Confused student EEG brainwave data

-K. Sai Krishna

Objective

- Binary classification model to find whether a student is confused or not.

About Dataset

- The EEG (electro-encephalo-gram) data is collected from 10 students while they were watching 10 MOOC(Massive open online course) videos each.
- Each MOOC video is roughly 2 minutes long.
- First 30 sec and last 30 sec data is removed and the remaining EEG data is used.
- Data is collected for every 0.5 seconds.
- The user-defined label is given by the student which says whether a student is confused or not.

Visualization of Dataset

	SubjectID	VideoID	Attention	Mediation	Raw	Delta	Theta	Alpha1	Alpha2	Beta1	Beta2	Gamma1	Gamma2	predefinedlabel	user-definedlabeln	age	ethnicity	gender
0	0.0	0.0	56.0	43.0	278.0	301963.0	90612.0	33735.0	23991.0	27946.0	45097.0	33228.0	8293.0	0.0	0.0	25	Han Chinese	M
1	0.0	0.0	40.0	35.0	-50.0	73787.0	28083.0	1439.0	2240.0	2746.0	3687.0	5293.0	2740.0	0.0	0.0	25	Han Chinese	M
2	0.0	0.0	47.0	48.0	101.0	758353.0	383745.0	201999.0	62107.0	36293.0	130536.0	57243.0	25354.0	0.0	0.0	25	Han Chinese	M
3	0.0	0.0	47.0	57.0	-5.0	2012240.0	129350.0	61236.0	17084.0	11488.0	62462.0	49960.0	33932.0	0.0	0.0	25	Han Chinese	M
4	0.0	0.0	44.0	53.0	-8.0	1005145.0	354328.0	37102.0	88881.0	45307.0	99603.0	44790.0	29749.0	0.0	0.0	25	Han Chinese	M
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
12806	9.0	9.0	64.0	38.0	-39.0	127574.0	9951.0	709.0	21732.0	3872.0	39728.0	2598.0	960.0	1.0	0.0	24	Han Chinese	F
12807	9.0	9.0	61.0	35.0	-275.0	323061.0	797464.0	153171.0	145805.0	39829.0	571280.0	36574.0	10010.0	1.0	0.0	24	Han Chinese	F
12808	9.0	9.0	60.0	29.0	-426.0	680989.0	154296.0	40068.0	39122.0	10966.0	26975.0	20427.0	2024.0	1.0	0.0	24	Han Chinese	F
12809	9.0	9.0	60.0	29.0	-84.0	366269.0	27346.0	11444.0	9932.0	1939.0	3283.0	12323.0	1764.0	1.0	0.0	24	Han Chinese	F
12810	9.0	9.0	64.0	29.0	-49.0	1164555.0	1184366.0	50014.0	124208.0	10634.0	445383.0	22133.0	4482.0	1.0	0.0	24	Han Chinese	F

12811 rows × 18 columns

Sample video (easy)

- VideoId = 0
- Video is about the Coulomb's law.
- Subjects with Subject Ids 6, 7, 9 marked this video as Confused.
- Remaining subjects marked this video as Not Confused.

Sample video (Tough)

- VideoId = 8
- Some key words used in this video are Belter Noise Immunity and Noise Margin.
- Subjects with Subject Ids 1, 7 marked this video as Not Confused.
- Remaining subjects marked this video as Confused.

ML models used

- Naive Bayes
- Logistic Regression
- SVM
- Ada-Boost
- **XGBoost**
- Bagging

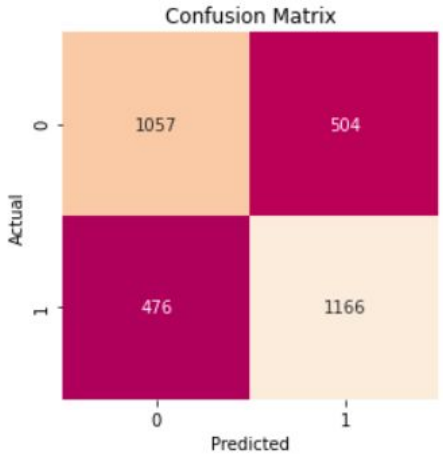
Model	ROC-AUC Score	F1-Score
Naive Bayes	0.6297430673519661	0.55
Logistic Regression	0.625830517150301	0.57
SVM	0.733440180527021	0.67
Ada-Boost	0.668655551229302	0.62
Bagging	0.6723806376655085	0.62
XGBoost	0.7428121983706064	0.68

Tried with different permutations of features for the XGBoost and SVM models.

```
inputs = Input(shape=(f,1))

Dense1 = Dense(64, activation = tf.nn.relu)(inputs)

lstm_1=  Bidirectional(LSTM(256, return_sequences = True, dropout=0.3))(Dense1)
lstm_2=  Bidirectional(LSTM(128, return_sequences = True, dropout=0.3))(lstm_1)

flat = Flatten()(lstm_2)

Dense_2 = Dense(128, activation = tf.nn.relu)(flat)
outputs = Dense(1, activation=tf.nn.sigmoid)(Dense_2)

model = tf.keras.Model(inputs, outputs)
```

Neural Networks

	precision	recall	f1-score	support
0.0	0.69	0.68	0.68	1561
1.0	0.70	0.71	0.70	1642
accuracy			0.69	3203
macro avg	0.69	0.69	0.69	3203
weighted avg	0.69	0.69	0.69	3203



Single Video Analysis

- Used the data of only a single video in both train and test data.
- Did this single video analysis for all the models.
- Among ML models, AdaBoost gave good results.
- Bidirectional LSTMs outperformed the other models in this analysis.

Video-Id	Neural Network (ALL) (F1 Score)
0	0.9405099153518677
1	0.9576271176338196
2	0.9341692924499512
3	0.940397322177887
4	0.9085545539855957
5	0.9675324559211731
6	0.9155405163764954
7	0.9491525292396545
8	0.9125000238418579
9	0.9844236969947815

Statistical Feature extraction

- Did statistical feature extraction using skewness and kurtosis.
- After performing this extraction we will get 100 data points(rows) in our dataset.
- Used 8 subjects for training (80 rows) and 2 subjects (20 rows) for testing.
- The accuracies are shown in the next slide.

Statistical Feature extraction

Kurtosis

Wave	SVM			XGBoost		
	Hyper (C=)	ROC-AUC	F1 Score	Hyper (n_estimators/lr)	ROC-AUC	F1 Score
Delta	0.5	0.7171717171717171	0.65	1000/0.0005	0.6565656565656566	0.60
Theta	4	0.46969696969696967	0.50	1000/0.5	0.5909090909090908	0.55
Alpha1	0.3	0.5252525252525253	0.45	1000/0.005	0.35858585858585856	0.40
Alpha2	6	0.5252525252525253	0.55	1000/0.0005	0.37878787878787884	0.40
Beta1	1	0.7373737373737373	0.40	1000/0.5	0.398989898989899	0.40
Beta2	0.1	0.8030303030303029	0.40	1000/0.5	0.5505050505050505	0.50
Gamma1	10	0.6464646464646465	0.65	1000/0.05	0.51010101010101	0.55
Gamma2	0.5	0.5454545454545454	0.45	1000/0.5	0.5707070707070707	0.55
All EEG waves	1	0.46464646464646464	0.60	2000/0.001	0.4949494949494949	0.50

Skew

Wave	SVM			XGBoost		
	Hyper (C=)	ROC-AUC	F1 Score	Hyper (n_estimators/lr)	ROC-AUC	F1 Score
Delta	1.5	0.7222222222222222	0.65	1000/0.005	0.5909090909090909	0.55
Theta	1	0.48484848484848486	0.45	2000/0.005	0.4545454545454545	0.45
Alpha1	1	0.5353535353535354	0.55	1000/0.0001	0.4242424242424242	0.40
Alpha2	2	0.494949494949495	0.50	1000/0.005	0.48989898989898994	0.45
Beta1	0.5	0.7878787878787878	0.35	1000/0.05	0.36363636363636365	0.40
Beta2	2	0.6464646464646464	0.35	1000/0.0005	0.24747474747474746	0.30
Gamma1	1	0.7373737373737373	0.70	1000/0.00005	0.5505050505050505	0.60
Gamma2	2	0.5151515151515151	0.50	1000/0.05	0.5858585858585859	0.55
All	1	0.6262626262626263	0.55	2000/0.0005	0.6666666666666666	0.55
All EEG waves	0.5	0.5656565656565655	0.60	1000/0.005	0.6161616161616161	0.50

T-test results

VideoId	Delta	Theta	Alpha1	Alpha2	Beta1	Beta2	Gamma1	Gama2	Description of each video
Video-0	2.60E-19	1.48E-09	3.35E-08	2.15E-15	1.31E-08	3.42E-18	3.99E-17	6.30E-17	Columbs law
Video-1	0.01917973021	0.2034150056	0.0903505782	2.23E-05	0.003523557699	1.40E-23	1.46E-14	1.52E-08	f(x) = (x-1)/(x-1)
Video-2	2.44E-05	1.55E-04	0.0002273851993	0.2220029939	0.02795505971	7.16E-13	2.52E-09	4.03E-05	g(x) = x^2 (x!=2), 1(x=2)
Video-3	2.16E-16	1.34E-16	2.39E-11	4.06E-20	4.51E-16	0.0007734567406	1.48E-05	4.16E-11	summation 1 to N
Video-4	0.003914280635	0.0009990733125	0.003993034803	0.01286590963	3.58E-11	5.06E-24	3.49E-13	4.12E-11	maths sigma
Video-5	3.48E-17	2.47E-13	7.31E-13	0.01562712624	0.1957588507	0.06030972578	0.2322098133	0.01298306039	Repulsion, attraction
Video-6	1.62E-13	6.03E-15	0.0001809823312	9.83E-11	1.19E-15	2.67E-15	9.62E-17	2.62E-12	annhilation, nuclear decal, radio active elements
Video-7	3.88E-05	0.06433494646	0.01818551969	1.53E-07	0.009865872274	2.05E-06	3.25E-09	0.001586300454	Circuts (Heat sensors, registers, voltage)
Video-8	7.45E-11	1.44E-16	5.13E-13	7.46E-17	7.85E-11	2.40E-08	8.32E-14	3.58E-06	Belter-noise immunity, Noise margin
Video-9	2.24E-25	2.95E-16	1.73E-10	7.01E-07	4.74E-09	0.6203719476	0.007864934358	1.65E-06	Low noise amplifier, Analog to digital

Thank You!