Proceedings of KFIS 2002 Spring Conference, 2002. 5. 25

사용자 독립적 특징 추출을 위한 연구

A Study on Subject Independent Feature Extraction

[1]방원철, [2]한정수, [1,2]변증남
[1]한국과학기술원 인간친화 복지 로봇시스템 연구센터, [2]한국과학기술원 전자전산학과

[1]Won-Chul Bang, [2]Jeong-Su Han and [1,2]Z. Zenn Bien
[1]HWRS-ERC, KAIST, [2]Dept. of EECS, KAIST
E-mail : bang@eeinfo.kaist.ac.kr

ABSTRACT

여러 사람에게서 생체신호를 측정하여 특징을 추출하는 경우 피실험자마다 다른 신체적 또는 생리학적 특징에 의해 같은 클래스로 분류하고 싶어도 다른 클래스로 잘못 분류되는 경우가 발생한다. 이와 같이 N 명의 사람에게서 얻은 생체신호로 M 개의 클래스를 분류하도록 훈련하여 새로운 사람의 생체신호를 M 개의 클래스로 분류하고자 할 때 발생하는 문제를 해결하기 위한 방법으로 피실험자 독립적인 클러스터링 방법을 제안하고자 한다. 이를 위한 수학적 기반으로 동치관계들의 교집합과 합집합에 근거한 새로운 연산자를 정의하고 이를 이용하여 최대 공통 클러스터(Largest Common Cluster; LCC)라는 새로운 개념을 정의한다. 이는 여러 사람에게서 얻은 정보에서 최대한 공통의 성질을 갖는 것들을 찾아내는 수학적이고 체계적인 방법이라 할 수 있다. 따라서 일단 LCC를 찾아내면 이를 특징(feature)으로 삼아 패턴분류기를 설계하면 여러 사람에게 적용가능한 생체신호 인식기를 설계할 수 있게 된다.

Key words : 특징 추출, 사용자 독립, 군집화, 패턴분류, 최대 공통 클러스터

I. Introduction

An effective way of extracting a feature set with good classification capability is first to generate as many induced feature vectors as possible from the bio-signals by proper clustering and determine those features that may reveal some common characteristics of most subjects. And then, a decision-making method, such as the rough set theory, is used to select a minimal feature set from all the features extracted.

One can easily think of a method in which the feature values from a subject are averaged and then clustering of the averaged values is performed to find borders between two clusters. This has been in fact a common practice in many earlier works. One may guess that finer clusters render better classification, but too many clusters result in procedural inefficiency due to redundant rules for classification. For the sake of practical solution, it would be essential to determine some appropriate number of clusters for each feature, but, again, this method of classification is only applicable to the subject from whom the features are extracted. To circumvent such difficulty of many conventional methods, we propose a method consisting of the following 4 steps:
(1) Increase the number of induced feature vectors until sufficiently many clusters are obtained by clustering the feature values,
(2) Determine some common

characteristics of clusters so as to provide with a property of subject-independency, (3) Minimize the number of induced feature vectors for effective classification, and (4) Select one set of features if there are multiple candidates for the minimal number of feature set. For subject-independency, we shall define a concept of the Largest Common Cluster in the next section. Using such a new concept, we shall show that one can extract features that are capable of classifying the bio-signals for many subjects. For minimal selection of features, we adopt the method of decision table in the framework of the rough set theory. For choosing one minimal feature set out of possibly many minimal feature sets, we use Bhattacharyya distance measure to determine the separation between any two clusters.

II. Set of the Largest Common Clusters

In this section, we introduce new concepts which enable to determine clusters common to a given number of subjects out of those clusters obtainable from a given feature.

Let U be a universe and E an equivalence relation over U. For each $x \in U$, we let $[x]_E$ denote the subset of U consisting of all elements which are equivalent to x with respect to E, i.e., $[x]_E = \{y \in U | xEy\}$. This set $[x]_E$ is referred to as the equivalence class determined by x based on E. Let E be a family of equivalence relations over U. Let $\bigcap \mathbf{E}$ denote the intersection of all equivalence relations belonging to E. Then $\bigcap \mathbf{E}$ is also an equivalence relation and its equivalence class based on $\bigcap \mathbf{E}$ is given by:

$$[x]_{\cap \mathbf{E}} = \bigcap_{E \in \mathbf{E}} [x]_E .$$

Also, the quotient set associated with $\bigcap \mathbf{E}$, $U/\bigcap \mathbf{E}$ is the family of all equivalence classes based on the equivalence relation $\bigcap \mathbf{E}$, i.e.,

$$U/\bigcap \mathbf{E} = \{ X | X = [x]_{\cap \mathbf{E}}, \forall x \in U \}.$$

Let $\bigcup \mathbf{E}$ denote the union of all equivalence relations belonging to E. Since $\bigcup \mathbf{E}$ is not an equivalence relation, neither its equivalence class nor the quotient set associated with $\bigcup \mathbf{E}$ can be defined. To make the case similar, however, we define

$$[x]_{\cup \mathbf{E}}{}^{*} = \bigcup_{E \in \mathbf{E}} [x]_E$$

and

$$U/\bigcup \mathbf{E}^{*} = \{ X | X = [x]_{\cup \mathbf{E}}{}^{*}, \forall x \in U \}.$$

The following example depicts the concepts of the intersection/union of all equivalence relations belonging to E.

Example 1 For

$$U = \{C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8\},$$

consider the family $\mathbf{E} = \{E_1, E_2, E_3, E_4\}$ of equivalence relations having the following equivalence classes:

$$U/E_1 = \{\{C_1\}, \{C_2, C_3, C_4\}, \{C_5, C_6, C_7, C_8\}\}$$
$$U/E_2 = \{\{C_1, C_2\}, \{C_3, C_4\}, \{C_5, C_6, C_7, C_8\}\}$$
$$U/E_3 = \{\{C_1, C_3\}, \{C_2, C_4\}, \{C_5, C_6, C_7, C_8\}\}$$
$$U/E_4 = \{\{C_1, C_2, C_3, C_4\}, \{C_5\}, \{C_6, C_7, C_8\}\}.$$

Then, we have

$$U/\bigcap \mathbf{E} = \{\{C_1\}, \{C_2\}, \{C_3\}, \{C_4\}, \{C_5\}, \{C_6, C_7, C_8\}\}$$

and

$$U/\bigcup \mathbf{E}^{*} = \{\{C_1, C_2, C_3, C_4\}, \{C_5, C_6, C_7, C_8\}\}.$$

We also need to define simple operations for sets as follows:

Definition 1 If X and Y are subsets of V, let

$$X \otimes Y = \{\{x, y\} \subset V \mid x \in X \text{ and } y \in Y\},$$

which implies the family of all sets which have two elements, one from X and the other from Y. Note that $X \otimes Y$ is different from the Cartesian product space, $X \times Y$.

Also, if X_i is a subset of V, for each $i \in \{1, 2, \cdots, I\}$, we denote

$$\coprod_{i \in \{1,2,\cdots,I\}} X_i = X_1 \otimes X_2 \otimes \cdots \otimes X_I$$
$$= \{\{x_1, x_2, \cdots, x_I\} \subset V \mid x_1 \in X_1 \text{ and } x_2 \in X_2 \text{ and} \cdots \text{and } x_I \in X_I\}$$

Now we are in the position to express the largest common clusters for a set of equivalence relations.

Definition 2 Let U be a universe, and E be a family of equivalence relations over U. The family of the *Largest Common Cluster sets* for E in U is defined by:

$$\coprod_{Y \in U/\bigcup \mathbf{E}^*} \{X \in U/\bigcap \mathbf{E} \mid \forall Z \in U/\bigcap \mathbf{E}, \operatorname{card}(X \cap Y) \geq \operatorname{card}(Z \cap Y)\}.$$

Example 2 Assume the universe of discourse U and the equivalence relations as in Example 1. Then the family of the largest common cluster sets for E in U is:

$$\coprod_{Y \in U/\bigcup \mathbf{E}^*} \{X \in U/\bigcap \mathbf{E} \mid \forall Z \in U/\bigcap \mathbf{E}, \operatorname{card}(X \cap Y) \geq \operatorname{card}(Z \cap Y)\}$$

$$= \{\{C_1\},\{C_2\},\{C_3\},\{C_4\}\} \otimes \{\{C_6,C_7,C_8\}\}$$
$$= \{\{\{C_1\},\{C_6,C_7,C_8\}\},\{\{C_2\},\{C_6,C_7,C_8\}\}, \{\{C_3\},\{C_6,C_7,C_8\}\},\{\{C_4\},\{C_6,C_7,C_8\}\}\}.$$

Thus,
$\{\{C_1\},\{C_6,C_7,C_8\}\}$, $\{\{C_2\},\{C_6,C_7,C_8\}\}$, $\{\{C_3\},\{C_6,C_7,C_8\}\}$ and $\{\{C_4\},\{C_6,C_7,C_8\}\}$
are the largest common cluster sets for E in U. In this case, we have four sets of the largest common clusters.

III. Conclusion

In this paper, we have proposed a method based on several soft computing techniques to extract induced feature vectors and determine a minimal feature set that shows high separability and is less subject-dependent. A new concept, the largest common cluster set consists of the clusters of which the union has the largest cardinality out of every set of clusters which are always able to be discerned by any equivalence relation.

According to the concept of the largest common cluster set, it should be further developed to construct a pattern classifier for signals from human which may easily cause subject dependency.

Acknowledgement: This work is partially supported by of Critical Technology 21 program of the Ministry of Science and Technology of Korea and Human-friendly Welfare Robot System Engineering Research Center (sponsored by KOSEF), KAIST.

IV. References

[1] Z. Pawlak, Rough sets: theoretical aspects and reasoning about data, Kluwer Academic Publishers, 1991.