

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/335095404>

Using EEG and Machine Learning to perform Lie Detection

Preprint · August 2019

CITATIONS

0

READS

5,146

1 author:



Yohan Kulasinghe

University of Moratuwa

3 PUBLICATIONS 3 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Kinect sensor for skeleton abnormality detection [View project](#)



Software Effort Estimation Model Based on Story Points for Agile Development [View project](#)

Using EEG and Machine Learning to perform Lie Detection

D.H. Yohan Kulasinghe (Undergraduate), Faculty of information technology
University of Moratuwa – Sri Lanka
yohan.16@itfac.mrt.ac.lk

Abstract— The electroencephalography (EEG) is a hot topic in the world. EEG signals read the electrical activity of the brain. Many researches are going on different areas to take advantage from this technology. Studying of EEG provides sound knowledge about how brain signals work in different emotions and activities. Lie detection is a developing technology which uses to act against crime. Traditionally, this is done using linguistic analysis, recognizing facial and body movement, Training observation and Voice stress analysis. The advancement of cognitive science and neuroscience EEG analysis gives a better understanding of brain function. Some machine learning techniques like SVM, k-Means, ANN, Linear Classifier helps to analyses EEG signals. Techniques like Fast Fourier Transform (FFT) uses to reduce the noise of signals which get from EEG. This study is to do literature survey on the above three connecting field. “Using EEG and machine learning to perform lie detection”. This review paper includes, Measuring EEG signals, EEG signal analysis, Feature extraction, how to use EEG in order to detection of lies, Algorithm for classification, Approaches and methods used to analyze EEG signal to perform lie detection, review and conclusion.

Key words – EEG, Lie detection, Machine Learning

I. INTRODUCTION

In the field of security and military there are many attempts to identify lying with high accuracy. Scientist and researches used many techniques for achieve that, Observing behavior, check sound stress, recognizing facial expression, analyzing heart rate, measuring skin conductance, analyzing the rate of breathing are some of the famous techniques among many [1]. But those attempts were pretty much outdated because of the expert liars can fool those previous techniques. With the advancement of technology and cognitive science, it is identified that, when a person doing activities, certain brain parts activated and according to activity [1] brain emits measurable electronic pulse. By analyzing those electronic pulse, it can say which part of the brain is activated at the movement. As a result of years of research done in cognitive science it was found that when people lying, the creativity generation part of the brain is activated. When someone telling a truth, the brain part which is identified to responsible for controlling memory and remembering is activated. On top of the finding about measurable electronic pulses emitting while brain works, a new technology call electroencephalography (EEG) evolved.

EEG is a technique which basically used for examining the brain activity in order to identify certain brain activities [2]. But

the limitless possibilities of EEG have been researching over centuries for now [2]. More applications in virtual simulation such as an object moving in a virtual environment. Controlling and using electronic devices, upload a mind to computer and recently this field extends to mind controlling of robots.

When considering lie detecting, experiment procedure of most researches is they blindfold the subject and ask them to answer the questions ask by researches. They have instructed to tell lies for the answers and collect dataset, which the readings of EEG. As well as the EEG results when they are telling a truth [1].

Machine learning is considered as one of the main branches of AI which provides machine an ability to learn and improve itself from experience and examples without being planly programmed. Process of training begins with observations of data and look for patterns in data to provide better decision [3]. Machine learning algorithms basically categorized as supervised and unsupervised. Supervised machine learning algorithms are applied to labeled datasets and if categories of data already known. In other hand unsupervised machine learning algorithms are used on unlabeled datasets. Other than that, there are two other machine learning methods called semi-supervised machine learning algorithms and reinforcement machine learning algorithms. Semi-supervised algorithm having both supervised and unsupervised characteristics, while reinforcement algorithm deals with actions and discovers on errors or rewards.

Most of the time researches collect data and use more portion of those sample data to train machine learning model to predict and use other half of data to test the application. Quality of the data which is collected using EEG is having noise and it's hard to identify patterns manually inside the dataset. So, machine learning is used to identify pattern and recognized / classify that a person is lying or telling a truth.

II. MEASURING EEG SIGNALS

Electroencephalography (EEG) is one famous topic among new trends in technology, since that carries many possibilities human ever needed. The human brain emits electrical signals called ‘brain signals,’ which can identify with a high-density electroencephalogram machine and sensors attached to the scalp [4]. Basically, EEG is the pattern of the current flows on a scalp

Because of the data which is captured by EEG is a complicated waveform, it is needed good signal processing methods to gain data from EEG. Data collected from EEG doesn't has much use when signals are in its original form. And EEG signals not process in its natural state [8]. Before entering to feature extraction process, it's better to pre-process EEG signals. Fourier Transform analysis on EEG signals allow the separation of signal and study different patterns of EEG, A method call Fast Fourier Transform(FFT) is applied to translate the complex EEG waveforms into a modest waveform [8]. This processing method can produce a waveform which is easier to analyze and distinguish from others. Since Fourier Transform is used to compare complex sign signals that extend through whole time domain. If an alternation occurs in boundaries, the whole Fourier features that cannot be determined with FFT.

These problems can be resolved using Gabor transform, also known as Short-Time Fourier Transform (SFTF) [5]. SFTF slicing the waves into short segments and execution analysis on each of these parts by using standard Fourier transform. These segments can be further simplify using FFT approach.

IV. LIE VS EEG

In cognition mainly alpha band of the EEG signal prominent over other. Alpha wave is prominent and significant in memory and logical related activities. Comparing power, RMS (Root mean Square) and variance of the EEG signal it can proved [7].

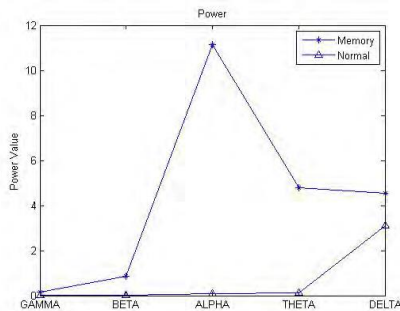


Fig 5 - Power variation of EEG sub bands while resting and thinking [7].

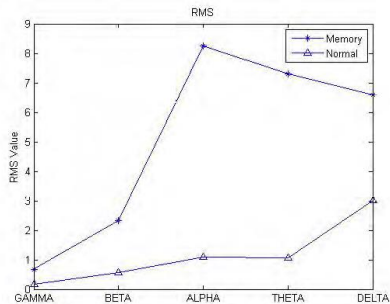


Fig 6 - RMS variation of EEG sub bands while resting and thinking [7].

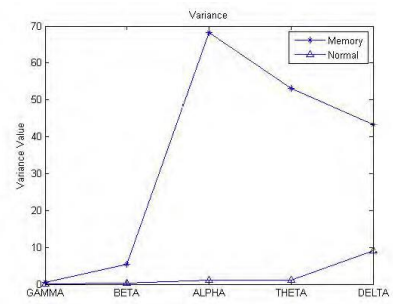


Fig 7 - Variance of EEG sub bands while resting and thinking [7].

And predominantly Frontal pole and Temporal pole is activated during activity like lying. So, when capturing EEG, it is mainly focused on electrode named FP1, FP2, T3 and T5. Moreover, the Alpha sub band of those captured signal form that electrodes are examined.

V. EXTRACTION OF RELEVANT ATTRIBUTE

After pre-processing was done, the third stage is, analysis of preprocessed EEG signals. In order to proceed further, feature extraction needed to be done [9]. Feature extraction means steps involved in getting the information related to feature. These features should describe unique brain activity. Feature should serve as input to the classification stage. In feature extraction, it is using some algorithm for catch dominant frequency or pattern [5]. Assume that a person telling lies, when the question asked, automatically brain will take time to remind the past event, duration of this thinking will be feature for the system. According to some approaches when a person telling a mix of truth and lie, brain needs to restructure, and that process takes some time, that gap of time can take as feature. In the study of EEG signals, signal amplitude, wavelength, frequency, voltage also considered as featuring for the classification model.

VI. CLASSIFICATION ALGORITHM

A. Classifiers

Classifiers is used to differentiate EEG signal [4], either the signal emits while telling a truth or lie. Most of the time above discussed features act as classifiers. Based on the characteristics of classifier, machine learning classification algorithm should use. Some of algorithm will discuss in following section of this paper.

i) Support vector machine (SVM)

It is supervised learning classification technique which considered as produces high accuracy classification [4]. Support Vector Machine algorithm's objective is to find hyperplane in a given dimensional plane according to the data distribution, algorithm work well in distinctly classified data points [10]. There are many possible hyperplanes for two classes of data points, objective of the algorithm is to find the plane has highest margin. The points that are closer to hyperplane determine the

position and orientation of hyperplane and they are the support vectors. These support vectors help to build SVM.

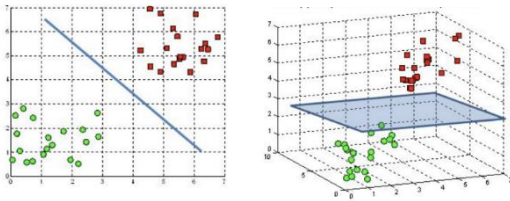


Fig 8 – left_ hyperplane in R^2 is a line, right_ hyperplane in R^3 is a plane [10].

ii) *K-means clustering*

It is Unsupervised learning clustering algorithm used to solve clustering problems [4] in unlabeled data. This considered as the simplest and popular unsupervised algorithm in machine learning. A collection of data points arranged together because of some similarities that makes clusters [11]. Similarities or the base of creating clusters define by machine. Data points are homogeneous according to the selected feature which creates the cluster of data, and outside of the cluster data points are heterogeneous to other clusters. During the clustering, number of iterations may happen. The algorithm starts with k centroid value which either randomly generated or randomly selected from dataset [12] and iterated between following two steps.

- Centroid defines clusters and each datapoint assign to the nearest centroid based on squared distance.
- Centroid recompute by taking mean of all datapoint assigned to that cluster.

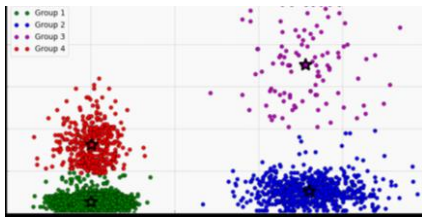


Fig 9 - clustering to centroid using speeding feature (y axis) against distance feature (x axis) [11].

iii) *Linear classifier*

Linear Classification belongs to the category of supervised learning, based on the value of a linear combination or characteristics [4] separation of the data point has been done. This algorithm also called as linear regression. If data points are having a linear connection or can be easily predict according to linear function, this algorithm works well in that type of situations. These combination of characteristics or features known as objects’ characteristics and typically presented in vector called feature vector in coding.

iv) *XGBoost*

Extreme Gradient Boosting is recently widely used algorithm in applied machine learning. XGBoost is very powerful model especially on large dataset. XGBoost is an implementation of gradient boosted decision trees for speed and performance [13].It supports Command Line Interface, C++, R, Python, Julia and Java. It mainly used for the regression, Classification and ranking process.

v) *Artificial neural network (ANN)*

ANN is a method of supervised learning. Based on the information pass through ANN changes its internal structure [14]. ANN internal structure contains connection between data points, and it is highly adaptive. classification is done based on the weight associates with the neural connections [4]. ANN can perform task by considering given example data. Basically, ANN consist with 3 type of layers such as input layer, output layer and hidden layer between input and output layer. when neural networks become more advance, number of hidden layers getting high. Neural network should train against sample data to adjust weight on connection to provide better classification.

VII. LIE DETECTING APPROACHES USING EEG

A. *Approach one*

Method

Brain signal emits when human lying is also essential for this approach. The waveforms are recorded after the FFT process during the training period. Then, the detection of the lie is carried out by comparing the detected waveform to the recorded waveform. As the similarity of the waveforms is high, it indicates the human is lying. Therefore, the result shows that this method can detect the lie. Telling a truth and then telling a lie takes 40 to 60 milliseconds because the brain must shift its data assembly strategies. Psychologists employed on the technology and believe it has 86 per cent accuracy [9].

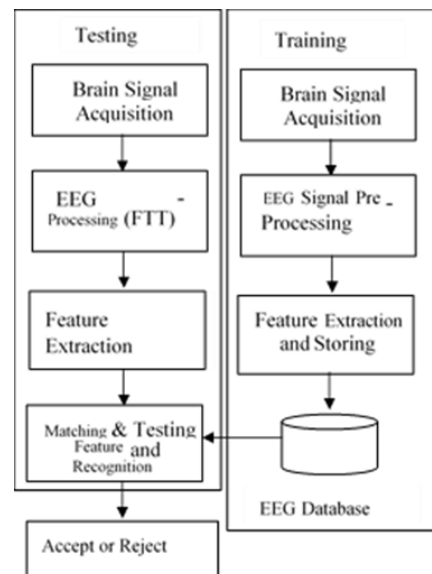


Fig 10 - Block Diagram for EEG Signal Comparison [9].

Problems

- Because of the signals are very noisy, accurate simplicity is need more than FFT.
- Brain signals strength and pattern changes according to mental status. So, it difficult to find the correct pattern.

Proposed Solution

- Using FFT and EEG classification / comparison discuss in following diagram.

B. Approach two

Method

Brain signals emit from various areas of brain. Catch the brain signal and determine active parts of the brain, and active pattern of EEG. Then simplify using FFT and get clean signal pattern form brain. When human telling lies the brain part responsible for creativity is activated, when human telling a truth the brain part responsible for the remembering is activated. However, the activation of the brain part is complex to identify due to emotions and cognitive status. So, it is needed to create a model to match. Collect data when people telling lies and create a model by compeering actual data with the model can give conclusion [4].

Problems

- Telling truth but mixed with lie is nearly impossible for detect, since both the brain areas are active.
- Tell a lie by remembering the past lie, can be miss identified.

Proposed Solutions

- Make a model of brain activity. And a way to filter emotions, then get classified result with more accuracy.

C. Approach Three

Method

Using untrained male between age 20 and 27 were involved in this experiment. They divided in to two groups, innocent and lying. collect signal analysis data using EEG ElectroCap [15]. Pre-processed signal using Independent Component Analysis (ICA) and feature extraction applied minimum, maximum, mode, median and mean amplitudes. For classification, SVM algorithm is used. This model used Matlab base program to detect a lie. Results are display using program [15].

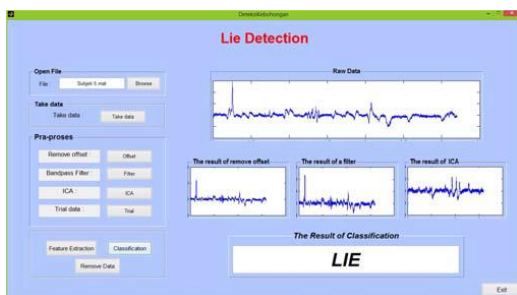


Fig 11 - Lie detector program [15].

This model reaches 70.83% accuracy and well known as P 300 project.

Problems

- Only use males and accuracy is 70%. In a field like lie detecting this accuracy is not enough.

Proposed solutions

- Find more sophisticated algorithms and more training with large dataset.

VIII. DISCUSSION

Traditionally lie detection was done by verbal as well as nonverbal methods which can describe psychologically, sociologically and behaviorally. There are careers in lie detection too, but trained people can pretend they are telling a truth even they are telling lies. Identifying lie affects in security and military areas. To get the advantage of knowing truth, most researches have been conducting continuously, and it is needed. There are various systems already created using many technologies such as observing behavior, checking sound stress, recognizing facial expression, checking heart rate, measuring skin conductance, rate of breathing, and those systems are continuously evolving. With the advancement of technology and cognitive science it was identified that when a person's brain involving activities certain brain parts activated according to activity and thinking, and the brain emits measurable electronic pulses. EEG is the study of that brain signals and those signal call EEG signal. Lots of studies carried out in the field of lie detection using EEG. But EEG signals contain too much noise, even an eye blink can create mess in all signals. Then it is needed to have proper signal analyzing methods. FTT, SFTF, ICA like analyzing techniques used for that. But only signal analysis is not enough to identify deception detection because it is a complex brain activity.

Some approaches to detect lie using EEG as well as Machine learning reviewed in this paper. Those approaches based on identifying, mainly classifying truth and lie. Most approaches asking questions from subjects. Blindfolded the subject before taking data, because eye blinking and emotion towards the person asking question can affect data. In that sense blindfolding is ok, but no studies done regarding voice. If the person asking question male or female, does it affect to EEG signal, those things need to consider in further studies. For analyzing signal most approaches use Fourier Series. Fourier Series simplify the signal in frequency domain. Experiments should continue signal pre-processing to identify more suitable techniques. Because still the features taken from EEG signals in experiment level, it only approach accuracy of 86% according to the most suitable feature to identify lies, the pre-processing technique should change. Need to do more experiment on feature extraction. According to the present situation, time deference between lie and truth taken as feature got the most accurate results. Machine learning classification algorithm take considerable weight, when talking accuracy. According to the study SVM is the most used

algorithm. Here in lie detection we use algorithm to differentiate truth and lie SVM works well on distinct datapoint. But signal generations when lying cannot be considered as distinct data. Then it is better to use K-Means Clustering to get more realistic results. Further studies need to do in order to prove this theory.

Comparison of basic method used		
		Accuracy
Approch one	Measuring the time gap between consecutive truth and lie.	86%
Approch Two	Evaluating the active part of the brain to determine weather the person telling a truth or lie	-
Approch Three	Create model with pre collected data using MATLAB base program and used SVM algorithm for classifying.	70%

Table 1 - Comparison between literature surveys

IX. CONCLUSION

Using describe approaches, lie detection systems implemented and as stated in ‘approach one’ it reaches 86% accuracy too. Using EEG and Machine Learning approach to lie detection mainly consist 4 steps according to the findings of this review paper. Those are 1. Collecting EEG, 2. Preprocessing EEG using Fourier Transform approach, 3. Feature extraction, and finally 4. Classify using ML. Further identified three ways of increasing accuracy of the results those are 1. collecting clean data for the training set without mixing with other emotions and actions, 2. Use good signal processing techniques, 3. Apply better and most suitable ML classification algorithm to differentiate truth or lie signal. As a conclusion for develop more sophisticated and reliable lie detecting system using EEG and Machine learning, it is needed to do more experiment in signal preprocessing and need to find the most suitable classification algorithm.

ACKNOLOGEMENT

Author must specially thank for Dr. Lochandaka Ranathunga for guidance and supervision on creating this paper and helping to understand the domain area. Author would specially thank all the lectures of the faculty of information technology, University of Moratuwa for their assistance for the system.

REFERENCES

- [1] J. Marsman, "EEG and Machine Learning to Perform Lie Detection," Channel 9, New Zealand , 2016.
- [2] Wikipedia, "Electroencephalography," [Online]. Available: <https://en.wikipedia.org/wiki/Electroencephalography>.
- [3] E. System, "What is Machine Learning? A definition," [Online]. Available: <https://www.expertsystem.com/machine-learning-definition/>.
- [4] S.Dhivya,A.Nithya, "A REVIEW ON MACHINE LEARNING ALGORITHM FOR EEG SIGNAL," in 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018), 2018.
- [5] V. Podgorelec, "Analyzing EEG Signals with Machine Learning for Diagnosing Alzheimer's Disease," ResearchGate, Maribor, 2012.
- [6] T. M. P. V. Lab, "Biomedical Signals Acquisition," [Online]. Available: http://www.medicine.mcgill.ca/physio/vlab/biomed_signals/EEG_n.htm.
- [7] P. Bhuvanewari, J. Satheesh Kumar, "A Note on Methods Used for Deception Analysis and Influence of Thinking Stimulus in Deception Detection," International Journal of Engineering and Technology (IJET).
- [8] Sim Kok Swee, Lim Zheng You, "Fast Fourier Analysis and EEG Classification Brainwave Controlled Wheelchair," in 2nd International Conference on Control Science and Systems Engineering, 2016.
- [9] Ganesh Janvale, Aaron M. Oirere, Vishal Waghmare, Ratnadeep Deshmukh, "Designing of new proposed technique for lie Detection using EEG Signals," in ResearchGate, 2011.
- [10] R. Gandhi, "Support Vector Machine—Introduction to Machine Learning Algorithms," Medium, 2018.
- [11] Dr. Michael J. Garbade, "Understanding K-means Clustering in Machine Learning," Medium, 2018.
- [12] A. Trevino, "Introduction to K-means Clustering," datascience.com, 2016.
- [13] J. Brownlee, "A Gentle Introduction to XGBoost for Applied Machine Learning," Machine Learning Mastery, 2016.
- [14] Rajni Bala Dr. Dharmender Kumar, "Classification Using ANN: A Review," International Journal of Computational Intelligence Research.
- [15] Artha Ivonita Simbolon, Arjon Turnip,Jeperson Hutahaean, Yessica Siagian, Novica Irawati, "An Experiment of Lie Detection based EEG-P300," in International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology, Indonesia., 2015.
- [16] S. E. Fienberg, "The Polygraph and Lie Detection," The National Academies' National Research Council, 2003.
- [17] S. McDonnell, "Chron," Hearst Newspapers, LLC , Careers in Lie Detection, [Online]. Available: <https://work.chron.com/careers-lie-detection-17143.html>.
- [18] J. N. M.A., "Psychology Today," The Truth About Lie Detection, 12 March 2012. [Online]. Available: <https://www.psychologytoday.com/us/blog/spycatcher/201203/the-truth-about-lie-detection>.