Rough Set Approach for Attributes Selection of Traditional Malay Musical Instruments Sounds Classification[1]

Norhalina Senan[1], Rosziati Ibrahim[1], Nazri Mohd Nawi[1],
Iwan Tri Riyadi Yanto[2] and Tutut Herawan[3]

[1]*Faculty of Computer Science and Information Technology*
Universiti Tun Hussein Onn Malaysia, Parit Raja, 86400 Batu Pahat,
Johor, Malaysia

[2]*Department of Mathematics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia*
Jalan Prof. Dr. Soepomo, Janturan Yogyakarta, Indonesia

[3]*Facultyof Computer System and Software Engineering, Universiti Malaysia Pahang*
Lebuh Raya Tun Razak, 26300 Gambang, Kuantan, Pahang, Malaysia

{halina,rosziati,nazri}@uthm.edu.my, iwan015@gmail.com, tutut@ump.edu.my

Abstract

Feature selection has become very vital aspect in musical instruments sounds for handling the problem of 'curse of dimensionality'. Various feature selection techniques have been applied in this domain focusing on Western musical instruments sounds. However, study on feature selection using rough sets of non-Western musical instruments sounds including Malay Traditional musical instruments is inadequate and still needs an intensive research. Thus, in this paper, an alternative feature selection technique using rough set theory based on Maximum Degree of dependency of Attributes (MDA) technique for Traditional Malay musical instruments sounds is proposed. The modeling process comprises eight phases: data acquisition, sound editing, data representation, feature extraction, data discretization, data cleansing, feature selection using proposed technique and feature validation via classification. The results show that the performance of the best 17 selected features is increase up to 99.82% and 98.03% with 1-NN and PART classifiers respectively.

Keywords: *Rough Set Theory; Dependency of Attribute, Feature selection; Traditional Malay musical instruments sounds dataset.*

1. Introduction

One of the most common problems encountered in many data mining tasks including music data analysis and signal processing is the issue of 'curse of dimensionality'. This problem deals with the high dimensional data with massive amount of attributes. Using the whole set of attribute is inefficient in term of time processing and storage requirements. In addition, it may difficult to interpret and may decrease the classification performance respectively.

---

[1]An early version of this paper appeared in the Proceeding of the 2nd International Conference Ubiquitous Computing and Multimedia Applications (UCMA) 2011, Daejeon, Korea, April 13-15, 2011, Communications in Computer and Information Science vol. 151, pp. 523–539, Springer-Verlag Berlin Heidelberg 2011.

It also has been proven that finding all possible reductions in an information system is NP-hard problem. For that, various feature selection algorithms have been proposed [1]. One of the potential techniques is based on rough set theory. The theory of rough set proposed by Pawlak in 1980s [2] is a mathematical tool for dealing with the vagueness and uncertainty data. In feature selection problem, rough set is implemented with the aim of finding the minimal subsets of attributes which sufficient to generate the same classification accuracy as the whole set of attributes. This minimal features set is known as *reduct*. Banerjee *et al.* [3] also claimed that the concept of *reduct* and *core* in rough set is relevant in feature selection to identify the essential features amongst the non-redundant ones. For that, several attempts on rough set as feature selection technique in musical instrument problems have been done in [4-6]. However, all these studies are conducted based on Western musical instruments. A study on rough sets for feature selection of non-Western musical instruments sounds is inadequate and still needs an intensive research.

Thus, in this paper, an alternative feature selection technique based on rough set theory for Traditional Malay musical instruments sounds is proposed. The technique is developed based on rough set approximation using maximum dependency of attributes proposed by [7]. The main contribution of our work is to select the most significant features by ranking the relevant features based on the highest dependency of attributes on the dataset. Then, the redundant features with the similar dependency value are deleted.

To accomplish this study, the quality of the instruments sounds is first examined. Then, the 37 features from two combination of features schemes which are perception-based and Mel-Frequency Cepstral Coefficients (MFCC) are extracted [8]. In order to employ the rough set theory, this original dataset (continuous values) is then discritized into categorical values by using equal width and equal frequency binning algorithm [9]. Afterwards, data cleansing process is done to remove the irrelevant features. The proposed technique is then adopted to rank and select the best feature set from the large number of features available in the dataset. Finally, the performance of the selected features based on the accuracy rate achieved is analyzed and compared. To do this, the open source of machine learning software known as Weka [10] with various classifier algorithms including Support Vector Machine (SVM), *k*-Nearest Neighbour (*k*-NN), Naïve Bayes, J48 and PART is utilized.

The rest of this paper is organized as follows: Section 2 discuss a related work on feature selection of musical instruments sounds classification problems. Section 3 presents the theory of rough set. Section 4 describes the details of the modeling process. A discussion of the result is presented in Section 5 followed by the conclusion in Section 6.

2. Related Works

An assortment of feature ranking and feature selection algorithms in musical instruments sounds classification have been applied [11-13]. Liu and Wan [11] carried out a study on classifying the musical instruments into five families (brass, keyboard, percussion, string and woodwind) using NN, *k*-NN and Gaussian mixture model (GMM). Three categories of features schemes which are temporal features, spectral features and coefficient features (with total of 58 features) were exploited. A sequential forward selection (SFS) is used to choose the best features. The *k*-NN classifier using 19 features achieves the highest accuracy of 93%. In [12], they conducted a study on selecting the best features schemes based on their classification performance. The 44 features from three categories of features schemes which are human perception, cepstral features and MPEG-7 were used. To select the best features, three entropy-based feature selection techniques which are Information Gain, Gain Ratio and Symmetrical Uncertainty were utilized. The performance of the selected features was

assessed and compared using five classifiers which are *k*-NN, naive bayes, SVM, multilayer perceptron (MLP) and radial basic functions (RBF). They found that the Information Gain produce the best classification accuracy up to 95.5% for the 20 best features with SVM and RBF classifiers. Benetos *et. al* [13] applied subset selection algorithm with branch-bound search strategy for feature reduction. A combination of 41 features from general audio data, MFCC and MPEG-7 was used. By using the best 6 features, the non-negative matrix factorization (NMF) classifier yielded an accuracy rate of 95.2% at best. They found that the feature subset selection method adopted in their study able to increase the classification accuracy.

As mentioned in Section 1, study of the rough set as feature selection technique in musical instrument problems can be found in [4-6]. Wieczorkowska [5] conducted a study on rough set for identifying the most important parameter for musical instruments sounds classification process with a limited number of parameters used. She found that the obtained 16 reducts out of 62 full features able to produce 90% accuracy rate. In [6], rough set and neural networks are applied to classify musical instrument sounds in the basis of a limited number of parameters. Their finding shows that rough set produce highest classification accuracy up to 97.98% compared to neural network with 0.18% more.

In overall, all these works demonstrate that the reduced features able to produce highest classification rate. On the other hand, Deng *et al.* [12] claimed that benchmarking is still an open issue in this area of research. This shows that the existing feature ranking approaches applied in the various sound files may not effectively work to other conditions. Therefore, there were significant needs to explore other feature ranking methods with different types of musical instruments sounds in order to find the best solution.

3. Rough Set Theory

In this section, the basic concepts of rough set theory in terms of data representation are presented.

3.1. Information System

Data are often presented as a table, columns of which are labeled by *attributes*, rows by *objects* of interest and entries of the table are *attribute values*. By an *information system*, a 4-tuple (quadruple) $S=(U,A,V,f)$, where U is a non-empty finite set of objects, A is a non-empty finite set of attributes, $V=\bigcup_{a\in A}V_a$, V_a is the domain (value set) of attribute a, $f:U\times A\rightarrow V$ is a total function such that $f(u,a)\in V_a$, for every $(u,a)\in U\times A$, called information (knowledge) function. In many applications, there is an outcome of classification that is known. This *a posteriori* knowledge is expressed by one (or more) distinguished attribute called decision attribute; the process is known as supervised learning. An information system of this kind is called a decision system. A *decision system* is an information system of the form $D=(U,A\cup\{d\},V,f)$, where $d\notin A$ is the *decision attribute*. The elements of A are called *condition attributes*.

3.2. Indiscernibility Relation

The notion of indiscernibility relation between two objects can be defined precisely.

Definition 3.1. *Let* $S=(U,A,V,f)$ *be an information system and let B be any subset of A. Two elements* $x,y\in U$ *are said to be B-indiscernible (indiscernible by the set of attribute* $B\subseteq A$ *in S) if and only if* $f(x,a)=f(y,a)$*, for every* $a\in B$.

Obviously, every subset of A induces unique indiscernibility relation. Notice that, an indiscernibility relation induced by the set of attribute B, denoted by $IND(B)$, is an equivalence relation. It is well known that, an equivalence relation induces unique partition. The partition of U induced by $IND(B)$ in $S=(U,A,V,f)$ denoted by U/B and the equivalence class in the partition U/B containing $x\in U$, denoted by $[x]_B$.

Given arbitrary subset $X\subseteq U$, in general, X as union of some equivalence classes in U might be not presented. It means that, it may not be possible to describe X precisely in an information system. A set X might be characterized by a pair of its approximations, called lower and upper approximations. It is here that the notion of rough set emerges.

3.3. Set Approximations

The indiscernibility relation will be used next to define approximations, basic concepts of rough set theory. The notions of lower and upper approximations of a set can be defined as follows.

Definition 3.2. *Let* $S=(U,A,V,f)$ *be an information system, let B be any subset of A and let X be any subset of U. The B-lower approximation of X, denoted by* $\underline{B}(X)$ *and B-upper approximations of X, denoted by* $\overline{B}(X)$*, respectively, are defined by*

$$\underline{B}(X)=\{x\in U \mid [x]_B\subseteq X\} \text{ and } \overline{B}(X)=\{x\in U \mid [x]_B\cap X\neq\phi\}.$$

The accuracy of approximation (accuracy of roughness) of any subset $X\subseteq U$ with respect to $B\subseteq A$, denoted $\alpha_B(X)$ is measured by $\alpha_B(X)=|\underline{B}(X)|/|\overline{B}(X)|$, where $|X|$ denotes the cardinality of X. For empty set ϕ, $\alpha_B(\phi)=1$ is defined. Obviously, $0\le\alpha_B(X)\le 1$. If X is a union of some equivalence classes of U, then $\alpha_B(X)=1$. Thus, the set X is *crisp* (precise) with respect to B. And, if X is not a union of some equivalence classes of U, then $\alpha_B(X)<1$. Thus, the set X is *rough* (imprecise) with respect to B [11]. This means that the higher of accuracy of approximation of any subset $X\subseteq U$ is the more precise (the less imprecise) of itself.

Another important issue in database analysis is discovering dependencies between attributes. Intuitively, a set of attributes D depends totally on a set of attributes C, denoted $C\Rightarrow D$, if all values of attributes from D are uniquely determined by values of attributes from C. In other words, D depends totally on C, if there a functional dependency between values of D and C. The formal definition of attributes dependency is given as follows.

Definition 3.3 *Let* $S=(U,A,V,f)$ *be an information system and let D and C be any subsets of A. Attribute D is functionally depends on C , denoted* $C\Rightarrow D$*, if each value of D is associated exactly one value of C.*

3.4. Dependency of Attributes

Since information system is a generalization of a relational database. A generalization concept of dependency of attributes, called a *partial dependency* of attributes is also needed.

Definition 3.4 *Let* $S = (U, A, V, f)$ *be an information system and let D and C be any subsets of A. The dependency attribute D on C in a degree k* $(0 \le k \le 1)$*, is denoted by* $C \Rightarrow_k D$*, where*

$$k = \frac{\sum_{X \in U/D} |\underline{C}(X)|}{|U|} \tag{1}$$

Obviously, $0 \le k \le 1$. If all set X are crisp, then $k = 1$. The expression $\sum_{X \in U/D} |\underline{C}(X)|$, called a lower approximation of the partition U/D with respect to *C*, is the set of all elements of *U* that can be uniquely classified to blocks of the partition U/D, by means of *C*. *D* is said to be fully depends (in a degree of *k*) on *C* if $k = 1$. Otherwise, *D* is partially depends on *C*. Thus, *D* fully (partially) depends on *C*, if all (some) elements of the universe *U* can be uniquely classified to equivalence classes of the partition U/D, employing *C*.

3.5. Reducts and Core

A *reduct* is a minimal set of attributes that preserve the indiscernibility relation. A *core* is the common parts of all reducts. In order to express the above idea more precisely, some preliminaries definitions are needed.

Definition 3.5. *Let* $S = (U, A, V, f)$ *be an information system and let B be any subsets of A and let a belongs to B. It say that a is dispensable (superfluous) in B if* $U/(B - \{b\}) = U/B$*, otherwise a is indispensable in B.*

To further simplification of an information system, some dispendable attributes from the system can be eliminated in such a way that the objects in the table are still able to be discerned as the original one.

Definition 3.6. *Let* $S = (U, A, V, f)$ *be an information system and let B be any subsets of A. B is called independent (orthogonal) set if all its attributes are indispensable.*

Definition 3.7. *Let* $S = (U, A, V, f)$ *be an information system and let B be any subsets of A. A subset* B^* *of B is a reduct of B if* B^* *is independent and* $U/B^* = U/B$.

Thus a reduct is a set of attributes that preserves partition. It means that a reduct is the minimal subset of attributes that enables the same classification of elements of the universe as the whole set of attributes. In other words, attributes that do not belong to a reduct are superfluous with regard to classification of elements of the universe. While computing equivalence classes is straighforward, but the problem of finding minimal reducts in information systems is NP-hard. Reducts have several important properties. One of them is a *core*.

Definition 3.8. *Let* $S = (U, A, V, f)$ *be an information system and let B be any subsets of A. The intersection off all reducts of is called the core of B, i.e.,*

$$\text{Core}(B) = \bigcap \text{Red}(B),$$

Thus, the *core* of *B* is the set off all indispensable attributes of *B*. Because the core is the intersection of all reducts, it is included in every reduct, i.e., each element of the core belongs to some reduct. Thus, in a sense, the core is the most important subset of attributes, for none of its elements can be removed without affecting the classification power of attributes.

4. The Modeling Process

In this section, the process of this study is presented. There are eight main phases which are data acquisition, sound editing, data representation, feature extraction, data discretization, data cleansing, feature selection using proposed technique and feature validation via classification. Figure 1 illustrates the phases of this process. To conduct this study, the pre-processing steps (from data representation to feature selection using proposed model) are implemented in MATLAB version 7.6.0.324 (R2008a). Weka [10] is used for feature validation via classification using various techniques. All these processes are executed on a processor Intel Core 2 Duo CPUs. The total main memory is 2 gigabytes and the operating system is Windows Vista. The details of the modeling process as follows:

4.1. Data Acquisition, Sound Editing, Data Representation and Feature Extraction

The 150 sounds samples of Traditional Malay musical instruments were downloaded from personal [14] and *Warisan Budaya Malaysia* web page [15]. The dataset comprises four different families which are membranophones, idiophones, aerophones and chordophones. The distribution of the sounds into families is shown in Table 1. This original dataset is non-benchmarking (real work) data. It is well-known that the quality of the data is one of the factors that might affect the overall classification task. For that, the dataset is firstly edited and trimmed. Afterwards, two categories of features schemes which are perception-based and MFCC features were extracted. All 37 extracted features from these two categories are shown in Table 2. The first 1-11 features represent the perception-based features and 12-37 are MFCC's features. The mean and standard deviation were then calculated for each of these features. The details of these phases can be found in [16].

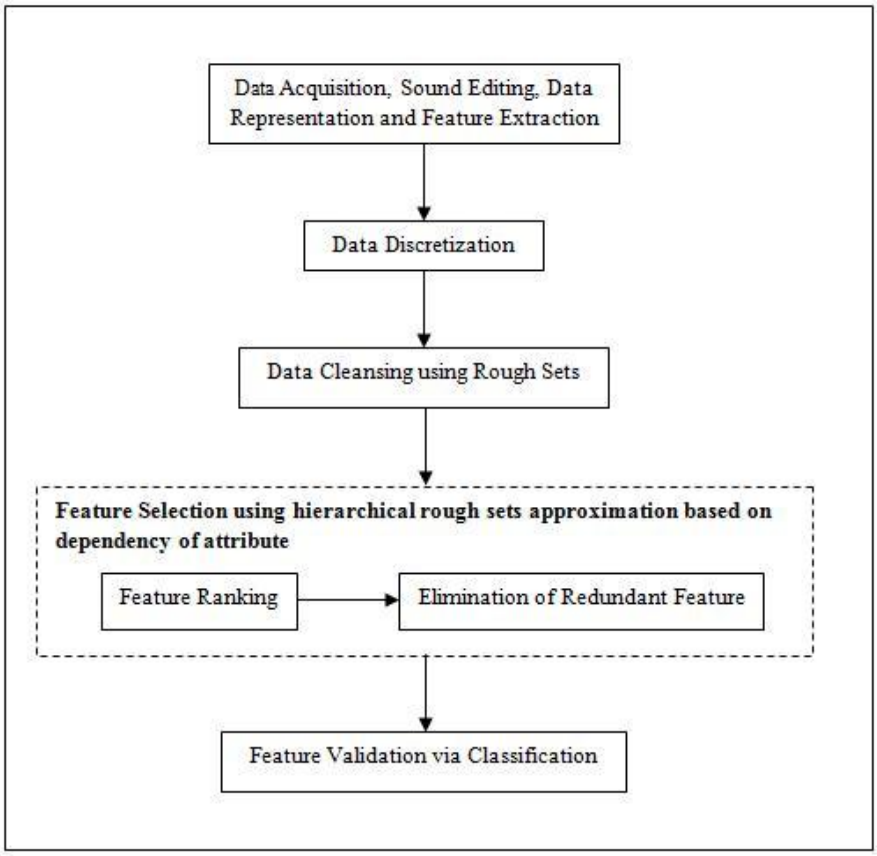

Figure 1. The modeling process for feature selection of the Traditional Malay musical instruments sounds classification.

Table 1. Data Sets

Family	Instrument
Membranophone	Kompang, Geduk, Gedombak, Gendang, Rebana, Beduk, Jidur, Marwas, Nakara
Idiophone	Gong, Canang, Kesi, Saron, Angklung, Caklempong, Kecerik, Kempul, Kenong, Mong, Mouth Harp
Aerophone	Serunai, Bamboo Flute, Nafiri, Seruling Buluh
Chordophone	Rebab, Biola, Gambus

Table 2. Features Descriptions

Number	Feature	Description
1	ZC	Zero Crossing
2	MEANZCR	Mean of Zero Crossings Rate
3	STDZCR	Standard Deviation of Zero Crossings
4	MEANRMS	Mean of Root-Mean-Square
5	STDRMS	Standard Deviation of Root-Mean-
6	MEANC	Mean of Spectral Centroid
7	STDC	Standard Deviation of Spectral Centroid
8	MEANB	Mean of Bandwidth
9	STDB	Standard Deviation of Bandwidth
10	MEANFLUX	Mean of Flux
11	STDFLUX	Standard Deviation of Flux
12	MMFCC1	Mean of the MFCCs #1
13	MMFCC2	Mean of the MFCCs #2
14	MMFCC3	Mean of the MFCCs #3
15	MMFCC4	Mean of the MFCCs #4
16	MMFCC5	Mean of the MFCCs #5
17	MMFCC6	Mean of the MFCCs #6
18	MMFCC7	Mean of the MFCCs #7
19	MMFCC8	Mean of the MFCCs #8
20	MMFCC9	Mean of the MFCCs #9
21	MMFCC10	Mean of the MFCCs #10
22	MMFCC11	Mean of the MFCCs #11
23	MMFCC12	Mean of the MFCCs #12
24	MMFCC13	Mean of the MFCCs #13
25	SMFCC1	Standard Deviation of the MFCCs #1
26	SMFCC2	Standard Deviation of the MFCCs #2
27	SMFCC3	Standard Deviation of the MFCCs #3
28	SMFCC4	Standard Deviation of the MFCCs #4
29	SMFCC5	Standard Deviation of the MFCCs #5
30	SMFCC6	Standard Deviation of the MFCCs #6
31	SMFCC7	Standard Deviation of the MFCCs #7
32	SMFCC8	Standard Deviation of the MFCCs #8
33	SMFCC9	Standard Deviation of the MFCCs #9
34	SMFCC10	Standard Deviation of the MFCCs #10
35	SMFCC11	Standard Deviation of the MFCCs #11
36	SMFCC12	Standard Deviation of the MFCCs #12
37	MFCC13	Standard Deviation of the MFCCs #13

4.2. Data Discretization

The features (attributes) extracted in the dataset is in the form of continuous value with non-categorical features (attributes). In order to employ the rough set approach in the proposed technique, it is essential to transform the dataset into categorical ones. For that, the discretization technique known as the equal width binning in [9] is applied. In this study, this unsupervised method is modified to be suited in the classification problem. The algorithm first sort the continuous valued attribute, then the minimum $x_{\min}$ and the maximum $x_{\max}$ of that attribute is determined. The interval width, *w*, is then calculated by:

$$w = \frac{x_{\max} - x_{\min}}{k^*},$$

where, $k*$ is a user-specified parameter for the number of intervals to discretize of each target class. The interval boundaries are specified as $x_{\min} + w_i$, where $i = 1,2,\cdots,k-l$. In this study, the *k* value is assigned to 3. Afterwards, the equal frequency binning method is used to divide the sorted continuous values into *k* interval where each interval contains approximately n/k data instances with adjacent values of each class.

4.3. Data Cleansing using Rough Set

As mentioned in Section 1, the dataset used in this study is raw data obtained from multiple resources (non-benchmarking data). In sound editing and data representation phases, the reliability of the dataset used have been assessed. However, the dataset may contain irrelevant features. Generally, the irrelevant features present in the dataset are features that having no impact on processing performance. However, the existence of these features in the dataset might increase the response time. For that, in this phase, the data cleansing process based on rough sets approach explained in sub-section 3.5 is performed to eliminate the irrelevant features from the dataset.

4.4. The Proposed Technique

In this phase, the construction of the feature selection technique using rough set approximation in an information system based on dependency of attributes is presented. The idea of this technique is derived from [7]. The relation between the properties of roughness of a subset $X \subseteq U$ with the dependency between two attributes is firstly presented as in Proposition 3.1.

Proposition 3.1. *Let* $S = (U, A, V, f)$ *be an information system and let D and C be any subsets of A. If D depends totally on C, then*

$$\alpha_D(X) \le \alpha_C(X),$$

for every $X \subseteq U$.

Proof. Let *D* and *C* be any subsets of *A* in information system $S = (U, A, V, f)$. From the hypothesis, the inclusion $IND(C) \subseteq IND(D)$ holds. Furthermore, the partition U/C is finer

than that U/D, thus, it is clear that any equivalence class induced by $IND(D)$ is a union of some equivalence class induced by $IND(C)$. Therefore, for every $x \in X \subseteq U$, the property of equivalence classes is given by

$$[x]_C \subseteq [x]_D .$$

Hence, for every $X \subseteq U$, we have the following relation

$$\underline{D}(X) \subseteq \underline{C}(X) \subset X \subset \overline{C}(X) \subseteq \overline{D}(X).$$

Consequently,

$$\alpha_D(X) = \frac{|\underline{D}(X)|}{|\overline{D}(X)|} \le \frac{|\underline{C}(X)|}{|\overline{C}(X)|} = \alpha_C(X).$$

The generalization of Proposition 3.1 is given below.

Proposition 3.2. *Let* $S = (U, A, V, f)$ *be an information system and let* $C_1, C_2, \cdots, C_n$ *and D be any subsets of A. If* $C_1 \Rightarrow_{k_1} D, C_2 \Rightarrow_{k_2} D, \cdots, C_n \Rightarrow_{k_n} D$, *where* $k_n \le k_{n-1} \le \cdots \le k_2 \le k_1$, *then*

$$\alpha_D(X) \le \alpha_{C_n}(X) \le \alpha_{C_{n-1}}(X) \le \cdots \le \alpha_{C_2}(X) \le \alpha_{C_1}(X),$$

for every $X \subseteq U$.

Proof. Let $C_1, C_2, \cdots, C_n$ and D be any subsets of A in information system S. From the hypothesis and Proposition 3.1, the accuracies of roughness are given as

$$\begin{gathered} \alpha_D(X) \le \alpha_{C_1}(X) \\ \alpha_D(X) \le \alpha_{C_2}(X) \\ \vdots \\ \alpha_D(X) \le \alpha_{C_n}(X) \end{gathered}$$

Since $k_n \le k_{n-1} \le \cdots \le k_2 \le k_1$, then

$$\begin{gathered} [x]_{C_n} \subseteq [x]_{C_{n-1}} \\ [x]_{C_{n-1}} \subseteq [x]_{C_{n-2}} \\ \vdots \\ [x]_{C_2} \subseteq [x]_{C_1} . \end{gathered}$$

Obviously,

$$\alpha_D(X) \le \alpha_{C_n}(X) \le \alpha_{C_{n-1}}(X) \le \cdots \le \alpha_{C_2}(X) \le \alpha_{C_1}(X).$$

Figure 2 shows the pseudo-code of the proposed technique. The technique uses the dependency of attributes in the rough set theory in information systems. It consists of five main steps. The first step deals with the computation of the equivalence classes of each attribute (feature). The equivalence classes of the set of objects U can be obtained using the indiscernibility relation of attribute $a_i \in A$ in information system $S = (U, A, V, f)$. The second step deals with the determination of the dependency degree of attributes. The degree of dependency attributes can be determined using formula in equation (1). The third step deals with selecting the maximum dependency degree. Next step, the attribute is ranked with the ascending sequence based on the maximum of dependency degree of each attribute. Finally, all the redundant attributes with similar maximum of dependency degree are deleted.

```
Algorithm: FSDA
Input: Data set with categorical value
Output: Selected non-redundant attribute
Begin
    Step 1. Compute the equivalence classes using
    the indiscernibility relation on each
    attribute.
    Step 2. Determine the dependency degree of
    attribute a_i with respect to all a_j, where
    i ≠ j.
    Step 3. Select the maximum of dependency
    degree of each attribute.
    Step 4. Rank the attribute with ascending
    sequence based on the maximum of dependency
    degree of each attribute.
    Step 5. Delete the redundant attribute with
    similar value of the maximum of dependency
    degree of each attribute.
End
```

Figure 2. The FSDA algorithm

The example to find the degree of dependency of attributes of an information system based on formula in equation (1) will be illustrated as in Example 4.1.

Example 4.1. To illustrate in finding the degree of dependency of attributes, the information system as shown in Table 3 is considered. From Table 3, based on each attribute, there are four partitions of U induced by indiscernibility relation on each attribute, i.e.

$$U/A = \{\{1,2,5\},\{3,4\}\},\ U/B = \{\{1\},\{2,3,4,5\}\},$$
$$U/C = \{\{1,2,3,4\},\{5\}\} \text{ and } U/D = \{\{1\},\{2,5\},\{3,4\}\}.$$

Based on formula in equation (1), the degree of dependency of attribute *B* on attribute *A*, denoted $A \Rightarrow_k B$, can be calculated as follows.

$$A \Rightarrow_k B, \quad k = \frac{\sum_{X \in U/B} |\underline{A}(X)|}{|U|} = \frac{|\{3,4\}|}{|\{1,2,3,4,5\}|} = 0.4 .$$

Using the same way, the following degrees are obtained

$$B \Rightarrow_k C, \quad k = \frac{\sum_{X \in U/C} |\underline{B}(X)|}{|U|} = \frac{|\{1\}|}{|\{1,2,3,4,5\}|} = 0.2 .$$

$$C \Rightarrow_k D, \quad k = \frac{\sum_{X \in U/D} |\underline{C}(X)|}{|U|} = \frac{|\{5\}|}{|\{1,2,3,4,5\}|} = 0.2 .$$

The degree of dependency of all attributes of Table 2 can be summarized as in Table 3.

Table 3. The degree of dependency of attributes of Table 2

Attribute (Depends on)	Degree of dependency			Maximum Dependency of Attribute
A	*B*	*C*	*D*	1
	0.2	0.2	1	
B	*A*	*C*	*D*	1
	0.4	0.2	1	
C	*A*	*B*	*D*	0.6
	0.4	0.2	0.6	
D	*A*	*B*	*C*	0.4
	0.4	0.2	0.2	

From Table 3, the attributes A, B, C and D are ranked based on the maximum degree of dependency. It can be seen that the attributes *A* and *B* have similar maximum degree of dependency. In order to select the best attributes and reduce the dimensionality respectively, only one of the redundant attributes will be chose. To do this, the selection approach in [7] is adopted where it is suggested to look at the next highest of maximum degree of dependency within the attributes that are bonded and so on until the bind is broken. In this example, attribute *A* is deleted from the list.

4.5. Feature Evaluation via Classification

The performance of the selected features generated from sub-section 4.4 is then further evaluated using three different classifiers which are SVM, *k*-NN and Naïve Bayes. Each classifier performs based on 10-fold cross validation. SMO algorithm in Weka[10] is applied for SVM classifier, with RBF kernels and a C value of 100. The kernel estimation during

training is utilized for Naïve Bayes classifier. The 1-NN is used in this study. All these classifiers are executed as multiple classifiers in Weka [10].

The accuracy rate achieved by the classifiers is then analysed to identify the effectiveness of the selected features. Achieving a high accuracy rate is important to ensure that the selected features are the best relevance features that perfectly serve to the classification architecture which able to produce a good result. At the end of this phase, the result is compared between full features and selected features. This is done in order to identify the effectiveness of the selected features in handling classification problem.

5. Results and Discussion

The aim of this study is to identify the best features using the proposed technique. Afterwards, the performance of the selected features is evaluated using five different classifiers which are SVM, 1-NN, Naïve Bayes, J48 and PART. The assessment of the performance is based on the accuracy rate obtained.

From the data cleansing step, it is found that {MMFCC1, SMFCC1} is the dispensable (irrelevant) set of features. It is means that the number of the relevant features is 35 out of 37 of original full features. After that, the proposed technique is employed to identify the best features set. As demonstrated in Table 4, all the 35 relevant features are ranked in ascending sequence based on the value of the maximum degree of attribute dependency. From the table, it is interesting to see that some of the features adopted in this study are redundant. In order to reduce the dimensionality of the dataset, only one of these redundant features is selected. It is revealed that the propose feature selection technique able to select the best 17 features out of 35 features available. The best selected features are given in Table 5.

Table 4. Feature Ranking using Proposed Method

Number of Features	Name of Features	Maximum Degree of Dependency of Attributes
3	STDZCR	0.826165
36	SMFCC12	0.655914
23	MMFCC12	0.52509
24	MMFCC13	0.52509
22	MMFCC11	0.237455
30	SMFCC6	0.208781
31	SMFCC7	0.208781
1	ZC	0.193548
37	SMFCC13	0.1819
32	SMFCC8	0.108423
33	SMFCC9	0.108423
34	SMFCC10	0.108423
35	SMFCC11	0.108423
27	SMFCC3	0.087814
29	SMFCC5	0.087814
11	STDFLUX	0.077061
21	MMFCC10	0.077061

20	MMFCC9	0.074373
6	MEANC	0.065412
19	MMFCC8	0.065412
18	MMFCC7	0.056452
28	SMFCC4	0.056452
7	STDC	0.042115
8	MEANB	0.042115
9	STDB	0.042115
13	MMFCC2	0.031362
16	MMFCC5	0.031362
17	MMFCC6	0.031362
5	STDRMS	0.021505
10	MEANFLUX	0.011649
2	MEANZCR	0
4	MEANRMS	0
14	MMFCC3	0
15	MMFCC4	0
26	SMFCC2	0

Table 5. The Best Selected Features

Number of Features	Name of Features	Maximum Degree of Dependency of Attributes
3	STDZCR	0.826165
36	SMFCC12	0.655914
23	MMFCC12	0.52509
22	MMFCC11	0.237455
30	SMFCC6	0.208781
1	ZC	0.193548
37	SMFCC13	0.1819
32	SMFCC8	0.108423
27	SMFCC3	0.087814
11	STDFLUX	0.077061
20	MMFCC9	0.074373
6	MEANC	0.065412
18	MMFCC7	0.056452
7	STDC	0.042115
13	MMFCC2	0.031362
5	STDRMS	0.021505
10	MEANFLUX	0.011649

After that, two datasets which consists of the full features and the selected features (generated from the propose technique) are used as an input to classify the Traditional Malay musical instruments sounds into four families which are membranophone, idiophone, chordophone and aerophone. This approach is meant to assess the performance of the best 17 selected features as compared to the 35 full features. For that, five different classifiers which are SVM, 1-NN, Naïve Bayes, J48 and PART are exploited.

From Table 6, the finding shows that for all three classifiers, the dataset with full features achieve classification accuracy above 95%. From these classifiers, 1-NN produces highest accuracy rate with 99.46% followed by SVM with 98.39% and Naïve Bayes with 94.98%. Even the classification accuracy decreases along with the number of features (selected features) in SVM and Naïve Bayes, the classification accuracy is still quite satisfactory with the overall performance achieved above 90%.

Table 6. The Comparison of Features Performance of Three Classifiers

Features	SVM	Naïve Bayes	1-NN	J48	PART
All 35	98.39%	94.98%	99.46%	98.30%	97.94%
Best 17	92.92%	93.73%	99.82%	98.30%	98.03%

However, by using J48 classifier, the accuracy rate of the selected features is sustained at 98.30% as compared to full features. Surprisingly, it is fascinating to see that the performance of the selected features is increases in 1-NN and PART with the accuracy rate of 99.82% and 98.03% respectively. Furthermore, the 1-NN classifier with the selected features produces the highest classification rate up to 99.82%. These promising results show that the propose feature selection technique able to select the best features for Traditional Malay musical instruments sounds.

6. Conclusion

In this study, the effectiveness of the propose feature selection technique based on rough set theory for the problem of Traditional Malay musical instruments sounds is examined. To perform this task, two categories of features schemes which are perception-based and MFCC which consist of 37 attributes are extracted. In order to utilize rough set theory, the equal width and equal frequency binning discretization technique is then employed to transform the original dataset with continuous-value (non-categorical features) into categorical form. The proposed technique is then adopted for feature selection through feature ranking and dimensionality reduction. Finally, five classifiers which are SVM, 1-NN, Naïve Bayes, J48 and PART in Weka is used to evaluate the performance of the selected features in terms of the accuracy rate obtained.

The finding shows that the propose technique able to generate 17 best features from 35 full features successfully. The result of low performance of the selected features in SVM and Naïve Bayes classifiers shows that it may not scale well to certain classifiers and dataset used. However, with the highest classification rate produce by 1-NN, J48 and PART classifiers, it shows that the propose feature selection technique able to identify the most relevant features and reduce the complexity process correspondingly. In addition, it also indicates that this propose technique might well perform with others classifiers and dataset. All these show that more studies is needed to overcome these problems. Thus, the future work will investigate the

effectiveness of the proposed technique towards other musical instruments sounds domain and apply different types of classifiers to validate the performance of the selected features.

Acknowledgements

This work was supported by the Universiti Tun Hussein Onn Malaysia (UTHM). Special thanks to Dr. Musa Mohd Mokji from Faculty of Electrical Engineering, Universiti Teknologi Malaysia for his kindness in teaching the Matlab programming for Digital Signal Processing.

References

[1] Guyo, I., and Elisseeff, A. An Introduction to Variable and Feature Selection. Journal of Machine Learning Research, 3, 1157–1182 (2003)

[2] Pawlak, Z.: Rough Sets. International Journal of Computer and Information Science, 11, 341–356 (1982)

[3] Banerjee, M., Mitra, S., and Anand, A.: Feature Selection using Rough Sets. In M. Banerjee *et al.*: Multi-Objective Machine Learning, Studies in Computational Intelligence Springer 16, 3–20 (2006)

[4] Czyzewski, A.: Soft Processing of Audio Signals. In Polkowski, L. & Skowron, A. (Eds.) Rough Sets in Knowledge Discovery: 2: Applications, Case Studies and Software Systems, 147-165, Heidelberg: Physica Verlag (1998)

[5] Wieczorkowska, A.: Rough Sets as a Tool for Audio Signal Classification. In Ras, Z.W. & Skowron, A. (Eds) Foundation of Intelligent Systems: Proceedings of the 11th International Symposium of Foundations of Intelligent Systems (ISMIS-99), 365-375, Springer-Verlaq Berlin Heildelberg (1999).

[6] Kostek, B., Szczuko, P., and Zwan, P.: Processing of Musical Data Employing Rough Sets and Artificial Neural Networks. In Tsumoto, S. *et al.* (Eds) RSCTC 2004, Lecture Notes in Artificial Intelligence 3066, 539-548, Springer-Verlag Berlin Heidelberg (2004)

[7] Herawan, T., Mustafa, M.D., and Abawajy, J.H.: Rough set approach for selecting clustering attribute. Knowledge Based Systems, 23 (3), 220–231, (2010)

[8] Senan, N., Ibrahim, R., Nawi, N.M, and Mokji, M.M.: Feature Extraction for Traditional Malay Musical Instruments Classification. In Proceeding of International Conference of Soft Computing and Pattern Recognition, SOCPAR '09, 454–459, (2009)

[9] Palaniappan, S., and Hong, T.K.: Discretization of Continuous Valued Dimensions in OLAP Data Cubes. International Journal of Computer Science and Network Security, 8, 116–126, (2008)

[10] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. H..: The WEKA Data Mining Software: An Update; SIGKDD Explorations, 11(1) (2009)

[11] Liu, M., and Wan, C.: Feature Selection for Automatic Classification of Musical Instrument Sounds. In Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '01, 247–248 (2001)

[12] Deng, J.D., Simmermacher, C., and Cranefield. S.: A Study on Feature Analysis for Musical Instrument Classification. IEEE Transactions on System, Man, and Cybernetics-Part B: Cybernetics 38 (2), 429–438, (2008)

[13] Benetos, E., Kotti, M., and Kotropoulus, C.: Musical Instrument Classification using Non-Negative Matrix Factorization Algorithms and Subset Feature Selection. In Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2006, 5, 221–224, (2006)

[14] Warisan Budaya Malaysia: Alat Muzik Tradisional, http://malaysiana.pnm.my/kesenian/Index.htm

[15] Shriver, R.: Webpage, www. rickshriver.net/hires.htm

[16] Senan, N., Ibrahim, R., Nawi, N.M., Mokji, M.M., and Herawan, T.: The Ideal Data Representation for Feature Extraction of Traditional Malay Musical Instrument Sounds Classification. In De-Shuang Huang et al. ICIC 2010, Lecture Notes in Computer Science 6215, pp. 345-353, 2010. © Springer-Verlag (2010)

Authors

Norhalina Senan received her B. Sc. and M.Sc. degree in Computer Science in Computer Science from Universiti Teknologi Malaysia. She is currently pursuing her study for Ph.D. degree in Feature Selection of Traditional Malay Musical Instruments Sounds Classification using Rough Set at Universiti Tun Hussein Onn Malaysia. She is a lecturer at Faculty of Computer Science and Information Technology, Universiti Tun Hussein Malaysia. Her research area includes data mining, multimedia and rough set.

Rosziati Ibrahim is a Professor at Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia. She obtained her PhD in Software Specification from the Queensland University of Technology (QUT), Brisbane and her MSc and BSc (Hons) in Computer Science and Mathematics from the University of Adelaide, Australia. Her research area is in Software Engineering that covers Software Specification, Software Testing, Operational Semantics, Formal Methods, Data Mining, Image Processing and Object-Oriented Technology.

Nazri Mohd Nawi is a Associate Professor at Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia. He received his B.S. degree in Computer Science from University of Science Malaysia (USM), Penang, Malaysia.His M.Sc.degree in computer science was received from University of Technology Malaysia (UTM), Skudai, Johor, Malaysia. He obtained his Ph.D.degree in Mechanical Engineering department, Swansea University, Wales. His research interests are in optimization, data mining techniques and neural networks.

Iwan Tri Riyadi Yanto received his B.Sc degree in Mathematics from Universitas Ahmad Dahlan Yogyakarta Indonesia. He obtained his M.IT from Universiti Tun Hussein Onn Malaysia. Currently, he is a lecturer at Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Ahmad Dahlan (UAD). He published more than 15 research papers in journals and conferences. His research area includes numerical optimization, data mining and KDD.

Tutut Herawan received his B.Ed and M.Sc degrees in Mathematics from Universitas Ahmad Dahlan and Universitas Gadjah Mada Yogyakarta Indonesia, respectively. He obtained his Ph.D from Universiti Tun Hussein Onn Malaysia. Currently, he is an assistant professor at Computer Science Program and head of Database and Knowledge Management (DBKM) research group at Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang (UMP). He published more than 50

research papers in journals and conferences. His research area includes data mining and KDD, rough and soft set theories.