An Application of Tucker Decomposition for Detecting Epilepsy EEG signals

Authors: Thao Nguyen Thieu[1], Hyung-Jeong Yang[2,*]

Abstract

Epileptic Seizure is a popular brain disease in the world. It affects the nervous system and the activities of brain function that make a person who has seizure signs cannot control and predict his actions. Based on the Electroencephalography (EEG) signals which are recorded from human or animal brains, the scientists use many methods to detect and recognize the abnormal activities of brain. Tucker model is investigated to solve this problem. Tucker decomposition is known as a higher-order form of Singular Value Decomposition (SVD), a well-known algorithm for decomposing a matric. It is widely used to extract good features of a tensor. After decomposing, the result of Tucker decomposition is a core tensor and some factor matrices along each mode. This core tensor contains a number of the best information of original data. In this paper, we used Tucker decomposition as a way to obtain good features. Training data is primarily applied into the core tensor and the remained matrices will be combined with the test data to build the Tucker base that is used for testing. Using core tensor makes the process simpler and obtains higher accuracies.

Key Words: Epileptic Seizure, Tucker Decomposition, Electroencephalography (EEG)

I. INTRODUCTION

Feature extraction and feature selection are the key criteria in classification and pattern recognition problems. This is greatly useful in the large dimension of input data cases. In recent years, interest in tensor decomposition is gotten more attention from many researchers from various fields. Tensor decomposition is a method to divide a tensor in multidimensionality into many smaller parts. Two well-known methods in this area are CANDECOMP/PARAFAC (CP) decomposition and TUCKER decomposition. Both CP decomposition and Tucker decompositions can be considered as higher-order generalization of Singular value decomposition (SVD) and Principle Component Analysis (PCA) [4]. These methods are always used to decompose data into simpler form, containing better features. Decompositions of tensor have applications in various fields: data mining, neuroscience, graph analysis, computer vision and elsewhere. This paper gives an application of tensor decomposition into EEG signals.

Electroencephalography (EEG) is a kind of the brain data. It is a measurement of many brain signals which are recorded from a number of electrodes placed along the scalp. More clearly, EEG is the recording of electrical activity along the scalp. It measures voltage fluctuations resulting from ionic current flows within the neurons of the brain. In clinical contexts. EEG refers to the recording of the brains spontaneous electrical activity over a short period of time, usually 20 to 40 minutes. An example of EEG signal is showed in figure 1.
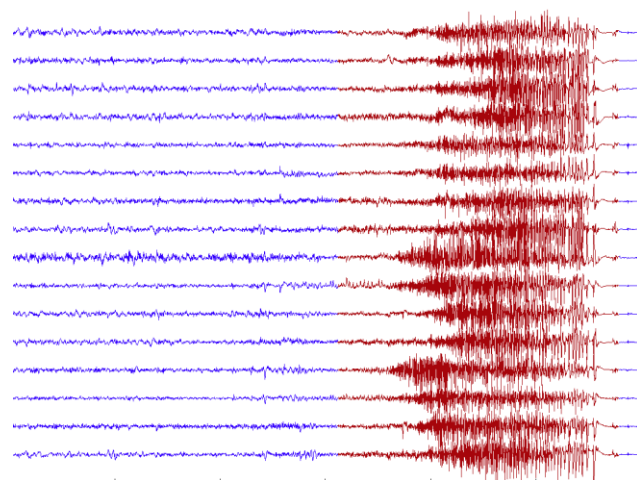

Fig. 1. EEG signal with 16 electrodes

EEG signals are widely used in current medical fields or signal processing. Recording brain signals of human, or even animal, is very important and useful in medical area. EEG records the abnormal activities of brain. Therefore,

the scientists could detect and diagnose some brain diseases based on this signal. One of the most popular brain diseases is Epilepsy.

Epilepsy is a sudden and recurrent brain malfunction and is a disease that reflects an excessive and hypersynchronous activity of the neurons within the brain [1]. It is probably the most prevalent brain disorder among adults and children. Over 50 million people worldwide are diagnosed with epilepsy, whose hallmark is recurrent seizures [2]. The prevalence of epileptic seizures changes from one geographic area to another [3]. This illness is very popular and it quite affects to human activities. It occurs frequently, repeated spontaneous, unpredictable and uncontrollable. For example, a man, who has seizure, could not control his actions. He could do some sudden actions which cannot be predicted and controlled. Nowadays, the scientists are constantly trying to find more effective methods to control this diseases. Figure 1 describes the Epilepsy EEG signals with 16 channels. The blue signals represent non-seizure segments and the red signals are seizure segments.

This paper proposes a method to diagnose and prevent epilepsy in human or animal successfully

In this paper, Tucker decomposition is used as a method to extract features or reduce the dimension of data. Naïve Bayes classifier is investigated as an algorithm to diagnose the epileptic seizure. More formally, we also show some notations related to tensor and tensor decomposition to support for this literature. It is presented in Section II. Section III is the way how to apply Tucker decomposition into third-order tensor. The description of EEG data and the result of classification process will be showed in Section IV. We conclude this paper in Section V.

II. DEFINITION OF TENSOR AND ITS CONCERNED NOTATIONS

1. Definition of tensor

A tensor is a multidimensional or N-way array. A tensor in N-dimension is called Nth-order tensor or N-way tensor. It is an element of the tensor product of N vector spaces [4].
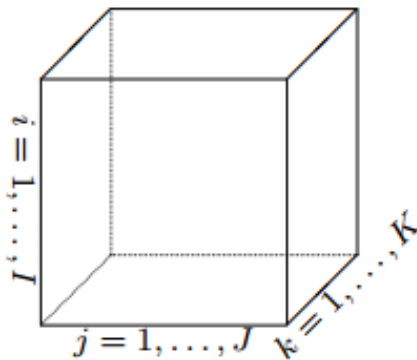

Fig. 2. An example of a third-order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ [4]

More clearly, a first-order tensor is a vector, a second-order tensor is a matric. Figure 2 represents a third-order tensor with three indices. The *order* of a tensor is the number of dimensions, also known as ways.

The definition of *fibers* in a tensor is for the higher order analogue of matrix rows and columns. When extracted from the tensor, fibers are always assumed to be oriented as column vectors.

Slices are two-dimensional sections of a tensor. A third-order tensor includes there slices: horizontal, lateral and frontal slices.

2. Simple notations
2.1. Norm of tensor

The norm of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is as following:

$$\|\mathcal{X}\| = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \dots i_N}^2}$$

2.2. Matricization

Matricization process of a tensor is the process of reordering the elements of an N-order tensor into a matrix.

Tensor elements $(i_1, i_2, \dots, i_N)$ maps to matrix element (i_n, j) where

$$j = 1 + \sum_{\substack{k=1 \\ k \neq n}}^{N} (i_k - 1) J_k$$

with $J_k = \prod_{\substack{m=1 \\ m \neq n}}^{k-1} I_m$

Note that, the mode-n matricization of tensor $\mathcal{X}$ is denoted by $\mathbf{X}_{(n)}$ and arranges the mode-n fibers to be the columns of the outcome matrix.

3. Tensor multiplication
3.1. The n-mode product

The n-mode product of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ with a matric $\mathbf{U} \in \mathbb{R}^{J \times I_n}$ denoted by $\mathcal{X} \times_n \mathbf{U}$ and is of size

3.2. The Kronecker product $I_1 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N$. We note that, J will be replaced by the position of I_n for the n-mode product.

Elementwise, we have

$$(\mathcal{X} \times_n \mathbf{U})_{i_1 \dots i_{n-1} j i_{n+1} \dots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 \dots i_N} u_{j i_n}$$

Or we can rewrite as following:

$$\mathcal{Y} = \mathcal{X} \times_n \mathbf{U} \Leftrightarrow \mathbf{Y}_{(n)} = \mathbf{U}\mathbf{X}_{(n)}$$

3.2. The *Kronecker* product

The Kronecker product of matrices $\mathbf{A} \in \mathbb{R}^{I \times J}$ and $\mathbf{B} \in \mathbb{R}^{K \times L}$ is denoted by $\mathbf{A} \otimes \mathbf{B}$ with the size $(IK) \times (JL)$ and defined by

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1J}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & & \\ \vdots & & \ddots & \\ a_{I1}\mathbf{B} & a_{I2}\mathbf{B} & & a_{IJ}\mathbf{B} \end{bmatrix}$$
$$= [\mathbf{a_1} \otimes \mathbf{b_1} \quad \mathbf{a_1} \otimes \mathbf{b_2} \quad \cdots \quad \mathbf{a_J} \otimes \mathbf{b_{L-1}} \quad \mathbf{a_J} \otimes \mathbf{b_L}]$$

Note that, $\mathbf{a}_i$ and $\mathbf{b}_j$ are the ith and jth column of matrices $\mathbf{A}, \mathbf{B}$, respectively.

III. THE TUCKER DECOMPOSITION

1. Tucker decomposition

Tucker decomposition (TD) was firstly introduced by Tucker in 1963 and refined in 1966. The 1966's version is the most comprehensive of the early literatures [4]. Tucker decomposition has many various names such as Higher-order SVD [5], *N*-mode SVD [6], three-mode factor analysis [7], *N*-mode principle components analysis [8].

The Tucker decomposition is the form of higher order singular value decomposition. This method decomposes a tensor into a main part, called core, and some matrices alone each mode. Therefore, Tucker decomposition is indicated in the three-way case where $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ as following:

$$\mathcal{X} \approx \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}$$
$$= \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} a_p \circ b_q \circ c_r \quad (1)$$

Here, $\mathbf{A} \in \mathbb{R}^{I \times P}, \mathbf{B} \in \mathbb{R}^{J \times Q}$, and $\mathbf{C} \in \mathbb{R}^{K \times R}$ are the factor matrices and the tensor $\mathcal{G} \in \mathbb{R}^{P \times Q \times R}$ is the *core tensor* after decomposing and its entries show the level of interaction between the different components.

Furthermore, P, Q and R are the number of components in factor matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$, respectively. In some cases, the storage for the decomposed version of the tensor can be significantly smaller than for the original tensor. The Tucker decomposition is also illustrated in figure 3.

The metricized form of Tucker decomposition in the three-way case are

$$\mathbf{X}_{(1)} \approx \mathbf{A}\mathbf{G}_{(1)}(\mathbf{C} \otimes \mathbf{B})^{\mathsf{T}}$$
$$\mathbf{X}_{(2)} \approx \mathbf{B}\mathbf{G}_{(2)}(\mathbf{C} \otimes \mathbf{A})^{\mathsf{T}}$$
$$\mathbf{X}_{(3)} \approx \mathbf{C}\mathbf{G}_{(3)}(\mathbf{B} \otimes \mathbf{A})^{\mathsf{T}}$$

$\otimes$ denotes the *Kronecker product*.

The Tucker model can be generalized to *N*-way tensors for third-order case as:

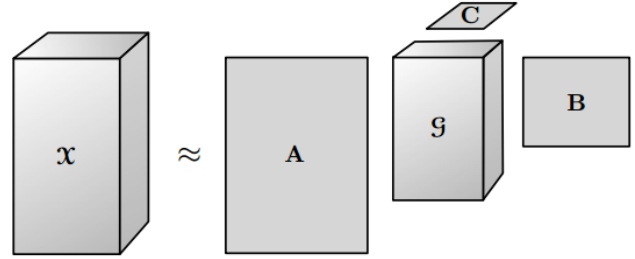$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}$$

Fig. 3. Tucker decomposition of a 3-way tensor [4]

2. Computing the Tucker decomposition

There are some methods to compute the Tucker decomposition. We also introduce one of them as in figure 5. It is called Tucker's "Method I" introduced by Tucker in 1966.
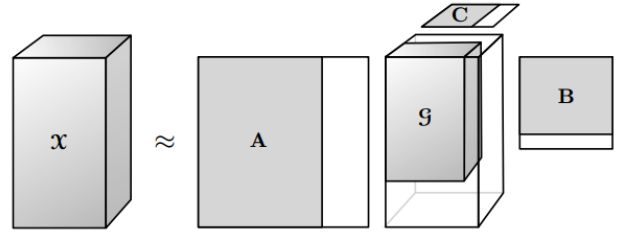

Fig. 4. Truncated Tucker decomposition of a three-way array [4]

```
procedure HOSVD(X, R₁, R₂, ..., R_N)
    for n = 1, ..., N do
        A⁽ⁿ⁾ ← R_n leading left singular vectors of X₍ₙ₎
    end for
    G ← X ×₁ A⁽¹⁾ᵀ ×₂ A⁽²⁾ᵀ ··· ×_N A⁽ᴺ⁾ᵀ
    return G, A⁽¹⁾, A⁽²⁾, ..., A⁽ᴺ⁾
end procedure
```

Fig. 5. Tucker's "Method I" for computing Tucker decomposition

Figure 4 shows a truncated Tucker decomposition. The truncated TD is not optimal in terms of giving the best fit as measured by the norm of the difference. However, it is a good starting point for an iterative alternating least squares algorithm.

4. The process for classification

The data signals, in this paper, can be represented as a third-order tensor. Following [9], firstly, applying Tucker decomposition into the training set. After decomposing a three-way tensor data, we obtained another tensor $\mathcal{G}$, called core tensor, and three matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$.

Reduced features are obtained by projecting the data tensor onto the feature subspace spanned by basis factors

A, B, C. In the other way, they are combined with the test part to build up the features for testing.

For more clearly, we consider a set (Train) of training samples and a set of test data (Test). The classification paradigm can be generally performed in these following steps:

1. Find the set of basis matrices and corresponding features for the training data.
2. Perform feature extraction for test data using the basis factor matrices found for the training data.
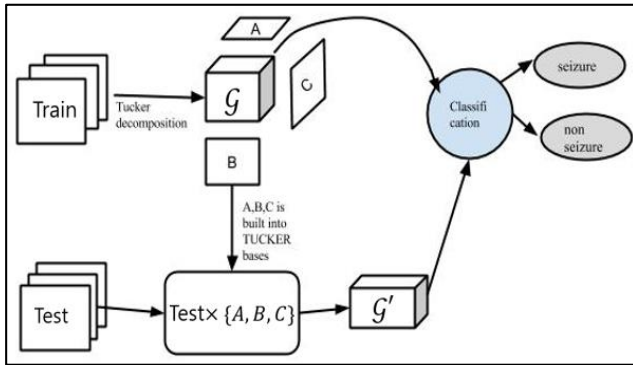3. Perform classification by comparing the test features with the training features.

Fig. 6. The flowchart of the process

The core tensor represents that the feature is much lower dimension than the raw data tensor. In other words, the reduced core tensor consists of features of the training samples in the subspace of the factor matrices.

The conceptual diagram illustrating a classification procedure based on Tucker decomposition of all sampling training data is clearly represented in figure 6.

IV. NAÏVE BAYES FOR CLASSIFICATION

Naïve Bayes classification is a well-known probabilistic classifier and is widely used in the world. Naïve Bayes classifier is produced based on Bayes theorem with strong independence assumptions between the features as following:

$$P(Y|X) = \frac{P(XY)}{P(X)} = \frac{P(X|Y)P(Y)}{P(X)}$$

Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values.

Applying for the classification problem, with $X = (x_1, ..., x_n)$ represents the training data which was vectorized. Using Bayes's theorem, the conditional probability can be decomposed as:

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

Where C_i is the ith possible outcomes or classes.
Therefore, we need to compute the probability of data over the outcome:

$$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i)$$

The algorithm of Naïve Bayes classifier can be summarized as following:

Step 1: Training Naïve Bayes based on the training data. Calculating $P(C_i)$ and $P(x_k|C_i)$

Step 2: Finding $\max_{C_i \in C}(P(C_i) \prod_{k=1}^n P(x_k|C_i))$

V. EXPERIMENT

1. Data description

The EEG dataset was recorded from four dogs with naturally occurring epilepsy using an ambulatory monitoring system. The dogs were enrolled in the project by their owners, who were seeking better treatment options for their pets during routine veterinary care. This data set is gotten freely at the International Epilepsy Electrophysiology portal, developed by the University of Pennsylvania and the Mayo Clinic [11].
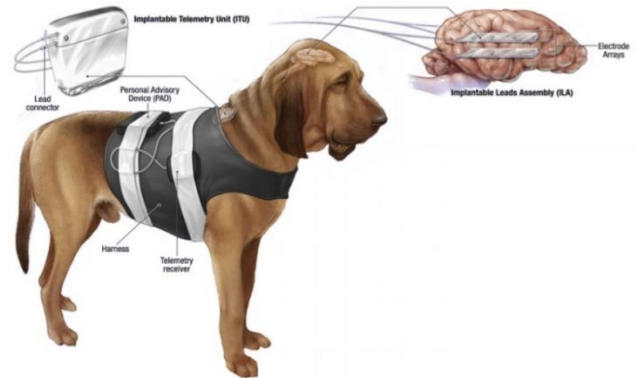

Fig. 7. A dog with its recording backpack [11][12]

All of signals in this dataset are sampled at 400Hz with 16 electrodes spreading on each dog head.

Figure 7 shows the dog with its recording backpack. Two strips of 4 electrodes were placed on either side of the cortex and the data was recorded continuously for a prolonged period of time

The table 1 shows the number of segments in four dogs. The data used in this paper are arranged sequentially.

Table 1. Number of samples

Subjects	Dog 1	Dog 2	Dog 3	Dog 4
Non-seizure data	418	1148	4760	2790

Seizure data	178	172	480	257

The figure 8 was plotted by the International Epilepsy Electrophysiology Portal's tool with 16 electrodes in the first 30 seconds.
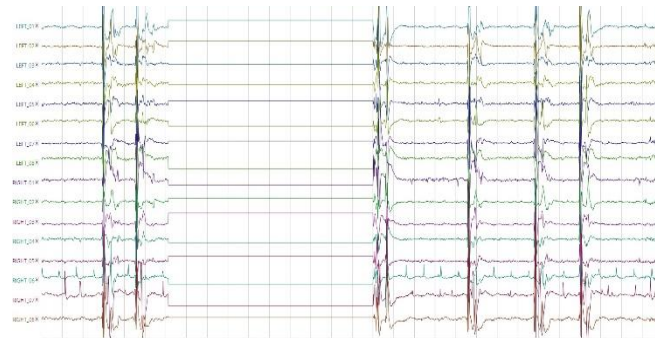

Fig. 8. The signals of a dog in the first 30 seconds.
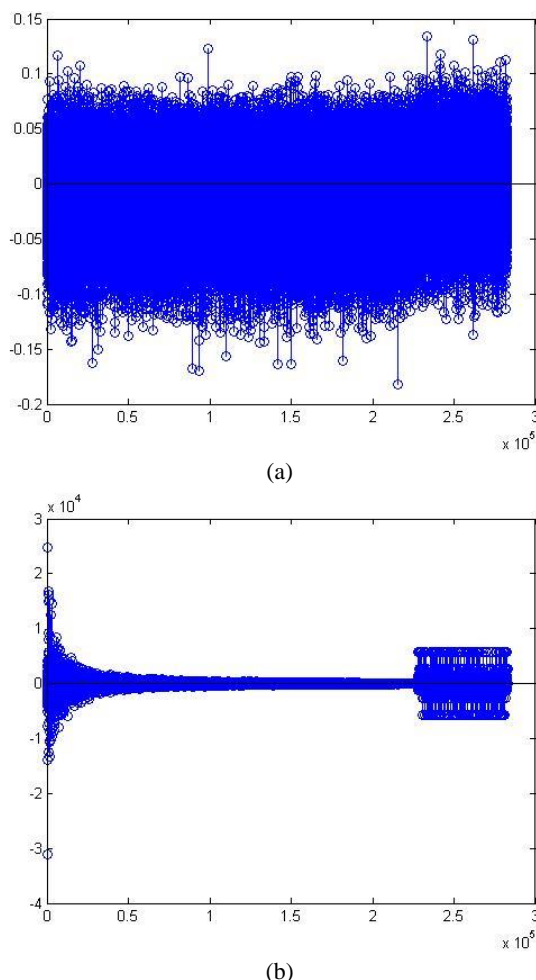

(a)

(b)

Fig.9. An example of the first channel of Dog2 (a) The original EEG data, (b) The signal of the core tensor after applying Tucker decomposition.

Dataset used in this paper is recorded to discriminate between seizure and non-seizure segments. Each segment is sampled in one second. The segments are arranged

sequentially, firstly seizure segments, then non-seizure segments. The data will be rebuilt as a third-order tensor of size $channels \times time \times segments$ to apply Tucker decomposition more easily.

2. Classification with Tucker decomposition

The training data was constructed as a three-dimension tensor of size $channels \times time \times segments$ signals.

The comparison of original signal and decomposed signal is represented in figure 9. The original signal (a) still contains much noise. In figure 9(b), the seizure and non-seizure segments are clearly discriminated by plotting the signal in the core tensor. Because Tucker model extracts the good information in the raw data and only retains the best features of the signal. Therefore, the signal in figure 9(b) is very clear and contains quite little noise information.

To classify the data, we trained a Naïve Bayes classifier. The accuracies of all dogs are always greater than 90 percent, even 95 percent. In the case, not using Tucker decomposition, the result of the first subject is quite bad, the accuracy is only approximately 79%. The other subjects get higher results, but not really good.

Table 2. The accuracy after classifying.

Subjects	Dog 1	Dog 2	Dog 3	Dog 4
Tucker + Naïve	0.91	0.93	0.95	0.92
Tucker + Neural	0.62	0.84	0.61	0.69
Non-Decomposition	0.79	0.93	0.90	0.88

We also trained dataset using another classifier, Neural Network, but the obtained results are not good. Naïve Bayes using in this case is better. All results are showed in Table 2. We also know that Naïve Bayes is a very simple classification method. Many studies are shown that NB can work with a large number of attributes and similarly for classes. The accuracy of NB is calculated by the probability relationship of each data and each class. And in the case, the difference of classes is clearly discriminated, Naïve Bayes of course gives very good accuracies. About Neural Network, this is still a good classifier but theoretically, it needs to be retrained many times to get a good weight (not optimal weight) for each arc in the network. It is also dependent many decisions about hidden layers, topology and variants. We note that the accuracies are also affected by the dividing between the number of non-seizure and seizure samples in each subject as the case of Dog1.

V. CONCLUSION

In this paper, a decomposition method is applied as a way to extract good features for EEG data in multidimensional space, specifically Tucker decomposition, or later known as Higher-order singular value decomposition (HOSVD).

233

The main goal of decomposition is "dividing" a multidimensional data into many parts which is smaller than original data but still contains full main information. A core tensor after decomposing a three-way data is a key feature for training. The remaining parts of this process will be contributed as the Tucker bases to build their test features.

Furthermore, the combination of Tucker decomposition and Naïve Bayes gave high accuracy. Naïve Bayes classifier is very useful and suitable for the case that has clear discrimination between two classes. Therefore, detecting a subject be a seizure or not becomes better.

REFERENCES

[1] Saeid Sanei, Jonathon Chambers *EEG signal processing*, pp. 161-191, 2007.

[2] Iasemidis, L. D., "Epileptic seizure prediction and control", *IEEE trans. Biomed Engng.*, 50, 2003.

[3] Annegers, J. F., "The epidemiology of epilepsy", *The treatment of Epilepsy*, pp. 157-164, 1993.

[4] Tamara G. Kolda, Brett W. Bader, "Tensor Decomposition and Applications", June 10, 2008.

[5] L. De Lathauwer, B. De Moor and J. Vandewalle, "A multilinear singular value decomposition", *SIAM Journal on Matrix Analysis and Applications*, pp. 1324-1342, 2000

[6] M.A.O.Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensor face", 7th *European Conference on Computer Vision*, 2002.

[7] L.R. Tucker "Some mathematical notes on three-mode factor analysis", *Psychometrika,* Vol. 31, pp. 279-311, 1966.

[8] P.M.Kroonenberg and J. De Leeuw, "Principal component analysis of three-mode data by means of alternating least squares algorithms", *Psychometrika,* Vol. 45, pp. 69-97, 1980.

[9] Anh Huy Phan, Andrzej Cichocki, "Tensor decompositions for feature extraction and classification of high dimensional datasets", *Nonlinear theory and its applications*, IEICE 1 (1), 37-68, 2010

[10] Marcel van Gerven, "Tensor decomposition for probabilistic classification", 2007

[11] The International Epilepsy Electrophysiology portal: https://www.ieeg.org/

[12] Upenn and Mayo Clinic's Seizure Detection Challenge: https://www.kaggle.com/c/seizure- detection

[13] Ali Shoeb, John Guttag, "Application of Machine Learning To Epilepsy Seizure". *Massachusettes Institute of Technology*, 02139, 2010

[14] Richard A.Johnson and Dean W.Wichern: *Applied Multivariate Statistical Analysis*, Sixth Edition.

[15] Zenglin Xu, Feng Yang, Yuan Qi, " Infinie Tucker Decomposition: Nonparametric Bayesian Models for Multiway data Analysis", *Proceedings of the 29 th International Conference on Machine Learning*, Edinburgh, Scotland, UK, 2012

[16] Bilian Chan, Zhening Li and Shuzhong Zhang, "On Tensor Tucker decomposition: the case for an adjustable core size", Technical report.

[17] A. Cichocki, D. mandic, A-H Phan, C. Caiafa, G. Zhou, Q. Zhao and L. De Lathauwer, "Tensor decomposition for Signal Processing Applications", *Signal Processing Magazine*, 2015

[18] Wikipedia, http://en.wikipedia.org/wiki

[19] N. Renard and S. Bourennane, "Dimensionality reduction based on tensor modeling for classification methods," IEEE Trans. Geosci. Remote Sens., vol. 47, no. 4, pp. 1123–1131, Apr. 2009.

[20] A. Cichocki, R. Zdunek, and A. H. Phan: Nonnegative Matrix and Tensor Factorizations. Chichester, U.K.: Wiley, 2009.

[21] Azam Karami, Mehran Yazdi, Gregoire Mercier: Compression of Hyperspectral Images Using Discrete Wavelet transform and Tucker decomposition, *IEEE Journal of selected topics in applied earth observations and remote sensing,* 2012.

[22] N. Kalouptsidis, "Signal Processing System, theory and design", John Wiley & Sons, INC, 1997.

[23] Rodrigo Quian Quiroga "Quantitative analysis of EEG signals: Time-frequency methods and Chaos theory", Institute of Physiology, Medical University Lubeck, 1998.

[24] Niedemeyer E. and da Silva F. L. "Electroencephalography: Basic Principles", Clinical Applications and Related Fields, 2004.

Authors

Thao Nguyen Thieu is studying Master's course at Chonnam National University. She graduated the bachelor at Ho Chi Minh University of Sciences. She research in to bio-signals as EEG, EMG.

H.J. Yang received her B.S., M.S. and Ph.D. degrees from Chonbuk National University, Korea. She was a Post-doc researcher at Carnegie Mellon University, USA. She is currently an associate professor at Dept. of Electronics and Computer Engineering, Chonnam National University, Gwangju, Korea. Her main research interests include multimedia data mining, pattern recognition, artificial intelligence, e-Learning, and e-Design