.

As shown in Figure 3, the calculation amount of cREAD is drastically reduced by the pruning operation, whereas the computational time is considerably larger compared to other methods, which complete the computation within 1 s. This is because the current version does not implement efficient algorithms, such as frequency counting or data compression. Some salient and excellent algorithms have been proposed in past studies [11], [12], [22]. In the next version, the above problem will be solved by incorporating such algorithms into cREAD.

VI. Conclusions

We proposed a new rule-based classification method (cREAD) for audio data based on exhaustive closed itemset mining algorithm. cREAD achieves efficient pruning of unnecessary closed itemsets by class label item. In this paper, by using a real audio sample dataset, we demonstrated that search space is drastically reduced by pruning unnecessary closed itemsets that have no class label items. In addition, we evaluated the classification accuracy by comparing the salient classifiers C4.5 and SVM. The results showed that cREAD presented the best accuracy among the three methods, although it required more computation time.

From these results, we can conclude that cREAD is a potentially useful method for audio classification. However, further technical improvements such as a parameter search and rule optimization as well as implementation of efficient frequency counting and data compression should be achieved. In future work, we will not only handle these technical issues, but also verify the performance by using various audio sample datasets.

Acknowledgment

The authors thank Dr. Mukai for his advice and cooperation. A part of this work was supported by A-STEP No.AS221Z01841F from JSTA and Grant-in-Aid for Young Scientists (B) No.21700233 from MEXT Japan.

References

[1] J. Liu and L. Xie, "SVM-Based Automatic Classification of Musical Instruments," *Intelligent Computation Technology and Automation (ICICTA)*, Vol. 3, pp. 669-673, 2010.

[2] L. Chen, S. Gunduz and M. T. Ozsu, "Mixed Type Audio Classification with Support Vector Machine," *International Congress on Mathematical Education 2006*, pp.781-784, 2006.

[3] R. Dannenberg,W. Birmingham, B. Pardo, C. Meek, N. Hu and G. Tzanetakis, "A comparative evaluation of search techniques for query-by-humming using the MUSART testbed," *J. Am. Soc. Inf. Sci. Technol.*, Vol. 58, pp. 687-701, 2007.

[4] I. Pollack, J. M. Pickett and W. H.Sumby, "On the Identification of Speakers by Voice," *J. Acoust. Soc. Am*, Vol. 26(3), pp. 403-412, 1954.

[5] L. Hou and J. Xie, "A New Approach to Extract Formant Instantaneous Characteristics for Speaker Identification," *International Journal of Computer Information Systems and Industrial Management Applications*, Vol. 1, pp.295-302, 2009.

[6] B. H. Juang and L. R. Rabiner, "Hidden Markov Models for Speech Recognition," *Technometrics*, Vol. 33(3), pp. 251-272, 1991.

[7] B. E. Boser, I. M. Guyon and V. N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers," *5th COLT*, pp. 144-152, 1992.

[8] R. L. Lawrence and A. Wrlght, "Rule-Based Classification Systems Using Classification and Regression Tree (CART) Analysis," *American Society for Photogrammetry & Remote Sensing*, Vol. 67(10), pp. 1137-1142, 2001.

[9] H. Ishibuchi, I. Kuwajima, and Y. Nojima, "Multiobjective Association Rule Mining," *Proceedings of PPSN Workshop on Multiobjective Problem Solving from Nature*, 2006.

[10] X. L. Li, Z. L. Du, T. Wang and D. M. Yu, "Audio Feature Selection Based on Rough Set," *International Journal of Information Technology*, Vol. 11(6), pp. 117-123, 2005.

[11] C. H. Lee, K. C. Yin, D. L. Yang and J. Wu "Efficient Mining of Frequent Closed Itemsets without Closure Checking," *International Journal of Computer Information Systems and Industrial Management Applications*, Vol. 1, pp. 58-67, 2009.

[12] T. Uno, T. Asai, Y. Uchida and H. Arimura, "An Efficient Algorithm for Enumerating Closed Patterns in Transaction Databases," *Lecture Notes in Artificial Intelligence,* 3245, pp. 133-141, 2004.

[13] R. Agrawal and R. Srikant, "Fast algorithm for mining association rules in large databases," *Proceedings of Very Large Data Bases 1994*, pp. 487-499, 1994.

[14] T. Uno, LCM (Linear time Closed itemset Minor), Available: http://research.nii.ac.jp/uno/code/LCM.htm

[15] X. Wang, T. Nagashima, K. Fukuta, Y. Okada and M. Sawai, "Statical method for classifying cries of baby based on pattern recognition of power spectrum," *International Journal of Biometrics*, Vol. 2(2), pp. 113–123, 2010.

[16] S. Mukai and K. Asaoka, "Congenital ankyloglossia with deviation of the epiglottis and larynx: symptoms and respiratory function in adults," *Annals of Otolaryxgol*, Vol. 100, pp. 1-20, 1991.

[17] P. A. Lachenbruch and M. Goldstein, "Discriminant Analysis," *Biometrics*, Vol. 35(1), pp. 69-85, 1979.

[18] J. R. Quinlan, C4.5 Programs for Machine Learning, Morgan Kaufmann, 1993.

[19] P. Winston, C4.5Tutorial, Available: http://www2.cs.uregina.ca/hamilton/courses/831/notes/ml/dtrees/c4.5/tutorial.html

[20] J. Thorsten, SVM-Light Support Vector Machine, Available: http://svmlight.joachims.org/

[21] C. C. Chang and C. J. Lin, LIBSVM–A Library for Support Vector Machines, Available: http://www.csie.ntu.edu.tw/ cjlin/libsvm/

[22] J. Han, J. Pei and Y. Yin, "Mining frequent patterns without candidate generation," *Proceedings of SIGMOD*, pp. 1-12, 2000.

Author Biographies

Yoshifumi Okada received his Ph.D. in engineering from Muroran Institute of Technology, Japan, in 2002. His current research interests are bioinformatics, Kansei engineering, data mining, statistical pattern recognition, and signal processing.

Kentaro Fukuta received his Ph.D. in engineering from Muroran Institute of Technology, Japan, in 2008. His current research interests are bioinformatics, Kansei engineering, and ontology engineering.

Takahiro Tada received his B.S. degree in engineering from Muroran Institute of Technology, Japan, in 2009. His current research interests are audio analysis, data mining, and machine learning.

Tomomasa Nagashima received his Ph.D. in Physics from Waseda University, Japan, in 1978. He is currently Director of Japanese Society of Kansei Engineering (JSKE) and Fellow of the Institute of Electronics, Information and Communication Engineers (IEICE), Japan. He is also the editor of IJBM. His research interests are Kansei engineering and bioinformatics.