task representing elements (like EEOPs and Facts) within the PS_E-Designer. Those elements will then be synchronized with the database of already existing task representing elements. On the other hand, the automatic introduction of the necessary interfaces within the reactive layer (skill server interfaces) as well as enhancements within the skill executor will be realized via interaction with the Rhapsody UML model. Finally, the programming interface for the abstract level, the PS_A-Designer, will be coupled to the PS_E-Designer so that a comprehensive Process Structure Design tool for MASSiVE will evolve.

References

[1] R. Bischoff, A. Kazi, and M. Seyfarth, "The MORPHA Style Guide for Icon-Based Programming," in *Proceedings of the 11th IEEE Int. Workshop on Robot and Human interactive Communication (ROMAN2002)*, (Berlin, Germany), pp. 482–487, Sep 25-27 2002.

[2] C. Martens, O. Prenzel, and A. Gräser, *The Rehabilitation Robots FRIEND-I & II: Daily Life Independency through Semi-Autonomous Task-Execution*. 2007. ISBN 978-3-902613-01-1, http://www.ars-journal.com/Reha.htm.

[3] R. Simmons, "Architecture, the backbone of robotic systems," in *Proceedings of the 2000 IEEE International Conference on Robotics and Automation*, (San Francisco, CA), Apr 2000.

[4] R. Bonasso, D. Kortenkamp, and D. Schreckenghost, D. Ryan, "Three tier architecture for controlling space life support systems," in *Proceedings of IEEE SIS'98*, (Washington DC, USA), 21 -23 May 1998.

[5] C. Schlegel and R. Wörz, "The software framework SmartSoft for implementing sensorimotor systems," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, vol. 3, pp. 1610–1616, Oct 1999.

[6] C. Martens, O. Prenzel, J. Feuser, and A. Gräser, "MASSiVE: Multi-Layer Architecture for Semi-Autonomous Service-Robots with Verified Task Execution," in *Proceedings of 10th Int. Conf. on Optimization of Electrical and Electronic Equipments OPTIM'06*, vol. 3, (Brasov, Romania), pp. 107–112, Transilvania University Press, Brasov, May 2006. ISBN 973-653-705-8.

[7] O. Prenzel, C. Martens, M. Cyriacks, C. Wang, and A. Gräser, "System-controlled user interaction within the service robotic control architecture massive," *Robotica*, vol. 25, no. 2, pp. 237–244, 2007.

[8] H. Kampe, "Verification of Petri-Net Based Process-Structures for the Programming of Service-Robots," tech. rep., University of Bremen, Institute of Automation (IAT), 2007. Student research project documentation, also presented on the 29th Colloquium of Automation, Salzhausen (2007), Germany.

[9] C. Tranoris and K. Thramboulidis, "An IEC-compliant Engineering Tool for Distributed Control Applications," in *Proceedings of 11th Mediterranean Conference on Control and Automation MED'03*, (Rhodes, Greece), June 18-20 2003.

[10] Fabien Chiron and Khalid Kouiss, "Design of IEC 61131-3 function blocks using SysML," in *Proc. of Mediterranean Conference on Control and Automation, 2007. MED '07.*, (Athens, Greece), pp. 1–5, 27 29 June 2007.

[11] T. Cao and A. C. Sanderson, "AND/OR Net Representation for Robotic Task Sequence Planning," *IEEE Transactions on Systems, Man, and Cybernetics - part C: Applications and Reviews*, vol. 28, May 1998.

[12] S. Russel and P. Norvig, *Artificial Intelligence - A Modern Approach*. Upper Saddle River, New Jersey: Prentice Hall, second ed., 2003.

[13] O. Prenzel, "Semi-Autonomous Object Anchoring for Service-Robots," in *Methods and Applications in Automation*, pp. 57–68, B. Lohmann (Ed.), A. Gräser, 2005.

[14] E. Gamma, R. Helm, R. E. Johnson, and J. Vlissides, *Design Patterns. Elements of Reusable Object-Oriented Software*. Amsterdam: Addison-Wesley Longman, 1 ed., 1995. ISBN 0201633612.

[15] B. P. Douglass, *Doing Hard Time: Developing Real-Time Systems with UML, Objects, Frameworks, and Patterns*. Addison-Wesley Professional, 1999.

[16] D. C. Schmidt, "Model-driven-engineering, guest editor's introduction," *IEEE Computer*, pp. 25–31, February 2006.

[17] Telelogic, "Telelogic web-page," 2008. http://www.telelogic.com.

[18] Trolltech, "Trolltech web-page," 2008. http://www.trolltech.com.

[19] Karl-Heinz John and Michael Tiegelkamp, *IEC 61131-3: Programming Industrial Automation Systems*. Springer, 2001. ISBN 3-540-67752-6.

[20] C. Martens, *Teilautonome Aufgabenbearbeitung bei Rehabilitationsrobotern mit Manipulator - Konzeption und Realisierung eines softwaretechnischen und algorithmischen Rahmenwerks*. PhD dissertation, University of Bremen, Faculty I Physics / Electrical Engineering, Nov 2003. (In German).

[21] I. Nassi and B. Shneiderman, "Flowchart Techniques For Structured Programming," in *ACM SIGPLAN (Special Interest Group on Programming Languages) Notices*, vol. 8, (Department of Computer Science, State University of New York), August 1973. http://www.geocities.com/SiliconValley/Way/4748/nsd.html.

Feature Subset Selection of Biosignals for Rehabilitation System

Z. Zenn Bien†, Jeong-Su Han†† and Sang Wan Lee†††

†,††† Department of Electrical Engineering and Computer Science, KAIST
373-1 Guseong-dong, Yuseong-gu, Daejeon, 305-701, Republic of Korea
Email: bien@kaist.ac.kr, bigbean@ctrsys.kaist.ac.kr
†† Digital Appliance Research Center, Samsung Electronics,
416 Maetan-dong, Yeongtong-gu, Suwon, 443-742, Republic of Korea
Email: jeongsu.han@samsung.com

Keywords : *Classifiability, Separability, Feature Selection, EMG signal, Bio Signal*

Abstract

Recently, bio-signals have been considered as means of communication and control of human-machine systems. The electromyogram (EMG) signal among them is a form of electric manifestation of neuromuscular activation associated with a contracting muscle, and is well known for its usage in wide variety of applications. Since characteristics differ from person to person, this user-dependency problem is handled by finding some characteristics of users which are as common as possible. To select such a feature, we define a new criterion function, called classifiability, which overcomes an inconsistency between separability and classification performance. Classifiability is defined by adopting the notion of separability index matrix (SIM), which provides relevant features with very low computational cost. Based on the proposed criterion function of classifiability, and in reference to the forward search technique, we propose a new feature selection algorithm (SIMF). In extensive experimental comparisons, the proposed algorithm is shown to outperform other filter-type feature selection methods for various criterion functions in terms of classification accuracy.

1. Introduction

Recently, bio-signals, such as EMG, EEG, and EOG have been considered as means of communication and control of human-machine systems. The effective recognition of these signals is a promising theme of study since it provides with convenient means for human-machine interaction such as the brain computer interface (BCI) that uses electroencephalogram (EEG), the prostheses controller using electromyogram (EMG), and the mouse using electroolfactogram (EOG).

Among them, EMG signal has been widely used in various applications. However, practical application is known to be difficult, because EMG signal is different from person to person. There are various factors that affect the physiological bio-signals such as EMG signals. These factors are both internal such as thickness of skins, tissues, and difference of the number of muscle fibers, and external such as electrode-electrolyte interface, and electrode configuration.

To handle this user-dependency problem, we may adopt two very different approaches. One direction is called "personalization" technique. This scheme admits the different characteristics of each user and tries to design personalized pattern recognition structures. To accomplish this task, it may utilize a user-dependent feature selection process and/or an adaptive classifier. In this approach, however, there are two main shortcomings. One is the necessity of an additional system such as a user identification recognition system to make the pattern recognition system know who he/she is. Because personalized recognition system is turned to a specific user, we must let the system know who the user is. The other is that a preprocessing step such as initial learning or feature subset selection is indispensable before using the system. This makes user tedious and the whole system inconvenient to use even if a preprocessing step is not much burden to users.

In this paper, we solve the same problem in a very different way. Our solution, which stands in opposite direction of the first method, is try to find as common characteristics of users as possible. Here common characteristics of users mean common feature subset which works well to all users. To select such a feature, we define a new criterion function, called classifiability. Conventional distance-based criterion functions give how separate feature distribution is but classifiability reveals classification capability of a specific feature for classes.

We propose a new criterion function, classifiability, based on separability index matrix (SIM), which provides relevant features with very low computational cost. Based on the proposed criterion function, classifiability, and in reference to the forward search technique, we propose a new feature selection algorithm (SIMF).

The organization of the remainder of this paper is as follows. In section 2, conventional criterion functions are introduced, and their limitation is discussed as well. Based on new criterion function which overcomes the limitation, a new feature selection algorithm is proposed in section 3.1 and 3.2, and then the strategy for solving user-dependency problem of bio signal is suggested in section 3.3. In section 4, performance evaluation for various datasets from UCI repository is conducted in terms of classification accuracy, followed by experiments for EMG dataset. Conclusion is provided in section 5.

2. Related Works and Problem

For effective feature subset selection (FSS), a number of algorithms have been reported, which are conventionally grouped into two types, namely, the filter method and the wrapper method [1]. The two methods differ from each other because of the criterion function used. The criterion function evaluates the goodness of a selected feature subset. In the filter method, the criterion function utilizes quantitative information such as the interclass distance of selected features [2]. In the wrapper method, on the other hand, the criterion function relies on classification accuracy in evaluating the goodness of a feature subset. This method involves a specific classifier in the process of selecting the feature subset.

Since the wrapper method directly relies on classification rate, it often performs better than the filter technique, but requires significantly higher computational cost since the fitness evaluation of a subset requires cross-validation or a bootstrapping procedure for error estimation of each subset [1, 3]. Furthermore, the solution of the wrapper method often lacks generality since the selected feature subset can be biased to a specific classifier. In the filter method, various criterion functions are already known and can be categorized into two groups; that is, distance-based measures, and relation-based measures. Fisher ratio [4], Mahalanobis distance [5], and Bhattacharyya distance [5] are typical examples of the distance-based measures which use a distance metric to measure class separability. These measures assign an average value of separability of classes for each feature; they may select features which are highly correlated with each other if the selected features have a big average separability value. The relation-based measures, represented by correlation or mutual information [3, 6], try to extract the relation between features and classes. In this approach, a feature is a good one if it is highly correlated with the classes, while uncorrelated with other features. A high computational cost is entailed, however, to acquire good features using these measures, especially when mutual information is involved, though there are some heuristics to reduce it [6, 7].

The distance-based measures, however, have low computational cost compared with the relation-based measures. It is difficult to remove redundant features, and there is a tendency for less relevant features be selected since these distance-based measures usually utilize their average (or sum) of separability for classes. For example, a selected feature shown in Fig.1 assumes a big distance-based measure if it has a relatively big separability between a certain pairs of classes. Most of separability measures in this group report that $\{x_1, x_2\}$ of case a) is a better feature combination than $\{x_1, x_3\}$ of case b) because of their high separability values. In view of classification, however, features set $\{x_1, x_3\}$ of case b) is a more relevant subset since they associate with a lower classification error than that of case a). Classification error usually occurs in the pairs of classes that have low separability.

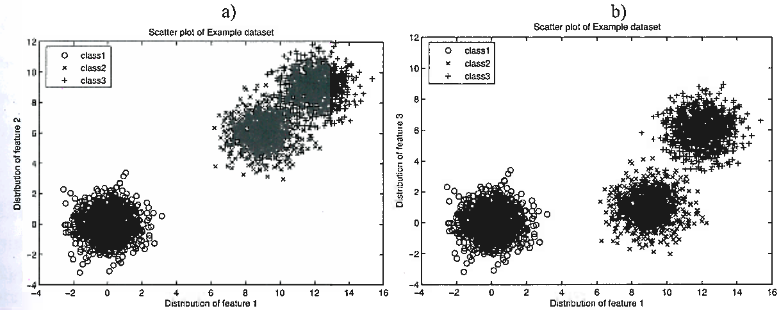

Fig.1. Scatter plot of Example dataset: (a) Scatter plot of feature x_1 and x_2; (b) Scatter plot of feature x_1 and x_3.

A highly separable state between class 1 and class 2 of feature x_1 in Fig. 1 can not give any help of decreasing the classification error occurred between class 2 and class 3. The goodness of a feature, therefore, must be evaluated by its ability to classify between every distinct pair of classes. We can judge classification capability of a certain feature by comparing separability values of every pair of classes of a certain feature with separability values of corresponding pairs of classes of all available features.

3. Proposed Feature Selection Algorithm (SIMF)

3.1 A New Criterion Function Based on SIM

In this section, we define a notion of "classifiability of a feature" as a new criterion function based on the concept of separability index matrix (SIM). For this, we introduce in the following some new notations and operations for matrix calculation.

Definition 1. (Element-wise operations of matrices)

Let C be a given integer, which denotes the total number of classes. For C x C matrices $A = [a_{i,j}]_{C\times C}$, $B = [b_{i,j}]_{C\times C}$, $P = [p_{i,j}]_{C\times C}$, *and* $Q = [q_{i,j}]_{C\times C}$ *where* $a_{i,j}, b_{i,j} \in \mathbf{R}$ *(the set of real numbers) and* $p_{i,j}, q_{i,j} \in \{0,1\}$, $1 \le i, j \le C$, *we define various element-wise operations of matrices denoted by* $\otimes$, $\oslash$, $\circledvee$, $\circledwedge$, $\ominus$, *and* $\circleddash$, *as follows:*

$A \otimes B = [r_{i,j}]_{C\times C}$ *where* $r_{i,j} = a_{i,j} \times b_{i,j}$ *(multiplication)*

$A \oslash B = [r_{i,j}]_{C\times C}$ where $r_{i,j} = a_{i,j} \div b_{i,j} \quad (b_{i,j} \neq 0)$ *(division)*

$P \circledvee Q = [r_{i,j}]_{C\times C}$ where $r_{i,j} = p_{i,j} \vee q_{i,j}$ *(logical OR operation)*

$P \circledwedge Q = [r_{i,j}]_{C\times C}$ where $r_{i,j} = p_{i,j} \wedge q_{i,j}$ *(logical AND operation)*

$\circleddash P = [r_{i,j}]_{C\times C}$ where $r_{i,j} = \sim p_{i,j}$ *(logical NOT operation)*

$A \circledgeq B = [r_{i,j}]_{C\times C}$ where $r_{i,j} = \begin{cases} 1, & a_{i,j} \geq b_{i,j} \\ 0, & a_{i,j} < b_{i,j} \end{cases}$ *(COMPARE operation)*

Definition 2. (Separability Degree Matrix (SDM))

Let C be the total number of classes. For $1 \leq i, j \leq C$, let w_i and w_j be the i-th and j-th class, respectively. Let there be given a criterion function $J(\cdot)$ of distance-measure type, such as the Bhattacharyya distance [5]. For a feature x_k, we denote $J(w_i, w_j; \{x_k\})$ as the separability (or, distance) value between class w_i and class w_j when the specific criterion function $J(\cdot)$ is applied and the feature x_k is used. Then the separability degree matrix (SDM) is defined as follows:

$$SDM_k = [J(w_i, w_j; \{x_k\})]_{C\times C} \tag{1}$$

Each element of SDM_k (separability degree matrix) represents a separability value of a pair of all distinct classes evaluated by using a feature x_k. As an example, we can use the generalized Fisher ratio.

Since SDM is symmetric, we shall omit the entries below the diagonal and show only the upper triangular elements of each matrix for notational simplicity.

Definition 3. (Separability Index Matrix (SIM))

For given SDM_k, $1 \leq k \leq N$, where N is the total number of features, let

$$SDM_{avg} = \frac{1}{C} \times \left(\sum_k SDM_k \right).$$

For $\Delta > 0$, let $SDM_\Delta = \Delta \times [1_{i,j}]_{C\times C}$. Here Δ is a threshold value given by the designer. Then the separability index matrix (SIM) of feature x_k is defined as follows:

$$SIM_k^{\Delta} = [r_{i,j}]_{C\times C} = (\ SDM_k \circledgeq SDM_{avg}\ ) \circledvee (\ SDM_k \circledgeq SDM_\Delta\ ) \tag{2}$$

Each element with '1' indicates that the corresponding classes are between-class separable and '0' means that they are not between-class separable. By virtue of the notion of SIM, we can easily distinguish relevant features from irrelevant and/or redundant features with respect to previously selected features.

Definition 4. (Classifiability)

For given SIM_k^{Δ}, $1 \leq k \leq N$ with N being the total number of features, let WM_k be a matrix given by

$$WM_k = SIM_k^{\Delta} \oslash \left(\sum_{i=1}^{N} SIM_i^{\Delta} \right).$$

Then the function defined as

$$C(x_k) = \sum_i \sum_j \left(SIM_k^{\Delta} \otimes WM_k \right) \tag{3}$$

is called the classifiability of a feature x_k and will serve as the new criterion function.

Note that a conventional distance-based criterion function tells how separate the feature distribution is, whereas the new criterion function of classifiability reveals classification capability of a specific feature for classes.

Definition 5. (Irrelevance, Relevance, and Redundancy of features)

Let SIM_k^{Δ} and SIM_V^{Δ} be SIM (Separability Index Matrix) of a feature x_k and SIM of previously selected feature subset $V = \{x_1, \cdots, x_n\}$ in the feature selection process.

(1) Feature x_k is said to be an irrelevant feature if $SIM_k^{\Delta} = \mathbf{0}$.

(2) Feature x_k is said to be a fully relevant feature if $SIM_k^{\Delta} \circledwedge SIM_V^{\Delta} = \mathbf{0}$ and $SIM_k^{\Delta} \neq \mathbf{0}$.

(3) Feature x_k is said to be a fully redundant feature if $SIM_k^{\Delta} \circledwedge SIM_V^{\Delta} = SIM_k^{\Delta}$ and $SIM_k^{\Delta} \neq \mathbf{0}$.

(4) Feature x_k is said to be a partially relevant (redundant) feature if $SIM_k^{\Delta} \circledwedge SIM_V^{\Delta} \neq SIM_k^{\Delta}$ and $SIM_k^{\Delta} \neq \mathbf{0}$.

A feature is considered as irrelevant if it has no between-class separable cases. A fully relevant feature has between-class separable cases which are not between-class separable by the previously selected feature subset V. We can discover such a feature from the remaining features by conducting simple Boolean AND operation between SIM_k^{Δ} and SIM_V^{Δ}. A redundant feature is a feature that provides between-class separable cases which are not needed in the previously selected feature subset V. Except for the above three cases, we regard them as partially relevant (redundant) features. Based on these concepts, we now present a feature selection algorithm which is of forward search type.

3.2 A Proposed Feature Selection Algorithm based on SIM

The proposed criterion function, classifiability based on SIM, provides an efficient way of evaluating the goodness of a feature. Our proposed feature selection method is based on a forward search paradigm, where, at each search, we add the feature that is most relevant among the remaining features. We remark that, without loss of generality, the proposed criterion function can be implemented with other types of search methods such as backward search method, floating search method, and so on.
The complete algorithm is described below, and flow chart is given in Fig.2.

Proposed feature selection algorithm (SIMF)

Let $V = \{\phi\}$, $V_I = \{\phi\}$, $S = \{x_1, x_2, \cdots, x_N\}$, and $S_I = \{1, 2, \cdots, N\}$.

Let V be the selected feature subset, V_I an index of the selected feature subset, S a given feature set, and S_I an index of given feature set.

1. Initialization

$SDM_V \leftarrow [0_{i,j}]_{C\times C}$ *where* $1 \leq i, j \leq C$

$SDM_k \leftarrow [J(w_i, w_j; \{x_k\})]_{C\times C}$ *for all* $k \in S_I$

$SIM_k^{\Delta} \leftarrow (\ SDM_k \circledgeq SDM_{avg}\ ) \circledvee (\ SDM_k \circledgeq SDM_\Delta\ )$ *for all* $k \in S_I$

$$WM_k \leftarrow SIM_k^{\Delta} \textcircled{D} \left(\sum_{i=1}^{N} SIM_i^{\Delta} \right) \; \textit{for all } k \in S_I$$

2. Search of the next best feature

Find $\arg\max_{k \in S_I} C(x_k)$ *where* $C(x_k) = \sum_i \sum_j (SIM_k^{\Delta} \otimes WM_k)$

$V_I \leftarrow V_I \cup \{k\}$

If $(\|V_I\| \neq 1)$

then $\arg\max_{k \in V_I} S(x_k)$ *where* $S(x_k) = \sum_i \sum_j (SDM_k \otimes WM_k)$

and $V_I \leftarrow \{\phi\}$

3. Update of the feature subset with the next best feature

$V \leftarrow V \cup \{x_k\}$, $S \leftarrow S - \{x_k\}$, $S_I \leftarrow S_I - \{k\}$

$SDM_V \leftarrow [J(w_i, w_j; V)]_{C \times C}$

$SIM_V^{\Delta} \leftarrow (\; SDM_V \textcircled{$\geq$} SDM_{avg} \;) \textcircled{V} (\; SDM_V \textcircled{$\geq$} SDM_{\Delta} \;)$

4. Update of the relevance of remaining features with respect to the selected feature subset

$SIM_k^{\Delta} \leftarrow (\; \textcircled{$\sim$} SIM_V^{\Delta} \;) \textcircled{$\wedge$} SIM_k^{\Delta}$ *for all* $k \in S_I$

$WM_k \leftarrow SIM_k^{\Delta} \textcircled{D} \left(\sum_{i=1}^{N} SIM_i^{\Delta} \right)$ *for all* $k \in S_I$

If "STOP" condition were satisfied, then go to 5; else go to 2.

5. Return V as the final subset

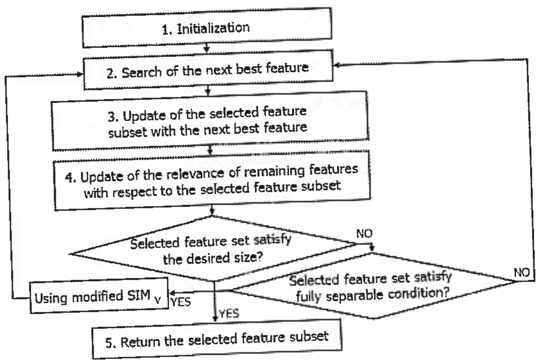

Fig.2. Flow Chart of SIMF

In this algorithm, "STOP' condition is met when either the number of selected features reaches to the maximum value or the "fully separable condition" is achieved. SIM_V^{Δ} is said to satisfy the fully separable condition if all elements of SIM_V^{Δ} except the diagonal terms is between-class separable in all distinct pairs of classes. This means that all elements of SIM_V is '1' except the diagonal terms being '0'.

3.3 Strategy for Solving User-Dependency Problem of Bio signal

As explained in introduction, EMG signal is different from person to person. To handle this user-dependency problem, we try to find as common characteristics of users as possible. Here common characteristics of users mean common feature subset which works well to all users. To select such a feature, we utilize our proposed criterion function with the following additional definition. The schematic diagram is shown in Fig. 3.

Definition 6. (Common Separability Index Matrix (cSIM))

Let SIM_k^p *be SIM of a feature* x_k *of user p. Here p means p-th user out of P users. For given* SIM_k^p *and* $1 \leq p \leq P$*, common separability index matrix,* $cSIM_k$*, of a feature* x_k*, is defined as follows:*

$$cSIM_k = SIM_k^1 \textcircled{$\wedge$} SIM_k^2 \textcircled{$\wedge$} \ldots \textcircled{$\wedge$} SIM_k^P . \tag{4}$$

Each element of cSIM of a specific feature is common between-class separable information over all *P* users. The proposed feature selection algorithm, SIMF, runs same way to select the best feature subset from EMG signals except adding cSIM process in initialization part.

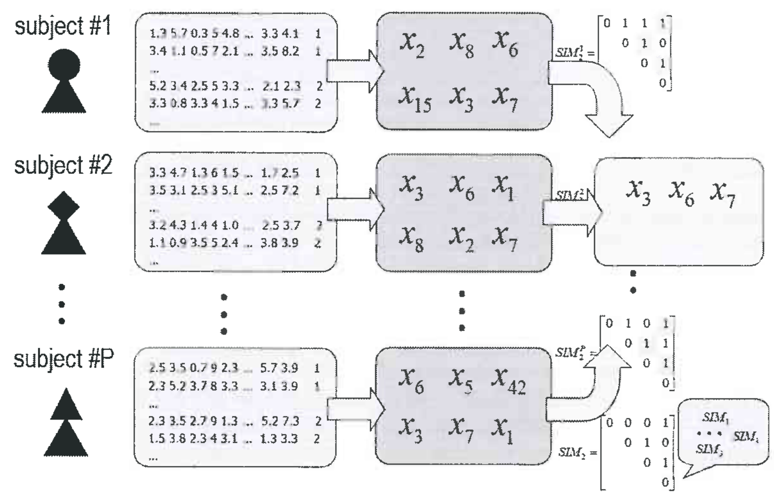

Fig.3. Schematic Diagram of SIMF with cSIM

4. Experimental Results

4.1 Test on UCI Benchmark Datasets

We have tested the proposed feature selection algorithm on 10 continuous data sets shown in Table 1. We have chosen these data sets from the UCI repository [8], which are well-known benchmark datasets with the following three conditions: (i) the number of features is 10 or greater, (ii) there should be no missing features, and (iii) all of the feature values are continuous. For the detailed account about the characteristics of datasets, one may refer to [8].

Table 1. Data Sets used for Experiment

Name of data set	Number of instances	Number of classes	Number of features	Feature size**
glass	214	7	10	small
vowel	990	11	10	small
wine	178	3	13	small
letter	20000	26	16	small
vehicle	846	4	18	small
segmentation	2310*	7	19	small
wdbc	569	2	30	medium
ionosphere	351	2	34	medium
satellite	6435*	6	36	medium
sonar	208	2	60	large

(Note: * Training data set and test data set are merged into one dataset for experiment.
** Feature size is decided upon the results of Kudo's study [13].)

For classification accuracy, we have considered various criterion functions shown in Table 2 in a feature selection algorithm along with the Branch and Bound search method which is known to give an optimal solution in the sense of the given criterion function.

Table 2. Criterion functions of feature selection algorithm used for experiment

Alias	Criterion functions
in-in	Inter-intra distance
maha-s	Sum of Mahalanobis distance
maha-m	Minimum of Mahalanobis distance
eucl-s	Sum of squared Euclidean distance
eucl-m	Minimum of squared Euclidean distance
bhatta	Bhattacharyya distance

To obtain classification accuracy of the selected feature subset by various feature selection algorithms, we have adopted three different classifiers – k-nearest neighbor classifier (shorten *knnc* for further reference), Parzen classifier (*parzenc*), and naive Bayes classifier(*naievebc*). Classification accuracy of the selected feature subset is affected by its size. Generally, if selected features are relevant, the classification accuracy improves as the number of features increases. We have set the maximum size of selected feature subset to 9 because no further improvement of classification accuracy is observed.

The original dataset was split into two parts: training dataset and test dataset by a random sub-sampling method. Several training datasets are made by taking 20%, 30%, 50% and 80% of the whole dataset. Classification accuracy of the selected feature subset is averaged over 30 independent trials. Therefore the total number of simulation for classification accuracy of all dataset is 252,000 (10 (dataset) × 7 (different feature selection algorithm, including proposed one) × 3 (different classifier) × 10 (cases of feature size, including all feature case) × 4 (different ratio of training dataset) × 30 (running iteration)).

Table 3 presents an averaged classification error with standard deviation of k-nearest neighbor classifier for each dataset at 50% training data ratio. We have omitted classification results of the other two classifiers, and for a different ratio of training dataset, since they show similar results of Table 3.

Table 3. Averaged classification error (%) and standard deviation for 30 trials of knnc for each dataset (note: Training data ratio is 50% of each dataset.)

Dataset	n	proposed	in-in/maha-s	maha-m	eucl-s	eucl-m	bhatta
Glass (N=10, C=7)	1	59.9±4.18	60.5±4.76	58.8±4.47	59.5±4.51	59.0±4.21	59.5±4.74
	2	43.9±3.80	39.4±3.74	44.2±4.39	50.7±4.11	44.0±3.49	45.8±4.47
	4	35.6±3.55	35.8±4.13	37.5±3.96	35.9±4.02	38.3±3.51	33.3±4.66
	6	31.3±3.79	40.5±3.39	35.1±3.88	31.6±4.30	32.1±4.15	35.4±4.02
	8	31.5±4.00	31.8±4.53	31.3±5.23	29.9±4.78	31.3±4.57	31.7±3.84
	All	31.8±4.39	31.9±4.18	30.6±3.85	32.1±4.81	31.5±4.38	32.0±4.15
Vowel (N=10, C=11)	1	68.0±1.29	66.3±2.46	80.1±2.96	66.3±2.91	80.4±2.95	69.3±1.91
	2	40.6±1.68	40.9±2.12	40.8±2.53	40.8±1.44	41.3±2.66	40.7±1.84
	4	13.3±1.28	14.3±1.78	14.2±1.63	13.3±1.38	19.5±1.73	14.1±1.54
	6	7.45±1.64	8.88±1.76	8.59±1.24	8.51±1.31	7.91±1.22	7.92±1.39
	9	5.08±1.50	5.25±1.39	5.26±1.35	5.26±1.24	5.37±1.74	4.24±0.88
	All	4.98±1.66	4.50±1.39	4.55±1.41	4.60±1.23	4.42±1.50	4.60±1.35
Wine (N=13, C=3)	1	22.8±4.09	22.1±3.55	29.0±4.55	33.0±3.49	34.4±3.24	23.6±4.83
	2	8.90±2.57	9.28±2.43	30.5±3.56	30.4±3.94	31.7±3.60	17.7±3.44
	4	31.2±3.27	31.7±4.40	30.9±3.66	30.9±4.72	30.2±3.05	29.3±4.11
	6	32.0±3.87	31.2±4.20	30.2±4.26	31.2±3.68	30.5±3.16	29.7±4.55
	9	30.4±4.40	30.1±3.89	29.9±4.66	31.2±3.43	29.7±3.48	30.8±2.83
	All	30.8±3.70	31.1±3.81	30.1±4.81	30.2±3.04	29.6±3.86	31.2±3.54
Letter (N=16, C=26)	1	86.2±1.00	86.3±1.11	86.3±0.88	86.2±1.15	86.3±0.81	86.0±1.04
	2	74.1±1.98	72.0±1.73	72.3±1.65	77.2±1.37	76.0±1.52	72.2±1.64
	4	45.9±1.66	45.5±1.30	57.7±1.62	48.6±2.02	57.4±1.60	45.5±1.29
	6	28.3±1.48	32.9±1.23	41.1±1.36	27.4±1.24	40.0±1.69	24.6±1.16
	9	19.5±1.30	17.7±1.25	40.5±1.50	17.6±1.20	28.7±1.43	17.3±1.12
	All	22.3±1.50	22.4±1.47	22.4±1.52	22.4±1.74	22.2±1.43	22.4±1.68
Vehicle (N=18, C=4)	1	61.0±2.61	63.3±1.66	61.0±2.52	54.8±2.56	54.4±2.33	53.7±2.67
	2	53.5±2.08	51.7±2.11	55.0±1.89	48.2±1.76	44.5±1.86	47.2±2.11
	4	42.7±1.51	44.1±1.84	45.8±1.69	45.6±2.22	41.8±1.65	42.6±1.72
	6	40.1±2.26	35.0±1.24	45.0±1.84	42.1±1.79	42.4±2.03	41.3±2.12
	9	37.4±1.88	33.4±1.89	45.2±1.52	41.1±2.17	40.0±1.68	39.4±2.01
	All	37.4±1.64	37.5±2.02	37.1±1.89	38.0±2.14	37.8±2.60	38.1±2.24
Segmen-tation (N=19, C=7)	1	53.2±2.34	85.7±0.00	85.7±0.00	55.0±2.30	55.1±2.68	48.1±1.12
	2	19.7±0.92	38.8±1.32	44.6±3.11	16.0±0.74	46.8±1.27	35.4±0.97
	4	8.70±0.94	19.2±1.05	19.5±1.40	12.5±0.80	11.9±0.86	20.6±1.10
	6	5.99±0.70	18.0±0.94	8.30±0.81	8.10±0.65	9.88±0.85	14.6±0.90
	9	6.46±0.79	6.69±0.60	6.12±0.73	7.43±0.62	6.31±0.76	6.55±0.63
	All	5.96±0.60	5.87±0.75	5.93±0.80	5.92±0.73	5.94±0.56	5.96±0.82
Wdbc (N=30, C=2)	1	9.14±1.46	9.54±1.23	9.54±1.23	8.78±1.70	8.78±1.70	8.91±1.23
	2	8.83±1.36	27.0±2.15	27.0±2.15	7.66±1.19	7.66±1.19	9.88±1.47
	4	8.65±1.69	9.58±1.62	9.58±1.62	7.70±1.37	7.70±1.37	9.12±2.10
	6	8.67±1.62	8.91±1.34	8.91±1.34	7.08±1.21	7.08±1.21	8.98±1.33
	9	7.48±1.34	8.83±1.26	8.83±1.26	7.37±1.56	7.37±1.56	7.14±1.15
	All	7.28±1.41	6.94±1.03	6.94±1.03	7.41±1.21	7.41±1.21	7.41±1.39
Ionos-phere (N=34, C=2)	1	24.9±1.47	17.9±1.67	36.0±0.00	18.4±3.02	18.4±3.02	25.4±1.61
	2	12.4±1.62	12.6±1.83	21.8±7.16	13.8±2.14	13.8±2.14	12.3±2.09
	4	11.2±2.61	10.8±2.06	18.1±2.06	15.2±2.40	15.2±2.40	11.7±2.09
	6	11.1±1.75	9.92±1.62	15.8±1.64	14.4±2.26	14.4±2.26	11.4±2.18
	9	13.2±1.92	11.5±1.74	14.4±2.33	14.4±2.69	14.4±2.69	14.2±2.44
	All	15.7±2.30	15.2±2.16	15.5±1.86	15.3±2.00	15.3±2.00	15.5±2.34
Sate-llite (N=36, C=6)	1	45.2±1.60	42.8±2.21	44.9±1.40	42.6±2.25	43.3±2.24	45.2±1.60
	2	23.1±2.89	28.8±4.14	25.2±3.66	42.7±4.46	20.9±4.84	23.1±2.89
	4	14.6±1.04	16.1±1.36	20.5±1.19	39.9±1.91	17.0±1.47	14.6±1.04
	6	14.0±1.16	14.6±1.40	14.8±1.29	38.6±1.94	14.6±1.22	14.0±1.16
	9	13.9±1.68	14.8±2.02	14.4±1.35	37.8±2.29	14.4±1.20	13.9±1.68
	All	13.5±1.75	13.5±1.67	13.3±1.63	13.8±1.70	13.3±1.66	13.5±1.75
Sonar (N=60, C=2)	1	27.0±4.87	25.5±4.21	25.5±4.21	37.0±3.12	37.0±3.12	25.5±4.21
	2	25.0±3.09	26.1±3.11	26.1±3.11	30.1±3.87	30.1±3.87	26.1±3.11
	4	24.5±3.40	21.4±3.86	21.4±3.86	31.0±3.94	31.0±3.94	21.4±3.86
	6	24.7±3.19	20.3±3.71	20.3±3.71	26.0±3.72	26.0±3.72	20.3±3.71

9	25.9 ± 3.09	21.3 ± 4.32	21.3 ± 4.32	21.6 ± 3.22	21.6 ± 3.22	21.3 ± 4.32
All	19.4 ± 3.30	20.7 ± 4.03	20.7 ± 4.03	20.7 ± 3.72	20.7 ± 3.72	20.7 ± 4.03

To analyze experimental results objectively, we have performed t-test [9] on classification results of other classifier and other dataset. In this study, the significance level was set to be 0.05, which made our confidence interval is 0.95. Table 4 summarizes t-test results according to various criterion functions for all the datasets considered.

Table 4. Results of paired t-test for all dataset according to various criterion functions

Dataset	Criterion function	Wins (W)	Ties	Loses (L)	W/(W+L)
Glass (N= 10, C= 7)	in-in/maha-s	19% (5/27)	74% (20/27)	7% (2/27)	71% (5/7)
	maha-m	26% (7/27)	70% (19/27)	4% (1/27)	88% (7/8)
	eucl-s	11% (3/27)	81% (22/27)	7% (2/27)	60% (3/5)
	eucl-m	15% (4/27)	78% (21/27)	7% (2/27)	67% (4/6)
	bhatta	4% (1/27)	89% (24/27)	7% (2/27)	33% (1/3)
Vowel (N=10, C= 11)	in-in/maha-s	17% (5/30)	73% (22/30)	10% (3/30)	63% (5/8)
	maha-m	47% (14/30)	53% (16/30)	0% (0/30)	100% (14/14)
	eucl-s	10% (3/30)	77% (23/30)	13% (4/30)	43% (3/7)
	eucl-m	43% (13/30)	57% (17/30)	0% (0/30)	100% (13/13)
	bhatta	10% (3/30)	73% (22/30)	17% (5/30)	38% (3/8)
Wine (N= 13, C= 3)	in-in/maha-s	10% (3/30)	87% (26/30)	3% (1/30)	75% (3/4)
	maha-m	23% (7/30)	73% (22/30)	3% (1/30)	88% (7/8)
	eucl-s	30% (9/30)	57% (17/30)	13% (4/30)	69% (9/13)
	eucl-m	27% (8/30)	67% (20/30)	7% (2/30)	80% (8/10)
	bhatta	23% (7/30)	67% (20/30)	10% (3/30)	70% (7/10)
Letter (N= 16, C=26)	in-in/maha-s	33% (10/30)	27% (8/30)	40% (12/30)	45% (10/22)
	maha-m	67% (20/30)	17% (5/30)	17% (5/30)	80% (20/25)
	eucl-s	40% (12/30)	33% (10/30)	27% (8/30)	60% (12/20)
	eucl-m	80% (24/30)	17% (5/30)	3% (1/30)	96% (24/25)
	bhatta	27% (8/30)	40% (12/30)	33% (10/30)	44% (8/18)
Vehicle (N= 18, C= 4)	in-in/maha-s	43% (13/30)	20% (6/30)	37% (11/30)	54% (13/24)
	maha-m	80% (24/30)	13% (4/30)	7% (2/30)	92% (24/26)
	eucl-s	63% (19/30)	23% (7/30)	13% (4/30)	83% (19/23)
	eucl-m	53% (16/30)	27% (8/30)	20% (6/30)	73% (16/22)
	bhatta	40% (12/30)	40% (12/30)	20% (6/30)	67% (12/18)
Segmentation (N= 19, C= 7)	in-in/maha-s	83% (25/30)	13% (4/30)	3% (1/30)	96% (25/26)
	maha-m	70% (21/30)	13% (4/30)	17% (5/30)	81% (21/26)
	eucl-s	63% (19/30)	17% (5/30)	20% (6/30)	76% (19/25)
	eucl-m	53% (16/30)	30% (9/30)	17% (5/30)	76% (16/21)
	bhatta	67% (20/30)	20% (6/30)	13% (4/30)	83% (20/24)
Wdbc (N= 30, C= 2)	in-in/maha-s/maha-m	30% (9/30)	40% (12/30)	30% (9/30)	50% (9/18)
	eucl-s/eucl-m	27% (8/30)	50% (15/30)	23% (7/30)	53% (8/15)
	bhatta	7% (2/30)	93% (28/30)	0% (0/30)	100% (2/2)
Ionosphere (N= 34, C= 2)	in-in/maha-s	7% (2/30)	60% (18/30)	33% (10/30)	17% (2/12)
	maha-m	87% (26/30)	13% (4/30)	0% (0/30)	100% (26/26)
	eucl-s/eucl-m	70% (21/30)	20% (6/30)	10% (10/30)	88% (21/24)
	bhatta	0% (0/30)	97% (29/30)	3% (1/30)	0% (0/1)
Satellite (N= 36, C= 6)	in-in/maha-s	63% (21/30)	17% (5/30)	20% (6/30)	76% (19/25)
	maha-m	63% (21/30)	30% (9/30)	7% (2/30)	90% (19/21)
	eucl-s	80% (21/30)	10% (3/30)	10% (3/30)	89% (24/27)
	eucl-m	47% (21/30)	20% (6/30)	33% (10/30)	58% (14/24)
	bhatta	0% (0/30)	100% (30/30)	0% (0/30)	-
Sonar (N= 60, C= 2)	in-in/maha-s/ maha-m/bhatta	3% (1/30)	47% (14/30)	50% (15/30)	6% (1/16)
	eucl-s/eucl-m	47% (14/30)	37% (11/30)	17% (5/30)	74% (14/19)

Based on the experimental results, we could make the following comments. First, as long as the classification accuracy is concerned, the proposed feature selection algorithm performs better than Branch and Bound search method with various criterion functions except for few cases (Table 4). It is remarked that the criterion functions 'in-in' and 'maha-s' always give the same feature selection results in the experiment.

4.2 Test on EMG Datasets

The objective of this experiment is to classify six basic motions (rest, up, down, right, left, and click motion) which are an elementary motion of a mouse cursor. 22000 instances were collected from 10 subjects, and the number of channels of EMG signals is 4. After collecting data, we convert them into 9 kinds of feature sets which are widely used for EMG-based applications. Total dimension of feature sets is 52. And, *k*-nearest neighbor classifier and parzen window classifier are used.

For fair comparison, we randomly split user group into two in each experiment, as shown in Table 5.

Table 5. Train and test groups

	Subject ID in Group A	Subject ID in Group B
Experiment #1	2,3,6,8,7,5	1,4,9,10
Experiment #2	3,4,5,6,1,7	8,2,9,10
Experiment #3	9,5,3,4,1,8	2,6,7,10
Experiment #4	8,9,1,7,6,2	10,3,4,5

Table 6 and 7 shows the user-independency of selected feature subsets. It is remarkable that exactly the same feature subsets are selected when cSIM is used. It means the proposed feature selection method with cSIM selects the common feature subset which works well to all users.

Table 6. User-independency of selected feature subsets for experiment #1, #2.

Feature selection methods	Feature subset of all users	Randomly chosen 6 subjects cases			
		Experiment #1		Experiment #2	
		Feature subset	overlap	Feature subset	overlap
fwd_inin	20,19,16,3,26, 49,17,40,51,18	20,14,16,27,51, 28,49,19,17,18	70%	20,16,27,19,2, 13,49,44,25,28	40%
w/o cSIM	12,11,10,28,3, 9,51,1,16,4	12,11,16,4,3, 10,15,23,27,19	60%	12,11,16,3,28, 27,31,4,24,15	60%
w/ cSIM	4,11,19,20,28, 16,24,18,12,15	16,11,18,20,28, 4,12,24,19,15	100%	28,11,19,20,16, 4,12,24,18,32	90%

Table 7. User-independency of selected feature subsets for experiment #3, #4.

Feature selection methods	Feature subset of all users	Randomly chosen 6 subjects cases			
		Experiment #3		Experiment #4	
		Feature subset	overlap	Feature subset	overlap
fwd_inin	20,19,16,3,26, 49,17,40,51,18	20,19,16,14,3, 49,17,44,18,25	70%	20,16,14,27,51, 49,19,17,18,28	70%
w/o cSIM	12,11,10,28,3, 9,51,1,16,4	12,11,10,20,3, 16,4,15,23,27	60%	12,11,26,28,16, 3,24,49,10,31	60%
w/ cSIM	4,11,19,20,28, 16,24,18,12,15	12,11,16,19,20, 28,4,24,18,15	100%	16,10,4,15,28, 20,12,24,19,18	90%

Also, it turns out that our strategy works well in terms of classification accuracy, as shown in table 8. Thus, we can conclude that the proposed method overcomes user-dependency problem.

Table 8. Results of paired t-test for four experimens

Classifier	Dataset	Wins (W)	Ties	Loses	W/(W+L)
knnc	Experiment #1	80% (32/40)	3% (1/40)	18% (7/40)	82% (32/39)
	Experiment #2	40% (16/40)	0% (0/40)	60% (24/40)	40% (16/40)
	Experiment #3	65% (26/40)	3% (1/40)	33% (13/40)	67% (26/39)
	Experiment #4	80% (32/40)	0% (0/40)	20% (8/40)	80% (32/40)
parzenc	Experiment #1	83% (33/40)	3% (1/40)	15% (6/40)	85% (33/39)
	Experiment #2	55% (22/40)	3% (1/40)	43% (17/40)	56% (22/39)
	Experiment #3	80% (32/40)	0% (0/40)	20% (8/40)	80% (32/40)
	Experiment #4	85% (34/40)	0% (0/40)	15% (6/40)	85% (34/40)

5. Conclusion

In this paper we have proposed a new criterion function, called "classifiability of a feature", based on separability index matrix (SIM). Classifiability represents classification capability of a certain feature and is one of effective measures to evaluate the goodness of a feature. Based on classifiability and SIM, we have proposed a new feature selection algorithm, which uses the relevance update to find the next best feature among the remaining features with little computation cost. We have performed extensive experimental comparison of the proposed algorithm with other filter-type feature selection methods with various criterion functions. The experimental results verified that the proposed algorithm outperforms other algorithms. Also, by adopting the notion of common separability index matrix, we suggested strategy for solving user-dependency problem which often arises in biosignal-based applications. The experiment results shows that our method selects subset of features which work commonly well for all users.

[1] Liu, H.; Motoda, H.: Feature Selection for Knowledge Discovery and Data Mining, Kluwer academic publishers, Boston, 1998.
[2] Kittler, J.: Feature selection and extraction in Handbook of Pattern Recognition and Image Proc., T.Y.Young and K.S.Fu, Eds. San Diego, CA: Academic, 1986, p.59-83.
[3] Peng, H.; Long, F.; Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 2005, p.1226–1238.
[4] Duda, R.; Hart, P.: Pattern Classification and Scene Analysis, Wiley, New York, 1973.
[5] Fukunaga, K: Introduction to Statistical Pattern Recognition, Second ed. Academic Press, 1990.
[6] Battiti, R.: Using mutual information for selecting features in supervised neural net learning, IEEE Transactions on Neural Networks 5, 1994, p.537–550.
[7] Kwak, N.; Choi, C.H.: Input Feature Selection by Mutual Information Based on Parzen Window, IEEE Tr. on Pattern Analysis and Machine Intelligence 24 (12), 2002, p.1667-1671.
[8] Murphy, P. M.; Aha, D. W.: UCI Repository of machine learning databases, http://www.ics.uci.edu/~mlearn/MLRepository.html, Irvine, CA: University of California, Department of Information and Computer Science, 1994.
[9] Rosner, B.: A generalization of the paired t-test, Journal of the Royal Statistical Society Series C, Applied statistics, 31(1), 1982, p.9-13.

Bremen Brain-Computer Interface Framework: Modular Signal Processing

T. LÜTH, A. GRÄSER

Institute of Automation, IAT, University of Bremen
Emails: lueth@, ag@iat.uni-bremen.de

Keywords: *Brain-computer interface (BCI), P300, Steady-state visual evoked potential (SSVEP), Event-related synchronization and desynchronization (ERD/ERS), Signal processing, Object oriented programming, Threads, UML diagram, Sequence diagram*

Abstract

In this work, the framework for the Bremen Brain-Computer Interface (BCI) will be described. A BCI converts electrical brain signals into control signals. Different brain patterns based on different strategies can be used for BCIs (e.g., focused attention or motor imagery). But most BCIs follow the same signal processing procedure - acquisition, preprocessing, feature extraction and classification. Most of these steps, and the communication between them, are independent of the selected BCI strategy. The classification, for example, will always classify a feature vector generated by a feature extraction method. But if the method for calculation of the feature vector changes, it should have not an effect on the classification. Therefore, the framework has to take into account that all signal processing steps have to be in "black boxes", where the communication between them is defined by standardized interfaces. Of course, every black box consists of several modules, which breaks down the signal processing steps into small parts. These parts, and thus the Bremen BCI framework, have to follow object oriented programming and software design patterns to make it easy to implement changes of the requirements on the BCI or the acquisition hardware. The result of this work is a general framework for Brain-Computer Interfaces (called the "Bremen BCI Framework"), where only one module has to be exchanged if a signal processing method varies. Therefore, only minimal changes are necessary for different BCI strategies or different signal processing methodologies. This feature is an important step towards adaptive and practical Brain-Computer Interfaces for research where such a flexible framework is required.

1 Introduction

The aim of Brain-Computer Interface (BCI) systems is to provide a communication system for severely disabled people. BCIs detect specific patterns in brain activity and translate them into control commands for soft- or hardware devices [1]. Not only the controlled applications but also the used brain activity patterns vary through the BCI research field. The main differentiation among BCIs is whether they are exogenous or endogenous. Exogenous BCIs use an external stimulus to evoke a response to that stimulus in brain activity. The experimental strategy for such BCIs is a focused or selective attention. Two typical brain patterns depending on an external stimulus will be described further in Section 2: Steady-State Visual Evoked Potentials and the P300 response. Endogenous