American Epilepsy Society Seizure Prediction Challenge

Team: Birchwood
Team member: Jialun He
Email: jialunhe@yahoo.com
Cambridge, Massachusetts, USA
28 Nov 2014

Introduction

The two major obstacles in this competition are over fitting and calibration prediction among subjects. The first one is probably due to very limited amount of training cases. For example, each patient data only has 3 independent seizure occurrences. For the first obstacle, a numerical experiment is carried out to determine how to split the data and select appropriate features. Support vector machine with RBF kernel provides best result among other common classifiers. Since the scoring matrix is based on AUC across subjects. Higher score of AUC in cross validation doesn't necessary lead to higher score in AUC across all subjects. Leader board (LB) score is used heavily as additional validation.

Feature Models

The dataset information is provided here. Each data file has about 10 minutes of time series data.

Dataset Information:

Subject	number of files: interictal	number of files: preictal	Ratio (int/pre)	channels	frequency	test
Dog1	480	24	20	16	399.9	502
Dog2	500	42	11.9	16	399.9	1000
Dog3	1440	72	20	16	399.9	907
Dog4	804	97	8.3	16	399.9	990
Dog5	450	30	15	15	399.9	191
Patient1	50	18	2.8	24	5000	195
Patient2	42	18	2.3	24	5000	150

The core features are the power spectral band. Other candidates of features are based on the common features used by the top finishers in the previous seizure detection competition. Such

PDF created with pdfFactory trial version www.pdffactory.com

as signal correlation between EEG channels and eigenvalue of the correlation matrix. Both in frequency domain and time domain.

Sample size and feature size experiment:

Sample size in second	Sample # in a file (in 4 x)	Dog1 preictal # (24x)	FFT magnitude node	FFT magnitude size	Feature size	Relative LB score
6	100	2400	6-266	48	1040	0.04909
30	20	480	30-1440	48	1040	0.08721
30	20	480	30-1440	24	656	0.09339
50	12	288	50-2400	18	560	0.10669
50	12	288	50-2400	24	656	0.10146
50	12	288	50-2400	12	464	0.10107
50	12	288	100-2400	18	560	0.09837
50	12	288	100-3000	18	560	0.08632
75	8	192	75-3600	12	464	0.09794
150	4	96	150-7200	12	464	0.10236
600	1	24	600-28800	48	1040	0.07325
600	1	24	600-28800	480	7952	0.0

The data can be split into equal size of various length to increase the number of training cases and reduce variance. A numerical experiment has been carried to determine the optimal way to split data, and determine the optimal feature size. The results are showed in the above table.

Single window model:

My best submission according to the LB score is based on single window model. All the data is first re-sample to 100Hz to reduce high frequency noise. Then every data file is split into 12 parts, about 50 seconds each. For each part of the split data, FFT is applied to transform the data to frequency domain. The power magnitudes in the frequency band from 1 to 50 Hz were selected and converted to logarithmic scale, then resample the frequency band to 18 bins to further reduce noise. The covariance and eigenvalues of the reduced frequency band across channels are also added as features, along with the covariance and eigenvalues in time domain.

Double window model:

I also tested other way to split and assemble data. One of the interesting method is the double window model. The model consist of two 50-second-window that is 5 minutes apart. Sequence number is used to combine data sequence between files. Therefore I can has approximate equal size of training cases comparing to single window model. When all the features in the single window model are used, the feature size is doubled, which leads to overfitting. Then I tried to remove some of the features and get similar LB score compared to single window model. But I

PDF created with pdfFactory trial version www.pdffactory.com

don’t have time to fine tune the classifier so it doesn’t represent the highest score in my submission.

Cross subject model:

Cross subject model has been tested. Using the same features as the individual subject. The results are worse than the results from individual subject. Also because of the number of channels and the sample frequency are not consistent across subjects. Special treatment is needed to extract common features across subjects/

Classifier Models

Several common classifiers in scikit-learn package have been tested, such as random forest, gradient tree boosting, support vector machine. Most of them had really good CV score for individual subject. But did not get good score in LB. The gaps between CV score and LB score were very big. One of the reason is that LB score is across all subject. Other possible reason is due to overfitting. Platt scaling was also added to calibrate the prediction across subjects. It improved LB score slightly.

My best submissions according to the LB score were based on support vector machine with RBF kernel, which produced better results because of more control in balancing bias and variance. The classifier gave an estimate for each 50-second window. An averaging method was used to combine the results and provide estimate for the whole 10-minute period. Several averaging methods, such as arithmetic average, geometric average and harmonic average, have been tested. Arithmetic average was best suited for evenly distributed estimate while harmonic average was best suited for oddly distribution. Therefore, a combined averaging method was used. A percentage projection was also used to align results across subjects based on the assumption that the test dataset was similar to the training dataset.

Software

The software is written in Python. The standard packages of numpy, scipy, scikit-learn and matplotlib are used extensively. The source code structure is based on the source code of the last competition winner. The structure is good at streamline batch processing but lack visualization tools to examine features and analytical results. Graphic output has been added to help understanding the data and analytical results. The repository is available at https://github.com/jlnh/SeizurePrediction. For single window model, simply run predict.py with appropriate setting in SETTINGS.json. For double window model, additional code modification required in the function of parse_input_data in task.py. For cross subject model, run predict_all.py and additional modification required to extract common features across subjects.

Dependency

PDF created with pdfFactory trial version www.pdffactory.com

_ Python 2.7

_ scikit learn-0.14.1

_ numpy-1.8.1

_ pandas-0.14.0

_ scipy

_ hickle (plus h5py and hdf5, see https://github.com/telegraphic/hickle for installation details)

Hardware and Runtime

The computations were done in a desktop with Intel i7 quad core CPU and 12GB RAM. The total computational time is about 2 hours when the classifier is set to use all the CPU threads.

How to Generate the Solution

- Modify SETTINGS.json file and put the data in the data dir. Sample SETTINGS.json is given here

  ```
  {
    "competition-data-dir": "seizure-data",
    "data-cache-dir": "data-cache",
    "submission-dir": "submissions",
    "figure-dir": "figure"
  }
  ```

- Run predict.py
- Check the submission file in submissions directory and the analytical graphic PDF file in figure directory

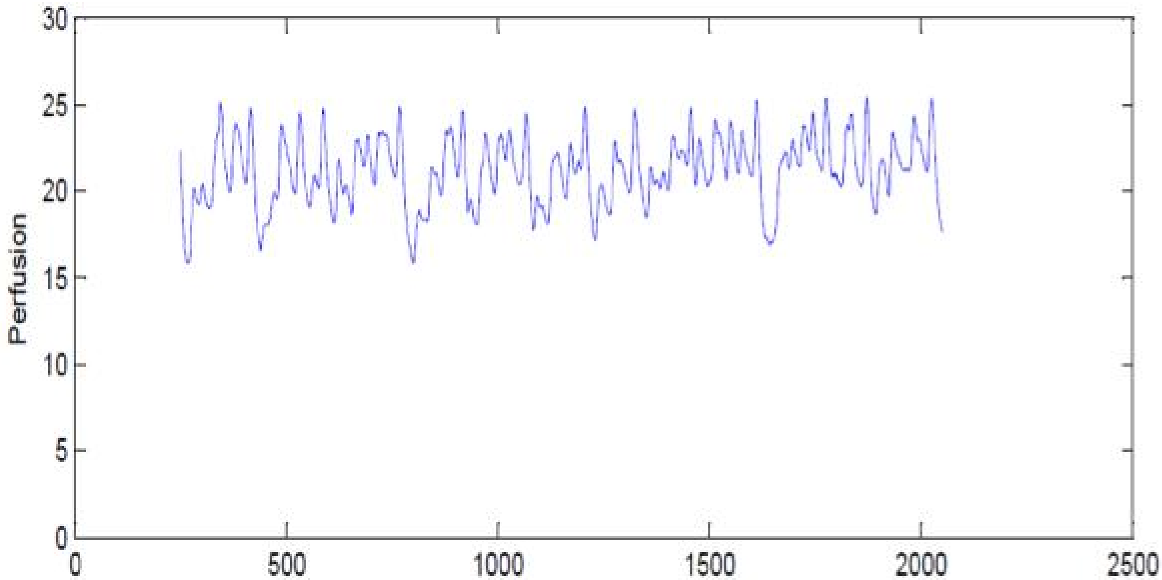
Additional Comments

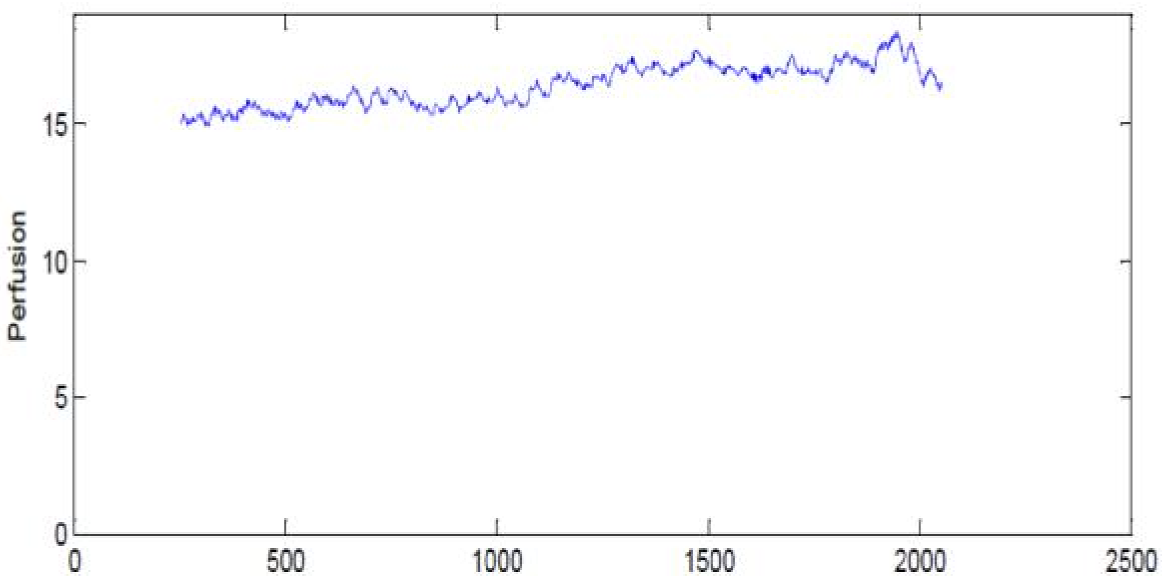
Due to time constraint, C and gamma in SVM was set for all subject. A quick visual inspection of the prediction histogram found that it would be better to select the parameters for individual subjects via grid search method.

Other Signals for Seizure Detection

PDF created with pdfFactory trial version www.pdffactory.com

Physiological signal of the brain other than EEG can also be used to detect seizure. I have worked on a seizure detection project using capillary blood flow rate (blood perfusion rate) in human brain a few years ago. The following two figures shows perfusion (ml/100g-min) vs. time (second) in brain monitoring 7 hours apart. The upper figure shows perfusion during seizure occurrence period while the bottom figure shows perfusion without seizure.





PDF created with pdfFactory trial version www.pdffactory.com

References

- Kaggle/UPenn and Mayo Clinic's Seizure Detection Challenge. May – August, 2014.
- Howbert JJ, Patterson EE, Stead SM, Brinkmann B, Vasoli V, Crepeau D, Vite CH, Sturges B, Ruedebusch V, Mavoori J, Leyde K, Sheffield WD, Litt B, Worrell GA (2014) Forecasting seizures in dogs with naturally occurring epilepsy. PLoS One 9(1):e81920.
- Cook MJ, O'Brien TJ, Berkovic SF, Murphy M, Morokoff A, Fabinyi G, D'Souza W, Yerra R, Archer J, Litewka L, Hosking S, Lightfoot P, Ruedebusch V, Sheffield WD, Snyder D, Leyde K, Himes D (2013) Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. LANCET NEUROL 12:563-571.
- Park Y, Luo L, Parhi KK, Netoff T (2011) Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. Epilepsia 52:1761-1770.
- Davis KA, Sturges BK, Vite CH, Ruedebusch V, Worrell G, Gardner AB, Leyde K, Sheffield WD, Litt B (2011) A novel implanted device to wirelessly record and analyze continuous intracranial canine EEG. Epilepsy Res 96:116-122.
- Andrzejak RG, Chicharro D, Elger CE, Mormann F (2009) Seizure prediction: Any better than chance? Clin Neurophysiol.
- Snyder DE, Echauz J, Grimes DB, Litt B (2008) The statistics of a practical seizure warning system. J Neural Eng 5: 392–401.
- Mormann F, Andrzejak RG, Elger CE, Lehnertz K (2007) Seizure prediction: the long and winding road. Brain 130: 314–333.
- Haut S, Shinnar S, Moshe SL, O'Dell C, Legatt AD. (1999) The association between seizure clustering and status epilepticus in patients with intractable complex partial seizures. Epilepsia 40:1832–1834.

PDF created with pdfFactory trial version www.pdffactory.com