From:	Lang, Brian
To:	Levin Kuhlmann
Cc:	Roman, Kelly; Lavery, Daniel; Dobson, Scott; Boxer, Alexander
Subject:	2nd Place Info for Melbourne University AES-MathWorks-NIH Seizure Prediction Challenge
Date:	Sunday, 4 December 2016 1:41:23 PM
Attachments:	EEG-kaggle_feature_breakdown.pptx

Good evening Levin,

Thank you for hosting a competition on such an important and interesting topic!

The final model utilized extremely randomized trees on both long (entire 10-minute file) and short (minute based) term generated features, about a total of 2200 in all.   It was trained and tuned, not at all exhaustively but rather crudely, for each patient.   It should be noted that deep-learning techniques (both CNNs and LSTMs) were quickly investigated but abandoned due to the fact of the limited data-set – the best result using such techniques was ~0.62.

Long term features included those easily implemented from literature (correlation coefficients between channels; spectral and temporal statistics on a downsampled (factor of 5) per channel basis).

The short term features were normalized summed energy in the standard spectral bands derived on a minute-by-minute basis with 30s of overlap.

For both long and short term feature extraction the data was whitened (i.e. mean-zero and standard deviation of unity) per channel. Additionally, all power spectrum features used the Welch method so as to help beat down the noise.

Feature optimization was briefly explored but not in-depth. Looking at some earlier submissions, it is worth noting that the short term features alone would have achieved a top 10 result while the long term features alone would have achieved a top 20 result. Lastly, a randomly selected hold-out set was used to optimize the various hyper-parameters on a per patient basis -- 10% for both ambient and target per patient sequences.

This was done in the Python/Anaconda suite. Feature generation took the longest (roughly a ½ hour to gather all features for all patients both training and testing) but needed to only occur once. Training took on the order of 1-2 minutes. In short, it was rather speedy.

Lastly, attached are three charts showing relative feature importance for each patient.

Also, if possible, could you forward us a copy of your briefing? We are rather interested to see the bigger picture of this work.
Please let me know if have any questions. Good luck with your briefing on Monday.

-Brian

Brian Lang, Ph.D.   Areté Associates

Senior Scientist   1550 Crystal Drive

Voice (703) 413-0290   Suite 703

Fax (703) 413-0295   Arlington, Virginia 22202

e-mail : blang@arete.com