TENSOR-BASED EPILEPSY PREDICTION SYSTEM WITH INCOMPLETE MULTI-WAY ELECTROENCEPHALOGRAM

HỆ THỐNG DỰ ĐOÁN ĐỘNG KINH DỰA TRÊN TENSOR VỚI ĐIỆN NÃO ĐỒ ĐA CHIỀU KHÔNG HOÀN CHỈNH

Nguyen Thi Ngoc Anh[1]*, Vo Trung Hung[2]

[1]*The University of Danang – University of Science and Education, VietNam*
[2]*The University of Danang – University of Technology and Education, VietNam*

*Corresponding author: ngocanhnt@ued.udn.vn

(Received: May 26, 2024; Revised: July 20, 2024; Accepted: July 23, 2024)

Abstract - In this paper, our overarching goal is to propose a novel technology that will facility to analyze multi-way arrays electroencephalogram (EEG) brain signals to forecast epileptic seizure activity in the presence of missing entries while most previous conventional techniques commonly are restricted to perform the forecast epilepsy through 2D-based noninvasive EEG with complete channels. As such, the proposed method can forecast future trends of epilepsy activity while simultaneously dealing with missing data within one framework automatically. The key novelty of the proposed method demonstrates (1) exploiting both latent states and time series dynamics for detecting patterns trends information in missing reconstruction as well as prediction, (2) preserving the nature of tensor structure of multiway EEG brain signals. The proposed method performs its robustness via demonstrating high seizure forecasting accuracy through the comparative study with other techniques on the real public dataset with different scenarios of corrupted data.

Key words - Electroencephalogram; missing data; epilepsy forecasting; tensor completion; time series analysis

Tóm tắt - Mục tiêu của bài báo là đề xuất công nghệ mới có khả năng phân tích các tín hiệu đa chiều của điện não đồ (EEG) để dự báo hoạt động co giật động kinh khi có sự hiện diện các giá trị bị thiếu khi mà hầu hết các kỹ thuật thông thường trước đây thường bị hạn chế với cấu trúc 2D và các kênh hoàn chỉnh. Do đó, phương pháp đề xuất dự báo xu hướng bệnh động kinh trong tương lai đồng thời tự động xử lý dữ liệu bị thiếu trong cùng một khung. Điểm mới lạ chính của phương pháp đề xuất gồm (1) khai thác cả trạng thái tiềm ẩn và động lực chuỗi thời gian để phát hiện xu hướng mẫu trong quá trình phục hồi dữ liệu bị mất cũng như dự đoán, (2) bảo toàn bản chất cấu trúc tensor của tín hiệu não EEG đa chiều. Phương pháp được đề xuất phát huy tính mạnh mẽ của nó thông qua việc so sánh với các kỹ thuật khác trên tập dữ liệu thực tế với các kịch bản dữ liệu bị mất khác nhau.

Từ khóa - Điện não đồ; dữ liệu bị mất; dự đoán động kinh; tensor hoàn chỉnh; phân tích chuỗi thời gian

1. Introduction

Time series forecasting involves detecting hidden patterns and trends in historical observations to predict future values over time. Many conventional studies have demonstrated promising results in forecasting with high-quality time series data and completed measurements. Unfortunately, real-world time series data-driven problems in prediction systems often faced with the challenge of missing values due to malfunctions of recording devices or factors related to human errors. To tackle the referred aspects, a straightforward solution for such data is commonly ignored or discarded because they are considered unsuitable for further analysis. Obviously, this does not make sense if the data size is small and not long enough sample observations for producing forecasting. Thus, the effectiveness of forecasting systems is required to handle missing values. The challenge could be done by recovering these missing values problems via the vector/matrix-based completion algorithms [1-12].

Another critical factor that affects time series forecasting is noise and tensor structure. Indeed, modern data recording mechanisms are more granular than traditional ones via expanding to multi-way representation (i.e., tensor) [13-15]. The conventional forecasting approaches have been performed forecasting with small collections of time series or individual. Specifically, linear autoregressive model (AR) is a well-known prediction for time series that performs its robustness on vector representation in prediction. However, this method cannot do forecasting when missing values occur and is restricted for large-scale forecasting in practice [16]. The probabilistic model refers to dynamic state space modeling can detect latent variables comprising by a state vector at each time point and forecast future trends on them. The original algorithm is called Kalman Filter (KF) that is suitable for dynamic forecasting applications via fitting noise and produce temporal modeling. However, this method cannot handle missing values and discards high-order property of data structure. It means that KF is only utility for the model of latent factor and measurements at every time point under a vector formulation. A solution for improving KF [19] can do forecasting and missing value reconstruction via identifying temporal - spatial characteristics of time series and handling noise property. However, this approach again limits for a short period forecasting over time and tensor structure of observations input [17-19]. Dynamic Context-ual Matrix Factorization (DCMF) is a matrix-based forecasting that enables to encode temporal characteristics in time series. However, it is restricted when time series data is modeled as tensor structure [20]. In summary, although these traditional vector/matrices-based techniques are quite general but fail to identify patterns in tensor data in term of forecasting tasks since they break the multi-way representation of tensor time series and lead to lose for capturing and interpreting the underlying tensor structure. Furthermore, they have been applied only on complete time series data.

In contrast with most matrices-based algorithms, many tensor-based approaches directly developed upon matrices-based in terms of completion and forecasting by preserving multi-way properties of tensor objects. The well-known tensor-based decomposition is the TUCKER approach. The method discovers the latent factors from a tensor form of data observations which can reveal the underlying components in each dimension of the multi-way arrays without unfolding multi-way arrays into matrix formulation to applying matrices-based factorization. However, the method cannot capture time series dynamics or deal with noise characteristic and thus cannot exhibit forecasting for future values in time series [21]. Dynamic tensor analysis (DTA) [22] Bayesian Probabilistic Tensor Factorization (BPTF) was developed in Ref. [23] has brought great solution for the tensor time series structure. However, they could not easily predict future values and cannot deal with arbitrary noise properties as well.

Motivated by studies on brain electrical activity, specifically in epileptic seizure prediction applications, we propose a robust model for tensor time series analysis, whereby there is one tensor per timestamp, to understanding the structural properties and patterns for appropriately reveal the interactions among multiple modes. Consequently, multi-way arrays EEG analyzing for epilepsy forecasting task are investigated by introducing the mathematical concept and modified models based on multi-modal data construction and analysis from the recent advanced tensor algebra [25-27].

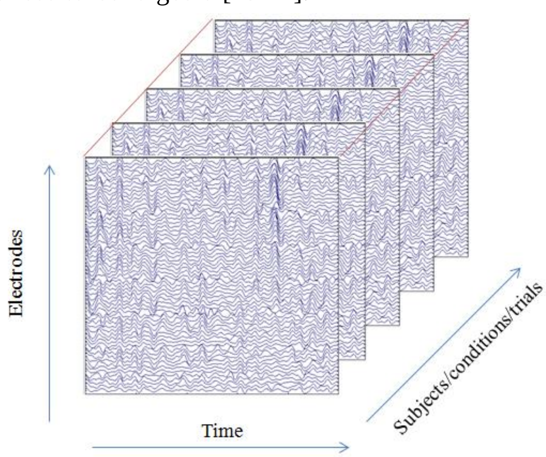

Figure 1. *An illustration of third-order structure of EEG epilepsy dataset: electrodes $\times$trials $\times$time samples*

In summary, the paper hinges on three critical factors in forecasting problems in tensor EEG time series-based epileptic seizure prediction, as given: (i) missing entries, long-term missing values is becoming big challenge for boosting forecasting accuracy in practice (ii) noise, in which time series forecasting task is affected by unwanted noise dramatically while analyzing because noise obscure underlying patterns or trends in a time series; (iii) multi-way representation: the noninvasive EEG-based epileptic seizure activity is naturally categorized by channel, epoch and time or trial, electrode and time frame, an illustrated 3-D tensor time series of EEG in Figure 1.

Problem definition: The general framework for tensor time series forecasting with occurrence of missing values is defined as follows; the overall graphics of the proposed forecasting formulation can be modeled as in Figure 2.

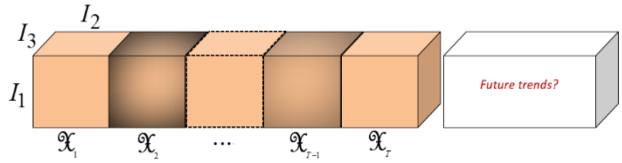

Figure 2. *Graphical illustration of the proposed forecasting model in 3-D tensor structure with the presence of missing values: $\mathcal{X}_1, \mathcal{X}_{T-1}$ are observed tensor while $\mathcal{X}_2, \mathcal{X}_T$ indicate partial observation tensor (missing tensor)*

Given an partially observed multi-way time series data $\mathcal{X}$ of finite collection *N-th* order tensor $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, ..., \mathcal{X}_T$ with time duration *T,* whereby each tensor $\mathcal{X}_T \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ has the dimension with others and it indicates as an partially observed tensor at every t^{th} time tick. Consequently, the index of $I_1 \times I_2 \times \cdots \times I_N$ ($1 \le n \le N$) denotes the dimension of the nth mode, each of the N dimension is defined the order of the observed/partial observed measurements $\mathcal{X}_T$. The last mode is the temporal mode with *T* dimensions that present the duration of the tensor time series.

An indicator tensor $\mathcal{W}_T \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N \times T}$ to define the missing values of observations of $\mathcal{X}_T$, given by:

$$\mathcal{W}_{i_1 i_2 \cdots i_N} = \begin{Bmatrix} 0 & \text{if } \mathcal{X}_{i_1 i_2 \cdots i_N} \text{ is a missing entry} \\ 1 & \text{if } \mathcal{X}_{i_1 i_2 \cdots i_N} \text{ is an observed entry} \end{Bmatrix} \quad (1)$$

The proposed method is an extension of original Kalman Filter to tensor-based formulation in general for multi-way EEG data for epilepsy forecasting with can handle missing values simultaneously. The contributions of the proposed method can be classified as follows: (i) successfully detecting the dynamics relational variables among temporal modes that boost the identification of strong patterns for predict future seizure activity; (ii) long-term period missing values can be supported while forecasting is performed; (iii) preserving the nature of the tensor time series structure of EEG epilepsy dataset. The proposed method proves its superiority in terms of long-term forecasting in the presence of missing values on empirical analysis of tensor time series from multi-way EEG compared with convention methods.

2. Proposed method

2.1. Notation and definition

Tensor is an generation of vector/matrices into higher-order form. A multi-dimensional array $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is called *N*-th order tensor [9, 13, 14, 15, 23]. Table 1 briefly introduces symbols used throughout the paper.

Table 1. *Symbol and definition*

Symbol	Definition and description
$\mathbf{a}$	Vector
$\mathbf{A}$	Matrix
$\mathcal{A}$	Tensor
$\odot$	Element-wise multiplication
$\otimes$	Kronecker product
$\circledast$	Contracted product
$\mathcal{X}$	Tensor time series with missing values
$\mathcal{X}_o$	The observed tensor in $\mathcal{X}$
$\mathcal{X}_m$	The partial observed tensor in $\mathcal{X}$
$\mathcal{X}_T$	Tensor time series at tth time tick
$\mathcal{W}_T$	Missing indicator tensor
$\mathcal{Z}$	Latent tensor
$\mathcal{B}$	The transition tensor
$\mathcal{D}$	The projection tensor
$\mathcal{Q}$	Transition tensor covariance
$\mathcal{Q}_o$	Initial tensor covariance
$\mathcal{R}$	Projection tensor covariance
$vec(\mathcal{X})$	Vectorization of tensor time series $\mathcal{X}$
$mat(\mathcal{X})$	Matricization of tensor $\mathcal{X}$

2.2. Proposed Tensor-based Kalman Filter

The key idea of the proposed method is to refine the standard Kalman Filter algorithm by extending indicator tensor missing value $\mathcal{W}$ and encoding multi-way EEG time series for forecasting task.

Firstly, the proposed approach initializes the latent tensor $\mathcal{Z}$ as a tensor normal distribution with the mean $\mathcal{U}_o$ and covariance $\mathcal{Q}_o$, formulated in (2).

$$\mathcal{Z}_1 \sim \mathcal{N}\left(\mathcal{U}_o, \mathcal{Q}_o\right) \tag{2}$$

Consequently, the latent tensor factor $\mathcal{Z}_t \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$ is formulated as multi-way Gaussian distribution with multi-way transition tensor $\mathcal{B}$ and the covariance $\mathcal{Q}$, in (3).

$$\mathcal{Z}_t \left| \mathcal{Z}_{t-1} \sim \mathcal{N}\left(\mathcal{B} \circledast \mathcal{Z}_{t-1}, \mathcal{Q}\right)\right. \tag{3}$$

The transition tensor $\mathcal{B}$ controls the temporal smoothness of evolving latent sequence tensor $\mathcal{Z}_1, \mathcal{Z}_{2,\cdots,}\mathcal{Z}_{T-1}, \mathcal{Z}_T$ that can be factorized into N factor matrices $\mathcal{B}^1, \mathcal{B}^2, \mathcal{B}^3, \cdots, \mathcal{B}^{N-1}, \mathcal{B}^N$, given in (4).

$$\text{mat}(\mathcal{B}) = \mathcal{B}^{\text{N}} \otimes \mathcal{B}^{\text{N}-1} \otimes \cdots \otimes \mathcal{B}^1 \tag{4}$$

The projection tensor $\mathcal{D}$ connects the sequence of observations and latent tensor $\mathcal{Z}_1, \mathcal{Z}_{2,\cdots,}\mathcal{Z}_{T-1}, \mathcal{Z}_T$. The conditional distribution of $\mathcal{X}_t$ is defined as a multi-way normal distribution with the covariance $\mathcal{R}$ with the mean is the contracted product of projection tensor and latent factor, given as equation (5).

$$\mathcal{X}_t \left| \mathcal{Z}_t \sim \mathcal{N}\left(\mathcal{D} \circledast \mathcal{Z}_t, \mathcal{R}\right)\right. \tag{5}$$

The projection tensor $\mathcal{D}$ also can be factorizable as in (6).

$$\text{mat}(\mathcal{D}) = \mathcal{D}^{\text{N}} \otimes \mathcal{D}^{\text{N}-1} \otimes \cdots \otimes \mathcal{D}^1 \tag{6}$$

Herein, our objective is to estimate the model probabilistic parameter $\theta \leftarrow \left\{\mathcal{U}_{\text{o}}, \mathcal{Q}_{\text{o}}, \mathcal{Q}, \mathcal{B}, \mathcal{R}, \mathcal{D}\right\}$ and search the optimal latent tensor factor $\mathcal{Z}_1, \mathcal{Z}_{2,\cdots,}\mathcal{Z}_{T-1}, \mathcal{Z}_T$ that able to maximize the following joint distribution of $\mathcal{X}_{\text{o}}, \mathcal{X}_{\text{m}}$ and $\mathcal{Z}$, as formulated in (7):

$$\begin{aligned} &\underset{\theta}{argmax}\, p\left(\mathcal{X}_{\text{o}}, \mathcal{X}_{\text{m}} \text{ and } \mathcal{Z}\right) \\ &= \underset{\theta}{argmax} \underbrace{\prod_{\text{t}=1}^{\text{T}} p\left(\mathcal{X}_{\text{t}} \middle| \mathcal{Z}_{\text{t}}\right)}_{\text{multi-way time series}} \cdot \underbrace{p\left(\mathcal{Z}_{\text{t}}\right) \cdot \prod_{\text{t}=2}^{\text{T}} p\left(\mathcal{Z}_{\text{t}} \middle| \mathcal{Z}_{\text{t}-1}\right)}_{\text{temporal dynamics}} \end{aligned} \tag{7}$$

The key idea for learning algorithm is to try to seek optimal values for maximizing equation (7). However, it is difficult because the modal exists the latent variables $\mathcal{Z}_1, \mathcal{Z}_{2,\cdots,}\mathcal{Z}_{T-1}, \mathcal{Z}_T$. Instead of dealing with this way, we investigate searching via expectation-maximization (EM) mechanism. With the fix set of parameters $\theta \leftarrow \left\{\mathcal{U}_{\text{o}}, \mathcal{Q}_{\text{o}}, \mathcal{Q}, \mathcal{B}, \mathcal{R}, \mathcal{D}\right\}$ that are first initialized randomly, EM mechanism is then presented until converging the required criterion. Specifically, with fixed observations and current parameters, the posteriors of the latent factor and their sufficient statistics via Kalman Filtering and Kalman smoothing procedure are calculated. With fixed both $\mathcal{X} = \mathcal{X}_{\text{o}} \cup \mathcal{X}_{\text{m}}$ and $\mathcal{Z}$, the new parameter of the proposed model $\theta \leftarrow \left\{\mathcal{U}_{\text{o}}, \mathcal{Q}_{\text{o}}, \mathcal{Q}, \mathcal{B}, \mathcal{R}, \mathcal{D}\right\}$ are then updated by maximizing the expectation of the distribution likelihood $\underset{\theta}{\text{argmax}}\, \text{E}\left[\log\left(\mathcal{X}_{\text{o}}, \mathcal{X}_{\text{m}} \text{ and } \mathcal{Z}\right)\right]$.

E-step:

Due to the drawback when derived the tensor normal distribution, thus with the fixed model parameters $\theta \leftarrow \left\{\mathcal{U}_{\text{o}}, \mathcal{Q}_{\text{o}}, \mathcal{Q}, \mathcal{B}, \mathcal{R}, \mathcal{D}\right\}$ and $\mathcal{X} = \mathcal{X}_{\text{o}} \cup \mathcal{X}_{\text{m}}$, we first unfold the tensor formulation by the vectorizations of $\mathcal{X}$ and $\mathcal{Z}$ and matricizations of $\mathcal{Q}_{\text{o}}, \mathcal{Q}, \mathcal{B}$, given following equation in (8,9,10).

$$\text{vec}\left(\mathcal{Z}_1\right) \sim \mathcal{N}\left(\text{vec}\left(\mathcal{U}_o\right), \text{mat}\left(\mathcal{Q}_o\right)\right) \tag{8}$$

$$\text{vec}\left(\mathcal{Z}_t\right) \middle| \text{vec}\left(\mathcal{Z}_{t-1}\right) \sim \mathcal{N}\left(\text{mat}\left(\mathcal{B}\right)\text{vec}\left(\mathcal{Z}_{t-1}\right), \text{mat}\left(\mathcal{Q}\right)\right) \tag{9}$$

$$\text{vec}\left(\mathcal{X}_t\right) \middle| \text{vec}\left(\mathcal{Z}_t\right) \sim \mathcal{N}\left(\text{mat}\left(\mathcal{D}\right)\text{vec}\left(\mathcal{Z}_t\right), \text{mat}\left(\mathcal{R}\right)\right) \tag{10}$$

The expectations and its sufficient statistics of latent tensor factors are then updated via Kalman Filtering (forward) and Kalman smoothing (backward) algorithm, as given as in (11,12,13):

$$E\left(\mathcal{Z}_t\right) = E\left(\text{vec}\left(\mathcal{Z}_t\right)\right) \tag{11}$$

$$E\left(\mathcal{Z}_t\left(\mathcal{Z}_t\right)'\right) = E\left(\text{vec}\left(\mathcal{Z}_t\right)\text{vec}\left(\mathcal{Z}_t\right)'\right) \tag{12}$$

$$E\left(\mathcal{Z}_t\left(\mathcal{Z}_{t-1}\right)'\right) = E\left(\text{vec}\left(\mathcal{Z}_t\right)\text{vec}\left(\mathcal{Z}_{t-1}\right)'\right) \tag{13}$$

M-step:

The new model parameter $\theta^{\text{new}} \leftarrow \left\{\mathcal{U}_{\text{o}}, \mathcal{Q}_{\text{o}}, \mathcal{Q}, \mathcal{B}, \mathcal{R}, \mathcal{D}\right\}$ can be divided into two kinds of formulation to update, namely non-multiway parameters and multiway parameters.

For case of non-multiway parameters, we take the derivation of the expected log-likelihood of the observation time sequences w.r.t the parameter to zero, defined as:

$$H(\theta)=E_{\mathfrak{X}_m,\mathfrak{Z}|\mathfrak{X}_o,\mathfrak{W}}\left(p\left(\mathfrak{X}_o,\mathfrak{X}_m,\mathfrak{Z}\right)\right) \tag{14}$$

By to obtain new non-multiway parameters, as follows:

$$\operatorname{vec}\left(\mathfrak{U}_o^{new}\right)=E\left(\operatorname{vec}\left(\mathfrak{Z}_1\right)\right) \tag{15}$$

$$\operatorname{mat}\left(\mathfrak{Q}_o^{new}\right)=E\left(\operatorname{vec}\left(\mathfrak{Z}_1\right)\operatorname{vec}\left(\mathfrak{Z}_1\right)'\right)-E\left(\operatorname{vec}\left(\mathfrak{Z}_1\right)\right)E\left(\operatorname{vec}\left(\mathfrak{Z}_1\right)'\right) \tag{16}$$

$$\operatorname{mat}\left(\mathfrak{Q}^{new}\right)=\frac{1}{T-1}\sum_{t=2}^{T}\left[\begin{array}{l} E\left(\begin{array}{l}\left(\operatorname{vec}\left(\mathfrak{Z}_t\right)\operatorname{vec}\left(\mathfrak{Z}_t\right)'\right)-\operatorname{mat}\left(\mathfrak{B}^{new}\right)\\ E\left(\operatorname{vec}\left(\mathfrak{Z}_{t-1}\right)'\operatorname{vec}\left(\mathfrak{Z}_t\right)'\right)\end{array}\right)\\ -E\left(\begin{array}{l}\operatorname{vec}\left(\mathfrak{Z}_t\right)\operatorname{vec}\left(\mathfrak{Z}_{t-1}\right)'\operatorname{mat}\left(\mathfrak{B}^{new}\right)'+\operatorname{mat}\left(\mathfrak{B}^{new}\right)\\ E\left(\operatorname{vec}\left(\mathfrak{Z}_t\right)\operatorname{vec}\left(\mathfrak{Z}_{t-1}\right)'\operatorname{mat}\left(\mathfrak{B}^{new}\right)\right)\end{array}\right)\end{array}\right] \tag{17}$$

$$\operatorname{mat}\left(\mathfrak{R}^{new}\right)=\frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{l} E\left(\begin{array}{l}\left(\operatorname{vec}\left(\mathfrak{X}_t\right)\operatorname{vec}\left(\mathfrak{X}_t\right)'\right)-\operatorname{mat}\left(\mathfrak{D}^{new}\right)\\ E\left(\operatorname{vec}\left(\mathfrak{Z}_t\right)\operatorname{vec}\left(\mathfrak{X}_t\right)'\right)\end{array}\right)\\ -\operatorname{vec}\left(\mathfrak{X}_t\right)E\left(\begin{array}{l}\operatorname{vec}\left(\mathfrak{Z}_t\right)'\operatorname{mat}\left(\mathfrak{D}^{new}\right)'+\operatorname{mat}\left(\mathfrak{D}^{new}\right)\\ E\left(\operatorname{vec}\left(\mathfrak{Z}_t\right)\operatorname{vec}\left(\mathfrak{Z}_t\right)'\operatorname{mat}\left(\mathfrak{D}^{new}\right)\right)\end{array}\right)\end{array}\right] \tag{18}$$

For case of multiway parameters of transition tensor $\mathfrak{B}$ and projection tensor $\mathfrak{D}$, they cannot update as non-multilinear parameters because they have specific form that can be factorized into N matrices. Consequently, we propose to maximize the following expected complete log-likelihood with respect to vector d and b via applying the gradient method to obtain projection tensor $\mathfrak{D}$ and transition tensor $\mathfrak{B}$, respectively.

$$L(d)=\operatorname{tr}\left[\begin{array}{l}\operatorname{mat}\left(\mathfrak{R}\right)^{-1}\operatorname{mat}\left(\mathfrak{D}\right)\\ \left[\sum_{t=1}^{T}E\left(\operatorname{vec}\left(\mathfrak{Z}_t\right)\operatorname{vec}\left(\mathfrak{Z}_t\right)'\right)\operatorname{mat}\left(\mathfrak{D}\right)'\right]\\ -2\left[\sum_{t=1}^{T}\operatorname{vec}\left(\mathfrak{X}_t\right)E\left(\operatorname{vec}\left(\mathfrak{Z}_t\right)'\right)\right]\end{array}\right] \tag{19}$$

$$L(b)=\operatorname{tr}\left[\begin{array}{l}\operatorname{mat}\left(\mathfrak{Q}\right)^{-1}\operatorname{mat}\left(\mathfrak{B}\right)\\ \left[\sum_{t=1}^{T}E\left(\operatorname{vec}\left(\mathfrak{Z}_t\right)\operatorname{vec}\left(\mathfrak{Z}_t\right)'\right)\operatorname{mat}\left(\mathfrak{B}\right)'\right]\\ -2\left[\sum_{t=1}^{T}\operatorname{vec}\left(\mathfrak{Z}_{t+1}\right)E\left(\operatorname{vec}\left(\mathfrak{Z}_t\right)'\right)\right]\end{array}\right] \tag{20}$$

In the last step, the future trend can be forecast easily based on the fixed model parameters and the update of latent tensors. Specifically, the next latent tensor $\mathfrak{Z}_{t+1}$ and new observations $\mathfrak{X}_{t+1}$ can be predicted as follows:

$$\mathfrak{Z}_{t+1}\leftarrow\mathfrak{B}\circledast\mathfrak{Z}_t \tag{21}$$

$$\mathfrak{X}_{t+1}\leftarrow\mathfrak{D}\circledast\mathfrak{Z}_{t+1} \tag{22}$$

During the forecasting tensor performance, the missing values can be imputed simultaneously. In particular, the missing values are first imputed by linear interpolation and then updated by the conditional expectation from the training data. The missing values of tensor $\mathfrak{X}_t$ can be inferred from the corresponding entries of vectorization of estimated tensor $\tilde{\mathfrak{X}}_t$, defined as follow:

$$\operatorname{vec}\left(\tilde{\mathfrak{X}}_t\right)=\operatorname{mat}\left(\mathfrak{D}\right)E\left[\operatorname{vec}\left(\mathfrak{Z}_t\right)\right] \tag{23}$$

Overall, the proposed tensor-based Kalman Filter for forecasting with the occurrence of missing values can be briefly summarized as in Figure 3:

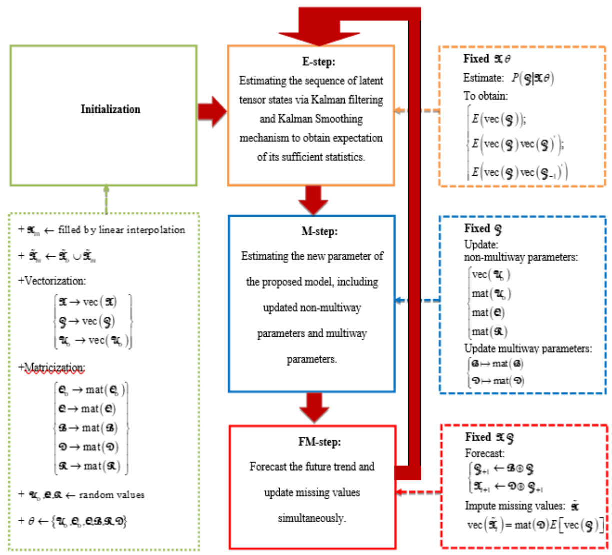

Figure 3. *The flowchart of the proposed tensor-based method for multi-way EEG-based epileptic seizure forecasting in the presence of missing values*

Model comprehension:

The proposed model is comprehensive in the sense that it simultaneously captures (1) the temporal smoothness along the time dimension, (2) noise: allow to model arbitrary noise, (3) tensor representation. As such, the proposed method includes several existing methods as its special cases, including:

- Kalman Filter: Relationship to KF model, the graphical representation of our HOKF becomes the traditional Kalman Filter if we set $N=1$ and the dataset is fully observed.

- DynaMMo: If we set $N=1$ and the data contain missing values of input observation, the model is similar to DynaMMo.

- Factor Analysis: If matricization of covariance tensor of transition tensor and projection tensor are setting to a diagonal matrix, the proposed model becomes Factor Analysis in case of $N = 1$.

- Probabilistic principal components analysis (PPCA): in case of matricization of covariance tensor is set to identity matrix, the proposed method is a special case of PPCA model.

3. Experimental results

3.1. Data description and experiment setting

The empirical performance is conducted on real epilepsy datasets, sourced from the Department of Epileptology, University of Bonn, Germany [28] to convince the capability of the proposed approach. Consequently, the third-order representation of multi-way EEG epilepsy can then be performed to discover the characteristics of underlying brain activity.

For generalizing the advantages of proposed algorithm for forecasting, experiments with different kinds of missing patterns including random missing entries and consecutive missing entries are performed, illustrated in Figure 4. Furthermore, the effectiveness of the proposed method is carried out with different time slices forecasting with fixed missing percentages. Specifically, some entries/portion of the data is deleted to mimic the data containing missing values with containing 20% of missing percentage. To reduce the random effect and to conclude a fair comparison with other methods, we did each simulation 20 times and recorded the average of MAE on all algorithms.

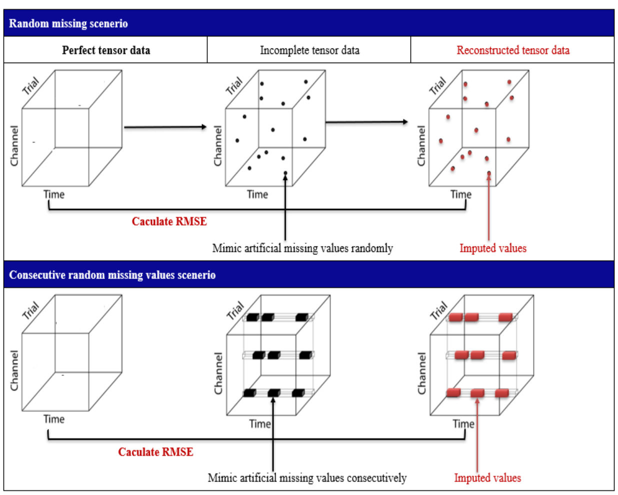

Figure 4. *Graphical illustration for generating two scenarios of artificial missing values patterns*

3.2. Evaluation criteria

To measure the quality of the proposed method, the root mean square error (RMSE) is defined in equation (26) and it used as a criterion to perform the effectiveness of the forecasting while containing missing values.

In formulation of matrix-based forecasting and reconstruction techniques, we define the RMSE between the actual data $\mathbf{X}$ and the imputed missing values data $\mathbf{X}$, as follows:

$$RMSE = \sqrt{\frac{\sum_{it}(1-\mathbf{W}_{it})\left(\mathbf{X}_{it}-\mathbf{X}_{it}\right)^2}{\sum_{it}(1-\mathbf{W}_{it})}} \qquad (26)$$

Our proposed tensor-based method is calculated as a similar way, the actual time sequence tensor $\mathfrak{X}$ and the reconstructed/forecasted tensor time series $\tilde{\mathfrak{X}}$ are flatted into the vectorization representation. Consequently, the eq. (26) is applied to computing the RMSE between them.

3.3. Experimental results

In our proposed model, the optimal of latent tensor factors throughout the following experiments is decided via using the heuristic rule, defined as in equation (25). The proposed method is trying to find the initialization of optimal number of latent factors for each mode by descending order of eigenvalues until satisfy the equation (25) with latent factor J = [10,15].

We first evaluate the efficiency and effectiveness of the proposed tensor Kalman Filter with perfect data (non-missing values occurrence) on different techniques in terms of forecasting. It is note that the proposed method (Tensor Kalman Filter) is differ with original Kalman Filter in way how the multiway parameter of transition tensor $\mathfrak{B}$ and projection tensor $\mathfrak{D}$ are updated. For a fair comparison, we define noise with matrices-based form for $\text{mat}(\mathfrak{Q})$, $\text{mat}(\mathfrak{Q}_o)$ and $\text{mat}(\mathfrak{R})$ as isotropic for both the standard Kalman Filter and the improved Tensor Kalman Filter. The common prediction algorithm in time series chosen for comparison is the AR model. Moreover, the robust matrix-based of DCMF technique is used since it presents the advantage via exploiting the latent factor and capture temporal characteristics for forecasting task.

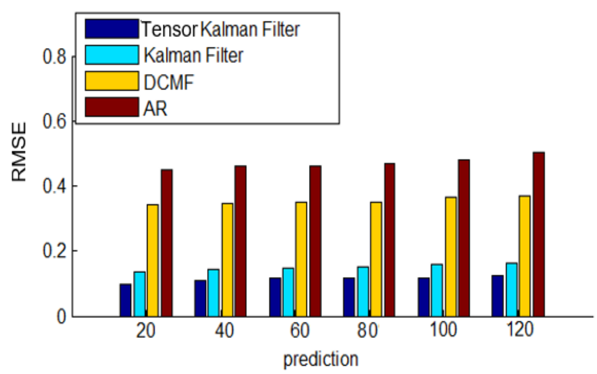

Figure 5. *RMSE comparison of proposed method with other techniques on complete tensor EEG-based epilepsy with different future time slices forecasting*

Figure 5 shows the RMSE error forecasting on four techniques where x-axis presents the duration time slices in forecasting, whereas the Y-axis performs the RMSE. It can be derived from the figure that the proposed method demonstrates slightly increase time slices for prediction with improvement up to 1.2x, 2x and 3x compared with Kalman Filter, DCME and AR, respectively. For all methods, the experiments were training with 90% original tensor dataset. The proposed method’s performance shows significant accuracy since it can successfully capture the temporal dynamics and correlation among modes of tensor time series.

It can be noted that the proposed method gives an impressive advantage for coping with the missing values naturally during the forecasting process performs. To confirm how effectiveness of the proposed method is, two types of missing patterns are considered for tensor time series structure, including missing entries at random where entries are randomly treated as missing values and consecutive missing values where a segment is occlusion.

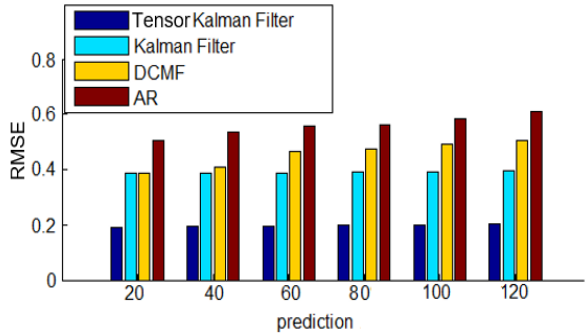

Figure 6. *RMSE comparison of proposed method with other techniques on incomplete tensor EEG-based epilepsy with different future time slices forecasting*

The experimental results have shown in Figure 6 present the forecasting while increasing different time slices prediction with fixed missing percentage up to 20%. As mentioned, the proposed model is able to discover the underlying relationship between modes of the recovered tensor whereas the Kalman Filter is neglected, which means that we can boost forecasting accuracy in the best way. Indeed, the proposed method, tensor-based Kalman Filter, demonstrates more accurate prediction than the original technique in imputation in order to boost the forecasting accuracy.

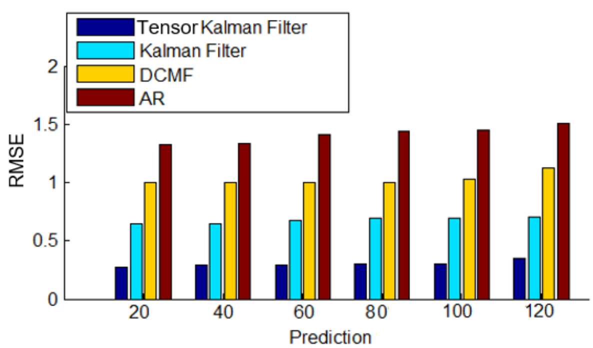

Figure 7. *RMSE comparison of proposed method with other techniques on 150 time slices consecutive missing pattern. It is noted that the Kalman Filter, DCMF and AR forecast on the reconstructed data from Kalman Filter*

As can be explicitly seen from the Figure 7, the experimental results show that the proposed technique has a lower RMSE performance even if the missing occlusion is consecutive lost with 150 time slices compared to the standard Kaman Filter imputation technique. Specifically, we implement by fixing 20% of missing percentage with 150 time slices occlusion and run Kalman Filter and proposed method imputation algorithms for reconstruction and forecasting. The experimental results from the bar chart of Figure 7 demonstrate that our proposed method again is visually superior to other competitors since the reconstructed signal is not affected much for the prediction. In particular, the AR, DCMF and Kalman Filter itself predict on the reconstructed data from the Kalman Filter with lower prediction accuracy up to 2x, 3x and 4x, respectively. To confirm how the proposed method can deal with occluded missing values effectively, Figure 8 performs the reconstruction of Kalman Filter and proposed method for comparison. Indeed, the proposed tensor-based Kalman Filter is outperforming compared with original techniques which is very close to the actual signal.

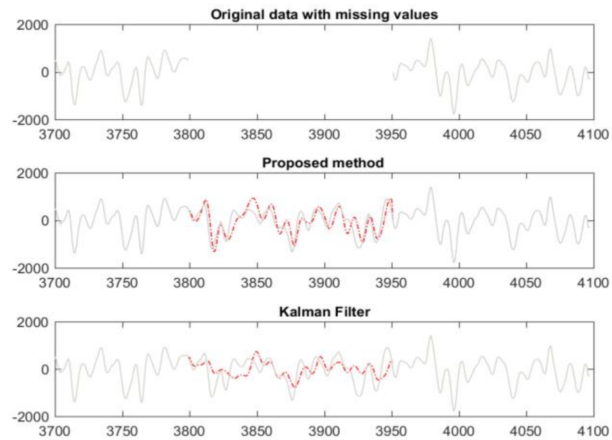

Figure 8. *An illustration of 150-time slices imputation while using 90% training data for forecasting: the grey dash indicates the ground truth values; from top to bottom: original data with occluded data; the reconstructed with proposed method; the last is for Kalman Filter method*

4. Conclusion

In this paper, the tensor-based Kalman Filter utilizing the forecasting with incomplete tensor time series has been proposed for the multi-way epilepsy EEG dataset. The main contribution of this research can be classified: (1) temporal dynamics: capturing the smoothness successfully among temporal modes that boost the discovery of key patterns for predict future seizure activity; (ii) incomplete data: handling missing values while forecasting process is demonstrated; (iii) tensor representation: preserving the original structure of nature multiway EEG EEG epilepsy dataset. The prediction performances on real multiway EEG epilepsy datasets demonstrated the effectiveness of our proposed approach that overwhelmingly outperforms compared with the matrices-based method in term of missing values imputation as well as forecasting tasks. Expectedly, the proposed method is a general framework in tensor time series with valuable outcomes that boost for extending and developing decision-making system for corrupted tensor time series on other real-life applications.

Acknowledgment: This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.01-2020.27.

REFERENCES

[1] C. H. Cheng, J. W. Wang, and C. H. Li, “Forecasting the number of outpatient visits using a new fuzzy time series based on weighted-transitional matrix”, *Expert system with applications*, vol. 4, no. 4,

pp. 2568 -2575, 2008.

[2] I. Rojas *et al.*, "Soft-computing techniques and ARMA model for time series prediction", *Neurocomputing*, vol. 71, issue 4-6, pp. 519-537, 2008.

[3] K. J. Blinowska, and M. Malinowski, "Non-linear and linear forecasting of the EEG time series", *Biological Cybernetics*, vol. 66, pp. 159–165, 1991.

[4] M. J. L. D. Hoon, T. H. J. J. Van der Hagen, H. Schoonewelle, and H. van Dam, "Why Yule-Walker should not be used for autoregressive modelling", *Annals of Nuclear Energy*, vol. 23, no. 15, pp. 1219–1228, 1996.

[5] S. M. Chen, and J. R. Hwang, "Temperature prediction using fuzzy time series", *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 30, no. 2, pp. 263–275, 2000.

[6] B. Brinkmann et al., "WE-120. Seizure forecasting and detection with wearable devices and subcutaneous EEG - A practical seizure gauge", *clinical neurophysiology*, vol. 141, pp. S62-S63, 2022.

[7] M. Mushtaq, U. Akram, M. Aamir, H. Ali, and M. Zulqarnain, "Neural Network Techniques for Time Series Prediction: A Review", *Int. J. Inform. Vis*., vol. 3, pp. 314–320, 2019.

[8] Z. C. Zhang, Y. Wang, J. J. Peng, and J. T. Duan, "An improved self-attention for long-sequence time-series data forecasting with missing values", *Neural computing and applications*, vol. 36, pp. 3921-3940, 2023.

[9] D. Wang, Y. Zheng, and G. Li, "High-Dimensional Low-Rank Tensor Autoregressive Time Series Modeling", *Journal of Econometrics*, vol. 238, no. 1, 2023.

[10] R. Chen, H. Xiao, and D. Yang, "Autoregressive models for matrix-valued time series", *Journal of Econometrics*, vol. 222, no. 1, part B, pp. 539-560, 2021.

[11] M. Maya, W. Yu, and X. Li, "Time series forecasting with missing data using neural network and meta-transfer learning", *In Proceedings of the 2021 IEEE Symposium Series on Computational Intelligence*, 2021, pp. 1–6.

[12] S. F. Wu, C. Y. Chang, and S. J. Lee, "Time series forecasting with missing values", *International Conference on Industrial Networks and Intelligent Systems*, 2015. DOI: 10.4108/icst.iniscom.2015.258269

[13] J. Deng, J. Deng, D. Yin, R. Jiang, and X. Song, "TTS-Norm: Forecasting Tensor Time Series via Multi-Way Normalization", *ACM Transactions on Knowledge Discovery from Data*, vol. 18, no. 1, pp 1–25, 2023, https://doi.org/10.1145/3605894.

[14] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Morup, "Scalable Tensor Factorizations with Missing Data", *Proceedings of the 2010 SIAM International Conference on Data Mining*, 2010, pp. 701-712, https://doi.org/10.1137/1.9781611972801.

[15] Q. Song, H. Ge, J. Caverlee, and X. Hu, "Tensor Completion Algorithms in Big Data Analytics", *ACM Transitions on Knowledge Discovery from Data*, vol. 13, no. 1, pp. 1- 48, 2019.

[16] K. J. Blinowska and M. Malinowski, "Non-linear and linear forecasting of the EEG time series", *Biological Cybernetics*, vol. 66, pp. 159–165, 1991.

[17] C. Teixeira, B. Direito, M. Bandarabadi, and A. Dourado, "Output regularization of SVM seizure predictors: Kalman Filter versus the "Firing Power" method", *Conference IEEE Engineering in Medicine and Biology Society*, 2012.

[18] V. P. Oikonomou, A. T. Tzallas, and D. I. Fotiadis, "A Kalman filter based methodology for EEG spike enhancement", *Computer Methods Programs Biomed*, vol. 85, no. 2, pp. 101-108, 2007.

[19] L. Li, J. McCann, N. Pollard, and C. Faloutsos, "Dynammo: Mining and summarization of coevolving sequences with missing values", *In KDD*, New York, NY, USA, ACM 2009.

[20] Y. Cai, H. Tong, W. Fan, and P. Ji, "Fast mining of a network of coevolving time series", *In SDM*, 2015. pp. 298-306, https://doi.org/10.1137/1.9781611974010.34

[21] S. Tora and K. Eguchi, "MPI/OpenMP hybrid parallel inference for Latent Dirichlet Allocation", *Proceedings of the Third Workshop on Large Scale Data Mining Theory and Applications*, 2011.

[22] R. Talmon, M. Ronen, Z. Stephane, C. Hitten, and R. Ronald, "Manifold Learning for Latent Variable Inference in Dynamical Systems", *IEEE Transactions on Signal Processing*, vol. 63, no. 15, pp. 3834-3856, 2015.

[23] J. Sun, D. Tao, and C. Faloutsos, "Beyond Streams and Graphs: Dynamic Tensor Analysis", *In KDD*, 2006. https://doi.org/10.1145/1150402.1150445

[24] F. Mansouri *et al*. "A fast EEG forecasting algorithm for phase-locked transcranial electrical stimulation of the human brain", *In: Frontiers in Neuroscience*, vol. 11, 2017.

[25] P. Peng, K. Zhang, and H. Wei, "Robust epileptic seizure prediction with missing values using an improved denoising adversarial autoencoder", *Proceedings of the 2022 4th International Conference on Image, Video and Signal Processing*, 2022, pp. 157-164, https://doi.org/10.1145/3531232.3531255, 2022.

[26] H. Daoud and M. A. Bayoumi, "Efficient epileptic seizure prediction based on deep learning", *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no, 5, pp. 804-813, 2019.

[27] B. H. Brinkmannn *et al*, "Seizure Diaries and Forecasting with Wearables: Epilepsy Monitoring Outside the Clinic", *Frontiers in neurology*, vol. 21, 2021. https://doi.org/10.3389/fneur.2021.690404

[28] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state", *Phys. Rev. E*, vol. 64, p. 061907, 2001. https://doi.org/10.1103/PhysRevE.64.061907