Determining Granger Causality in multi-channel EEG recordings with Support Vector Regression

Shoshana Gottesman

DREU – Summer 2009

1. Introduction

Epilepsy is a neurological disorder in which patients suffer from seizures, or brief periods of excessive electrical activity in the brain. The majority of patients can have their seizures held in check by anti-seizure medication; in more extreme cases, surgery is performed to remove the area where the seizures originate. For a minority of patients, however, none of these measures are effective, and they suffer from frequent seizures which can be debilitating and cause permanent brain damage. [1] For patients such as these, and indeed all epileptics, an early warning system to predict the onset of a seizure and perhaps provide appropriate intervention before the seizure begins would greatly improve their quality of life.

Any successful seizure prediction algorithm must be able to detect patterns of interaction between different areas of the brain at and before the onset of a seizure. Seizures generally start in a particular part of the brain, different in each patient, known as the epileptogenic area. From there, the abnormal activity spreads through the brain as neurons fire in tandem, setting off neurons in other areas and sending a wave of electricity rippling across the brain. Determining whether neurons firing in one area of the brain affect the activity of another region can be crucial in charting the development pattern of a seizure. One method for measuring causal relationships is Granger Causality. This index measures the degree of correlation between two time-series. While Granger Causality is not absolute causality, it is a promising method for determining the influence of one area of the brain on another.

In our work, we used Support Vector Regression and Granger Causality to measure the likelihood of electrical activity in one part of the brain (represented by a particular channel in an EEG recording) affecting brainwave patterns in another. Our results are incomplete and therefore inconclusive, but seem promising.

2. Related Work

A significant amount of work has been done on using Granger Causality to detect interaction between different regions of the brain. Baccala et al [2] used Granger Causality along with directed coherence to obtain a clearer picture of cortical interactions in mice. Hess, et al [3] used a variant of Granger Causality (time-variant GC) to successfully detect interdependencies between brain regions. Winterhalder et al. [4] compared Granger Causality to partial coherence, partial directed coherence, and directed transfer function. Roebroeck et al. [5] used Granger Causality mapping to study directed influence between neurons. Granger Causality has also been successfully used in economics [6,7,8,9].

In light of the successful use of Granger Causality to detect interactions between brain regions, we decided to choose this method for use in our causation analysis of a multichannel

EEG. While this has been done before, we wanted to determine whether using support vector regression in our work would yield positive results.

3. Methods

3.1. Support Vector Regression

Support Vector Regression (henceforth SVR) is a method of using Support Vector Machines, a common classification algorithm, to solve linear regression problems. In SVR, the goal is to find the line that best separates the data points in the series into a specified number of classes, while maximizing the margin between the classes [10, 11, 18].

Often, such a line is not easily drawn in a linear plane, so a kernel function is used to convert the data into a higher-dimension plane where linear regression can be easily performed. [12] The cost parameter c determines the tightness of the margin and the tradeoff between accuracy and generalizability. [13]

3.2 Granger Causality

Granger Causality (GC) is a method for measuring causality between time series developed by Clive W. J. Granger.

To calculate the Granger Causality for two data series x and y; that is, whether y Granger-causes x, the prediction accuracy of x is calculated. It is then compared to the prediction accuracy of the same time series x with values from time series y added to it. If the new series, x + y, produces more accurate predictions than x alone, y can be said to Granger-cause x. Prediction accuracy is measured by the variance of the error of time series x, as compared to that of the new series x + y, using the equation

$$F_{x \to y} = \ln \left(\frac{var(x)}{var\ (y \to x)} \right)$$

where x and y are time series, var(x) is the variance of the error of x alone, and var(y-> x) is the variance of the error of the series x + y. [3]

It should be noted that Granger Causality is not absolute causation. There are myriad factors that can influence the improvement in prediction error; the addition of y may have no impact. However, strong Granger Causality values do generally imply a high likelihood of correlation, if not causation, between the two time series [14, 3, 15, 16].

3.3 Dataset

The dataset used was that of the University of Freiburg, Germany. [17] (See appendix A for details on prior unsuccessful attempts with other data sets.) Data from a 15-year old female patient with simple partial and complex partial seizures originating in the frontal lobe was used in our work. While ictal (before and during a seizure) and inter-ictal (between seizures, normal) recordings were available, the volume of the data was so large that we used only one set of ictal recordings for our very preliminary testing. The patient's data consisted of several sets of files containing recordings from seizures, as well as at least 50 minutes of pre-ictal (immediately preceding the onset of a seizure) recording. Each set had 6 files, corresponding to the 6 channels of the EEG recording. Each file contained 921,600 data values, recorded at a sampling rate of 256 Hz.

3.4 Computation and Calculation of Granger Causality

The program used to process the data was written in Java and makes extensive use of the open source Machine Learning code of Weka, a Java ML program, version 3.6 [18]. The program read the data values from text files and converted them to the .arff files required by Weka, with 100 data values in each sample. Then it used the various classification classes provided by Weka to perform support vector regression on the time-series in the file. (See Appendix B for details.) The regression generated a numeric prediction for the 100^{th} value in each sample, then calculated the error for the sample; that is, the difference between the actual and predicted values.

We then calculated the variance of the error over the entire data set, using the formula

$$V = \frac{\sum e^2}{n}$$

where V is the variance, e is the error for each instance, and n is the total number of instances in the data set. Note that we did not normalize the data; hence the simplified equation.

This process was repeated for every file in the dataset. Then, for every pair of files x and y, corresponding values of y were interspersed between the values in x, and every 100^{th} value in x predicted using the expanded time series x +y. Then the program repeated the process, this time interspersing x-values between the values of y to produce the series y + x.

To illustrate, consider the two time series x = {0,2,4,6,8,10,12}and y = { 1,3,5,7,9,11,13}.

First, support-vector regression is used on the first 6 values of x (0, 2, 4, 6, 8, 10) to try to predict the 7^{th} value, (12), and the difference between the actual (12) and predicted values is recorded as the error. Then, the regression is repeated to attempt to predict the 7^{th} value of y (13) and the resulting error recorded.

The next step is constructing the merged time series x + y. This yields {0,1,2,3,4,5,6,7,8,9,10,11,12}, with the 7th value of y discarded. SV- regression is now performed on the newly joined time series to predict the last value, 12, and the prediction error recorded. The same thing is done to form the series y +x {1,0,3,2,5,4,7,6,9,8,11,10,13} and attempt to predict the last value, 13 (in this case, it is the 12, the last value of x, that is dropped).

If the prediction error of x+y is lower than that of x alone, the Granger Causality measure for x + y is positive. The larger the positive difference between the two, the higher the GC-value is and the more likely it is that y Granger-caused x. The same holds true for y +x.

The above procedure is a rough explanation of the methods we used to derive Granger Causality values for each pair of files. Since each file contains 921,600 values, the time series is divided into 100-values slices and the regression repeated for each one; the error for each is calculated and used to derive the variance of the prediction error over the entire time series.

We then used the equation for Granger Causality given in section 3.2 to calculate the Granger causality for x + y and y + x. This process was repeated for each pair of files in the set of 6 channels, 36 pairs in all, with the resulting Granger Causality values forming a 6 x 6 matrix. Obviously, determining whether a single channel impacts itself is nonsensical, so cells in the matrix with the same row and column number, that is, representing the impact of a said channel on itself, were filled in with a 0. (Incidentally, due to the nature of the equation for Granger Causality, actually calculating the value of x + x would yield a 0 regardless.)

4. Results

The program was run once for each of the following C – values: .001, .01, .1, 1.0, 10.0, and 100.00. sampling of the results can be found below.

	1	2	3	4	5	6
1	0	0.026	-0.009	-0.01	-0.016	-0.008
2	-1.057	0	-0.008	-0.007	-0.001	0.009
3	1.801	1.816	0	-0.004	-0.014	-0.004
4	0.241	0.238	0.25	0	0	-0.012
5	0.143	0.153	0.165	0.167	0	0.024
6	-0.057	-0.073	-0.044	-0.028	-0.034	0

Table 1. Granger Causality, C-value 0.1

The 0's on the diagonal are placeholders, as described above; all other 0's are the result of the regression. Values above 0 are highlighted.

	1	2	3	4	5	6
1	0	0.068	0.193	0.013	0.16	0.103
2	-1.098	0	0.086	0.214	0.067	0.088
3	1.665	1.668	0	-0.045	-0.07	-0.088
4	0.168	0.208	0.153	0	-0.053	-0.021
5	0.086	0.098	0.091	0.081	0	-0.027
6	-0.013	-0.03	-0.018	-0.011	-0.016	0

Table 2. Granger Causality, C-value 1.0

The 0's on the diagonal are

placeholders, as described above. Values above 0 are highlighted.

Several things are immediately evident from the results. First, there is very little difference between the values obtained using different C-values. This is more easily seen in the scatter plots below, where the red squares (C = 1.0 and 0.01, respectively) and blue diamonds (C= 0.1 and 0.001) closely overlap in all but three cases. While there is some fluctuation over the y-axis in some places, the values obtained using different C-values are generally quite close to each other.

Figure 1. Granger Causality values for each pair of files; C-values 0.1 and 1.0.

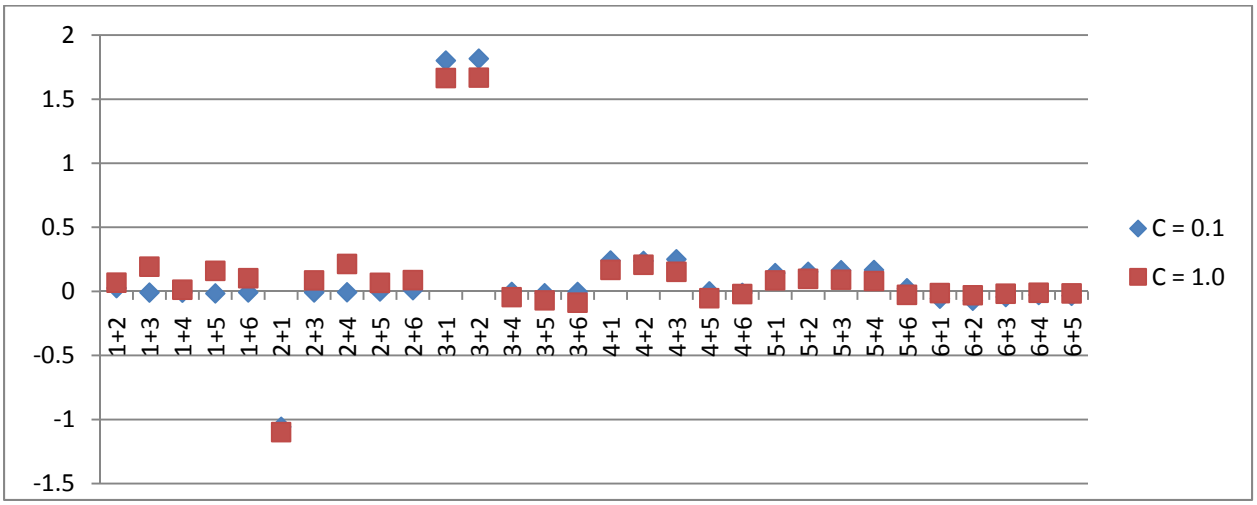
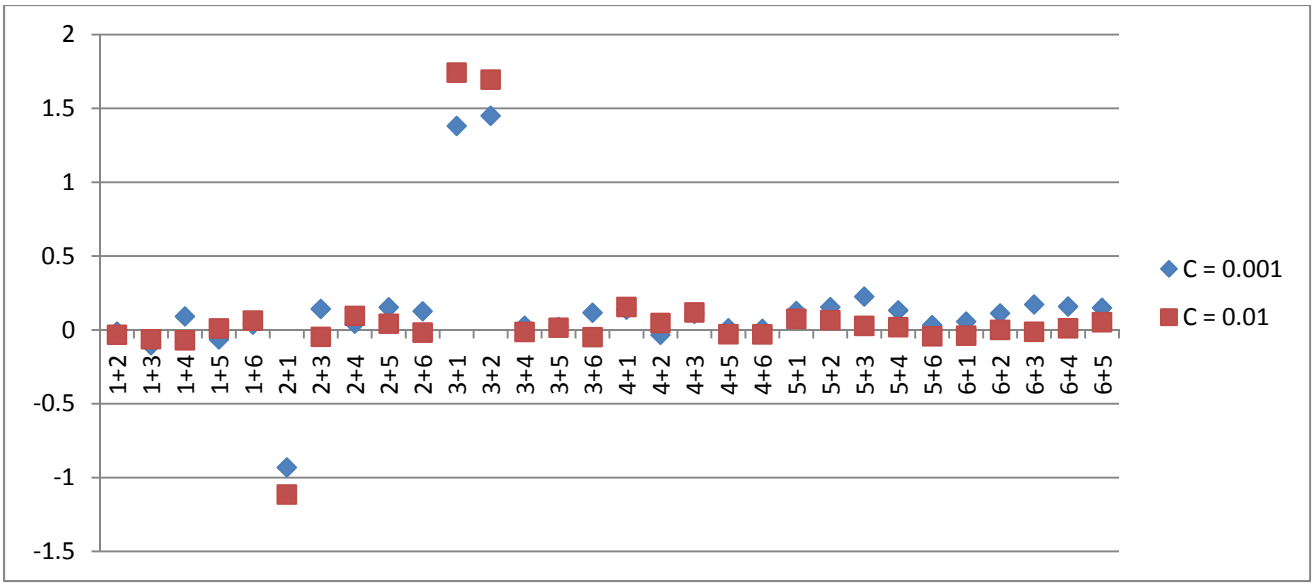


Figure 2. Granger Causality values for each pair of file; C-values .001 and .01.



Most of the values are relatively small, suggesting little or no correlation between the two channels tested. Some pairs, however, have values higher than 1, suggesting that a causative relationship exists between the two. In particular, the pairs 2 + 1 and 3 + 1 have values higher than 1.5, while the inverse pairs (1 + 2 and 1 + 3) have low or even negative values. This seems to indicate that channel 1 influenced both 2 and 3, but the impact did not occur in the reverse direction.

5. Conclusions

The summer ended long before our program finished testing even one patient, so our results are severely limited. However, as discussed above, several of the channel pairs had high Granger-Causality values, suggesting that electrical activity in one of the channels influenced that of the other. Of course, such methods are limited by the fact that they only allow for bivariate analysis. Though we circumvented this limitation somewhat by testing all possible pairs, it is likely that the actual patterns of causation in epileptogenesis are significantly more complex than can be represented by only bivariate relationships. While using only Granger Causality lacks the sophistication and finer control of truly multivariate methods, our results show that using support vector regression with Granger Causality may be a promising direction for future research in the genesis of epileptic seizures.

References

1) “About Epilepsy.” The Epilepsy Foundation. http://www.epilepsyfoundation.org/about/

2) L.A. Baccala, K. Sameshima, G Ballester, A. C. Do Valle, and C. Timolaria. Studying the interaction between brain structures via directed coherence and granger causality. Applied Signal Processing, 1998.

3) Wolfram Hesse, Eva Moller, Matthias Arnold, and Barbel Schack. The use of time-variant eeg granger causality for inspecting directed interdependencies of neural assemblies. Journal of Neuroscience Methods, 124(1):27{44, 2003.

4) Matthias Winterhalder, Bjrn Schelter, Wolfram Hesse, Karin Schwab, Lutz Leistritz, Jens Timmer, and Herbert Witte. Detection of directed information flow in biosignals. Biomedizinische Technik (Biomed Tech), 2006.

5) Alard Roebroeck, Elia Formisano, and Rainer Goebel. Mapping directed influence over the brain using granger causality and fmri. NeuroImage, 25(1):230{242, 2005.

6) Clive W.J. Granger, Bwo-Nung Huang, and Chin Wei Yang. A bivariate causality between stock prices and exchange rates: evidence from recent Asian flu. The Quarterly Review of Economics and Finance, 4(3), 2000.

7) Craig Hiemstra and Jonathan D. Jones. Testing for linear and nonlinear granger causality in the stock price-volume relation. The Journal of Finance, 49(5), 1994.

8) C.W.J. Granger. Investigating causal relations by econometric models and cross-spectral methods. Econometrica, 1969.

9) Norman R. Swanson. Money and output viewed through a rolling window.Journal of Monetary Economics, 41:455{474, 1997.

10) Vapnik, V. “The Nature of Statistical Learning Theory”. Springer, New York, 1998.

11) A.J Smola and B. Scholkopf. A Tutorial on Support-Vector Regression. For Statistics and Computing, 2003

12) K.R Muller, A.J. Smola, G. Ratsch, B. Scholkopf, J. Kohlmorgen, and V. Vapnik. Predicting Time Series with Support Vector Machines. Artificial Neural Networks – ICANN ’97.1997.

13) Sherrod, Phillip, “Introduction to Support Vector Machine (SVM) Models”, DTREG.com, <http://www.dtreg.com/svm.htm>

14) "Linear Regression." *Wikipedia, The Free Encyclopedia*. 17 May 2009, 20:43 UTC. 1 Sep 2009 <http://en.wikipedia.org/w/index.php?title=Linear_Regression&oldid=290579919>.

15) Lion, Guy. “Granger Causality” (slideshow). December 2005.

< http://www.slideshare.net/gaetanlion/granger-causality-presentation>

16) Sorenson, Brent E. “Granger Causality”. 1 Mar 2005. <http://www.uh.edu/~bsorense/gra_caus.pdf>

17) University of Freiburg Epilepsy Data Set. <https://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database>

18) Parella, Francesco. “An Introduction to Support Vector Machines.” Online Support Vector Regression. 7 Sep 2009. < http:// onlinesvr.altervista.org >

Appendix A

The University of Freiburg dataset that we used was not our original choice. We began our research using the brainwave dataset of the University of Bonn (http://www.epileptologie-bonn.de/cms/front_content.php?idcat=193). The dataset we worked with contained 100 files, each containing 4097 data values. We were under the misconception that the files contained simultaneous recordings from 100 channels of a multichannel EEG. Since each file contained far fewer values than those of the Freiburg dataset (4097 instead of 921,600) the program completed the testing of each file very quickly, and we were able to test all 10,000 pairs of files in a relatively short period of time. However, after working with this data for some time, we discovered that the files were not, as we had thought, simultaneous recordings from different channels of an EEG, but rather random recordings from different patients at various times. Obviously, this meant that the dataset was completely unsuited for our research, and we had to begin again using the Freiburg dataset described in the paper.

Appendix B

The program we used to perform support-vector regression on the time series and calculate the Granger Causality values for the results was written in Java and made extensive use of the Weka open-source Machine Learning software. The program used standard Java file processing classes to write the data values in each large file into a .arff file (the file type accepted by Weka) in groups of 100 (when performing the regression on a single time series) or 199 (when performing the regression on two merged time series; see section 3.4 above) values each.[1]

[1] Note: the program can actually be customized to work with any size slice of values, but in our case we used 100.

The resulting .arff files were processed as follows: the DataSource and Instances classes[2] were used to read the data from the .arff file and transform each group of values into an instance of the dataset, with the last value as the class label. The instances were then split into training and test sets, with a 70-30 training-test split. The SMOreg and RegSMOImproved classes[3] were used to build and train the classifier using the training set, and then to classify each instance in the test set. The error for each instance was calculated as the difference between the actual and predicted value of the class label, and the variance of the error taken over the entire test set. This was repeated for every pair of files in the original dataset, as described in section 3.4. Granger Causality was calculated for each pair using the equation in section 3.2, and the results output to a text file.

[2] weka.core.converters.ConverterUtils.DataSource and weka.core.Instances, respectively.
[3] weka.classifiers.functions.SMOreg and weka.classifiers.functions.supportVector.RegSMOImproved, respectively.