Audio Feature Selection Based on Rough Set

Xiao-Li Li, Zhen-Long Du, Tong Wang, Dong-Mei Yu
lixiaoli@lut.cn, duzhl@cad.zju.edu.cn, zjxywt@163.com , yudm@lut.cn

School of Computer and Communication,
Lanzhou University of Technology, Lanzhou, 730050
P.R. China

Abstract

Keeping audio features is important for audio index. However, in most cases the features number is huge, thus direct processing is time-consuming. Feature selection, as a preprocessing step of data mining, has turned to be very efficient in reducing dimensionality and removing irrelevant data. In this paper, we propose a feature selection algorithm based on Rough Set theory, which could find out the feature subset from audio stream. The definition of *discernibility of ordinal attributes set* is introduced to discover the subsets containing implicit features. Moreover, based on the discernibility definition, we consider the discernibility of two and three *ordinal attributes set*, together with the discernibility of individual attribute, thus the extracted *reduct* is more complete and meaningful, which is consistent with the experimental evaluations.
Keyword: audio feature, Rough Set theory, ordinal attributes set, frame discernibility

I. Introduction

With the advance of digitizing technique, more and more media archives are produced. The media is helpful for people to readily acquire the information. How to efficiently index or reduce the audio archives is urgent to people's usage. To our knowledge, no one has applied rough set theory to stream data processing. In this paper, we focus on the construction of optimal feature selection from audio media based on rough set theory.

Audio media has its traits. Firstly, the common auditory sound is the keynote syntheses of different frequencies. For instance, our enjoying music is the sound combination of musical instruments, background music and singer voice etc. The sound with long duration is meaningful, and the instantaneous sound is hardly hearable. Ordinarily speaking, the audio datum shows the high dimensionality characteristic, the low dimensionality model is poor to represent the audio. Finally, audio data explicitly and/or implicitly contain many auditory features, such as accent, tone, style, tempo, rhythm, and so on. All of these require peculiar methods to be applied to the audio.

Feature selection [2–4, 8, 9, 11, 12] is a process to find an optimal feature subset from the given data set according to the given goal and criterion. Feature selection should diminish the cardinality of the feature subset and ensure that the classification accuracy does not significantly decrease. Existing methods of feature selection can be categorized into two classes [9, 12]: filter approach and wrapper approach. The filter-based feature selection algorithm is performed as a preprocessing step for information induction, and the resultant feature subset is accurate with extra running spend. The wrapper-based one performs iteratively, the optimal accuracy may be set manually for controlling the computational cost. The existing feature selection algorithms [3, 4, 9] mainly examine the time-less information. But if consider the multiple ordinal objects along with individual object, we would discover more knowledge from the same given data set.

Audio stream is composed of ordinal audio elements, whose concerned attributes also show the ordinal characteristic. How to extract a compact reduct from ordinal data set is discussed in this paper. We exploit twelve **MFCC** (Mel Frequency Cepstral Coefficients) along with **Energy** value of each audio frame as data set, examine the discernibility of individual attribute, two ordinal attributes and three ordinal attributes, then extract reduct from the gained discernible attributes sets. Experiments show the proposed algorithm could find out the reduct, which could maximally preserve the feature of stream data.

The paper is organized as follows: Section two is some rough sets concepts related to feature selection. The following one introduces the optimal feature selection algorithm from multiple ordinal attributes. Next presents the experimental evaluation. Final gives the concluding remarks.

II. Feature Selection Based on Rough Set

Rough set theory was introduced by Pawlak [1] in early 1980s as a mathematical tool to deal with uncertainty problem. In rough set theory, data is stored in a table, which may be called decision table. Rows of the decision table stand for objects, and columns show attributes. And a decision table is denoted as $T=(U, A, C, D)$, where U is a non-empty finite object/instance set, A is a finite set of attributes. The attributes in A is further classified into two disjoint subsets, condition attribute set C and decision attribute set D. $A=C\cup D$, and $C\cap D=\varnothing$.

A. Rough Set Attribute Reduction

Rough Set Attribute Reduction (RSAR) employs simple set operations for extracting condense knowledge. RSAR is particularly domain independent, requiring no human intervention and no additional parameters. RSAR assumes that the data is time-independent and immutable clusters in datasets, which is not the case in the real world. Moreover, the reduced attributes is regarded as a significant omission.

Central to RSAR is the concepts of *reduct* and *core*. The equivalence relation R is employed for classification of U. The pair $apr = (U, R)$ is called an approximation space. R partitions U into disjoint subsets, which is denoted by U/R.

$$IND(R) = \bigcap [x]_R \tag{1}$$

The equivalence relation may be regarded as the available knowledge for the considering objects. For an arbitrary set $X\subseteq U$, it may be impossible to describe X precisely using the equivalence classes of R. The *lower bound* and *upper bound* approximation are employed for rough representation of X. An element in the *lower bound* approximation necessarily belongs to X, while an element in the *upper bound* approximation possibly belongs to X. The *boundary set* **BND** is composed of the element which belongs to *upper bound* but not to *lower bound* approximation.

$$\underline{R}X = \{x\in U \mid [x]_R \subseteq X\} \tag{2}$$

$$\overline{R}X = \{x\in U \mid [x]_R \cap X \neq \varnothing\} \tag{3}$$

$$BND(X) = \overline{R}X - \underline{R}X \tag{4}$$

Positive Region $POS_R(X)$ actually equals to the lower bound approximation $\underline{R}X$, while *negative region* $NEG_R(X)$ is the complement set of $POS_R(X)$. For an element of $x\in POS_R(X)$, it certainly belongs to X, while one of $x\in NEG_R(X)$, it is uncertain to decide whether the element belongs to X.

Attribute dependency is important issue to attribute reduction. Attribute reduction techniques eliminate superfluous attributes and create a minimal sufficient subset of attributes of considering knowledge. Such minimal sufficient subset of attributes, called a *reduct*, is an essential part of knowledge. The including information of *reduct* is the same as the original. The set of attributes common to all *reducts* of set C is called the *core* of C. The dependency degree of two attribute sets of P and Q is defined by introducing the cardinality $|\cdot|$:

$$\kappa = \frac{|POS_P Q|}{|U|} \tag{5}$$

B. Rough Set Ordinal Attribute Reduction

The purpose of feature selection is to identify the significant features, eliminate the irrelevant or dispensable features to the knowledge. The benefits of feature selection are twofold: it considerably decreases the running time of the reduction algorithm, and increases the accuracy of the resulting model.

Discernibility

Let $x, y \in U$ be any two distinct objects. The *discernibility* between x and y is defined as:

$$\alpha(x, y) = \{q \in A \mid f(x, q) \neq f(y, q)\} \tag{6}$$

$\alpha(x, y)$ is composed of attributes set which can discern object x and y.

Discernibility matrix could distinguish groups of objects. It is the collection of *discernibility* of any two objects, which entirely reflects the discernibility of U.

Discernibility Function for $<U, R>$ can be defined as:

$$\square = \wedge_{1<j<i<n} \vee \alpha(x, y) \tag{7}$$

where $\wedge$ and $\vee$ are boolean conjunction and disjunction operations.

U	x_1	x_2	x_3	x_4
μ_1	2	0	1	1
μ_2	2	1	0	1
μ_3	1	0	0	2

(a) Data set

	1	2	3
1			
2	$\{x_2, x_3\}$		
3	$\{x_1, x_3, x_4\}$	$\{x_1, x_2, x_4\}$	

(b) Discernibility matrix

Table 1 Example of Discernibility Matrix

Take Table 1(a) as example data, *discernibility matrix* is given in Table 1(b), and *discernibility function* is derived as bellow [1, 7]:

$$\begin{aligned} \square &= (x_2 \vee x_3) \wedge (x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee x_2 \vee x_4) \\ &= x_2 x_3 \vee x_3 x_4 \end{aligned}$$

The *reduct* and *core* can be found out by exploiting the *discernibility function*. Each entry of *discernibility matrix* only shows the discernibility of pairs object. As we know, some features involve several objects, for example, multiple temporal consecutive objects. In this case, the induced *reduct* and *core* by direct reduction are sufficient but not necessary to original information, especially for the continuous stream. Because the induction may cause the information lost, and the generated *reduct* could not faith to the original information. Though the removed object is unnecessary when the reduction only considers individual object, it may be useful to represent the latent knowledge. In this paper, the examine data is audio media, then the focus of feature selection is on multiple ordinal objects. We exploit the discernibility of the ordinal consecutive object to complete the lost knowledge of reduct, make the *reduct* hold as maximal as possible features. Hence we introduce the *discernibility of of Ordinal Attributes Set* for reduction procedure.

Compatible

Düntsch [10] ever presented the *compatible* concept for restricted relation analysis, inspired by his work, in this paper, we make some modification to *compatible* for purpose of stream data reduction. We suppose that the attribute set is order, denoted as $\leq_p$, p is the given predicate or condition.

For $x_i, x_j \subseteq X$, and let

$$x_i \leq_p^+ x_j \quad \text{if and only if } (\forall a \in x_i, \forall b \in x_j)(a \leq_p b)$$

Then two classification $V_i \subseteq U/x_i$ and $V_j \subseteq U/x_j$ is *compatible* if either

$$V_i \leq^+ V_j \;\text{ or }\; V_i \geq^+ V_j$$

The compatible classification is regarded as indiscernible in the sense of data variation trends. Such as, our heard sound involves the repetitively varying acoustic intensity, which relates less to the audio features, and should remove from *reduct*. Hence the *compatible* introduction is necessary to keep the reduct compact when reducing the ordinal attributes data.

Discernibility of Ordinal Attributes Set

For any two compatible classification A and B, if they are *Compatible*, they would be indiscernible. The dependency degree is defined as that:

$$\gamma_{A\leq^+ B} = \frac{\bigcup |M|}{|U|} \quad \text{and} \quad M \subseteq A \leq^+ B$$

III. Feature Selection Based on Rough Set

U	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	v
μ_1	-11.40	14.54	14.15	3.13	-1.03	0.83	-4.17	-0.32	0.62	-4.09	0.62	-0.04	-41.10
μ_2	-12.35	18.29	13.94	3.17	-0.18	-1.10	-4.39	0.81	0.62	-4.10	-2.13	0.61	-41.03
μ_3	-12.74	18.83	11.40	2.59	0.48	0.31	-4.14	-0.73	-0.30	-2.04	-0.98	-2.50	-42.21
μ_4	-16.94	13.90	10.34	5.36	1.23	0.84	-3.40	-3.75	-1.74	-1.85	0.16	-0.82	-46.57
μ_5	-36.65	-0.77	-0.79	-0.18	-1.97	-0.96	0.87	-0.52	0.47	-0.88	1.40	2.42	-49.58
μ_6	-16.46	9.06	-6.87	-9.37	1.05	1.18	4.09	-1.93	-0.63	-3.14	-0.31	1.06	-42.61
μ_7	-1.12	11.24	1.47	-18.73	3.03	3.78	-3.57	-1.25	-1.51	-2.19	-1.77	1.75	-31.90
μ_8	-4.45	3.28	0.38	-16.28	7.76	6.70	-6.70	1.19	-1.67	-2.52	-2.11	-3.99	-29.76
μ_9	-5.10	5.07	-1.03	-11.61	12.09	3.35	-5.96	1.71	-5.87	0.49	-0.41	-5.73	-32.42
μ_{10}	-3.40	12	2.70	-2.77	1.92	2.66	-2.13	-0.69	-3.98	-1.72	-0.17	-4.08	-38.30

Table 2. Audio Frame Data Characterized by *MFCC* and *Energy*

In Table 2, we choose 10 representative audio frames $\{\mu_1, \mu_2, \cdots, \mu_9, \mu_{10}\}$ as illustrative example of stream data, and make use of 12 *MFCC* $\{x_1, x_2, \cdots, x_{11}, x_{12}\}$ and *Energy* ν as frame attributes.

In audio reduction, the actual need is the *reduct* of audio frame, so we view rows as attribute set and the columns as objects, that is, the data of Table 2. after anticlockwise rotating 90^0 is then fed to reduction procedure. We name the proposed algorithm as *Rough Set Ordinal Attribute Reduction* (RSOAR).

During the RSOAR Implementation, the below logical conditions might be encounted for deciding the discernibility.

(1) $\mu_{\mathrm{i}} = \mu_{\mathrm{i}-1}$

(2) $\mu_{\mathrm{i}} \leq^{+} \mu_{\mathrm{i}-1}$ and $\dfrac{\min(\mu_{\mathrm{i}-1}, \mu_{\mathrm{i}})}{\max(\mu_{\mathrm{i}-1}, \mu_{\mathrm{i}})} > \eta$

(3) $\mu_{\mathrm{i}} \geq^{+} \mu_{\mathrm{i}-1}$ and $\dfrac{\min(\mu_{\mathrm{i}-1}, \mu_{\mathrm{i}})}{\max(\mu_{\mathrm{i}-1}, \mu_{\mathrm{i}})} > \eta$

(4) $\mu_{\mathrm{i}-1} \leq^{+} \mu_{\mathrm{i}} \leq^{+} \mu_{\mathrm{i}+1}$ and $\dfrac{\min(\mu_{\mathrm{i}-1}, \mu_{\mathrm{i}}, \mu_{\mathrm{i}+1})}{\max(\mu_{\mathrm{i}-1}, \mu_{\mathrm{i}}, \mu_{\mathrm{i}+1})} > \eta$

(5) $\mu_{\mathrm{i}-1} \geq^{+} \mu_{\mathrm{i}} \geq^{+} \mu_{\mathrm{i}+1}$ and $\dfrac{\min(\mu_{\mathrm{i}-1}, \mu_{\mathrm{i}}, \mu_{\mathrm{i}+1})}{\max(\mu_{\mathrm{i}-1}, \mu_{\mathrm{i}}, \mu_{\mathrm{i}+1})} > \eta$

The RSOAR algorithm is outlined as follows.

Input: Stream Data Set: *U*; Attributes Set: *A*

Output: Decision Data set: *D*

Step 1:

Generate discernibility matrix M_1, M_2 and M_3, which separatively examine the discernibility of individual attribute, two ordinal attributes and three ordinal attributes.

Step 2:

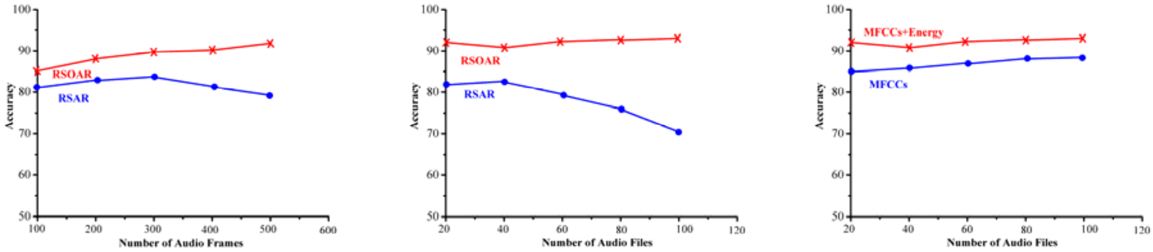
Generate the reduct $reduct_3$ of M_3, for Each element $r_i \in reduct_3$, it is kept in $reduct_2$ when $reduct_2$ is evaluated, and the final required *reduct* is the intersection of $reduct_2$ and $reduct_1$.

IV. Experiments Evaluation

Such audio features as musical structure, tempo, rhythm, melody, chord, and so on, could be employed for discriminating audio. In [5, 6] 12 MFCC coefficients were used for characterizing the audio spectral feature. Pye [6] concluded that MFCC has strong classifying ability and independent of music compression scheme. Hence we choose MFCC and Energy to characterize audio data.

We design three experiments for evaluating the compactness and completeness of gained reduct by the proposed algorithm. 50 pieces of folk music are chosen as experiment data, their playing time varies from 30 seconds to about 5 minutes, and about 25 pieces is sung by individual musical instrument, the reminder is by two or three instruments. We reduce all of them by RSAR and RSOAR separatively, The extracted reduct is used for the index pattern. The index *accuracy* is the ratio of the number of accurately indexing pieces to total number of pieces.

The first experiment is for test the influence of audio frame number to indexing accuracy, Fig. 1(a) shows that as frame number increases, the index accuracy of either RSAR or RSOAR increases; however, the index accuracy of RSAR decreases when it reaches a maximum, while the index ccuracy of RSOAR keeps increase. From the first experiment, we could conclude that reduct gained by RSOAR is more complete than reduct by RSAR. The second experiment is for comparison the index accuracy of RSAR and RSOAR, in Fig. 1(b), it is obvious that RSOAR is more effective than RSAR in reducing the stream data. The last experiment proves that selection of MFCC along with Energy as attributes set could preserve more knowledge than only employing MFCC, as illustration in Fig. 1(c).



(a) (b) (c)

Figure 1 Comparison of RSAR and RSOAR

V. Conclusion

In this paper, we present a novel algorithm to generate the minimal reduct from ordinal attribute sets. Conventional discernibility only considers the distinction of individual attribute, which cause the reduction of ordinal attribute sets to generate incomplete reduct. So we introduce the discernibility of ordinal attributes set, which fully reflects the discernibility of data. By examining the discernibility of individual attribute, as well as ordinal attributes set, our algorithm not only regains the lost features by RSAR, but also re-eliminates the redundant features.

Experiments show that in some cases redundant attributes determined by RSAR become necessary when the attribute set is ordinal. The extracted reduct by RSOAR is even meaningful, which is proved by higher audio indexing rate. Meanwhile, the gained reduct may be exploited for more intensive mining from stream data. The proposed algorithm is preference for the reduction of ordinal attributes set, but may be applied to the signal reduction.

References

[1] Z. Pawlak, “Rough Sets,” *Information Journal of Computer and Information Science*, 1982, vol. 11, no. 5, pp. 341–356.

[2] R. Jensen, Q. Shen, “Semantics-preserving dimensionality reduction: Rough and Fuzzy-Rough-Based approaches,” *IEEE Trans. on Knowledge and Data Engineering*, 2004, vol. no.12, pp. 1457–1471.

[3] A. An, Y. Huang, X. Huang, N. Cercone, “Feature selection with Rough Sets for web page classification,” *LNCS 3135*, Nov. 2004.

[4] M. Boussouf, M. Quafafou, “Scalable feature felection using Rough Set theory,” *LNCS 2005*, June, 2003.

[5] L. Lu, H. You, H.-J. Zhang, “A new approach to query by humming in music retrieval,” In *Proc. of International Conference on Multimedia and Expo (ICME’2001)*, 2001, pp. 776–779.

[6] D. Pye, “Content-Based methods for the management of digital music,” In *Proc. of ICASSP’2000*, 2000, vol. 22, no. 5, pp. 2437–2440.

[7] G.Y. Wang, “Attribute core of decision table,” In *Proc. of the Third International Conference on Rough Sets and Current Trends in Computing*, 2002.

[8] Y.Y. Yao, “Information granulation and rough set approximation,” *International Journal of Intelligent Systems*, 2001, vol. 16, no. 1, pp. 87–104.

[9] A. Chouchoulas, “Incremental feature selection based on rough set theory,” *PhD Disseration of the University of Edinburgh*, 2001.

[10] I. Düntsch, G. Gediga, “Relation restricted prediction analysis,” In *Proc. of the 15th IMACS World Congress*, 1997, pp. 619–624.

[11] H. Wang, D. Bell, F. Murtagh, “Axiomatic approach to feature subset selection based on relevance,” *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 1999, vol. 21, no. 3, pp. 271–277.

[12] M. Dash, H. Liu, “Feature selection for classification,” *Intelligent Data Analysis*, 1997, vol. 1, no. 3. pp. 131–156.

Xiao-Li Li received BSc from Northwest Normal University, China, in 1994, and MEng from Lanzhou University of Technology in 2005, China, respectively. Her research interests are machine learning, audio process etc.

Zhen-Long Du is Ph.D. candidate of State Key Lab of CAD&CG, Zhejiang University, China. His research focus is on multimedia process, computer vision, and computer graphics.