Learning Tensor-based Representations from Brain-Computer Interface Data for Cybersecurity

Md Lutfor Rahman*, Sharmistha Bardhan*, Ajaya Neupane, Evangelos Papalexakis, Chengyu Song

University of California Riverside
mrahm011@ucr.edu, sbard002@ucr.edu, ajaya@ucr.edu, epapalex@cs.ucr.edu, csong@cs.ucr.edu

Abstract. Understanding, modeling, and explaining neural data is a challenging task. In this paper, we learn tensor-based representations of electroencephalography (EEG) data to classify and analyze the underlying neural patterns related to phishing detection tasks. Specifically, we conduct a phishing detection experiment to collect the data, and apply tensor factorization to it for feature extraction and interpretation. Traditional feature extraction techniques, like power spectral density, autoregressive models, and Fast Fourier transform, can only represent data either in spatial or temporal dimension; however, our tensor modeling leverages both spatial and temporal traits in the input data. We perform a comprehensive analysis of the neural data and show the practicality of multi-way neural data analysis. We demonstrate that using tensor-based representations, we can classify real and phishing websites with accuracy as high as 97%, which outperforms state-of-the-art approaches in the same task by 21%. Furthermore, the extracted latent factors are interpretable, and provide insights with respect to the brain's response to real and phishing websites.

1 Introduction

Phishing is a type of social engineering attacks, where attackers create fake websites with the look and feel similar to the real ones, and lure users to these websites with the intention of stealing their private credentials (e.g., password, credit card information, and social security numbers) for malicious purposes. Because phishing attacks are a big threat to cybersecurity, many studies have been conducted to understand why users are susceptible to phishing attacks [11,35,36,42], and to design automated detection mechanisms, e.g., by utilizing image processing [40], URL processing [7, 38], or blacklisting [37]. Recently, Neupane et al. [25–27] introduced a new detection methodology based on the differences in the neural activity levels when users are visiting real and phishing websites. In

* Both authors contributed equally.

this paper, we advance this line of work by introducing tensor decomposition to represent phishing detection related brain-computer interface data.

With the emergence of the Brain-Computer Interface (BCI), electroencephalography (EEG) devices have become commercially available and have been popularly used in gaming, meditation, and entertainment sectors. Thus, in this study, we used an EEG-based BCI device to collect the neural activities of users when performing a phishing detection task. The EEG data are often analyzed with methodologies like time-series analysis, power spectral analysis, and matrix decomposition, which consider either the temporal or spatial spectrum to represent the data. However, in this study, we take advantage of the multi-dimensional structure of the EEG data and perform tensor analysis, which takes into account spatial, temporal and spectral information, to understand the neural activities related to phishing detection and extract related features.

In this paper, we show that the tensor representation of the EEG data helps better understanding of the activated brain areas during the phishing detection task. We also show that the tensor decomposition of the EEG data reduces the dimension of the feature vector and achieves higher accuracy compared to the state-of-the-art feature extraction methodologies utilized by previous research [26].

Our Contributions: In this paper, we learned tensor representations of brain data related to phishing detection task. Our contributions are three-fold:

- We show that the multi-way nature of tensors is a powerful tool for the analysis and discovery of the underlying hidden patterns in neural data. To the best of our knowledge, this is the first study which employs the tensor representations to understand human performance in security tasks.
- We perform a comprehensive tensor analysis of the neural data and identify the level of activation in the channels or brain areas related to the users' decision making process with respect to the real and the fake websites based on the latent factors extracted.
- We extract features relevant to real and fake websites, perform cross-validation using different machine learning algorithms and show that using tensor-based representations can achieve the accuracy of above 94% consistently across all classifiers. We also reduce the dimension of the feature vector keeping the features related to the highly activated channels, and show that we can achieve better accuracy (97%) with the dimension-reduced feature vector.

The tensor representations of the data collected in our study provided several interesting insights and results. We observed that the users have higher component values for the channels located in the right frontal and parietal areas, which meant the areas were highly activated during the phishing detection task. These areas have been found to be involved in decision-making, working memory, and

memory recall. Higher activation in these areas shows that the users were trying hard to infer the legitimacy of the websites, and may be recalling the properties of the website from their memory. The results of our study are consistent with the findings of the previous phishing detection studies [26,27]. Unlike these studies, our study demonstrates a tool to obtain the active brain areas or channels involved in the phishing detection task without performing multiple statistical comparisons. On top of that, our methodology effectively derives more predictive features from these channels to build highly accurate machine-learning based automated phishing detection mechanism.

2 Data Collection Experiments

In this section, we describe details on data collection and preprocessing.

2.1 Data Collection

The motivation of our study is to learn tensor based representations from the BCI measured data for a phishing detection task, where users had to identify the phishing websites presented to them. We designed and developed a phishing detection experiment that measured the neural activities when users were viewing the real and fake websites. We designed our phishing detection study inline with the prior studies [11, 25–27]. Our phishing websites were created by obfuscating the URL either by inserting an extra similar looking string in the URL, or by replacing certain characters of the legitimate URL. The visual appearances of the fake websites were kept intact and similar to the real websites. We designed our fake webpages based on the samples of phishing websites and URLs available at PhishTank [33] and OpenPhish [29]. We choose twenty websites from the list of top 100 popular websites ranked by Alexa [3] and created fake versions of the 17 websites applying the URL obfuscation methodology. We also used the real versions of these 17 websites in the study. We collected data in multiple sessions and followed the EEG experiments like prior studies [22, 41]. In each session of the experiment, the participants were presented with 34 webpages in total.

We recruited fifteen healthy computer science students after getting the Institutional Review Board (IRB) approval and gave them $10 Amazon gift-card for participating in our study. We had ten (66.66%) male participants, and five (33.33%) female participants with the age-range of 20-32 years. The participants were instructed to look at the webpage on the screen and give response by pressing a 'Yes'/'No' button using a computer mouse. We used commercially available, light-weight EEG headset [1] to simulate a near real-world browsing experience. EmotivPro software package was used to collect raw EEG data. We presented with the same set of (randomized) trials to all the participants. All participants performed the same tasks for four different sessions. There was a break of approximate 5 minutes between two consecutive sessions. We collected all sessions data in the same day and same room. We have total 2040 (Participants(15) X Number of sessions (4) X Number of events per session(34)) responses. We discarded 187 wrong responses and only considered 1853 responses for our analysis.

2.2 Data Preprocessing

The EEG signals can be contaminated by eye blink, eyeball movement, breath, heart beats, and muscles movement. They can overwhelm the neural signals and may eventually degrade the performance of the classifiers. So, we preprocess the data to reduce the noise before modeling the data for tensor decomposition. Electrooculogram (EOG) produced by eye movements and Electromyography (EMG) produced by muscles movement are the common noise sources contaminating the EEG data. We used the AAR (Automatic Artifact Removal) toolbox [14] to remove both EOG and EMG [18]. After removing the EOG and EMG artifacts, EEG data were band pass filtered with the eighth-order Butterworth filter with the pass-band 3 to 60 Hz to remove other high frequency noises. The band pass filter keeps signals within the specified frequency range and rejects the rest. The electrical activities in the brain are generated by billions of neuron and the raw EEG signals we collected using sensors of Emotiv Epoc+ device had received signals from a mixture of sources. So we applied the Independent Component Analysis (ICA) [17], a powerful technique to separate independent sources linearly mixed in several sensors, to segregate the electrical signals related to each sensor. Our EEG data pre-processing methodology is similar to the process reported in [24].

3 Problem Formulation & Proposed Data Analysis

Tensor decomposition method is useful to capture the underlying structure of the analyzed data. In this experiment, the tensor decomposition method is applied to the EEG brain data measured for a phishing detection task.

One of the most popular tensor decomposition is the so-called PARAFAC decomposition [15]. In PARAFAC, by following an Alternating Least Square (ALS) method we decompose the tensor into 3 factor matrices. The PARAFAC decomposition decomposes the tensor into a sum of component rank-one tensors. Therefore, for a 3-mode tensor where $X \in R^{IXJXK}$, the decomposition will be,

$$X = \sum_{r=1}^{R} a_r \circ b_r \circ c_r \tag{1}$$

Here, R is a positive integer and $a_r \in R^I$, $b_r \in R^J$ and $c_r \in R^K$ are the factor vectors which we combine over all the modes and get the factor matrices. Figure Figure 1 is showing the graphical representation of PARAFAC decomposition. However, PARAFAC model assumes that, for a set of variables the observations are naturally aligned. Since, in our phishing experiments, this is not guaranteed, we switched to PARAFAC2 model which is a variation of PARAFAC model.

The dimension of the feature matrix varies in dimension 68 X N, where 68 is for the number of event and N indicates the number of components or features. We have selected different number of features for our experiment to test what number of features trains a better model.

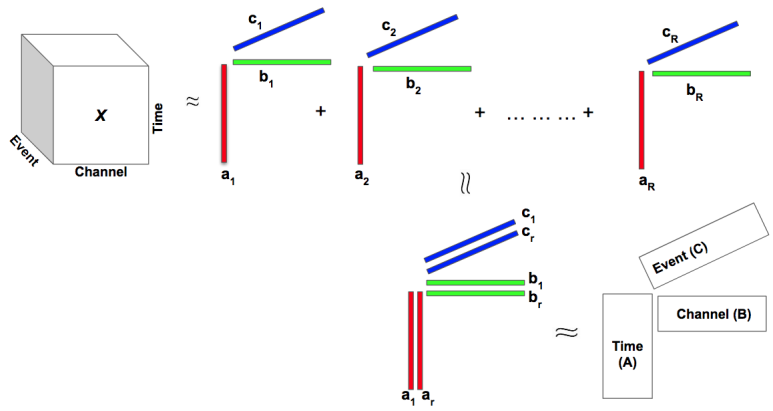

Fig. 1: PARAFAC decomposition with 3 factor matrices (Time, Channel and Event). Event matrix (blue colored) is used as features.

3.1 PARAFAC2 Decomposition

In real life applications, a common problem is the dataset is not completely aligned in all modes. This situation occurs for different problems for example, clinical records for different patients where patients had different health problems and depending on that the duration of treatments varied over time [32]. Moreover, participants response record for phishing detection where each of them took a variable amount of time to select and decide whether the website presented is a real one or phishing one. In these examples, the number of samples per participant does not align naturally. The traditional models (e.g., PARAFAC and Tucker) assume that, the data is completely aligned. Moreover, if further preprocessing is applied in the data to make it completely aligned it might be unable to represent actual representation of the data [39] [16]. Therefore, in order to model unaligned data, the traditional tensor models need changes. The PARAFAC2 model is designed to handle such data.

The PARAFAC2 model is the flexible version of the PARAFAC model. It also follows the uniqueness property of PARAFAC. However, the only difference is that the way it computes the factor matrices. It allows the other factor matrix to vary while applying the same factor in one mode. Suppose, the dataset contains data for K subjects. For each of these subjects (1, 2, ..., K) there are J variables across which I_k observations are recorded. The I_k observations are not necessarily of equal length. The PARAFAC2 decomposition can be expressed as,

$$X_k \approx U_k S_k V^T \tag{2}$$

This is an equivalence relation of Equation 1. It only represents the frontal slices X_k of the input tensor X. Where, for subject k and rank R, U_k is the factor matrix in the first mode with dimension I_k X R, S_k is a diagonal matrix with dimension R X R and V is the factor matrix with dimension J X R. The S_k is the

frontal slices of S where S is of dimension R X R X K and also $S_k = diag(W(k,:))$. Figure 2 shows the PARAFAC2 decomposition.

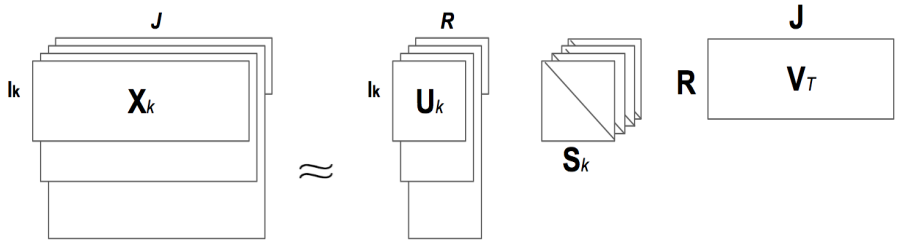

Fig. 2: PARAFAC2 decomposition of a mode - 3 tensor.

PARAFAC2 can naturally handle sparse data or dense data [19]. However, this statement was true only for a small number of subject [6]. The SPARTan algorithm is used for PARAFAC2 decomposition when the dataset is large and sparse [32].

3.2 Formulating Our Problem using PARAFAC2

In order to apply different tensor decomposition method, at first we need to form the tensor. We form the initial tensor by considering all participants phishing detection brain data. The tensor for this experiment is of three dimensions, time X channel X events.

In this experiment, the participants were given the option to take the necessary time to decide whether the current website is phishing or not. Since, the participants were not restricted to take a decision within a particular time-frame, it has been found that for each event different participants took variable amount of time. Therefore, it is not possible to apply general tensor decomposition algorithm and even form a general tensor.

In order to solve the above problem, the PARAFAC2 model is used in this experiment. The SPARTan [32] algorithm is used to compute the PARAFAC2 decomposition. This algorithm has used the Matricized-Tensor-Times-Khatri-Rao-Product (MTTKRP) kernel. The major benefit of SPARTan is that it can handle large and sparse dataset properly. Moreover, it is more scalable and faster than existing PARAFAC2 decomposition algorithms.

3.3 Phishing Detection & Tensor

In this project, each participant was shown the real and phishing website and during that time, the brain EEG signal was captured. The participants were given the flexibility to take the required amount of time to select whether the

website is real or not. Therefore, the observations for a set of variables do not align properly and the PARAFAC2 model is used to meaningfully align the data.

In order to create the PARAFAC2 model, the EEG brain data for all user for both real/phishing website was merged. The 3-mode tensor was then formed as Time X Channel X Events. In events, both the real and the phishing website are considered. Therefore, the tensor formed from this dataset consists of 1853 events, 14 channels (variables) and a maximum of 3753 observations (time in seconds). Figure 3 shows the PARAFAC2 model of the phishing experiment.

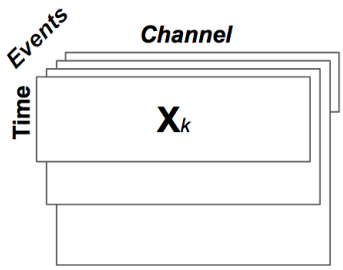

Fig. 3: PARAFAC2 model representing the brain EEG data across different events.

The 3 factor matrices obtained from the decomposition are U, V and W. These factor matrices representing the mode Time, Channel and Events respectively. In this experiment, we analyzed the V and W factor matrices to see which channels capture the high activity of brain regions and also distinguish between real and phishing events respectively.

In the SPARTAN algorithm [32], a modified version of the Matricized-Tensor-Times-Khatri-Rao-Product (MTTKRP) kernel has been used. It computes a tensor that is required in the PARAFAC2 decomposition algorithm. For a PARAFAC2 model, if our factor matrices are H, V and W and of dimension RXR, JXR, and KXR respectively, then for mode 1 with respect to K MTTKRP is computed as,

$$M^{(1)} = Y_{(1)}(W \odot V) \tag{3}$$

The computation here is then parallelized by computing the matrix multiplication as the sum of outer products for each block of $(W \odot V)$. The efficient way to compute the specialized MTTKRP is, first computing $Y_k V$ for each row of the intermediate result and then computing the Hadamard product with W(k,:). Since Y_k is column sparse, it reduces the computation of redundant operations. For this project, we have computed the factor matrices in Channel mode and Events mode using the above method.

Brain Data vs Tensor Rank In exploratory data mining problems, it is really important to determine the quality of the results. In order to ensure a good quality of the decomposition, it is important to select a right number of components as the rank of the decomposition. In this experiment, we used the AutoTen [30] algorithm to assess the performance of the decomposition with different ranks.

The application of AutoTen algorithm is not straightforward for the phishing experiment, since the observations for a set of variables do not align properly. Therefore, a number of additional operations are performed to bring the tensor of the whole dataset into a naturally aligned form. From equation(2), if we decompose U_k as $Q_k H$, then we can rewrite equation 2 as,

$$X_k \approx Q_k H S_k V^T \tag{4}$$

Where Q_k is with dimension $I_k X R$ and H is with dimension R X R. Q_k has orthonormal columns. Now, if both sides of the above equation is multiplied by Q_k^T, then we get,

$$Q_k^T X_k \approx Q_k^T Q_k H S_k V^T \approx H S_k V^T \tag{5}$$

Therefore, we can write,

$$Y_k \approx H S_k V^T \tag{6}$$

Where Y_k is the outer product of Q_k^T and X_k. The above equation is now same as the PARAFAC decomposition with consistency in all the modes. Y_k is also a tensor and is used in the AutoTen algorithm as input. The AutoTen algorithm was run for maximum rank 20 and it has been found that 3 is the rank for which the model can perform better. Therefore, for the PARAFAC2 decomposition using SPARTan, rank 3 is used.

4 Classification Performance

In this section, we discuss our classification performance for detecting the real and phishing page based on neural data. We merge all the data across all the sessions and across all the users. We extracted features from brain data using tensor decomposing with rank 3 computed by our modification of AutoTen as discussed in section 3.3. We then applied the different type of machine learning algorithms for distinguishing the real and fake website based on brain data and checked their performance. We tested with Bayesian type BayesNet(BN), Function type Logistic Regression and MultilayerPerceptron, Rules type JRip and DecisionTable, Lazy type KStar and IB1 and Tree type J48, RandomTree, Logistic Model Tree (LMT), and RandomForest (RF). We present the best one (BayesNet, Logistic Regression, JRip, IB1, RandomForest) from each type of machine learning algorithms. We use 15-fold cross validation because we have 15 users data in our dataset. Here, the dataset is divided into 15 subsets where 14 subsets will be in training set and rest one subset will be in the testing subset.

Table 1: Classification Performance: In this table, we present the classification results of the five classifiers. Here, we have classification results for two scenarios. One for considering all channels for features extraction and another for considering only top 6 channels based on their activation. We have highlighted the accuracy of the best performing classifier in grey.

Metric	Accuracy		Recall		Precision		F-measure	
Algorithm	All	Top 6	All	Top 6	All	Top 6	All	Top 6
BayesNet	84.83	92.49	84.83	92.49	85.86	92.92	84.74	92.48
Logistic Regression	94.98	95.08	94.98	95.08	95.00	95.16	94.98	95.08
JRip	91.90	97.07	91.90	97.03	91.92	97.05	91.90	97.03
IB1	94.44	97.57	94.44	97.57	94.44	97.57	94.44	97.57
RandomForest	93.41	97.62	93.43	97.63	93.41	97.63	93.41	97.62

We tested our model using several metrics: accuracy, precision, recall, F1 score and Area Under the Curve (AUC). We compared our classification performance in two different cases.

- **All Channels**: In this setting, we consider all 14 channel's data as feature vectors.
- **Top 6 Channels**: In this setting, we consider only top 6 highly activated channel's data as feature vectors. Details discussion for this can be found in section 5.

The summary of classification performance for different metrics(Accuracy, Recall, Precision, and F-measure) can be found in Table 1. We have seen that for considering all channels logistic regression algorithm gives 94% accuracy. We get 97% accuracy for considering top 6 highly activated channels using Random Forest algorithm. We achieved improved performance than the prior study which reported 76% accuracy of their phishing detection model built using neural signals when the participants were asked to identify real and fake websites under fNIRS scanning [26].

We also validated our classification performance by plotting the ROC curve in Figure 4 using the Random Forest algorithm which gives the best accuracy among all the algorithms. In an ideal scenario, the AUC should be 100%. The baseline for AUC is 50%, which can be achieved through purely random guessing. Our model achieved 97.32% AUC for when considering all channels data and 99.22% when considering only top 6 highly activated channels data. We have seen that our True Positive Rate is 79.04 in case of all channels data and True Positive Rate is 94.91 in case of top 6 channels data while keeping False Positive Rate less than 1%. Reducing the channels gives us better phishing detection accuracy.

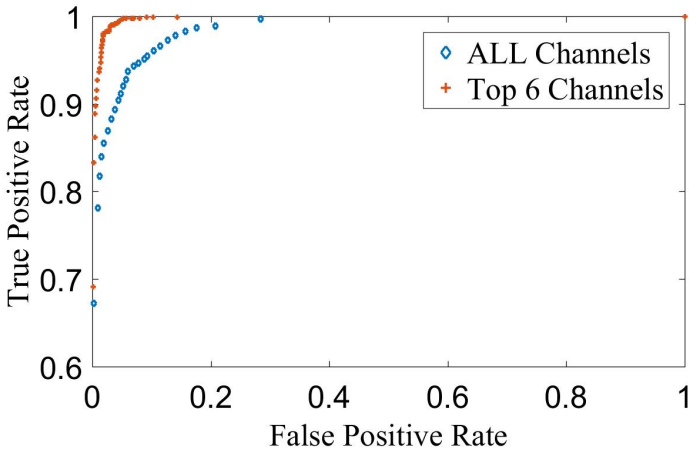

Fig. 4: AUC curve for All channels vs Top 6 channels using the Random Forest algorithm. Here, we observed that TPR for all channels is 79.04% and 94.91% for top 6 channels when FPR is $< 1\%$

5 Discussion

In this section, we answer why we are getting good accuracy in classifying real and fake websites using brain data. We highlight the several key points for getting the good accuracy. First, we show that certain brain areas are highly activated during the phishing detection task. Second, we show that there is a statistically significant difference between the real and fake components.

5.1 Phishing Detection vs Brain Areas

In this section, we provide a concise neuro-scientific insight of the brain data measured for the phishing detection. We discuss the relationship between the brain activities and phishing detection task. In our experiments, we collected brain data from human scalp using a commercially available non-invasive brain computer interface device. The data we collected using Emotiv Epoc+ device come from fourteen (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4) different sensors as shown in Figure 5. These sensors are placed on different regions according to the International 10-20 system. Two sensors positioned above the participant's ears (CMS/DRL) are used as references. Sensors location and functionality of each region is given below:

- *Frontal Lobe*, located at the front of the brain and associated with reasoning, attention, short memory, planning, and expressive language. The sensors that are placed in those area are AF3, F7, F3, FC5, FC6, F4, F8, and AF4.
- *Parietal Lobe*, located in the middle section of the brain and associated with perception, making sense of the world, and arithmetic. The censors P7 and P8 belongs to this area.

– *Occipital Lobe*, located in the back portion of the brain and associated with vision. The sensors from this location are O1 and O2.
– *Temporal Lobe*, located on the bottom section of the brain and associated with sensory input processing, language comprehension, and visual memory retention. The sensors of this location are T7 and T8.

Based on the factor analysis in channel dimension, we observed that mostly Frontal lobe and Parietal lobe sensors (AF3, F3, FC5, F7, P7, and P8) are highly activated for the phishing detection task. In Figure 5 a), we present the channel activity based on channel factor data. Here, we consider all phishing detection events and get the factor matrix data in channel dimension using rank 3. We consider the first component data for drawing this graph. We have found that same subset of channels while considering the second and the third component data. In Figure 5 b) we show the corresponding brain mapping for phishing detection task. Higher the red is the higher brain activity for phishing detection task. Our findings are aligned with the prior fMRI [27] and fNIRS [26] studies.

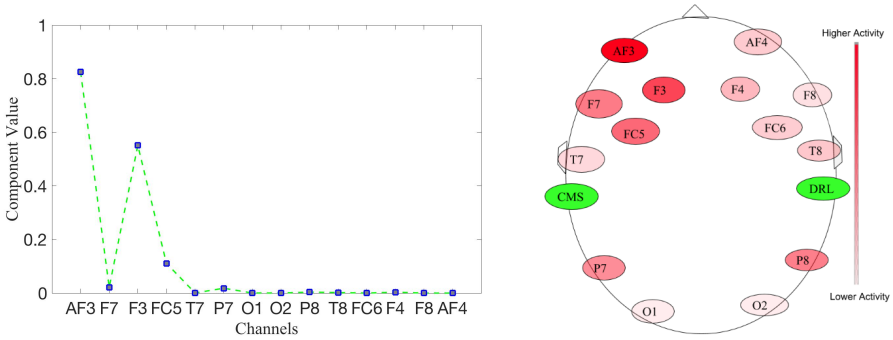

Fig. 5: a) shows the channel activity after the application of SPARTan decomposition on the tensor. The channel data for the first component is plotted in this figure to determine which channels have high activity. b) shows the corresponding brain region activation.

5.2 Statistical Analysis: Real vs Fake Events

In this subsection, we present the statistical analysis of the components obtained from the tensor analysis. First, we performed the Kolmogorov-Smirnov (KS) test to determine the statistical distribution of the first component values of the real and fake factor matrix. In KS test we observed that the distribution of the real and fake samples was non-normal ($p < .0005$). We then applied Wilcoxon Singed-Rank Test, a non-parametric test comparing two sets of scores that come from the same participants, to measure the difference between real and fake components. We observed that there was statistically significantly high differences between the real and fake components ($Z = 6.8$, $p < .0005$).

5.3 Feature Space Reduction

One of the primary application of our study is the reduction of the dimension of the feature vector by keeping the features related to highly activated frontal and parietal channels. We observed that the prediction accuracy of the machine learning model trained on the features belonging to the top 6 highly activated channels was better than the prediction accuracy of the models better trained on features related to all channels. Our model achieved 97% of accuracy while applying reduced features vector. From the ROC curve in Figure 4, we can see that our true positive rate increases from 79% to 94% when we use reduced feature vector in classification while keeping false positive rate $< 1\%$.

6 Related Works

Phishing attacks usually come in different forms or structures. In the case of the phishing website, the front-end structure of the website or URL is changed which is sometimes difficult to distinguish from the real website. There are a number of tools that are considering different features to detect a phishing website automatically. However, different studies show that these tools should consider the behavioral aspect of the user as well [12]. In different experiments, participants were tested to identify the features of a website. For example, evaluating the website URL, identifying icons or logos and past web experiences. It has been found that participants who know about phishing are less likely to fall for a phishing website.

In order to make the user aware of phishing website, proper education on this topic is required. There are several works that discuss how to identify phishing website from URLs [23]. These works show that, by looking at the lexical and host-based (IP-address, domain name, etc.) features of the URL, it can be easily found out whether the website is phishing or not. In this work, the accuracy obtained in classifying the phishing and the real webpage is 95-99%. Furthermore, it has been found that if appropriate education is provided, the user will be more efficient in avoiding phishing website [4]. Moreover, it has also been studied that what type of browser phishing warnings works better for the user and the performance of active warnings outperform the passive ones [13].

Apart from understanding user behavior while browsing the internet, it is also possible to prevent phishing by focusing on tracking the hacker's behavior. The hybrid feature selection method is applied to capture the phishing attacker's behavior from email header [2] and they achieved an accuracy of 94%. In these methods, both the content of email header and behavioral basis of it is considered for feature selection.

Automated Phishing Detection Method: In order to automatically detect phishing website, the pattern of the URL is considered as the primary method, and with the aid of machine learning algorithms it can protect the user from a phishing attack. However, these models do not perform well due to the lack in the number of features. Moreover, the domain top-page similarity based method

is also used for phishing detection [34] where they obtained maximum AUC of 93%.

There are few more automated phishing detection system that use density based spatial clustering techniques to distinguish phishing and real website [21] with the accuracy of 91.44%. Linear classifiers are also used for phishing detection problem, and phishing domain ontology is also used for this task [43]. The content of a webpage is analyzed and based on their linguistic feature, an accuracy of 97% is achieved.

Tensor Decomposition and Phishing Detection: Tensor is useful for EEG brain data representation and visualization as well. It provides a compact representation of the brain network data. Moreover, it is useful to use tensor decomposition method to capture the underlying structure of the brain data. In Cichocki et al. [9], a brain computer interface system is used where tensor decomposition is applied in EEG signals. Tensor decomposition has already been applied for feature extraction in different problems involving EEG data. In P300 based BCIs, tensor decomposition is used to extract hidden features because of its multi-linear structures [28]. Unlike the general Event-related Potentials(ERP) based BCI problems, tensor can consider both temporal and spatial structure for feature extraction instead of only temporal structure which ensures better accuracy [10] [8]. Tensor decomposition method has also been used for the classification of Mild and Severe Alzheimer's Disease using brain EEG data [20].

Tensor decomposition has been used for brain data analysis as well. GEBM is an algorithm that models the brain activity effectively [31]. SEMIBAT is a semi-supervised Brain network analysis approach based on constrained Tensor factorization [5]. The optimization objective is solved using the Alternating Direction Method of Multipliers (ADMM) framework. The proposed SEMIBAT method showed 31.60% improved results over plain vanilla tensor factorization for graph classification problem in EEG brain network.

Tensor decomposition methods have been applied for a variety of problems related to the analysis of brain signal. However, the idea of applying tensor decomposition methods in an automated system where the main task is to classify phishing and real websites based on brain EEG data is novel. In our case, we achieved the classification accuracy of real and phishing websites as high as 97% using neural signatures.

7 Conclusion

In this paper, we show that the tensor representation of brain data helps better understanding of the brain activation during the phishing detection task. In this scheme, owing to tensor representation on multi-modes of channel, time, and event, different characteristics of EEG signals can be presented simultaneously. We observed that right frontal and parietal areas are highly activated for participants during the phishing website detection task. These areas are involved in decision making, reasoning, and attention. We use the AutoTen algorithm to

measure the quality of the result and also to choose a proper rank for the decomposition. We reduce the dimension of feature vectors and achieve a maximum 97% of classification accuracy while considering only highly activated brain area sensor's data. Our results show that the proposed methodology can be used in the cybersecurity domain for detecting phishing attacks using human brain data.

References

1. Emotiv eeg headset. `https://www.emotiv.com` (2017), accessed: 5-17-2017
2. A. Hamid, I.R., Abawajy, J.: Hybrid feature selection for phishing email detection. In: Xiang, Y., Cuzzocrea, A., Hobbs, M., Zhou, W. (eds.) Algorithms and Architectures for Parallel Processing. pp. 266–275. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)
3. Amazon.com, Inc: Alexa skill kit. `https://developer.amazon.com/alexa-skills-kit` (2027)
4. Arachchilage, N.A.G., Love, S.: Security awareness of computer users: A phishing threat avoidance perspective. Computers in Human Behavior **38**, 304 – 312 (2014)
5. Cao, B., Lu, C.T., Wei, X., Philip, S.Y., Leow, A.D.: Semi-supervised tensor factorization for brain network analysis. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 17–32. Springer (2016)
6. Chew, P.A., Bader, B.W., Kolda, T.G., Abdelali, A.: Cross-language information retrieval using parafac2. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 143–152. KDD '07, ACM (2007)
7. Chu, W., Zhu, B.B., Xue, F., Guan, X., Cai, Z.: Protect sensitive sites from phishing attacks using features extractable from inaccessible phishing urls. In: Communications (ICC), 2013 IEEE International Conference on. pp. 1990–1994. IEEE (2013)
8. Cichocki, A., Washizawa, Y., Rutkowski, T., Bakardjian, H., Phan, A.H., Choi, S., Lee, H., Zhao, Q., Zhang, L., Li, Y.: Noninvasive bcis: Multiway signal-processing array decompositions. Computer **41**(10), 34–42 (Oct 2008)
9. Cichocki, A., Washizawa, Y., Rutkowski, T., Bakardjian, H., Phan, A.H., Choi, S., Lee, H., Zhao, Q., Zhang, L., Li, Y.: Noninvasive bcis: Multiway signal-processing array decompositions. Computer **41**(10) (2008)
10. Cong, F., Lin, Q.H., Kuang, L.D., Gong, X.F., Astikainen, P., Ristaniemi, T.: Tensor decomposition of eeg signals: a brief review. Journal of neuroscience methods **248**, 59–69 (2015)
11. Dhamija, R., Tygar, J.D., Hearst, M.: Why phishing works. In: Proceedings of the SIGCHI conference on Human Factors in computing systems. pp. 581–590. ACM (2006)
12. Downs, J.S., Holbrook, M., Cranor, L.F.: Behavioral response to phishing risk. In: Proceedings of the Anti-phishing Working Groups 2Nd Annual eCrime Researchers Summit. pp. 37–44. eCrime '07, ACM (2007)
13. Egelman, S., Cranor, L.F., Hong, J.: You've been warned: An empirical study of the effectiveness of web browser phishing warnings. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1065–1074. CHI '08, ACM (2008)
14. Gómez-Herrero, G., De Clercq, W., Anwar, H., Kara, O., Egiazarian, K., Van Huffel, S., Van Paesschen, W.: Automatic removal of ocular artifacts in the eeg without

an eog reference channel. In: NORSIG, Signal Processing Symposium. pp. 130–133. IEEE (2006)
15. Harshman, R.A.: Foundations of the parafac procedure: Models and conditions for an" explanatory" multimodal factor analysis (1970)
16. Ho, J.C., Ghosh, J., Sun, J.: Marble: High-throughput phenotyping from electronic health records via sparse nonnegative tensor factorization. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 115–124. KDD '14, ACM (2014)
17. Hyvärinen, A., Oja, E.: Independent component analysis: algorithms and applications. Neural networks **13**(4), 411–430 (2000)
18. Joyce, C.A., Gorodnitsky, I.F., Kutas, M.: Automatic removal of eye movement and blink artifacts from eeg data using blind component separation. Psychophysiology **41**(2), 313–325 (2004)
19. Kiers, H.A.L., ten Berge, J.M.F., Bro, R.: Parafac2 - part i. a direct fitting algorithm for the parafac2 model. Journal of Chemometrics **13**, 275–294
20. Latchoumane, C.F.V., Vialatte, F.B., Jeong, J., Cichocki, A.: Eeg classification of mild and severe alzheimer's disease using parallel factor analysis method. Advances in Electrical Engineering and Computational Science pp. 705–715 (2009)
21. Liu, G., Qiu, B., Wenyin, L.: Automatic detection of phishing target from phishing webpage. In: 2010 20th International Conference on Pattern Recognition. pp. 4153–4156 (Aug 2010)
22. Luck, S.J.: Ten simple rules for designing erp experiments. Event-related potentials: A methods handbook **262083337** (2005)
23. Ma, J., Saul, L.K., Savage, S., Voelker, G.M.: Beyond blacklists: Learning to detect malicious web sites from suspicious urls. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1245–1254. KDD '09, ACM (2009)
24. Neupane, A., Rahman, M.L., Saxena, N.: Peep: Passively eavesdropping private input via brainwave signals. In: International Conference on Financial Cryptography and Data Security (2017)
25. Neupane, A., Rahman, M.L., Saxena, N., Hirshfield, L.: A multi-modal neurophysiological study of phishing detection and malware warnings. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. pp. 479–491. ACM (2015)
26. Neupane, A., Saxena, N., Hirshfield, L.: Neural underpinnings of website legitimacy and familiarity detection: An fnirs study. In: Proceedings of the 26th International Conference on World Wide Web. pp. 1571–1580. International World Wide Web Conferences Steering Committee (2017)
27. Neupane, A., Saxena, N., Kuruvilla, K., Georgescu, M., Kana, R.: Neural signatures of user-centered security: An fMRI study of phishing, and malware warnings. In: Proceedings of the Network and Distributed System Security Symposium (NDSS). pp. 1–16 (2014)
28. Onishi, A., Phan, A.H., Matsuoka, K., Cichocki, A.: Tensor classification for p300-based brain computer interface. In: Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. pp. 581–584. IEEE (2012)
29. OpenPhish: Phishing url. `https://openphish.com/feed.txt` (2017), accessed: 05-10-2017
30. Papalexakis, E.E.: Automatic Unsupervised Tensor Mining with Quality Assessment. ArXiv e-prints (Mar 2015)

31. Papalexakis, E.E., Fyshe, A., Sidiropoulos, N.D., Talukdar, P.P., Mitchell, T.M., Faloutsos, C.: Good-enough brain model: Challenges, algorithms and discoveries in multi-subject experiments. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 95–104. KDD '14, ACM, New York, NY, USA (2014)
32. Perros, I., Papalexakis, E.E., Wang, F., Vuduc, R., Searles, E., Thompson, M., Sun, J.: Spartan: Scalable parafac2 for large & sparse data. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 375–384. KDD '17, ACM (2017)
33. Phishtank: Join the fight against phishing. `https://www.phishtank.com/` (2017), accessed: 05-10-2017
34. Sanglerdsinlapachai, N., Rungsawang, A.: Using domain top-page similarity feature in machine learning-based web phishing detection. In: 2010 Third International Conference on Knowledge Discovery and Data Mining. pp. 187–190 (Jan 2010)
35. Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L.F., Downs, J.: Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 373–382. ACM (2010)
36. Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L.F., Hong, J., Nunge, E.: Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In: Proceedings of the 3rd symposium on Usable privacy and security. pp. 88–99. ACM (2007)
37. Sheng, S., Wardman, B., Warner, G., Cranor, L.F., Hong, J., Zhang, C.: An empirical analysis of phishing blacklists (2009)
38. Thomas, K., Grier, C., Ma, J., Paxson, V., Song, D.: Design and evaluation of a real-time url spam filtering service. In: Security and Privacy (SP), 2011 IEEE Symposium on. pp. 447–462. IEEE (2011)
39. Wang, Y., Chen, R., Ghosh, J., Denny, J.C., Kho, A., Chen, Y., Malin, B.A., Sun, J.: Rubik: Knowledge guided tensor factorization and completion for health data analytics. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1265–1274. KDD '15, ACM (2015)
40. Whittaker, C., Ryner, B., Nazif, M.: Large-scale automatic classification of phishing pages. In: Proceedings of the Network and Distributed System Security Symposium (NDSS) (2015)
41. Woodman, G.F.: A brief introduction to the use of event-related potentials in studies of perception and attention. Attention, Perception, & Psychophysics **72**(8), 2031–2046 (2010)
42. Wu, M., Miller, R.C., Garfinkel, S.L.: Do security toolbars actually prevent phishing attacks? In: Proceedings of the SIGCHI conference on Human Factors in computing systems. pp. 601–610. ACM (2006)
43. Zhang, J., Li, Q., Wang, Q., Geng, T., Ouyang, X., Xin, Y.: Parsing and detecting phishing pages based on semantic understanding of text **9**, 1521–1534 (06 2012)