Exploring SVM-based Epileptic Seizure Prediction

Ruosi (Moira) Feng
Jacobs School of Engineering
University of California, San Diego
La Jolla, CA 92093, US
r2feng@eng.ucsd.edu

Wendy Yunqi Yu
Jacobs School of Engineering
University of California, San Diego
La Jolla, CA 92093, US
yuy031@eng.ucsd.edu

Abstract—Epileptic seizure is one of the most common brain disorders in the world. Reliable seizure prediction could potentially aid patients in their daily lives. Machine learning models have been proven to be useful in epileptic seizure prediction. In this work, different support vector machine (SVM) classifiers are applied to classify interictal and preictal intervals in a patient to predict seizure onset by alerting at seizure horizon. The effect of hyperparameter gamma in the RBF kernel of a nonlinear SVM on model performance is also evaluated.

I. Introduction

More than fifty million people around the world are diagnosed with epilepsy. About three million patients have been affected by epilepsy in the United States alone, making epilepsy the third most common brain disorder. Although the main cause of epilepsy remains unknown, early diagnosis can be useful for treating epilepsy. Early prediction of epileptic seizures can alert patients before it actually occurs so that patients could take action such as intaking seizure suppressing drugs. Research has shown promising seizure prediction using a variety of methods including machine learning and deep learning models [1-6]. Features used in models include spectral power, spike rate, energy, mean phase coherence and so on. For a given model with a given set of features, results may vary depending on data, data handling and model parameter tuning. We examined how hyperparameter affects model performance using the SVM model and a set of spectral power features on the dataset collected from the CHB-MIT Scalp EEG Database that was used in A. H. Shoeb's work [1].

II. Methods

A. Dataset

The dataset used is the CHB-MIT Scalp EEG Database collected at the Children's Hospital in

Boston. The 24 subjects were continuously monitored for around 22 hours. EEG signals were recorded using the International 10-20 system positions at 256 samples per second with 16-bit resolution. The start and end times of seizure onset events and the number of seizures recorded were labeled by professional clinicians and included as a separate text file.

As the scope of our project is limited to patient-specific seizure prediction, we chose to use the recording numbered "chb24" in the dataset, which contains the most number of seizure onset. This choice is made to maximize the number of training data for cross-validation as we plan to use a balanced number of training data for two preictal and interictal classes.

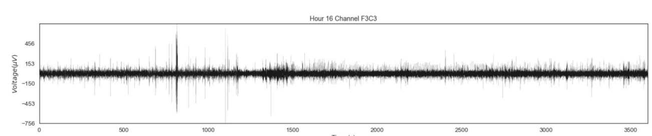


Figure 1. Example EEG signal recording of one channel

B. Signal Pre-processing

EEG data are subject to artifacts such as line noise and movement artifacts. Power line hums at 60 Hz have been removed by a notch filter. Unfortunately, there is no information on the movement artifacts associated with the recordings therefore they might affect model performance to an extent.

C. Data Selection & Labeling

Since cross-validation is used to evaluate how hyperparameters affect the classifier performance, suitable training and test data for preictal class (foreshadowing an onset) and interictal class (non-seizure signals) need to be selected. Preictal and interictal intervals of the EEG signals were identified. Interictal windows are at least 30 minutes before or 60 minutes after an epileptic seizure event, as shown in Figure .

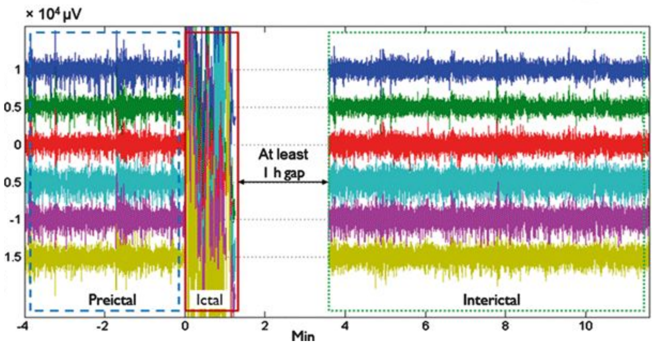


Figure 2. Preictal, ictal & interictal EEG signals diagram as in [5]

This temporal segmentation ensures that non-seizure signals do not include seizure-implying information, to minimize the similarity of these two classes. Preictal windows are defined as the 15 minutes right before an epileptic seizure event. Thus, some preictal data was discarded because it falls within an hour of previous seizure onset. This selection resulted in 11 intervals of usable preictal EEG signals of this patient. For channels, we started from a set of 6 channels that are spatially symmetric.

D. Feature Extraction: Spectral Power

According to the literature review[2], the spectral power of EEG signals in a sliding time window is a commonly used feature method for epileptic seizure detection that gives a decent performance. Thus, we adopted spectral powers of different frequency bands as the features of our dataset. We implemented the moving window analysis [6] by using half-overlapped 20-second-long windows providing a prediction of a seizure every 10 seconds. Relative spectral power in 7 frequency bands of 6 channels was calculated and used as a feature set for each 20s-long window. Spectral bands are split to be delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), gamma1 (30-57 Hz), gamma2 (63-128 Hz) and total (0.5-128 Hz) band. Each window's spectral power for a specific band is normalized by the window's total power.

This feature extraction results in a total of 42 features for a given 20-second window for every 10 seconds of EEG signals in the preictal & interictal intervals we identified. The following are the total number of samples for each class:

- Preictal data size = 858
- Interictal data size = 1444
- Total number of samples = 2302

E. Cross-validation

A "leave-one-out" 11-fold cross-validation is then used and the testing results were averaged to ensure the generalizability of the classifier. Since there are 11 intervals of preictal data, we leave one set out to be used in testing while the other 10 sets for training an SVM classifier. The same number of interictal samples were randomly drawn without replacement from the interictal dataset. In a real-world application, the prior of the non-seizure class is often much higher than the seizure onset events for most patients, and the EEG of normal activities also bears a wide variety. Thus, randomly sampling from the interictal data makes the samples used in training more representative of non-seizure signals.

F. SVM training & testing

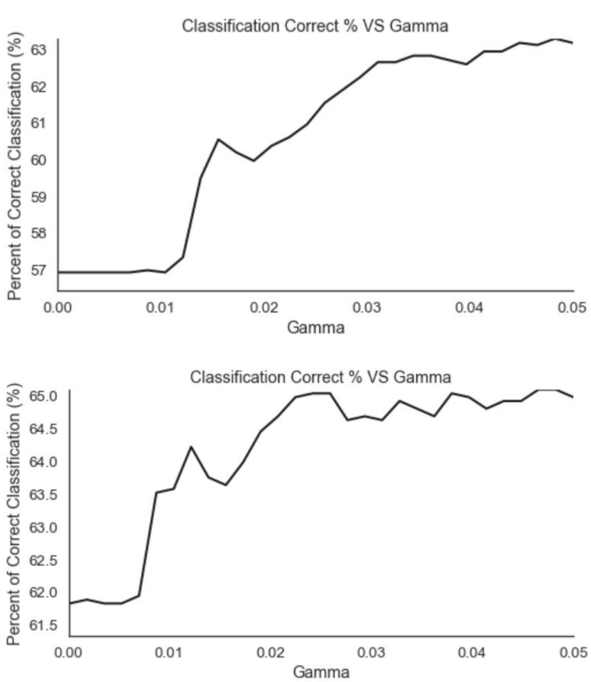
For model training and classification, SVM classifiers are trained with linear and non-linear kernels and hyperparameters to explore how the classification performance is affected. SVM is used for several reasons. First, it is a versatile discriminative model with a kernel function, meaning that the hyperparameters used in the kernel function can be adjusted to accustom to the specific dataset. It is also more computationally efficient than deep learning methods such as long short-term memory networks (LSTM), requiring less training time and data size. For nonlinear SVM, an radial basis function (RBF) kernel is used, which is described as the following equation:

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma\|\mathbf{x} - \mathbf{x}'\|^2)$$

RBF is adopted as it is a popular kernel function used in SVM, which measures the exponential of euclidean distance between two vectors scaled by a parameter γ. γ is swept as a hyperparameter to control how far the influence of one data point goes in the SVM model. After fitting the SVM model, cross-validation testing is performed and the results are evaluated based on the confusion matrix.

III. Results

We experimented with using different channel sets of EEG signals. The following plot shows the percent of correct classification VS γ values 4 different sets of selected channels are used. When Channel set 3 is used, the highest accuracy, 66.43%, is reached when γ = 0.048.



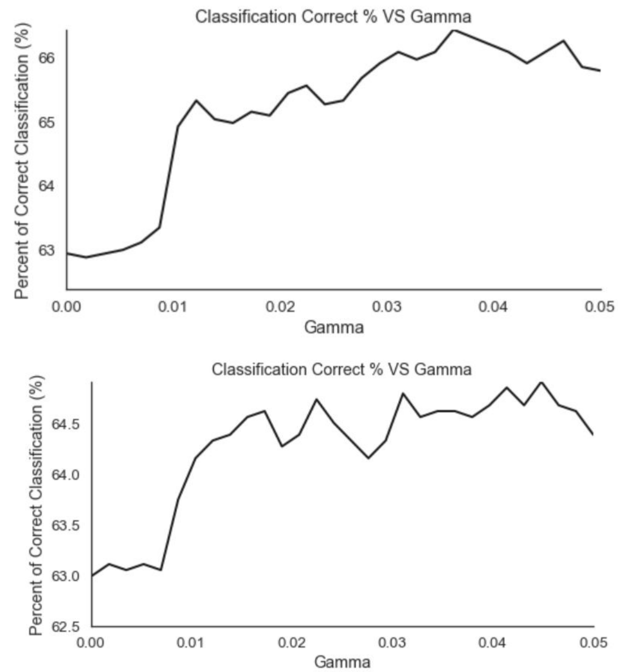


Figure 3: SVM with RBF kernel classification performance VS γ using different channels of data. Channel set 1: {FP1-F3, F3-C3, C3-P3, P3O1} (4). Channel set 2: {FP1-F3, F3-C3, C3-P3, FP2-F4, F4-C4, C4-P4} (6). Channel set 3: {F3-C3, C3P3, F4C4, C4P4, CZPZ, P7T7} (6). Channel set 4: {FP1F3, F3C3, C3P3, PEO1, FP2F4, F4C4, C4P4, P4O2} (8).

Using γ = 0.05 with 4 channels, the setting that gives one of the best classification correctness, the average confusion matrix of 11-fold cross validation is obtained.

		prediction outcome	
		p	**n**
actual value	**p'**	57.69%	42.31%
	n'	26.92%	73.07%

Figure 4: Averaged confusion matrix with best γ value

To compare the performance of RBF and polynomial kernels, models are trained using 6 channels. In addition to RBF kernel, SVM using a polynomial is also trained with different values of degree. The best classification result, 60.66%, occurs when degree = 4. When degree = 1, the classifier is equivalent to a linear SVM. For a RBF kernel, the best result is 60.14%.

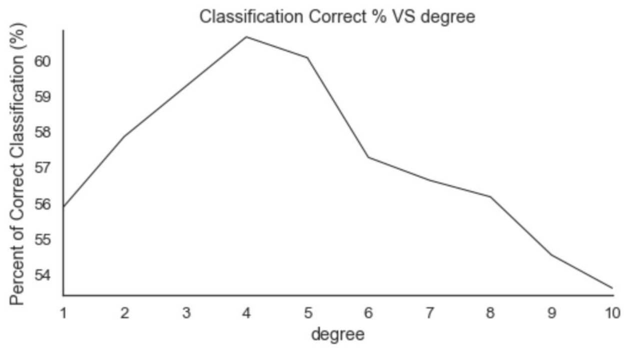


Figure 5: SVM with polynomial kernel classification performance VS degree

IV. Discussions & Conclusions

For a RBF kernel, our results show that performance as in percentage of correct classification increases until at about γ equals to the reciprocal of the number of features, and then starts to decrease as γ keeps increasing. Further increase in the gamma results in little improvement in performance.

By comparing polynomial and RBF kernels, we can see that the peak performance of these two methods are similar, and likely limited by features and data used. Both behave better than a linear kernel, which is expected because the high dimensional features are not linearly separable.

The overall performance is lower than ideal. According to the normalized confusion matrix for mean performance, interictal intervals are more likely to be classified as preictal when the classification outcomes are wrong. The limited number of available data certainly plays a role, as the results are highly polarized in the cross-validation results. We observed the following extreme examples of confusion matrices for two of the classifiers in the 11-fold cross-validation shown in figure 6. The top subplot suggests this classifier always predicts “interictal”, and the bottom one shows another classifier that always predicts “interictal”.

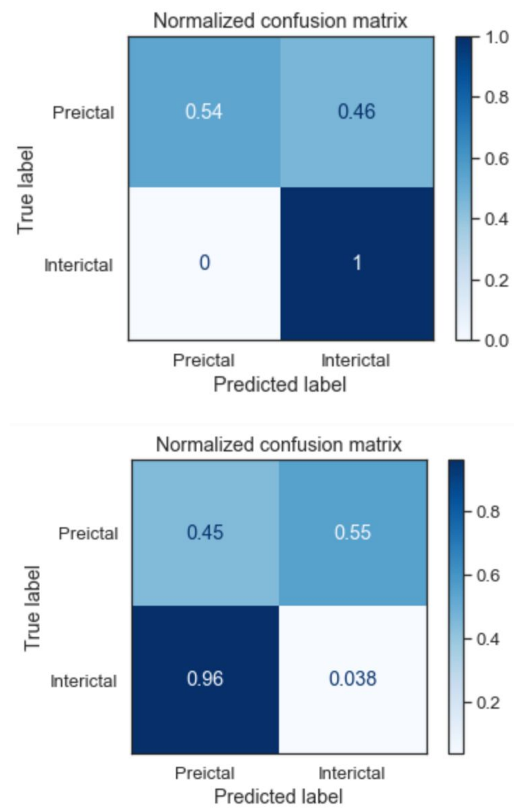


Figure 6: Two extreme examples of confusion matrix for one fold of testing in cross-validation

To resolve the observed issue, we tried different ways to segment the data, and limit the number of frequency bands used in feature extraction, but neither seems to fix this issue. Other possible reasons include the imbalanced training data between the interictal and the preictal intervals, the interictal and preictal intervals selection, possible movement artifacts throughout the data and feature calculation and selections. In addition, if we knew the nature of the seizure type and location of this particular patient, this information may help us to effectively choose features.

V. Future Work

Our average performance suggests that there is a lot of room for improvement whereas tuning the hyperparameter gamma could only help to an extent. Possible improvement for this work includes better selections of channels based on the fact if they are statistically different from each other and more features such as mean spike rate and mean phase coherence. If we had time, we may use T-test or ANOVA to investigate which channels are the best in terms of containing useful and non-redundant information. The optimal interictal and preictal intervals selection could be further evaluated by varying time before and after seizure onset intervals as another hyperparameter.

Contributions

Moira Feng contributed to the data segmenting & labeling, writing the code for cross-validation using the SVM model, and relevant sections in the report. Wendy Yu contributed to the signal pre-processing and feature extraction implementation and their respective sections for the report.

References

[1] A. H. Shoeb, "Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment", PhD diss., Massachusetts Institute of Technology, 2009.

[2] K. Gadhoumi, et al., Seizure prediction for therapeutic devices: A review, Journal of Neuroscience Methods, Volume 260, 2016, Pages 270-282, ISSN 0165-0270, https://doi.org/10.1016/j.jneumeth.2015.06.010.

[3] C. PR, et al., Seizure prediction: methods. Epilepsy Behav. 2011;22 Suppl 1(Suppl 1):S94–S101. doi:10.1016/j.yebeh.2011.09.001

[4] I. Kiral-Kornek, et al., Epileptic Seizure Prediction Using Big Data and Deep Learning: Toward a Mobile System, EBioMedicine, Volume 27, 2018, Pages 103-111, ISSN 2352-3964, https://doi.org/10.1016/j.ebiom.2017.11.032.

[5] Y. Park, et al., (2011), Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. Epilepsia, 52: 1761-1770. doi:10.1111/j.1528-1167.2011.03138.x

[6] F. Mormann, et al., Seizure prediction: the long and winding road, Brain, Volume 130, Issue 2, February 2007, Pages 314–333, https://doi.org/10.1093/brain/awl241

[7] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.