Cognitive State Predicting of Pilots with Machine Learning of Physiological Data

IE7275 Data Mining in Engineering

Jeff Johnston

Introduction to Problem

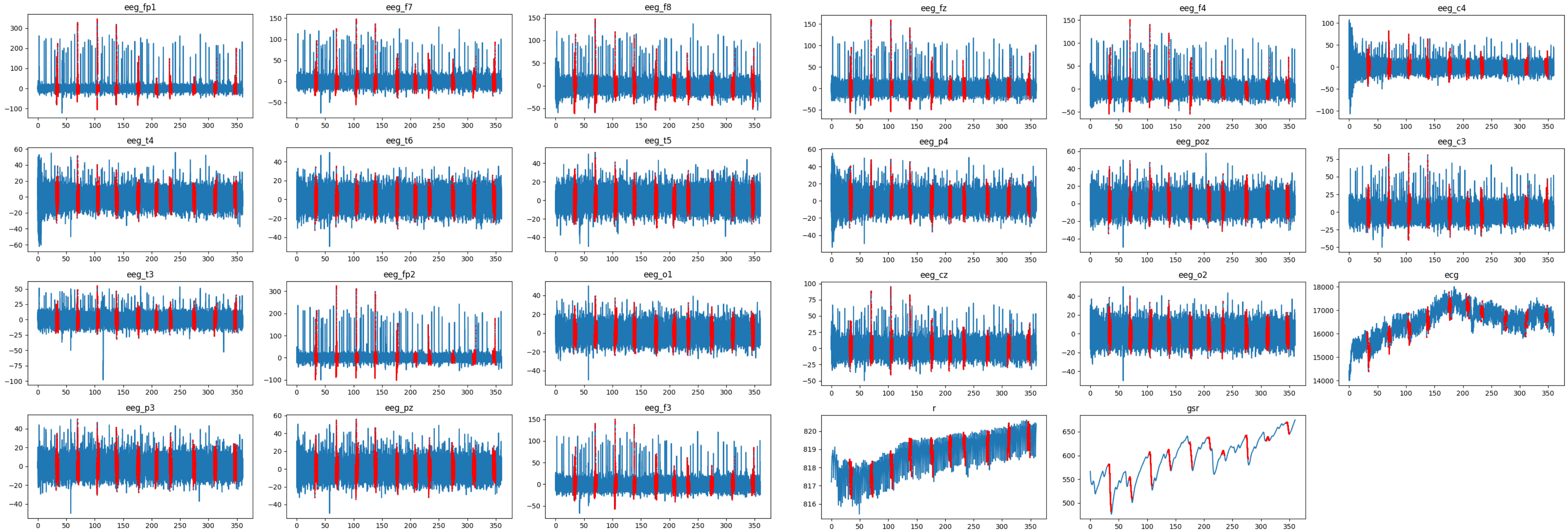
- 2019 competition from Booz Allen Hamilton with overall intent of reducing commercial aviation fatalities.
- Real physiological data from 18 pilots during experimental conditions outside of the aircraft.
- Pilots faced intentional distractions to induce one of the following cognitive states:
  - **Channelized Attention** (CA) – state of focus on one task to the exclusion of all others
  - **Diverted Attention** (DA) – state of having one’s attention diverted by actions or thought processes associated with a decision
  - **Startle/Surprise** (SS) – i.e. jump scare

Exploratory Data Analysis

- 4,867,421 rows x 28 columns
- Sensors sample rate of 256 Hz (~256 samples per second)
- Columns:
  - ‘crew’: (1, 2, 3, 4, 5, 6, 7, 8, 13)
  - ‘experiment’: (CA, DA, SS)
  - ‘time’
  - ‘seat’: (0, 1)
  - 20 channels of EEG (electroencephalogram) data
    - 'eeg_fp1', 'eeg_f7', 'eeg_f8', 'eeg_t4', 'eeg_t6', 'eeg_t5', 'eeg_t3', 'eeg_fp2', 'eeg_o1', 'eeg_p3', 'eeg_pz', 'eeg_f3', 'eeg_fz', 'eeg_f4', 'eeg_c4', 'eeg_p4', 'eeg_poz', 'eeg_c3', 'eeg_cz', 'eeg_o2’
  - EGC (electrocardiogram) data – ‘ecg’
  - Respiration – ‘r’
  - GSR (galvanic skin conductance) – ‘gsr’
  - ‘event’: (A, B, C, D)

Exploratory Data Analysis

- Raw data plots over time
- Crew 3, Pilot 1, Diverted Attention



Exploratory Data Analysis

- Grouped the data by 'crew', 'seat', 'event'
- Broke the timeseries into windows of time
- Extracted features throughout the timeseries from each signal
  - Ordinal patterns, permutation entropy and complexity
- New data frame
  - 1149 rows x 49 columns

	crew	seat	event	pe_eeg_fp1	comp_eeg_fp1	pe_eeg_f7	comp_eeg_f7	pe_eeg_f8	comp_eeg_f8	pe_eeg_t4	...
0	1	0	A	0.962960	0.047118	0.978569	0.027709	0.979022	0.027234	0.983132	...
1	1	0	A	0.971943	0.035798	0.983263	0.022061	0.974494	0.033304	0.986097	...
2	1	0	A	0.962349	0.048035	0.979061	0.027628	0.975439	0.032203	0.982220	...
3	1	0	A	0.965927	0.043520	0.983061	0.022143	0.975629	0.032015	0.989625	...
4	1	0	A	0.966704	0.043051	0.980912	0.024946	0.973151	0.034919	0.980093	...

Modeling

- Modeled the data as four outcomes
  - Baseline, Channelized-Attention, Diverted-Attention, Startle/Surprise
- Modeled the data as binary outcomes
  - Baseline, Non-baseline

- Logistic Regression
- Neural Network Multilayer Perceptron
- Decision Tree

Modeling

- Logistic Regression
  - Quaternary
- Time: 0.0514 sec

	precision	recall	f1-score	support
Baseline	0.72	0.84	0.78	131
Channelized-Attention	0.73	0.67	0.70	84
Diverted-Attention	0.00	0.00	0.00	12
Startle/Surprise	0.00	0.00	0.00	3
accuracy			0.72	230
macro avg	0.36	0.38	0.37	230
weighted avg	0.68	0.72	0.70	230

- Logistic Regression
  - Binary
- Time: 0.0697 sec

	precision	recall	f1-score	support
baseline	0.72	0.82	0.77	131
non-baseline	0.71	0.59	0.64	99
accuracy			0.72	230
macro avg	0.72	0.70	0.70	230
weighted avg	0.72	0.72	0.71	230

Modeling

- MLP
  - Quaternary
- Time: 0.8474 sec

	precision	recall	f1-score	support
Baseline	0.83	0.92	0.87	131
Channelized-Attention	0.86	0.85	0.85	84
Diverted-Attention	0.67	0.17	0.27	12
Startle/Surprise	0.00	0.00	0.00	3
accuracy			0.84	230
macro avg	0.59	0.48	0.50	230
weighted avg	0.82	0.84	0.82	230

- MLP
  - Binary
- Time: 0.7974 sec

	precision	recall	f1-score	support
baseline	0.83	0.90	0.86	131
non-baseline	0.85	0.76	0.80	99
accuracy			0.84	230
macro avg	0.84	0.83	0.83	230
weighted avg	0.84	0.84	0.84	230

Modeling

- Decision Tree
  - Quaternary
- Time: 0.1183 sec

	precision	recall	f1-score	support
Baseline	0.77	0.76	0.76	131
Channelized-Attention	0.70	0.74	0.72	84
Diverted-Attention	0.12	0.08	0.10	12
Startle/Surprise	0.00	0.00	0.00	3
accuracy			0.70	230
macro avg	0.40	0.39	0.39	230
weighted avg	0.70	0.70	0.70	230

- Decision Tree
  - Binary
- Time: 0.1118 sec

	precision	recall	f1-score	support
baseline	0.71	0.73	0.72	131
non-baseline	0.63	0.60	0.61	99
accuracy			0.67	230
macro avg	0.67	0.66	0.67	230
weighted avg	0.67	0.67	0.67	230

Hyperparameter tuning

Gridsearch CV method

- Logistic Regression, Quaternary
  - Best parameters: C = 29.76351, penalty = l2, solver = liblinear
  - Best accuracy score: **0.726**
  - Time: 459.77 seconds
- Logistic Regression, Binary
  - Best parameters: C = 11.28838, penalty = l2, solver = liblinear
  - Best accuracy score: **0.719**
  - Time: 156.14 seconds

Hyperparameter tuning

Gridsearch CV method

- MLP, Quaternary
  - Best parameters: activation = relu, alpha = 0.0001, hidden layer size = (100, 50), learning rate = adaptive, solver = sgd
  - Best accuracy score: **0.848**
  - Time: 2817.18 seconds (~47 min)
- MLP, Binary
  - Best parameters: activation = relu, alpha = 0.001, hidden layer size = (100, 50), learning rate = adaptive, solver = sgd
  - Best accuracy score: **0.833**
  - Time: 2701.63 seconds (~45 min)

Hyperparameter tuning

Gridsearch CV method

- Decision Tree, Quaternary
  - Best parameters: criterion = gini, max depth = 15, min samples per leaf = 4, min samples split = 2
  - Best accuracy score: **0.729**
  - Time: 37.60 seconds
- Decision Tree, Binary
  - Best parameters: criterion = gini, max depth = none, min samples per leaf = 1, min samples split = 10
  - Best accuracy score: **0.742**
  - Time: 54.95 seconds

Results

- Best method: baseline MLP – 84.8% accuracy

- Not much difference between quaternary and binary, difficulty with Diverted-Attention and Startle/Surprise (small sample sizes)

- Future work
  - Focus on ecg, r, and gsr data
  - Further feature selection
  - Adjust feature extraction statistics
  - Apply additional modeling methods
  - Include more samples

References

- Amie Johnson, Jack Sarle, Maggie, Paul Terwilliger, Shannon Walker, Will Cukierski, William Espenschied. (2018). Reducing Commercial Aviation Fatalities. Kaggle. https://kaggle.com/competitions/reducing-commercial-aviation-fatalities
- MNE-python https://mne.tools/stable/auto_tutorials/intro/10_overview.html
- IE6400 Foundations of Data Analytics, Project 3. Jeffrey Johnston, Sanidhya Karnik, Himanshu Randad, Digvijay Ashish Raut, John Ayush Xavier
- Radhakrishnan, S. (2018). Permutation-entropy/p_entropy.py . https://github.com/srk-srinivasan/Permutation-Entropy/blob/master/p_entropy.py