

## Using Higher Order Nonlinear Operators for SVM Classification of EEG Data

I. Martisius, R. Damasevicius, V. Jusas, D. Birvinskas

Software Engineering Department, Kaunas University of Technology,

e-mails: ignas.martisius@stud.ktu.lt, robertas.damasevicius@ktu.lt, vacius.jusas@ktu.lt, darius.birvinskas@stud.ktu.lt

**crossref** <http://dx.doi.org/10.5755/j01.eee.119.3.1373>

### Introduction

Brain-computer interface (BCI) is a communication system that translates brain activity into commands for a computer or other digital devices [1]. The major goal of BCI research is to develop systems that allow disabled users to communicate with other persons, to control artificial limbs, or to control their environment. Other applications include multimedia communication, Augmented Reality applications, robot control and game development.

The majority of BCI systems work by reading and interpreting cortically evoked electro-potentials across the scalp via an electro-encephalogram (EEG). The EEG signal has become the main data source of BCI study due to its low cost and non-invasive nature. The EEG data is inherently complex and difficult to analyze. Oscillatory activity in the EEG is classified into different frequency bands or rhythms: delta (0.5–3.5 Hz), theta (4–8 Hz), alpha 1 (8–10.5 Hz), alpha 2 (10.5–13 Hz), beta 1 (13–21 Hz), beta 2 (20–32 Hz), and gamma (36–44 Hz) [1]. Because EEG signals are non-stationary and nonlinear, and normally interfered by eye movements and muscle noises, it is difficult to differentiate the classes of mental tasks from EEG [2]. Different features can be extracted from the EEG data such as: time domain features related to changes in the amplitude of neurophysiologic signals, occurring time-locked to the presentation of stimuli or time-locked to actions of the user of a BCI, frequency domain features related to changes in oscillatory activity, and spatial domain features extracted and combined from several electrodes [1].

BCI systems require correct classification of signals interpreted from the brain for useful operation. After acquiring the EEG data, the pre-processing (filtering, denoising), feature extraction, and dimensionality reduction is performed, before machine learning algorithms can be applied to learn from a training dataset how to classify the signals into classes, where each class corresponds to a specific action of the user. A variety of machine learning and artificial intelligence methods are used for EEG data

classification such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), Bayesian Linear Discriminant Analysis (BLDA), Support Vector Machine (SVM), Fisher Linear Discriminant Analysis (FLDA), etc. [3]. Usually, the raw EEG data is pre-processed using the DSP methods such as Fourier analysis or wavelet transform for denoising and the filtered signal is used as input data for the classification method.

Recently, the non-linear operators such as the Teager-Kaiser energy operator (TKEO) [4, 5] have attracted the attention of researchers in the BCI domain. The TKEO so far has been applied in speech recognition [6], for image enhancement [7] and in EEG data analysis for detecting high frequency oscillations [8], mental task classification [9], and sleep spindle detection [10].

The structure of the remaining parts of the paper is as follows. Section II discusses the Teager-Kaiser energy operator. Section III proposes and describes the proposed higher order nonlinear operator. Section IV describes data classification using Support Vector Machine. Section V presents a case study. Section VI presents conclusions and outlines future work.

### Teager-Kaiser Energy Operator

Nonlinear models are systems where either the additivity or the scalability properties do not hold in general

$$H[f(t)+g(t)] \neq H[f(t)]+H[g(t)], \quad H[\alpha \cdot f(t)] \neq \alpha \cdot H[f(t)], \quad (1)$$

where  $f(t), g(t)$  are signals,  $H$  is a nonlinear model (operator), and  $\alpha$  is a constant.

The Teager-Kaiser Energy Operator (TKEO), proposed by Teager [4] and further investigated by Kaiser [5], is a special case of nonlinear models. For a continuous real-valued signal  $x(t)$ , the TKEO  $\Psi[x(t)]$  is defined as follows

$$\Psi[x(t)] = \left( \frac{\partial x}{\partial t} \right)^2 - x(t) \cdot \frac{\partial^2 x}{\partial t^2}. \quad (2)$$

An approximation of the derivatives by one-sample differences provides the definition of the TKEO for the discrete-time signal [5]

$$\Psi[x(n)] = x^2(n) - x(n-1)x(n+1). \quad (3)$$

Moore *et al.* [11] propose a generalization of the Teager operator as 1-D Volterra filter

$$\Psi^m[x(n)] = [x(n)]^{2/m} - [x(n-1)x(n+1)]^{1/m}. \quad (4)$$

Tomar *et al.* [12] introduce two generalizations of TKEO. A variable length TKEO (VTEO) is defined as

$$\Psi_i[x(n)] = x^2(n) - x(n-i)x(n+i). \quad (5)$$

The Summed-over Variable length Teager Energy Operator (S-VTEO) is defined as

$$\xi_i[x(n)] = \sum_{k=1}^u \Psi_k[x(n)]. \quad (6)$$

A combination of (4) and (5) operators is proposed in [13]

$$\Psi_i^m[x(n)] = [x(n)]^{2/m} - [x(n-i)x(n+i)]^{1/m}. \quad (7)$$

A generalization of the continuous TKEO as the higher-order energy operator (HOEO)  $\Psi_k$  is proposed in [14]

$$\Psi_k[x(t)] = \frac{\partial x}{\partial t} \frac{\partial^{k-1} x}{\partial t^{k-1}} - x(t) \cdot \frac{\partial^k x}{\partial t^k}. \quad (8)$$

For discrete-time series, the HOEO can be rewritten as the discrete energy operator (DEO) [Maragos]:

$$\Psi_{km}[x(n)] = x(n)x(n+k) - x(n-m)x(n+k+m). \quad (9)$$

The advantage of the TKEO family of operators over the traditional DSP analysis methods such as Fourier Transform or wavelet analysis is the ability of the TKEO to discover high-frequency low-amplitude components [8] in analyzed data. The TKEO unlike conventional energy takes into account the frequency component of the signal as well as the signal amplitude.

### Proposed Nonlinear Operators

In a general case, the TKEO operator can be generalized to the Homogeneous Multivariate Polynomial Operator (HMPO)  $\Psi_m^k[x(n)]$ , where the 2<sup>nd</sup> order HMPO is defined as

$$\Phi_m^2[x(n)] = \sum_{i=-z}^z \sum_{j=-z}^z A_{ij} x(n+i)x(n+j), \quad (10)$$

where  $z = \lfloor m/2 \rfloor$ , and  $A$  is the coefficient matrix.

The 3<sup>rd</sup> order HMPO is defined as follows

$$\Phi_m^3[x(n)] = \sum_{i=-z}^z \sum_{j=-z}^z \sum_{k=-z}^z A_{ijk} x(n+i)x(n+j)x(n+k). \quad (11)$$

For example, TKEO (2) can be written as follows

$$\Psi[x(n)] = \Phi_3^2[x(n)], \text{ where } A = \begin{pmatrix} 0 & 0 & -1/2 \\ 0 & 1 & 0 \\ -1/2 & 0 & 0 \end{pmatrix}. \quad (12)$$

Such operator also can be seen as a special case of the 2D Volterra system as noted by Kvedalen [13].

The properties of the  $\Phi_m^k[x(n)]$  operator are as follows:

- *Symmetry*. Reversing the signal in time does not change the resulting value;
- *Robustness*. The operator is robust even if the signal passes through zero, i.e.,  $x(n)=0$ , i.e. there is no division operation;
- *Complexity*. Complexity of the operator is  $\Theta(m^k)$ .

### Classification using Support Vector Machine

Support Vector Machines (SVM) [15] is a binary classification algorithm based on structural risk minimization. First, the SVM implicitly maps the training data into a (usually higher-dimensional) *feature space*. A *hyper-plane* (decision surface) is then constructed in this feature space that bisects the two categories and maximizes the margin of separation between itself and those points lying nearest to it (the *support vectors*). This decision surface can then be used as a basis for classifying vectors of unknown classification.

Consider an input space  $X$  with input vectors  $x_i \in X$ , a target space  $Y = \{1, -1\}$  with  $y_i \in Y$  and a training set  $T = \{(x_1, y_1), \dots, (x_N, y_N)\}$ . In SVM classification, separation of the two classes  $Y = \{1, -1\}$  is done by the *maximum margin* hyper-plane, i.e. the hyper-plane that maximizes the distance to the closest data points and guarantees the best generalization on new examples. In order to classify a new point  $x_j$ , the classification function  $g(x_j)$  is used

$$g(x_j) = \text{sgn} \left( \sum_{x_i \in SV} \alpha_i y_i K(x_i, x_j) + b \right), \quad (13)$$

where  $SV$  are the support vectors,  $K(x_i, x_j)$  is the kernel function,  $\alpha_i$  are weights, and  $b$  is the offset parameter.

If  $g(x_j) = +1$ ,  $x_j$  belongs to the *Positive class*, if  $g(x_j) = -1$ ,  $x_j$  belongs to the *Negative class*, if  $g(x_j) = 0$ ,  $x_j$  lies on the decision boundary and can not be classified.

Therefore, here we have a binary classification problem in which the outcomes are labelled either as positive ( $P$ ) or negative ( $N$ ) class. There are four possible outcomes from a binary classifier. If the outcome from a prediction is  $P$  and the actual value is also  $P$ , then we have a *true positive (TP)*; however if the actual value is  $N$  then we have a *false positive (FP)*. Conversely, a *true negative (TN)* has occurred when both the prediction outcome and the actual value are  $N$ , and *false negative (FN)* is when the prediction outcome is  $N$  while the actual value is  $P$ . To evaluate the precision of classification the following metrics are commonly used:

- 1) *Precision* is a measure of how well a binary classification test correctly identifies the true positives against all the positive results;
- 2) *Recall* is a measure of how well a binary

classification test correctly identifies the positive cases;

3) Accuracy is the proportion of true results (both true positives and true negatives) in the test data;

4) *F-measure* ( $F$ ) evaluates the accuracy of classification as the harmonic mean of specificity and recall;

5) Area Under Curve (AUC) is the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one.

### Case study

For experiments, Data set Ia (Tübingen, «self-regulation of SCPs», subject 1) [16] from the BBCI competition datasets (<http://bbci.de/competition/>) was used. The datasets were taken from a healthy subject. The subject was asked to move a cursor up and down on a computer screen, while his cortical potentials were taken. During the recording, the subject received visual feedback of his slow cortical potentials (SCPs). The dataset consists of 135 trials belonging to class 0 and 133 trials belonging to class 1. Each trial consists of 896 samples from each of 6 channels. The sampling rate of 256 Hz and the recording length is 3.5s. The dataset was randomly partitioned into 5 parts, and 5-fold cross-validation was used to evaluate the classification results.

The following nonlinear operators were applied to the raw EEG data: TKEO  $\Psi[x(t)]$  (3) [5], TKEO-Volterra  $\Psi^2[x(t)]$  (4) [11], VTEO  $\Psi_2[x(t)]$  (5) [12], VTEO-Volterra  $\Psi_2^2[x(t)]$  (7) [13], and the proposed HMPO  $\Psi_3^3[x(t)]$  (11), where the non-zero elements of the 3D matrix  $A$  are:  $a_{-1,-1,-1}=1$ ,  $a_{-1,-1,0}=3$ ,  $a_{-1,-1,1}=-1$ ,  $a_{-1,0,0}=2$ ,  $a_{-1,0,1}=-2$ . The matrix values were set based on the results of grid-based search using all possible combinations of integer numbers from the range  $\{-3, 3\}$

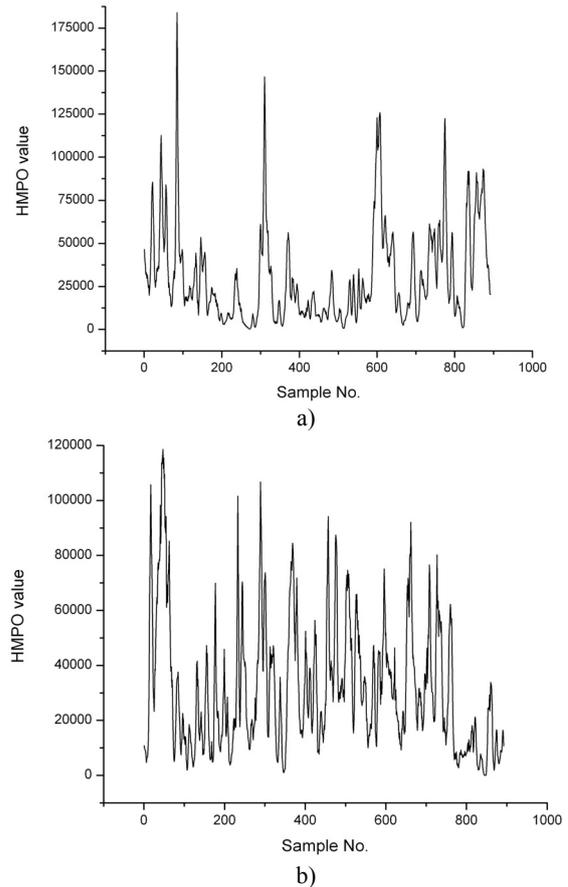
**Table 1.** Experimental results

Operator applied	Classification metric				
	<i>Acc.</i>	<i>Prec.</i>	<i>Rec.</i>	<i>F</i>	<i>AUC</i>
None (original EEG data)	0.7800	0.7462	0.8501	0.7891	0.9018
TKEO $\Psi[x(t)]$ (3)	0.4740	0.4805	0.7527	0.5849	0.4670
TKEO-Volterra $\Psi^2[x(t)]$ (4)	0.5931	0.5657	0.7889	0.6583	0.5428
VTEO $\Psi_2[x(t)]$ (5)	0.5635	0.5415	0.8051	0.6466	0.6500
VTEO-Volterra $\Psi_2^2[x(t)]$ (7)	0.4813	0.4207	0.5330	0.4214	0.1381
DEO $\Psi_{1,1}[x(t)]$ (9)	0.5262	0.5160	0.7202	0.6063	0.5851
HMPO $\Psi_3^3[x(t)]$ (11)	<b>0.8283</b>	<b>0.8042</b>	<b>0.8727</b>	<b>0.8349</b>	<b>0.8450</b>

Classification of data was performed using the SVMPerf [17] implementation of Support Vector Machine (available at <http://svmlight.joachims.org/>) with linear kernel. Kernel parameters were optimized using the Nelder-Mead algorithm based method described in [18]. The

results of experiments are summarized in Table 1 (best results in each category are shown in bold).

Evaluating the experimental results, we can claim that the 3<sup>rd</sup> order nonlinear operators such as the proposed HMPO can demonstrate better results for feature identification of the EEG data than traditional 2<sup>nd</sup> order operators such as TKEO or its generalizations as demonstrated by Table 1. Even visually inspecting the graphs of positive and negative series in Figure 1, one can see that the HMPO operator allows for better identification of significant features (slow cortical potential signals of  $\sim 1$  Hz frequency).



**Fig. 1.** Samples of raw EEG data after application of HMPO operator: a – positive instance; b – negative instance

### Conclusions and future work

In this paper, we proposed a novel nonlinear operator based on the generalization of the Teager-Kaiser Energy Operator, called Homogeneous Multivariate Polynomial Operator (HMPO). The applicability of the proposed operator is demonstrated for classification of the EEG signals. The experimental results obtained using a Support Vector Machine demonstrate an improvement of the classification results.

The proposed operator can be used for developing new EEG signal processing algorithms, which can be used in Brain-Computer Interface applications, e.g., for robot control in the noisy environment.

Future work will focus on the integration of higher-order nonlinear operators with the DSP-based filtering techniques to improve the classification accuracy of the EEG data.

## References

1. **Hoffmann U., Vesin J. M., Ebrahimi T.** Recent advances in brain-computer interfaces // IEEE 9th Workshop on Multimedia Signal Processing (MMSP'2007), 2007. – P. 17–19.
2. **Guo L., Wu Y., Zhao L., Cao T., Yan W., Shen X.** Classification of Mental Task From EEG Signals Using Immune Feature Weighted Support Vector Machines // IEEE Trans. on Magnetics, 2011. – Vol. 47. – No. 5. – P. 866–869.
3. **Selim A. E., Wahed M. A., Kadah Y. M.** Machine Learning Methodologies in Brain-Computer Interface Systems // Proc. of Cairo Int. Biomedical Engineering Conf. (CIBEC'2008). – P. 1–5.
4. **Teager H. M.** Some observations on oral air flow during phonation // IEEE Trans. Acoust., Speech, Signal Processing, 1980. – Vol. 28. – P. 599–601.
5. **Kaiser J. F.** On a simple algorithm to calculate the 'energy' of a signal // Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'90), 1990. – Vol. 1. – P. 381–384.
6. **Nehe N. S., Holambe R. S.** Power Spectrum Difference Teager Energy Features for Speech Recognition in Noisy Environment // Third Int. Conf. on Industrial and Information Systems (ICIIS'2008), 2008. – P. 1–5.
7. **Huang M.-Y., Tseng D.-C., Liu M. S. C.** Wavelet image enhancement based on Teager energy operator // Proc. 16th Int. Conf. on Pattern Recognition, 2002. – Vol. 2. – P. 993–996.
8. **Nelson R., et al.** Detection of High Frequency Oscillations with Teager Energy in an Animal Model of Limbic Epilepsy // 28th Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBS'06), 2006. – P. 2578–2580.
9. **Kaleem M. F., et al** Application of Empirical Mode Decomposition and Teager energy operator to EEG signals for mental task classification // Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC'10), 2010. – P. 4590–4593.
10. **Ahmed B., Redissi A., Tafreshi R.** An automatic sleep spindle detector based on wavelets and the Teager energy operator // Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC'2009), 2009. – P. 2596–2599.
11. **Moore M., Mitra S., Bernstein R.** A Generalization of the Teager Algorithm // IEEE Workshop on Nonlinear Signal Processing. – Ann Arbor, Michigan, 1997.
12. **Tomar V., Patil H. A.** On the development of variable length Teager energy operator (VTEO) // 9th Annual Conf. of the Int. Speech Communication Association (ISCA'08), 2008. – P. 1056–1059.
13. **Kvedalen E.** Signal processing using the Teager energy operator and other nonlinear operators. – MSc thesis. – University of Oslo, 2003. – 20 p.
14. **Maragos P., Potamianos A.** Higher order differential energy operators // IEEE Signal Processing Lett., 1995. – Vol. 2. – No. 8. – P. 152–154.
15. **Vapnik V.** Statistical Learning Theory. – Wiley-Interscience, New York, 1998.
16. **Birbaumer N., et al.** A Brain-Controlled Spelling Device for the Completely Paralyzed // Nature magazine, 2000. – No. 398. – P. 297–298.
17. **Joachims T.** A Support Vector Method for Multivariate Performance Measures // 22nd Int. Conf. on Machine Learning (ICML), 2005. – P. 377–384.
18. **Damaševičius R.** Optimization of SVM Parameters for Recognition of Regulatory DNA Sequences // TOP, 2010. – Springer, 2010. – Vol. 18. – No. 2. – P. 339–353

Received 2011 12 13

Accepted after revision 2012 01 14

### **I. Martišius, R. Damaševičius, V. Jusas, D. Birvinskas. Using Higher Order Nonlinear Operators for SVM Classification of EEG Data // Electronics and Electrical Engineering. – Kaunas: Technologija, 2012. – No. 3(119). – P. 99–102.**

Brain-Computer Interface (BCI) systems require application of complex analysis, signal processing, denoising, feature extraction, dimensionality reduction and classification methods on acquired raw electroencephalogram (EEG) data to allow for useful operation. In this paper, we consider application of nonlinear operators such as Teager-Kaiser Energy Operator (TKEO) and its multiple generalizations on the EEG signals and evaluate the efficiency of the operators using a Support Vector Machine (SVM) classifier with linear kernel. We propose a new generalization of TKEO, called Homogeneous Multivariate Polynomial Operator (HMPO), and compare the efficiency of the 3rd order HMPO with other nonlinear operators. Experimental results show that the 3rd order HMPO operator allows for better identification of significant features representing slow cortical potentials in the EEG data. III. 1, bibl. 18, tabl. 1 (in English; abstracts in English and Lithuanian).

### **I. Martišius, R. Damaševičius, V. Jusas, D. Birvinskas. Netiesinių aukštesnės eilės operatorių taikymas EEG duomenims klasifikuoti naudojant SVM klasifikatorių // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2012. – Nr. 3(119). – P. 99–102.**

Tinkamam rezultatui užtikrinti smegenų ir kompiuterio sąsajoms reikalingi sudėtingi elektroencefalogramos (EEG) signalų apdorojimo, triukšmų šalinimo, požymių išskyrimo ir klasifikavimo algoritmai. Šiame straipsnyje nagrinėjamas netiesinių operatorių, tokių kaip Teager-Kaiser energijos operatorius (TKEO) ir keletas jo atmainų, naudojimas EEG signalams apdoroti. Šių operatorių efektyvumas vertinamas klasifikavimui taikant atraminių vektorių metodą (SVM) su tiesine funkcija. Siūlomas naujas TKEO algoritmo apibendrinimas, pavadintas multivariantiniu homogeniniu polinomų operatoriumi (HMPO), ir 3-ios eilės HMPO lyginamas su kitais netiesiniais operatoriais. Eksperimento rezultatai rodo, kad 3-ios eilės HMPO operatorius leidžia lengviau atpažinti lėtų žievinų potencialų pasikeitimo požymius. II. 1, bibl. 18, lent. 1 (anglų kalba; santraukos anglų ir lietuvių k.).