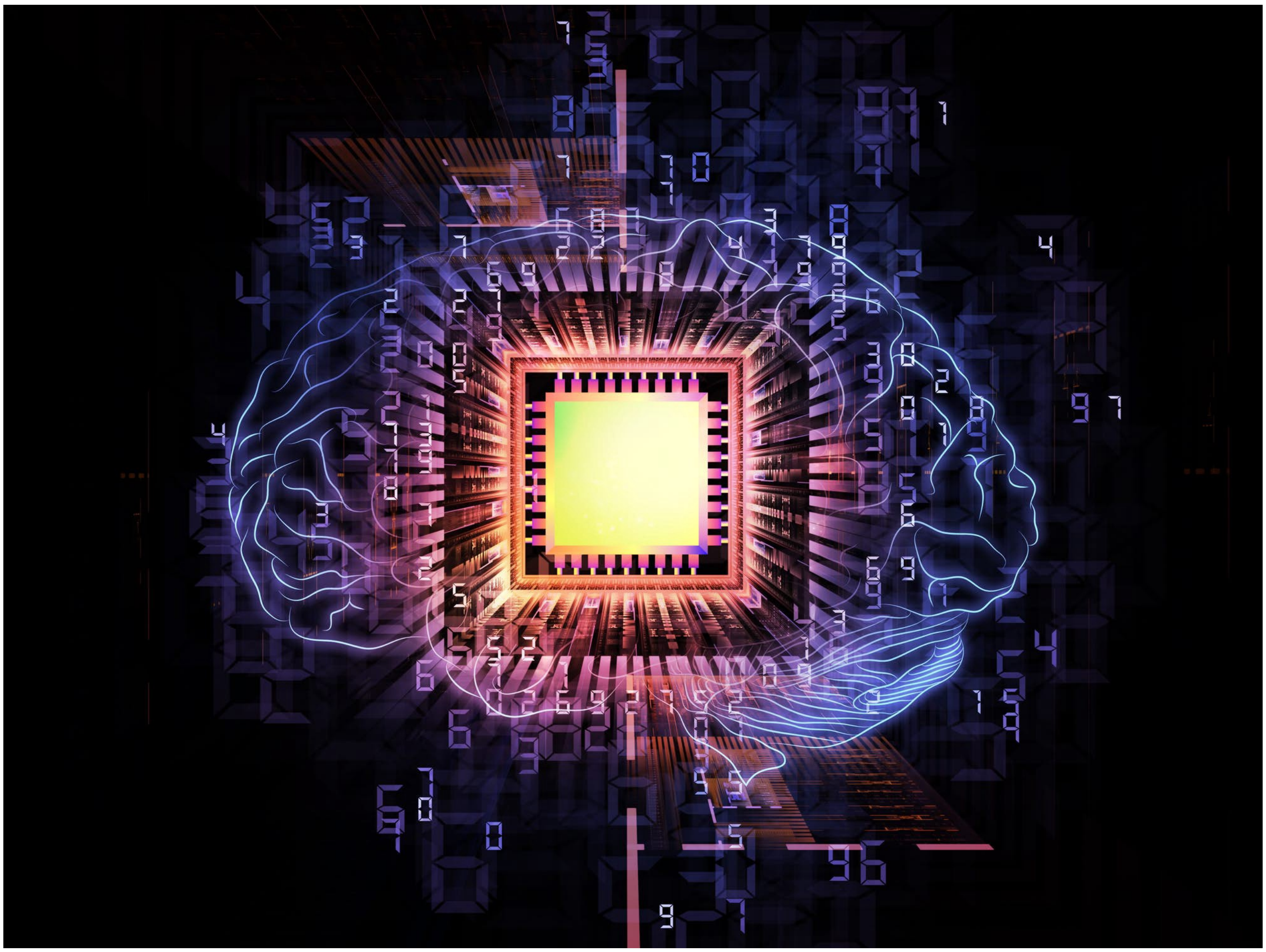
Building an AI that can read your mind

towardsdatascience.com/building-an-ai-that-can-read-your-mind-8b22ad5a7f05

November 5, 2019

Machine learning for Mental State Classification using EEG data



(Source: Shutterstock)

AI that can read your mind

This might sound like the plot from a dystopic future of a sci-fi movie. However, with recent progress, this technology is now taking the leap from science fiction to merely science.

Autonomous non-invasive detection of brain activity is potentially useful in multiple domains such as human robot interaction and mental healthcare. It can provide an extra dimension of interaction between user and device, as well as enabling tangible information to be derived that does not depend on verbal communication.

Such innovations also mean better brain-computer interfacing. This would open up whole new platforms for man-machine communication, including in helping out people with physical or mental disorders. A brain-computer interface might allow a paralyzed person to move a robotic arm, or a person with a spinal cord injury to control a motorized wheelchair.

With the increasing availability of low-cost electroencephalography (EEG) devices, brainwave data is becoming affordable for the consumer industry as well as for research, introducing the need for autonomous classification without the requirement of an expert on hand.

In this article, I will cover an example case of how machine learning can be used to analyze brain activity. Through EEG recordings from commercially available devices, I show how machine learning models can be used to predict the corresponding mental state of the subject.

Machine Learning for Mental State Classification

Recording high quality EEG data is not straight foreward unless you are affiliated with a lab performing such experiments. However, I recently came across an interesting article "A Study on Mental State Classification using EEG-based Brain-Machine Interface" by Jordan J. Bird, Luis J. Manso, Eduardo P. Ribiero, Anikó Ekárt and Diego R. Faria. Luckily, they have shared the data used in their research openly for others to experiment with. What is particulary interesting, in my opinion, is the use of consumer grade devices which you can simply order on Amazon for a couple of hundred USD. The methods for data recording and processing in their study is covered in the below section.

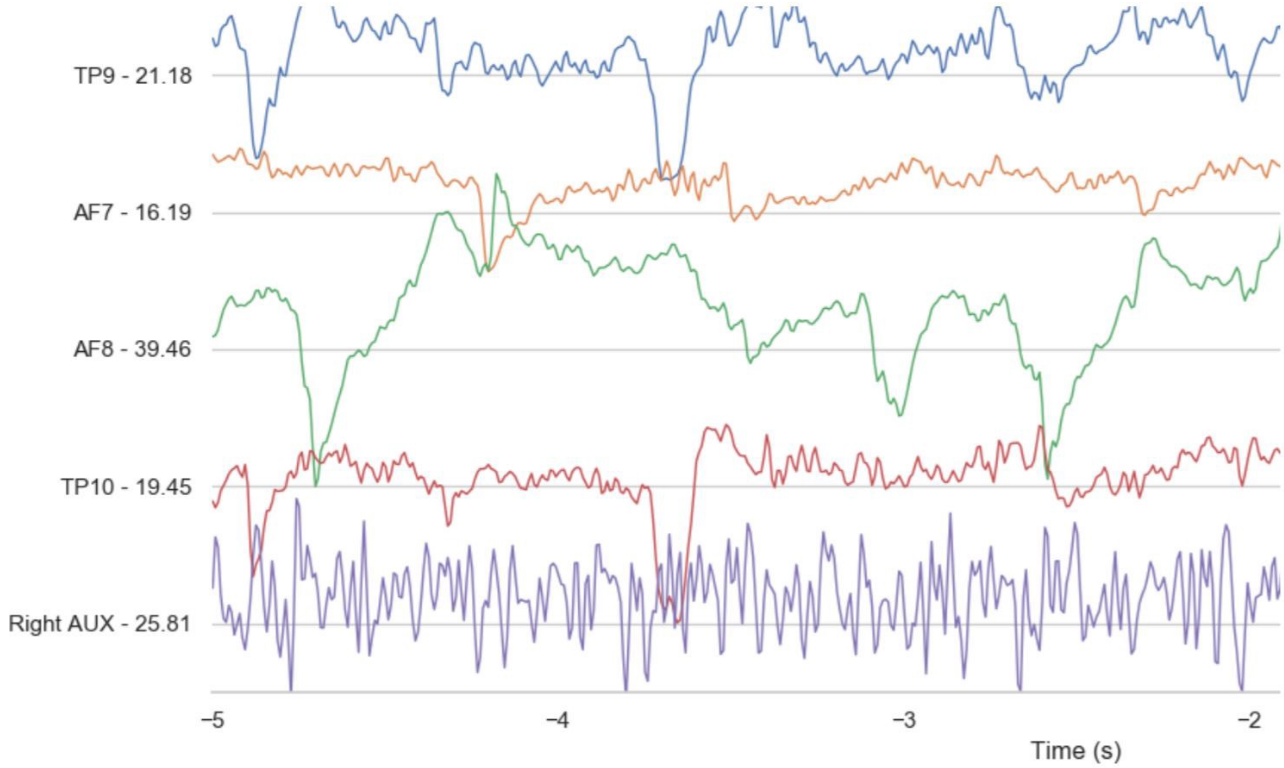
Experimental details

> The study employs four dry extra-cranial electrodes via a commercially available MUSE EEG headband. This is a wearable brain sensing device which measures brain activity via 4 electroencephalography (EEG) sensors.
>
> To evoke different mental states, the experiment utilized a selection of film clips shown in the below table, representing positive and negative valence

Stimulus	Valence	Studio	Year
Marley and Me	Neg	Twentieth Century Fox, etc.	2008
Up	Neg	Walt Disney Pictures, etc.	2009
My Girl	Neg	Imagine Entertainment, etc.	1991
La La Land	Pos	Summit Entertainment, etc.	2016
Slow Life	Pos	BioQuest Studios	2014
Funny Dogs	Pos	MashupZone	2015

Film Clips used as Stimuli for EEG Brainwave Data Collection.Source

> Sixty seconds of data were recorded from two subjects (1 male, 1 female, aged 20–22) for each of the 6 film clips found in the table below, producing 12 minutes of brain activity data (6 minutes for each emotional state). Six minutes of neutral brainwave data were also collected resulting in a grand total of 36 minutes of EEG data recorded from subjects. (Three minutes of data were collected per day to reduce the interference of a resting emotional state). With a variable frequency resampled to 150Hz, this resulted in a dataset of 324,000 data points.
>
> **Proposed Set of Features for EEG signals**
>
> Feature extraction and classification of EEG signals are core issues in brain computer interface (BCI) applications. One challenging problem when it comes to EEG feature extraction is the complexity of the signal, since it is non-linear, nonstationary, and random in nature. The signals are considered stationary only within short intervals, that is why the best practice is to apply short-time windowing technique to meet this requirement. However, it is still considered an assumption that holds during a normal brain condition.



Example of a live EEG stream of the four Muse sensors, Right AUX did not have a device and was discarded due to it simply being noise. This live feed graph has a Y-Axis of measured microvolts at t=0 on each sensor, and an X-axis detailing the time reading. Source

> This subsection describes the set of features considered to adequately discriminate different classes of mental states. These features rely on statistical techniques, time-frequency based on fast Fourier transform (FFT), Shannon entropy, max-min features in temporal sequences, and others. All features proposed to classify the mental states are computed in terms of the temporal distribution of the signal in a given time window. This slide window is defined as a period of 1 second, i.e. all features are computed within this time instant. For a more detailed introduction to the feature extraction, have a look at the original article

Machine learning algorithm

Following the above approach of feature extraction from the raw EEG data, we are left with a dataset containing 2548 features. For each row of the dataset, we have the corresponding target variable: '*Neutral*', '*Negative*' or '*Positive*'. Our goal then, is to train a machine learning model, based on this set of features, to succesfully predict the corresponding mental state.

In this example case, I started out with one of my "go-to" algorithms, the Random forest classifier, as it is simple to set up and often works quite well "out-of-the-box", without much hyperparameter tuning.

As a side-note: It would also be interesting to try a convolutional neural network approach on the raw time-domain data (rather than the set of extracted features which include various frequency characteristics of the signal). Since convolutions applied in the time-domain is intimately connected to the frequency characteristics of the signal through the convolution theorem, this might be a promising approach to reduce the amount of pre-processing and feature extraction.

Cross-validation

One of the first things you learn about in applying ML is the importance of cross-validation: evaluating the performance of your model on a portion of your dataset separate from what you used to train your model. One way of doing this, is to holdout parts of the dataset when training your model, and estimate performance of the trained model using e.g the following approach:

- ***Train your model on 70% of your labeled data***
- ***Evaluate the trained model on the remaining 30%***

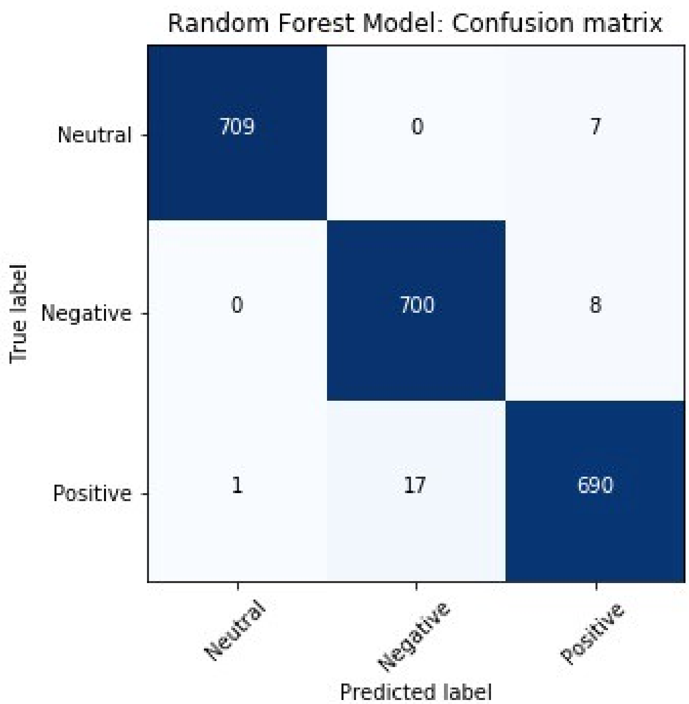
K-fold cross-validation improves on this by letting you do this multiple times so you can see whether the test performance varies based on which samples you used to train / test.

Source: http://karlrosaen.com/ml/learning-log/2016-06-20/

By running through the train/test comparison several times you will get a better estimate of model performance, and sanity check that the model is not performing wildly differently after being trained on different segments of your labeled data, which in itself could indicate instability in your model or too small a sample set.

In my case, I performed a 10-fold cross validation when training the model, and calculated the accuracy evaluated over different segments of the data. The final performance of the model can be visualized through the confusion matrix shown below.

In the field of machine learning and specifically the problem of statistical classification, a **confusion matrix**, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm. Each row of the matrix represents the instances in the actual class while each column represents the instances in the predicted class. The name stems from the fact that it makes it easy to see if the system is confusing the classes (i.e. commonly mislabeling one as another).



Confusion matrix: True vs. predicted emotional states

The final results obtained when evaluating model predictions over the 10-folds during cross-validation, was a quite impressive accuracy of 0.987 (+/- 0.01). This means that, based on the set of extracted features from the raw EEG data, we can with close to 99% accuracy predict the mental state of the person as either ‘Neutral’, ‘Negative’ or ‘Positive’.

Summary of results

Although the results from this example case are actually very impressive when it comes to predicting emotional states from EEG recordings, there is still some way to go when it comes to more advanced applications. Also, the limited sample size of only two subjects during the experimental recordings raises the question of generalization to new individuals. Still, as an example case showing very promising results, it represents a good starting point for further investigations.

Outlook

Are we then approaching the dystopic sci-fi future of total mind control, where we are giving up the privacy of even our own thoughts?

The ability of AI to read a person's brain activity raises significant ethical issues around privacy and security that research leaders need to be cognizant of. This technology open up a Pandora's box of new malicious uses, which can e.g. include stealing sensitive information from your brain after manipulating your mind to think about them. An article published in the journal *Nature* describes an imminent future where "it will be possible to decode people's mental processes and directly manipulate the brain mechanisms underlying their intentions, emotions and decisions."

Still, breakthroughs within this field is a further, important step along the pathway of decoding the brain patterns that underlie thinking. Such innovations also mean better brain-computer interfacing, which would open up whole new platforms for man-machine communication. This might allow a paralyzed person to move a robotic arm, or a person with a spinal cord injury to control a motorized wheelchair. Applications such as "smart prostetics", as shown in the below video, could represent a huge step forward for disabled people all over the world.

"Beyond bionics", The Guardian

Although I consider myself being a "tech optimist", I still believe that there is a need for rules and regulations in order to make sure that this technology helps those who need it without disastrous consequences.

What the future holds is anyone's guess, but what is certain is that we'll find more ways to wittingly or unwittingly interact with computers. The promises of AI-powered interfaces are great, but so are the challenges.