Improving Seizure Forewarning using Graph Dissimilarity and Layout Analysis

Lewis R. Williams
Fayetteville State University

Mentors
Dr. Lee M. Hively
Oak Ridge National Laboratory

Dr. J. Todd McDonald
University of South Alabama

Improving Seizure Forewarning using Graph Dissimilarity and Layout Analysis

Abstract

The goal of this research is to improve prediction of anomalous states in physiological and electronic-based physical systems. Approximately 3 million Americans and 50 million people worldwide have epileptic seizures. Epileptic seizures are an example of a physiological process that is difficult to predict. With the detection and forewarning of epileptic seizures, one can reduce or avoid the consequence of various medical crises, decrease patient hospitalization, and lower costs of health care and loss of productive time. Researchers currently analyze observations of electroencephalography (EEG) data to help predict when seizures will occur.

The major contribution of the research will involve analyzing the differences in phase space graphs using graph-theoretic dissimilarity measures. Such measures include calculations such as degree, node difference, link difference, algebraic connectivity, and topological invariants. By applying and measuring dissimilarity among phase space graphs, we expect to improve the accuracy of predictive statistics which forewarn of events such as seizures.

Based on results from this study, the research team will apply these analytical techniques to other areas of study such as hardware failure prediction and cyber anomaly detection.

Section 1: Introduction

Epilepsy is a neurological disorder in which a person has repeated seizures. The onset of epilepsy ranges from infants to when one is late in their years. The general cause of this physiological abnormality is when a person's brain tissue is damaged or destroyed from illness or injury. Research shows that 3 million Americans and 50 million people worldwide suffer from epileptic seizures [1].

The aim of this research is to improve the accuracy of epilepsy forewarning predictive statistics by studying various graph visual layout algorithms and different graph properties. By utilizing this approach, we can improve detection and forewarning of epileptic seizures and thereby reduce or avoid medical crises, decrease patients' hospitalization time, and increase or maintain productive time.

The remainder of this paper is organized as follows. In section 2 the background of epilepsy forewarning will be reviewed briefly. In section 3, we explain the analysis approach for constructing a phase space graph. We discuss graph layout algorithms in section 4 followed by graph properties in section 5. Finally, we present our observations and results in Section 6 and reiterate our research goal, impact, and future work in Section 7.

Section 2: Background of Epilepsy Forewarning

For decades scientists have been trying to accurately detect and forewarn of epileptic seizures. Seizures typically are not dangerous within themselves, but their randomness can be life threatening because of concomitant injuries. For example, people with epilepsy could be operating a vehicle, surrounded by sharp objects, or engaging in an activity that is otherwise not

dangerous until they have a seizure. It has been recorded by epileptologists that many patients are aware when their seizure is likely to occur [5].

Early work in the 1970s by Viglione and Walsh [6] not only created the first electronic classifier for epilepsy forewarning but also created a device that was tested on patients with epilepsy. The research was still in its infancy so Viglione and Walsh were only able to forewarn the seizure by 6 seconds.

The advancement of science, technology, engineering, and math (STEM) has contributed immensely to the growth of epileptic seizure prediction. The electroencephalography (EEG) is a device that records electrical activity (brain waves and fluctuations) from the scalp. Viglione and Walsh used EEG data as the basis of their classifier and, like many medical devices since the 1970s; EEG devices have improved in accuracy and precision. On the mathematical side of this problem, advances in nonlinear dynamics, chaos theory, and time series data analysis have proven that epileptic seizures are events that progress over time from an initial stage of onset. Continued advances and improvements in the STEM fields will strengthen the accuracy and reliability of seizure prediction methods.

Typical quantifying measures for seizure prediction include traditional nonlinear measures (TNM) and phase-space dissimilarity measures (PSDM). TNM have historically yielded inconsistent detection and event forewarning for this problem [7]. PSDM have shown better consistency and sensitivity for epileptic event detection and forewarning [3].

Section 3: Phase Space Graph Construction

Our research into improving prediction of epileptic events uses a theorem-based, data driven technique using raw EEG data collected from human observations. Our analysis approach

uses the differences between phase-space graphs during an observation period to determine whether a seizure should be predicted or not. A phase-space is based on an underlying dynamical system which is itself a representation of a real life state or event. For example, if one wanted to predict if there will be rain or snow then the dynamical system will be the weather. For this research the dynamical system is the brain. The next step would be to gather data from the brain.

We can quantify brain activity with EEG data that measures voltage fluctuations at electrodes attached to the scalp, as illustrated in Figure 1. After data are recorded for time periods ranging from two to ten hours, time series analysis is used to see if patterns are present that can be used for seizure prediction. Eye blinks and other muscular artifacts complicate such analysis because they introduce unnecessary noise in the data. We can reduce the impact of artifacts in the data by using differences between adjacent channels (a bipolar channel) and by applying filters. In our phase-space analysis experiments, we use a single channel of bipolar data and a quadratic fit filter to reduce this noise. We partition a stream of time-serial data taken from one EEG observation into analysis windows (cutsets) and the subsequent phase-space analysis proceeds at the cutset level.

Given artifact filtered data from a particular cutset, we then symbolize each data point by converting them into a discrete set of values: 0 or 1 for two symbols; 0, 1, or 2 for three symbols, etc. The symbolized data allows construction of unique, discrete, dynamical states.

Once the data is symbolized, we apply Taken's time delay embedding theorem (TDE) [3]. This theorem allows conversion of time-serial data into a geometric object, which is referred to as a phase-space graph. The reconstruction of the data by this theorem faithfully preserves the dynamical system's topology and directionality. The symbolized data are converted to unique

states that represent vertices in a graph. The state-to-state transitions form links that represent edges in a graph. Graph theorems guarantee that topological measures of the graph do not depend on the labeling of the nodes or links.[1]

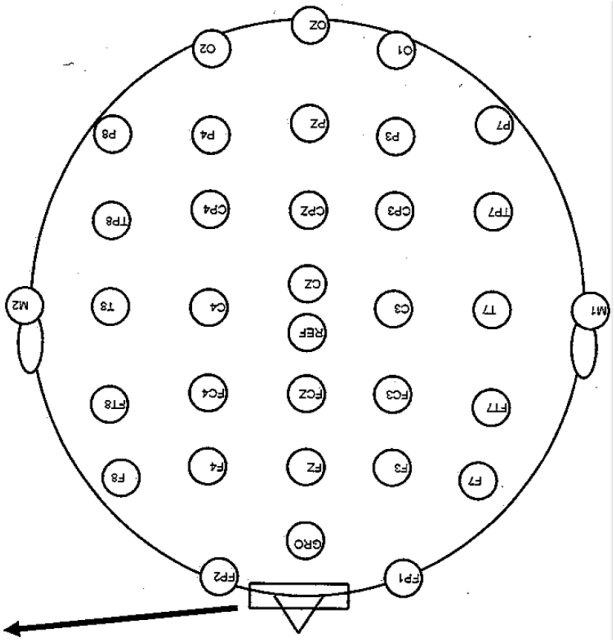

Figure 1: Illustration of an electroencephalogram on a patient.

Section 4: Graph Layout Algorithm

One main objective of this research was to analyze phase-space graphs based on their visual properties in order to improve our existing forewarning algorithms. In particular, we wanted to know whether some graph visual layout algorithms might highlight the differences between graphs better than others and whether the visualization itself would provide insight into the dynamics of seizure events themselves. To aid in visualization of the graphs, we used to graph diagramming applications (yEd and Cytoscape) and experimented with various graph layout algorithms.

[1] http://www.yworks.com/en/index.html
http://www.cytoscape.org/

Graph layout algorithms organize a network visually according to specific rules. Layout algorithms are found in various graph diagramming applications and some of these applications provide unique proprietary layout options. Both yEd and Cytoscape have attributes that were acceptable and unacceptable during this study.

Graphs loaded into yEd were visually easier to look at, compared to Cytoscape. The graphs in yEd were bigger and because of that the analysis of an edge or node of a graph was less complex. Although, the graphs were easier to see in yEd, not all of its graphing layouts could support the data intensity of each graph compared to Cytoscape. Given the complexity, size, and properties of the graphs that were studied, certain layout algorithms were too time and resource expensive to be of benefit. It is also fair to say that the Hierarchical layout style could not support the data in both diagramming software. Because of the high computational intensity of the majority of the phase-space graphs, yEd had a higher running time than Cytoscape.

One important feature that both diagramming applications had was geometric transformation. Geometric transformations made it possible to view a graph in the same layout by rotating it, translating it, and using other geometric transformations. Cytoscape could not reflect a graph about the X and Y-axis like yEd. Each diagramming software had features that were useful in this study but because of their vastness many are left unmentioned.

The main graphing layouts used for this research was the organic layout and the circular layout. The organic layout shows the clusters of a graph. A cluster is a group that contains objects that are more similar to each other than objects in other clusters. Nodes were the object for this layout. The Organic layout also position nodes so that their edges are about the same size. The edges could also be manipulated if one chooses to.

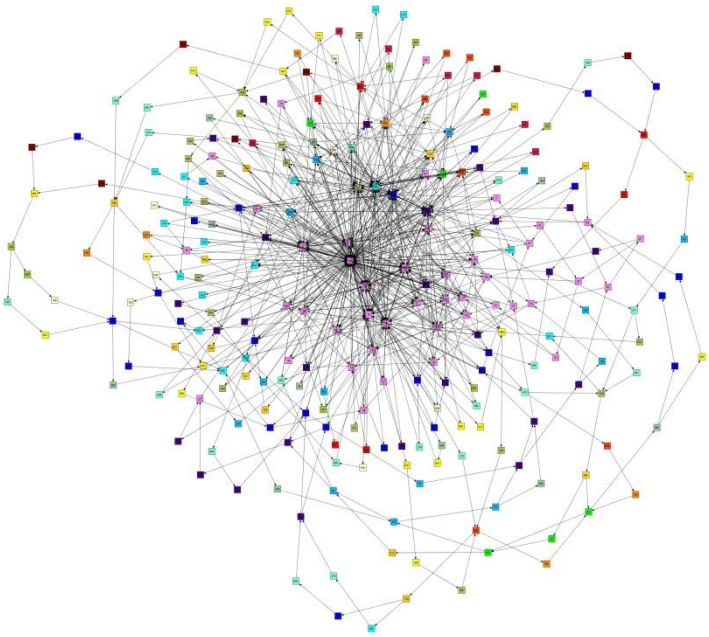

Figure 2: Organic layout algorithm

The circular layout creates nodes partition (separate circles) by analyzing the graphs connectivity structure. The Circular layout can arrange a graph into one single circle or into a representation where nodes that are reachable by two disjoint paths have their own circle. The main purpose of using this graphing layout was to analyze the graphs edges and see if some major dissimilarity were found. With this layout different features could be used to further manipulate the graphs overall look.

The Circular layout algorithm and the Organic layout algorithm could be used in Cytoscape and yEd. Other layouts such as Hierarchical, Tree, Grid, etc. were potential assets for this research but because of the graphs' complexity and computational intensity they were unable to upload in the softwares and many were visually complex to look at, so they had to be created using a programming language. The event data set were typically the graphs that could not upload.

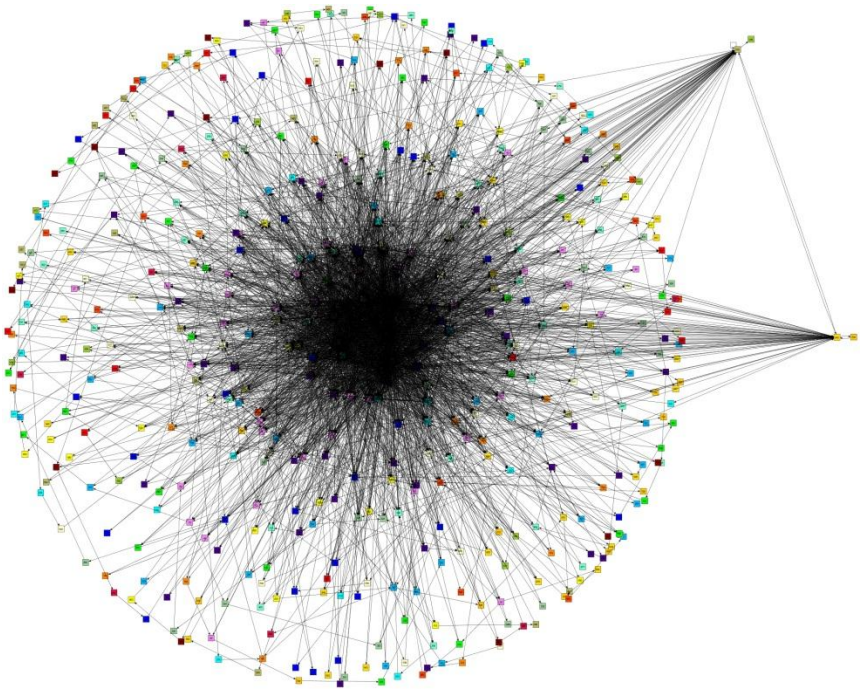

Figure 3: Circular layout algorithm

Section 5: Graph Dissimilarity Visually and Computationally

In this study of graph layout algorithms, the EEG data that were gathered from 40 different observation periods were utilized. Of these 60 observation data sets, 40 observations have a seizure event and 20 observations have no seizure event.

After the node difference and link difference were computed for each cutset, a forewarning algorithm that makes predictions on whether a seizure will occur or not occur was created. The algorithm used the node and link difference dissimilarity measures. There are four categories that the prediction for each of the 60 observation data sets can fall into: true negative (TN), true positive (TP), false negative (FN), and false positive (FP). The best results produced a true rate of 57/60.

This study focused attention on the observations that produced FN and FP results. The interest was to know whether there were graph properties or visual features of the graphs that might help us understand why we got false results and how we might improve the forewarning algorithm. In this study Observations 45, 50, and 57 were considered. Observations 45 and 50 were false positives, meaning an event did not occur, but the forewarning was indicated. Observation 57 was a false negative, in which an event did occur, but the forewarning was not indicated. The hope is that by analyzing these observations, a dissimilarity between one phase-space graph and the next could be identified and thus the bettering of the forewarning process.

The first observation being considered is observation 45. The forewarning which took place at cutset 137 was analyzed as well as several cutsets before and after the forewarning. When compared to the base cases the test cases show some apparent differences. Since the considered graphs are unlike the base cases, identifying similarities between the base cases and test cases is important to identify a phase shift. In the Circular and Organic layout the graphs are heavily dense in the middle. That is because different partitions (which are nodes that are reachable by two edge disjoint paths) are packed in the middle.

In observation 50, a false positive, the majority of its base cases and test cases were all identical. Observation 50 is a false positive because change was enough between the test case and base case to forewarn of an epileptic seizure but none had occurred. Therefore by definition similarities were still being considered for the phase shift. No strong similarities were found.

Observation 57 was a false negative meaning a seizure occurred but there were subtle changes that an event was not forewarned. In each observation an event was forewarned by a sequence of at least eleven 2s or 1s from the threshold vale. In order to identify a phase-shift in a false negative the first sequence of 2s or 1s were considered; in observation 57 it was ones.

Because these graphs look the same and they were false negatives, dissimilarities between the base cases and test cases were considered. Cutsets 25-28 were analyzed and no dissimilarity could be visually seen.

The two graph dissimilarities used in this research were node difference (U_N) and link difference (U_L). Each base case was compared to one another but with no redundancy. For instance base case 1 could not be compared to itself. For each comparison a node difference and link difference were obtained. A standard deviation and average were computed for every U_N and U_L. Then, each test case cutset were compared to each base case individually and produced a U_N and U_L from each comparison. Once an average was found for every node difference and link difference they were normalized. By optimizing through the entire data set a threshold was obtained. This threshold will give the number of successive occurrences which gives the forewarning.

Other graph properties such as biconnectedness, which is the graph property that states the graph will remain connected if any vertex were removed, were considered. Planar which is a graph property that is defined by when the edges of a graph intersect only at a graph's end points, was also considered. Another graph invariant that was briefly considered was strong connectivity. A strongly connected graph is described as having a path from and to every vertex to all other vertices. For instance, there will be a directed edge going into a specific node and another leaving that node. Not every phase-space graph is strongly connected, planar, nor biconnected.

The three graph invariants were analyzed for observations 45, 50, and 57. The observations show high biconnectedness, strong connectivity, and very little planarity.

Observation 45 shows high fluctuation between biconnectedness and strong connectivity between each cutsets; so it should still be looked into for future studies.

Section 6: Results

One method used to find a phase-shift, was to visually see each cutset using graphics. The complexity of the majority of the graphics made them visually hard to analyze and to identify a property that would have been beneficial to indicate a phase-shift. Also, because of the complexity and computational intensity of some graphics, especially with an event data set, some graphing layout algorithms could not produce an image. Therefore, identifying certain dissimilarities or similarities between cutsets was visually impossible.

At the start of this research, with a total of 60 observations, 57 were accurately predicted for epileptic seizures with a prediction distance of .10. The only dissimilarity measures that were implemented at the start of this research and during the summer were node difference and link difference. With the improvement of the forewarning analysis and dissimilarity measures, the true rate improved from 57/60 to 58/60, with a new prediction distance of 0.05. As the true rate improves the prediction distance moves closer and closer to zero.

Section 7: Discussion and Conclusion

The goal of improving the forewarning analysis has been met. The improvement of the forewarning analysis will not only have an impact on its future research but also in many different areas of research. This forewarning analysis approach can be used for any physiological change such as sepsis, ventricular fibrillation, etc. and many industrial machine changes such as machine failure amongst many [3]. This forewarning analysis greatly impacts a person’s

hospitalization time and medical cost. The forewarning analysis could also be implemented in cyber security for cyber event detection. Anomalies detected within a computer system can be stopped, thus saving companies millions of dollars in damages and exposure.

The future work for this research would be to analyze and find dissimilarity measures for the graph properties that were found using Cytoscape. The graph properties that Cytoscape computes may help to identify a phase-shift that the current dissimilarity measures did not pick up. This will hopefully improve the forewarning analysis.

Acknowledgement

I would like to thank my mentors Dr. Lee Hively and Dr. J. Todd McDonald for their time and patience. I would also like to thank Mrs. McCoy, my fellow RAMS scholars for their support and encouragement. Also, without the support from the Computational Science and Engineering division, UT-Battelle, U.S. Department of Energy, and many others this internship would not have been a success.

References

1. "Epilepsy Foundation of America" http://www.epilepsyfoundation.org/aboutepilepsy/. June 14, 2012.
2. Michael Smoot, Keiichiro Ono, Johannes Ruscheinski, Peng-Liang Wang, Trey Ideker Cytoscape 2.8: new features for data integration and network visualization Bioinformatics. 2011 February 1; 27(3): 431–432.Published online 2010 December 12.
3. Phase-space dissimilarity measures of nonlinear dynamics: Industrial and biomedical applications" V. Protopopescu and L M.Hively.
4. Baumgartner C, Serles W, Leutmezer F, et al. Preictal SPECT in temporal lobe epilepsy: regional cerebral blood flow is increased prior to electroencephalography-seizure onset. J Nucl Med. 1998;39(6):978-982.
5. "Seizure Prediction: Its Evolution and Therapeutic Potential. David Krieger and Brian Litt.http://www.us.elsevierhealth.com/media/us/samplechapters/9781416061717/Chapter%2001.pdf. June 16, 2012.
6. "The First International Collaborative Workshop on Seizure Prediction: summary and data description." Klaus Lehnertza and Brian Litt. Copyright 5 January 2005.
7. L.M. Hively, V.A. Protopopescu, and N.B. Munro *Epilepsy Forewarning via Phase-Space Dissimilarity*, J. Clin. Physiol., Vol. 22 (December 2005) pgs. 402-409.