UNIVERSITY OF TARTU

Institute of Computer Science
Computer Science Curriculum

Jevgeni Savostkin

Improving accuracy of brain-computer interface with multiple trials

Master's Thesis (30 ECTS)

Supervisor: Ilya Kuzovkin, MSc
Supervisor: Raul Vicente, PhD

Tartu 2017

Improving accuracy of brain-computer interface with multiple trials

Abstract:

Brain-computer interface (BCI) is a computer system for extracting brain electronic neural signals and using them to control computer applications. Besides measuring, BCI converts raw signals to digital data and maps the data to the exact computer commands. Unfortunately, the probability of the right command prediction usually is below 100% and therefore it could be improved.

This is a problem for BCI systems, since they will not be widely trusted and used, while the prediction accuracy is relatively low. There are many existing solutions, which provide increase of the prediction accuracy, mainly based on trying out different classification techniques and algorithms. Existing methods focus mostly on a training part of the system. Our approach is to try to improve the accuracy upto 99%, while experimenting with test part algorithms, and to determine how many trials are required during the test mode to reach the desired provided accuracy.

The solution described in the thesis is based on Condorcet's jury theorem. It means that if we have single events with probability of more than 50%, then by combining them, the total probability for these events would rise. This work shows the actual results (prediction accuracies) and provides their comparison using the voting techniques based on the Condorcet's jury theorem. It also describes the dependency of the number of test mode measurements and accuracies.

The BCI technology is relatively young direction. In order to fully integrate it into our ordinary life, the contribution from the scientist and engineers is required for composing and choosing the most reliable system with the components. The following work represents a contribution to the brain-computer interface field.

Keywords: Brain-computer interface, Random Forests

Contents

1 Introduction

Motivation

Many health issues can disrupt the neuromuscular channels, which brain uses to communicate with different parts of organism. Channels are used to control muscles and pass the feelings. With these controls, a human can successfully participate in an ordinary life, controlling the surrounded environment. Amyotrophic lateral sclerosis (ALS), brainstem stroke, brain or spinal cord injury, cerebral palsy, muscular dystrophies, multiple sclerosis and other diseases cause problems in neural channels or the muscle control performance. There are three ways how to restore muscle disabilities. The first is increasing capability of the existing neural channels. That means using existing well-functioning muscles to fulfill suffering ones (e.g use of hand movements to produce artificial speech). The second is the use of control signal measurement systems (electromyography) in order to record signals sent to muscles, translate them and repeat the action in a prosthesis. And the latter is attaching a non-muscular communication module as a control channel to a brain, which is BCI.[1]

Unfortunately, BCI systems have relatively low prediction accuracy, which nowadays makes their implementations too "raw". There many ways how BCI data is being handled in order to gain better results, but none of them are perfectly accurate.

Scope

This work consists of creating a BCI application, which will communicate with Emotiv EPOC headset and try to predict user distinct thoughts about the targets provided by the application. The application will work in two modes. The first, learning mode, for obtaining test data samples to "teach" an algorithm and give better output results in future. This is necessary step to perform for every new user (subject). The second, testing mode, is for checking the accuracy of prediction by the algorithm. When the predictions based on a single measurements will give a relatively high accuracy, a multi-measurement session is going to be run. A multi-measurement mode considers taking into account several classification results made in a row and choose the most suitable with voting algorithms. Expected to get higher results with increase of the number of classification samples participating in voting. The results from single and multi-trial sessions will be recorded and compared to the expected. The application will be available to work in offline (classification will be done after the dataset is recorded) and online modes (classification will be done instantly after sensors data is recorded).

Research problem

Multi-measurement could bring better prediction result than single. The objective is to determine how multi-measurement mode's accuracies differ from single-measurement's and how the number of classification samples would improve the statistics. The results will be compared to the ones theoretically calculated.

Contribution

Implement an application to get measurements first of all from a single trial. Select and teach the system to process the results. Calculate the theoretical prediction accuracy for multi-trial sessions based on a single try session results. Run of multi-trial sessions with validating results and compare them to the theoretical values.

Structure
Current thesis structure is as follows:

- Background / State of the Art - significant technologies used in current work are described along with comparison to similar projects

- Contribution - detailed overview of the idea, list of finished work, grouped by the domains of the system and followed with meaningful details

- Experimental results - test sessions with output results described

- Discussion - difference with the theoretical model and limitations of the system are provided

- Conclusion - sum up of the goal, expected and actual results with brief explanations and future adaptation prepositions

2 Background / State of the Art

2.1 Condorcet's jury theorem

The fundamental theorem for the current work is Condorcet's jury theorem. The rule states that in majority vote with two options available (*Option*1, *Option*2), if voters (jurors) have independent probability p for voting for *Option*1, then:

- If p is higher than 50%, then the more voters is participating, the higher probability for the majority decision for *Option*1 will be

- If p is lower than 50%, then the more voters is participating, the lower probability for the majority decision for *Option*1 will be

The Condorcet's jury theorem is defined by the following expression:

$$\mu = \sum_{i=m}^{N} (\frac{N!}{(N-i)!i!})(p)^i(1-p)^{N-i}$$

where N is the number of jurors, p is the probability of the individual juror giving the true result, μ is the probability that a jury gives the true result. Based on the formula, the higher initial probability (single juror's probability) is, the less jurors it is required to reach the absolute probability.

Our main idea's aspect is to produce multiple so-called jurors to be used in a voting system to increase their common probability. Their probability, in our case, means the increase in quality of the BCI system performance. In a role of a single juror will be a single prediction trial (i.e concentration attempt on a target) with it's accuracy. Accordingly, multiple trial approach considers with use of several prediction trials and their accuracies, what on the other hand, will require more time on a whole prediction task.
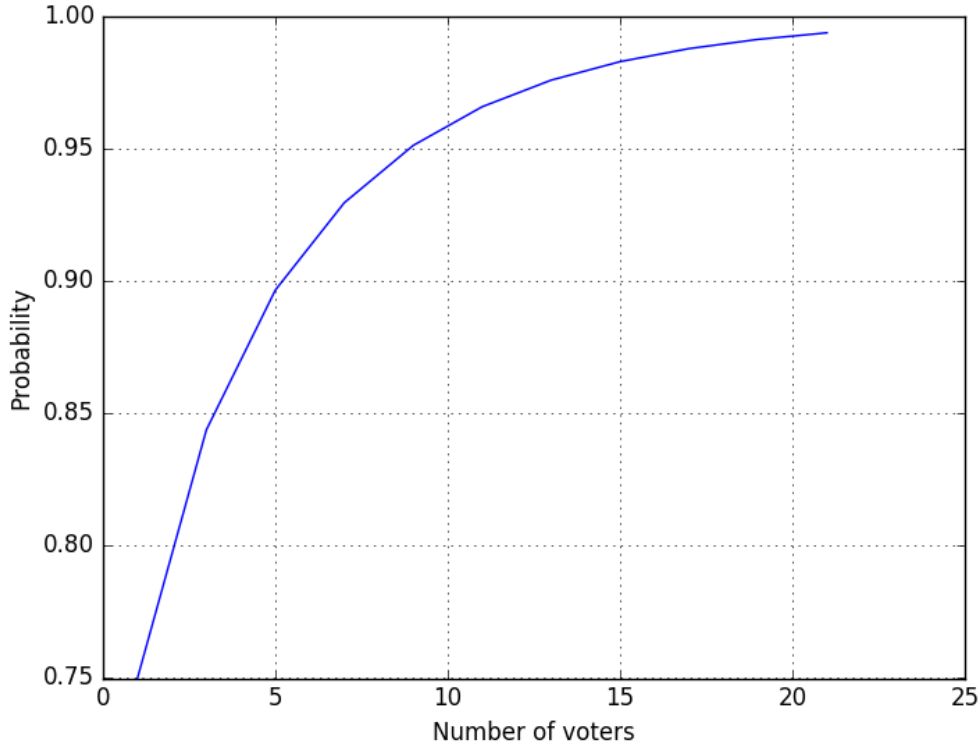
Figure 1: Plot showing the dependency between number of voters and probability result when a single voter probability is 75%

Figure 2 shows that with a single juror's probability of 75% it is required approximately 21 jurors to reach the near maximum probability.

2.2 Brain-computer interface

Brain-computer interface (BCI) is an interface that does not require muscle control from user to communicate with a device. It requires user to think about the distinct target. The interface records electroencephalographic (EEG) signals from the scalp surface which represents our brain activity. These signals have low amplitude (usually measured in microvolts), whereas frequencies above 30 Hz have especially low values which tends to zero.[2]

The signals could be translated into a control commands for the certain devices, what is especially useful for the people suffering from lock-in (e.g. Brainstem stroke, severe polyneuropathy) or muscle control diseases. BCI systems could give

such people a possibility to control the environment, perform word processing or even operate a neuroprosthetics or orthosis. There are two types of BCI available: one way and two way. In case of one way type, a computer is accepting signals from the measuring device. Two way system deals with exchange of information between both sides.[3]

BCI system structure could be divided into the four modules[3]:

1. Source Module This module digitizes and saves signals coming from brain without handling them. This component knows how to obtain data from the sensors and store them to the specifically formatted file. These data samples are usually mapped to the sensor names (every sensor located on the scalp has it's own name determined by location) and event classificator.

2. Signal Processing Module This module is responsible for conversion of raw data signals into something more meaningful for controlled machine or the commands. Conversion is divided in two stages: feature extraction and feature translation. The extraction considers receiving data from the source module and preparing them for translation module, which means obtaining the signal properties, like frequency domain values for the given sensors. The feature translation is an algorithm, which determines the identity of the control signal sent with a given signal data.

3. User Application Module: In addition to signal processing module, an application module takes the control signal to perform operations in the application. Usually the application has its own graphical interface, which allows the user to select and think about some sort of targets, like letters, images, icons or directions. The user could also give his feedback about the prediction validity through the application. The feedback could be given orally or tactilely.

4. Operator Module It is a module, which defines system constants and parameters, like learning mode length, targets or any kind of signal processing variables. In addition, this module could plot the information on graphics without knowing of the input data nature. This allows a user to see real-time feedback about happening events.

BCI use is a skill, which requires practicing. An algorithm, which translates the signal features to the control signal, should "learn" to output with more accuracy. Learning is performed, based on the feedback provided by the user. That means, the user should participate in the algorithm teaching for many sessions. In addition, during the sessions, the user should try to think in the same way

he normally does. Otherwise, it might leave a negative impact, when the subject experiences some sort of distractions. These exercises require concentration and it takes time to get used to it.

2.3 Emotiv EPOC

In order to obtain raw signal data from the scalp, we use Emotiv EPOC BCI headset. It is a multi-channel wireless (communicates using Bluetooth) headset with 14 channels (sensors) for the following International locations: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4. The device converts an analog signal to a digital with a 14 bits resolution and 128 Hz sampling rate. The bandwith is 0.2 - 45 Hz. This suite can monitor the user's emotional state in real time.[4]

It comes with the out of the box software Control Panel and TestBench, which visualize the features based on the signal. The Control Panel software outputs recognized emotional states, facial expressions and mental commands. With Test-Bench it is possible to see a raw or EEG signal regarding distinct channels. In addition it provides a signal quality for the sensors and the connection status between the headset and Bluetooth receiver.

2.4 Short Time Fourier Transform

In a features extraction level a raw complex signal wave should be decomposed to the subwaves. Subwaves helps to construct a frequency domain representation of the complex signal where frequencies and respective amplitudes are mapped together. Although in terms of spectral analysis Fourier Transform is dominating, in case of nonstationary signals where EEG signal belongs better to use short time Fourier transform (STFT).[5]

In STFT, a signal is split on datasets (frames) with N samples, where N represents a window length. This frames overlap with 50% between each other. Before the Fourier transform a Hanning window is applied to reduce aliasing of the signal. Finally, after Fourier transform within STFT the result is outputted.

2.5 Classification task

Despite the fact that relationships between some brainwaves and subject's mental states has been established, these mental states are too common and non-descriptive. For example they can help to distinguish if the subject is relaxing or concentrating on something. For tracking more specific mental states like concentrating on a distinct target, a unique model should be applied for each subject. Features got from a BCI signal should be divided on groups defined by target type.

The other similarities (patterns) within the groups should be found and used to predict new input data. To define similarities a classification should be executed.

Classification is the task of learning a target function f that maps each attribute set x to one of the predefined class labels y[6]. A model received as a result of a classification could help distinguish between different target classes. Attribute set x(also known as features or key characteristics) could contain continuous (e.g real numbers) as well as discrete values (e.g labels).

2.6   Random Forests algorithm

As described above, the goal of translation phase is to understand what control signal has been described with signal features received from the extraction phase. That means we should classify our data samples, where the classes would be a set of targets a user should deal with. For generating a classification model a machine learning algorithm should be applied.

A machine learning algorithm is a data-driven algorithm, that predicts in which data group a new input value belongs (classification) or which continuous output the input data maps (regression). These decisions are made based on the existing data samples which are grouped by some property. There exist two major learning types of the algorithms[7]:

- Supervised - creates a model with labeled (classified) input data samples, so that groups of data have own class label

- Unsupervised - the algorithm does not know anything about the data as well as the classes

Unsupervised algorithm is a good way to analyse the data without knowing how to use it and on which potential groups it could be split. However, in our case we know that we should split data according to targets and thus, chose a supervised learning machine algorithm.

We will use Random Forest as a machine algorithm which is the one of the most precise for the work with EEG data, according to the [8]. It shows better classification accuracy and performance than the other modern algorithms. It's accuracy dominates over the other algorithms in case of parameter optimization, which we plan to do as well.

A Random Forests is an adaptation of Decision Tree algorithm developed by Leo Breiman and Adele Cutler, where instead of using single tree, a bunch of trees are used. Every one of these trees is generated by using randomly selected subsets of the existing data samples. Finally, each of every tree is handled separately to find out it's predicted class and with a majority vote from every tree a final result is defined. Random Forests model does not overfit. Using the right kind

of randomness brings accuracy in solving classification and regression problems, however regression problems have lower accuracy.[9]

2.7 Collaborative Brain-Computer Interface

Yijun Wang *et.al* describes in [10] a technique which has similar approach to this work. The main idea of their work to use collaborative EEG input data for predictions. They made a decision-making experiment using multiple users (subjects) thinking about the same targets simultaneously. Subjects must make Go (target) or NoGo (non-target) decisions in scope of their application.

The application shows them images where are animal images (target) and others (non-target). With a period of 20ms one of the shown images is flashing and the subjects must make a decision about does it belong to a target group or not. The decision is done by pressing or releasing a button, thus a motor inhibition process was invoked (movement-related signal).

Every subject had to train the algorithm and test it in a single user mode. Subjects managed to reach mean classification accuracy of 75.8% with using mean response time (reaction time) 377 ms. Already this pointed on reliable prediction with single-trial usage. A collaborative classification was tried considering 5,10,15 subjects simultaneously which resulted 91.4%, 97.6% and 99.1% accuracy respectively. This clearly shows improvement over the single-trial classification.

In case of multi-user approach a weighted voting system was used, where a subject with a better prediction statistics got more weight and influenced the output result more in the future classifications. This is the technique what our solution will use along with multi-trial classification. Our goal is to use multiple trials of a single user instead of single trial of several users as the related work tends to do. Our way is to use optimized Random Forests classification algorithm which according to [8] could bring the more precise result, than support vector machine (SVM) algorithm which is used in animal classifications. Finally, a single user approach has wider fields of use and less complex in implementation as a multi-user technique.

Maybe to find some articles about ML results enhancing using Condorcet's ?

3   Contribution

3.1   Overview

The main idea of the given work is to show the effectiveness of a multiple trial approach in a BCI perspective. A regular way of prediction considers a single prediction trial for a single prediction task. If any noise has been recorded due to technical issues or a subject was distracted and could not properly concentrate on a target during the trial, then it will directly reflect on a classification result and dramatically on the accuracy. However, if we know that our system in the most cases works properly and gives the right answers, then we could compensate appearing of noisy and faulty predictions with the others with a better quality.

In multiple trial approach the final prediction result is calculated using intermediate prediction results which every trial gives. Multiple trial way takes more time to give the answer than a single, however the accuracy should be improved in case when single trial accuracy gives more than 50%. Our goal is to understand how many trials it is required to get the accuracy 99% in theory and check it against the reality.
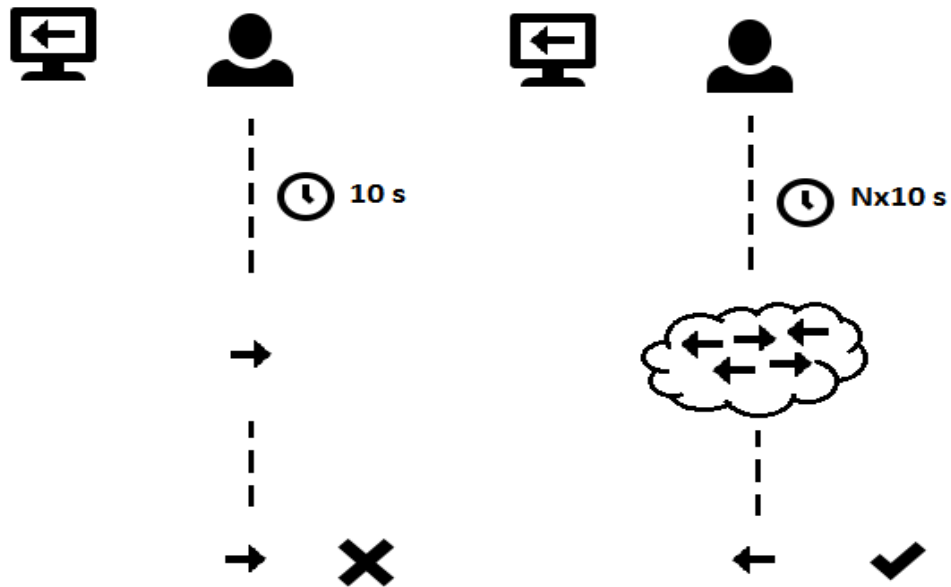

Figure 2: A main concept of multiple trial approach. If single trial duration is 10s and it brings only one and the final result, then applying N trials will give several intermediate results where from the final will be selected

3.2 Experimental design

To provide an experimental platform for the research question a special system was designed and implemented along with the experimental flow. The general plan is to use a two-class (two-target) BCI system to collect the data necessary for the analysis. The flow for the experiment could be described as follows:

1. Populate the training data. This includes a lot of 10-second sessions for measurements of the brainwave signals with extracting their features and storing them to the system. Our purpose is to get as much as samples as we can for the reasonable amount of time. It is crucial to switch between targets in order to get more balanced training dataset.

2. Determine the baseline accuracy. That means performing machine learning techniques on the given training data. It involves training of a classifier and it's validation. In addition to that, applying voting systems with the main goal - improve the accuracy of prediction. Accordingly to Condorcet's jury theorem it is wise to find the method with the highest accuracy and use it as a baseline, since this will reduce the number of samples required further. Upon the best baseline accuracy will be found it will be calculated how many trials it will be required to get the desired accuracy.

3. Populate the test data. Similarly to the first point it is required to measure a lot of brainwave signals and populate them into persistence storage. However, at that point the length of the session will depend on the one discovered using Condorcet's jury theorem. Likewise in the training data collecting, it is important to keep the set balanced. Needs to be noted that the actual target labels will be stored to the dataset as well to be used in the followed accuracy calculation.

4. Analyse the multiple trial test data. In this stage the accuracy of multiple trial approach will be determined. The given results will observed and compared to the baseline. The dependency between the accuracy and different number of voters will be plotted for the better overview.

Before each training or testing session a signal quality check is performed in order to ensure that the headset sensors are well placed on a head and the signal quality is high. This check is done using Emotiv Epoc Control Panel software. The training data is collected multiple times within several days which means that there is no guarantee that headset sensor locations were all the time on the same place what is more close to realistic scenario.

Random Forest classifier is set up to use 100 trees and the random seed is manually specified in order to avoid different results within the different classification

session runs. The classifier instance is set to return the probabilities of targets instead of the predicted targets. Processing continuous probabilities instead of discrete targets gives flexibility in voting systems applied in the current work.

3.3   Targets

Every prediction task is an activity considered with a subject's concentration on a target. A target could be a tangible physical object or a picture. In that case concentration on the target will most probably mean staring it. But more common is that a concentration task (i.e mental activity) is done without having any visible objects, but rather by thinking or imaging the given target.

For the targets in the current system two classes were selected:

- Relax - the mental task considering a subject to focus sight on a fixation cross and trying to meditate.

- Excitement - the high performance mental task considering a subject to try focusing on a fixation cross and calculate the mathematical expressions displayed by the application.

In the current context there are used graphical objects for the supporting reasons. Fixation cross is used in order to avoid target's eyes movement, what could cause undesirable brainwaves. Mathematical expressions are generated and shown as a high concentration task which needs to be solved with the subject's brain. The example of graphics seen by the target during the measurement session is as follows:
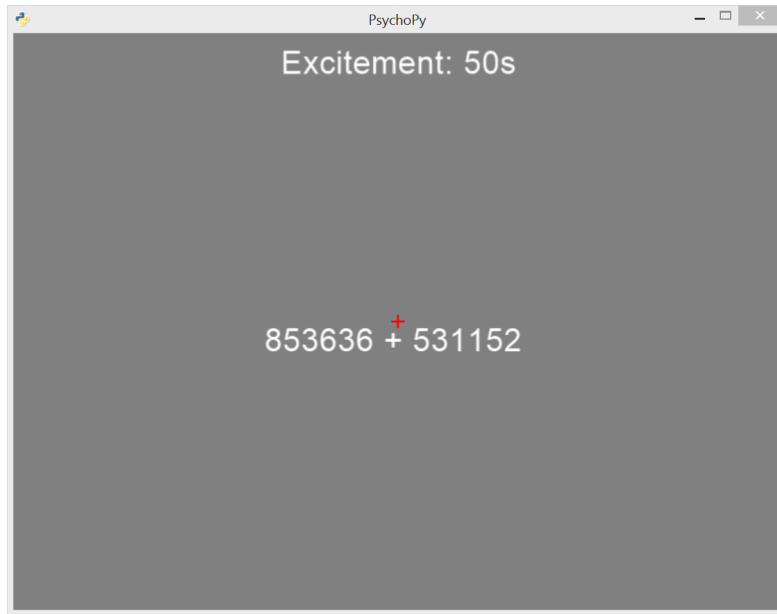
Figure 3: A dialog window displaying what a subject sees during a test for both target tasks. A red fixation cross is located at the center of the window. Mathematical expressions are printed as an Excitement task.

At the very beginning of the experiments when it was not really clear what targets produce better classification results the motor activity signals were used as targets:

- Left - the motor task considering imaging movement of the left hand.

- Right - the motor task considering imaging movement of the right hand.

After some constant unsuccessful (near to random) results using the motor tasks it was decided that mental task could be divided on a high and low performance types. Hence, *Excitement* and *Relax* targets were selected and showed better statistics.

With the several measurements and validations it was determined that 10 seconds is a suitable length for a measurement session. Taking into account that the short-time Fourier transformation extends the number of the input samples (since it uses sliding window), the single measurement session is containing 18 signal data samples.

3.4 Methods

During the baseline establishment phase a classifier was trained and validated using 5-fold cross-validation, which means that a classifier has been trained and validated 5 times with a different validation sets. Every prediction answer from a classifier within 5 trials was recorded into one result set. So in total the prediction result set is containing the same number of samples as the whole training data set has.

For getting the accuracies there are used different methods for the calculation. As described before, different voting types are used within the methods to determine the intermediate answers (the answers from measurement sessions) before definition of the final average. The voting types used in the subsequent methods explained:

- Majority voting - all the voters have the same influence on the final result.

- Voting with a threshold - voting result depends on whether the score returned for the target exceeds the minimum threshold value. In our project the score is the number of occurrences for the class in the defined set. It will be defined for a target which usually appears more times in a set and consequently has higher accuracy. The threshold will be chosen dynamically, trying out available values and measuring their results.

The methods used to calculate the final accuracies and validate the final ensemble results could be divided on two groups:

- Sample based (voting-free) - in this case the accuracy is found with the number of correct predictions divided on the the total amount of the prediction attempts. This is applied on the whole training dataset, so neither grouping nor finding intermediate results is done before.

- Measurement session based (voting is applied) - in this case the accuracy is found with the number of correct intermediate predictions divided on the the total amount of the intermediate prediction attempts. The intermediate predictions are the voting results from single measurement sessions (in our case the result from 18 samples). There are different ways used in order to determine those answers, which will be covered in the next subsections.

In general we use the probabilities returned with a RF (Random Forests) classifier to choose the right class. In our binary classifier case, the standard case is that if the probability for a given class exceeds 50%, then it will be considered as an answer. On the other hand, we have a different method, which labels the targets using the probability threshold for a highly recognizable class. The desired

threshold is calculated iterating through all the possible threshold values from 0%
to 100% (with a step of 1%) and comparing the prediction accuracies, to select
the best at the end.

In summary we will apply the following 5 methods to determine the accuracies:

1. Voting-free

2. Voting-free using RF probability threshold

3. Majority voting

4. Majority voting using RF probability threshold

5. Voting with a threshold

3.5 Implementation

For fulfilling goals in the related topic a custom application was written using
Python 2.7 programming language and open-source scientific libraries SciPy[12]
and NumPy[13]. The application contains the following components:

- Raw data reader - connects to a BCI headset over Bluetooth and reads raw
  data

- Preprocessor - does Short-time Fourier transform of a bulk raw data and
  determines signal amplitudes for desired frequencies (features)

- Data storer - stores preprocessed data to a file system (e.g CSV file)

- Classificator - a machine learning algorithm which teaches classificator and
  uses it to predict preprocessed a target based on features

- Voting handler - in case of multi-trial sessions it is used to select the single
  answer among the given set

- Main application - provides a command-line user interface to select applica-
  tion modes and run necessary components

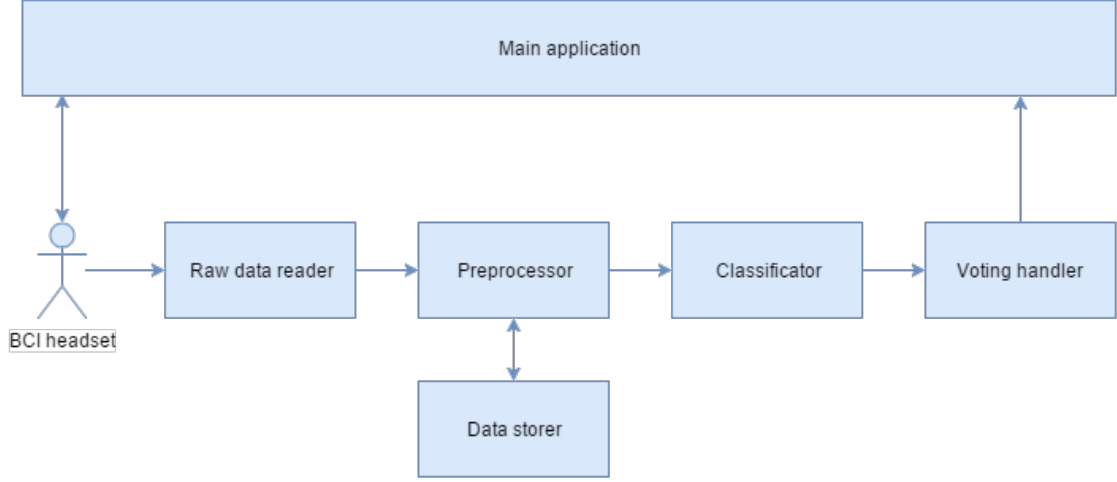Some of the major components are described in the following subsections.

Figure 4: Application structure model

3.5.1 Raw data reader

This component is required to poll the Emotiv EPOC headset and store sent data to a temporary buffer. The communication between computer and headset is established over a Bluetooth Smart network through a USB dongle provided from the manufacturer. Emokit[11] open-source library with minor changes is used to read the data coming to USB dongle from a headset, decrypt and encode it. Polling is done in an infinite loop. Taking into account sampling frequency of the device - 128 Hz, every 8 ms a new signal is received from the headset and stored in a queue.

3.5.2 Preprocessor

During preprocessing a set of raw samples is converted to a frequency domain representation. To do this the Short Time Fourier Transform is used with a sliding window. Sliding window technique allows to capture more precisely non-constant signals. Sliding window length equals to 1 s with 0.5 s overlapping. These factors increase the number of samples to $2 * N - 2$ where N is the original number of samples. For example a BCI session with a duration of 1 minute will output 119 samples.

According to Emotiv EPOC specifications[4] 45 Hz is the maximum bandwidth value for the signal. In our implementation we use 1 - 45 Hz for the frequency range.

3.5.3 Classificator

Classificator component is responsible for determination of a target (class) from the number of samples given by preprocessor component. Each sample (record) contains amplitudes for all the sensors and different frequencies which makes $14 * 45 = 630$ features. In addition to them we add a timestamp and the actual target label to each record. The common timestamp is assigned to the samples coming from one measurement session. This records compose a dataset which is fed up to a classifier (machine learning algorithm) to train it.

A 5-fold cross-validation is done in the given component. Technically this means selecting proportional validation data chunks (with a size of 20% of the training dataset) from the whole dataset without overlapping and using the rest of the samples as training data. This will result in training several classifiers. These classifiers prediction outputs will be concatenated and used further as a training data prediction.

3.5.4 Application modes

Describe here the application modes?

4 Experimental results

The experimental results for the single and multiple trial approaches are described in the followed subsections. The result subsections are split by the techniques of measuring the final accuracies described in section 3.4. Single trial statistics is based on the training data, while the multiple trial approach results are adjusted with the test dataset.

4.1 Single trial approach

4.1.1 Voting-free

The average prediction accuracy for a testing data is determined as **67.1%**. Based on the prediction results the following confusion matrix was constructed:

Table 1: Confusion matrix showing classifiers prediction performance used in cross-validation without vo

	Predicted Relax	Predicted Excitement	Total
Actually Relax	597	570	1167
Actually Excitement	263	1105	1368
Total	860	1675	2535

This table shows that *Excitement* is being better recognized and has 0.24 error rate, meanwhile the target *Relax* has almost 50% accuracy.

4.1.2 Voting-free using RF probability threshold

This technique is used in order to balance the prediction accuracies between the targets. Threshold or the minimum probability for the most frequent target is set up as a border between the two classes. Threshold is compared to the labels probabilities returned by Random Forests algorithm, which are calculated inside the algorithm implementation using ensemble voting among several decision trees.

Comparing the prediction accuracies using different thresholds and selecting the best combination gave the probability threshold 0.54 for the *Excitement* target. Below are the prediction statistics regarding usage of it. The prediction accuracy is **68.4%**.

Table 2: Confusion matrix showing classifiers prediction performance using voting-free strategy and 0.54 threshold for the Excitement class classification

	Predicted Relax	Predicted Excitement	Total
Actually Relax	754	413	1167
Actually Excitement	387	981	1368
Total	1141	1394	2535

4.1.3 Majority voting

In the following technique a majority voting was applied among 18 prediction samples, to get the intermediate results. After finding single voter answers, their average accuracy was determined as **78.5%**.

Table 3: Confusion matrix showing classifiers prediction performance applying a majority voting within measurement sessions data

	Predicted Relax	Predicted Excitement	Total
Actually Relax	37	27	64
Actually Excitement	3	73	76
Total	40	100	140

4.1.4 Majority voting using RF probability threshold

Using the same threshold as in voting-free strategy: 0.54 for the *Excitement* target. The prediction accuracy is **83.6%**.

Table 4: Confusion matrix showing classifiers prediction performance using 0.54 threshold for the Excitement class classification with applying a majority voting within measurement sessions data

	Predicted Relax	Predicted Excitement	Total
Actually Relax	49	15	64
Actually Excitement	8	68	76
Total	57	83	140

Place here ROC curve? Describe AUC?

4.1.5 Voting with a threshold

The following approach is based on calculating the threshold which bring the higher accuracy. The threshold is defined as the minimum number of occurrences for the exact target to make this target win in a vote. In our case a threshold is applied to the *Excitement* target since it is a high accuracy target. All the dependencies between the threshold values and accuracies is shown below.
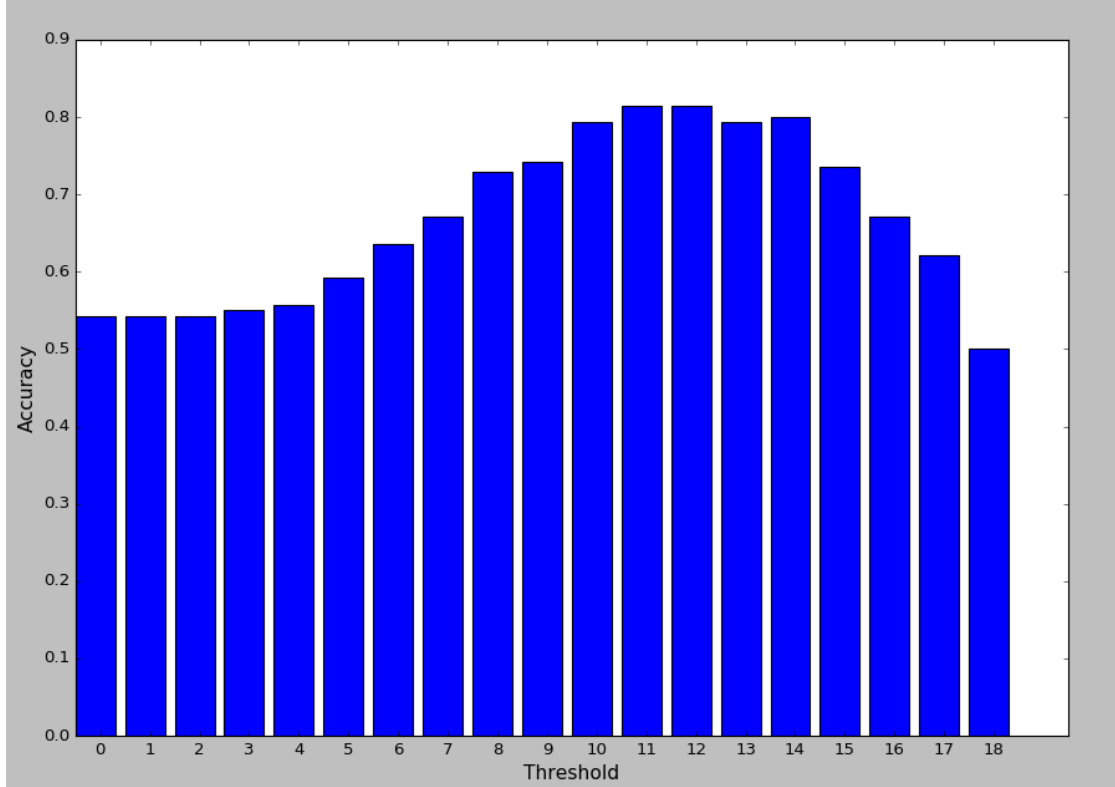

Figure 5: Excitement threshold value dependency on the prediction accuracy

The threshold for a voting for *Excitement* with the best classification accuracy of **81.4%** is 11 and 12. In our implementation we take the first value, which is 11.

Table 5: Confusion matrix showing classifiers prediction performance using 11 as a threshold for the Excitement class classification with applying a majority voting within measurement sessions data

	Predicted Relax	Predicted Excitement	Total
Actually Relax	45	19	64
Actually Excitement	7	69	76
Total	52	88	140

Usage of threshold has balanced False Negatives and True Negatives, however the error rate for the *Excitement* has been grown compared to the majority voting strategy.

4.1.6 Condorcet's rule based results

The number of required measurement trials to reach the minimum average accuracy of 99% is adjusted using Condorcet's jury theorem. The comparative table for the applied methods is shown below. The table is ordered by the single trial accuracies:

Table 6: Different methods accuracies and the required trials number to get 99%

Method	Single trial accuracy (%)	Trials required
Majority voting using RF probability threshold	83.6	9
Voting with a threshold	81.4	11
Majority voting	78.6	15
Voting-free using RF probability threshold	68.4	37
Voting-free	67.1	43

Based on the given results the majority voting using RF probability threshold is the best among the methods and using it will required 9 measurement sessions to get the desired accuracy.
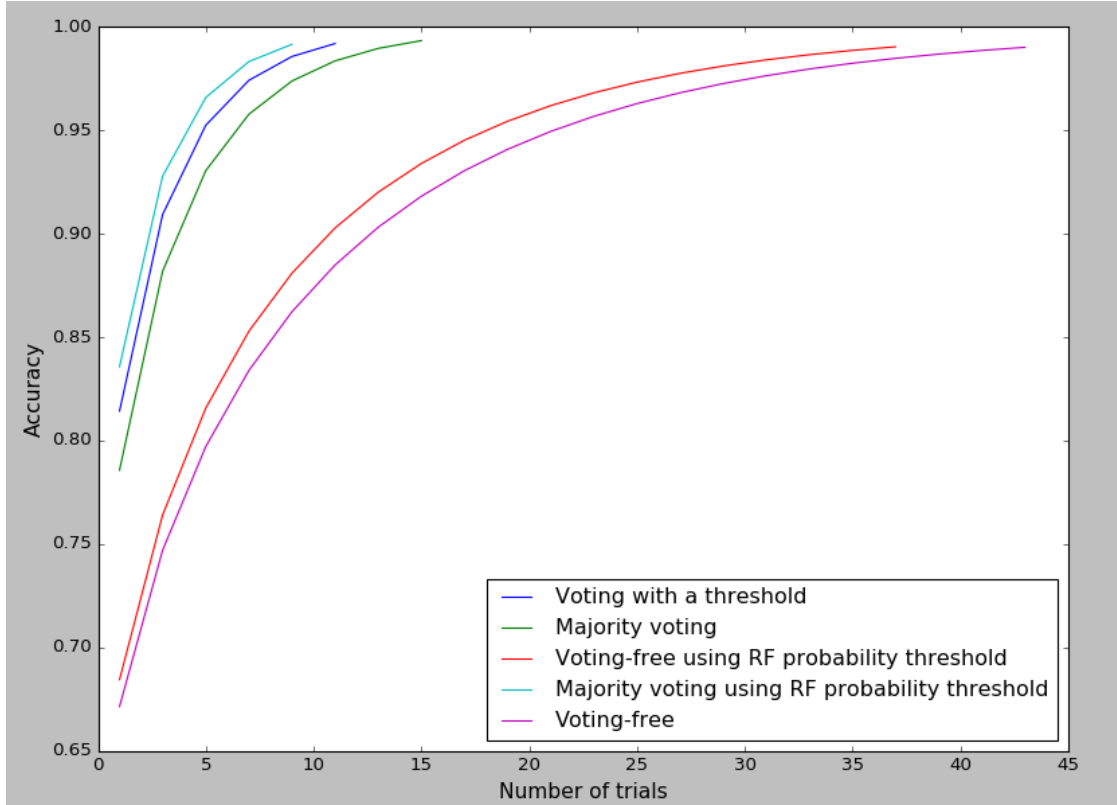
Figure 6: Correlation between the number of trials and the expected accuracy for different methods

4.2 Multiple trial approach

The test data for the given approach was measured with another measurement session length, which is defined using the number of required trials from section 4.1.6. Since brainwave signal measurements is an energy and time consuming activity it was decided to set 9 trials (90 seconds) for the test session length. It will be enough to fully validate the majority voting using RF probability thresholds and partially validate the other methods.

Place here nearly the same structure as in the single approach?

Or choose only top N methods and apply only them ?

5 Discussion

Compare condorcet dependency curves expected vs actual?

6 Conclusion

About a one-page conclusion text.

References

[1] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. Clin Neurophysiol. 2002 Jun;113(6):767-91. Review.

[2] Vidal JJ. Toward direct brain-computer communication. Annu Rev Biophys Bioeng. 1973;2:157-80. Review.

[3] Shivangi Miglanim, Surbhi Gupta. International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-2, Issue-8) Brain Computer Interface. August 2013

[4] Emotiv EPOC specifications. https://www.emotiv.com/epoc/

[5] Al-Fahoum AS, Al-Fraihat AA. Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains. ISRN Neuroscience. 2014;2014:730218.

[6] Pang-Ning Tan, Michael Steinbach, Vipin Kumar. Introduction to Data Mining. March 25, 2006. ch. 4, pp. 146-149.

[7] Russell, Stuart; Norvig, Peter (2003) [1995]. Artificial Intelligence: A Modern Approach (2nd ed.). Prentice Hall. ISBN 978-0137903955

[8] Madis Masso, Empirical Comparison of Machine Learning Algorithms Based on EEG Data. 2016

[9] Breiman, L. Machine Learning (2001) 45: 5.

[10] Wang, Yijun, and Tzyy-Ping Jung. "A Collaborative Brain-Computer Interface for Improving Human Performance." Ed. Pedro Antonio Valdes-Sosa. PLoS ONE 6.5 (2011): e20422. PMC. Web. 10 Nov. 2016.

[11] Emokit SDK. https://github.com/openyou/emokit

[12] SciPy open-source software for mathematics, science, and engineering SDK. https://www.scipy.org/

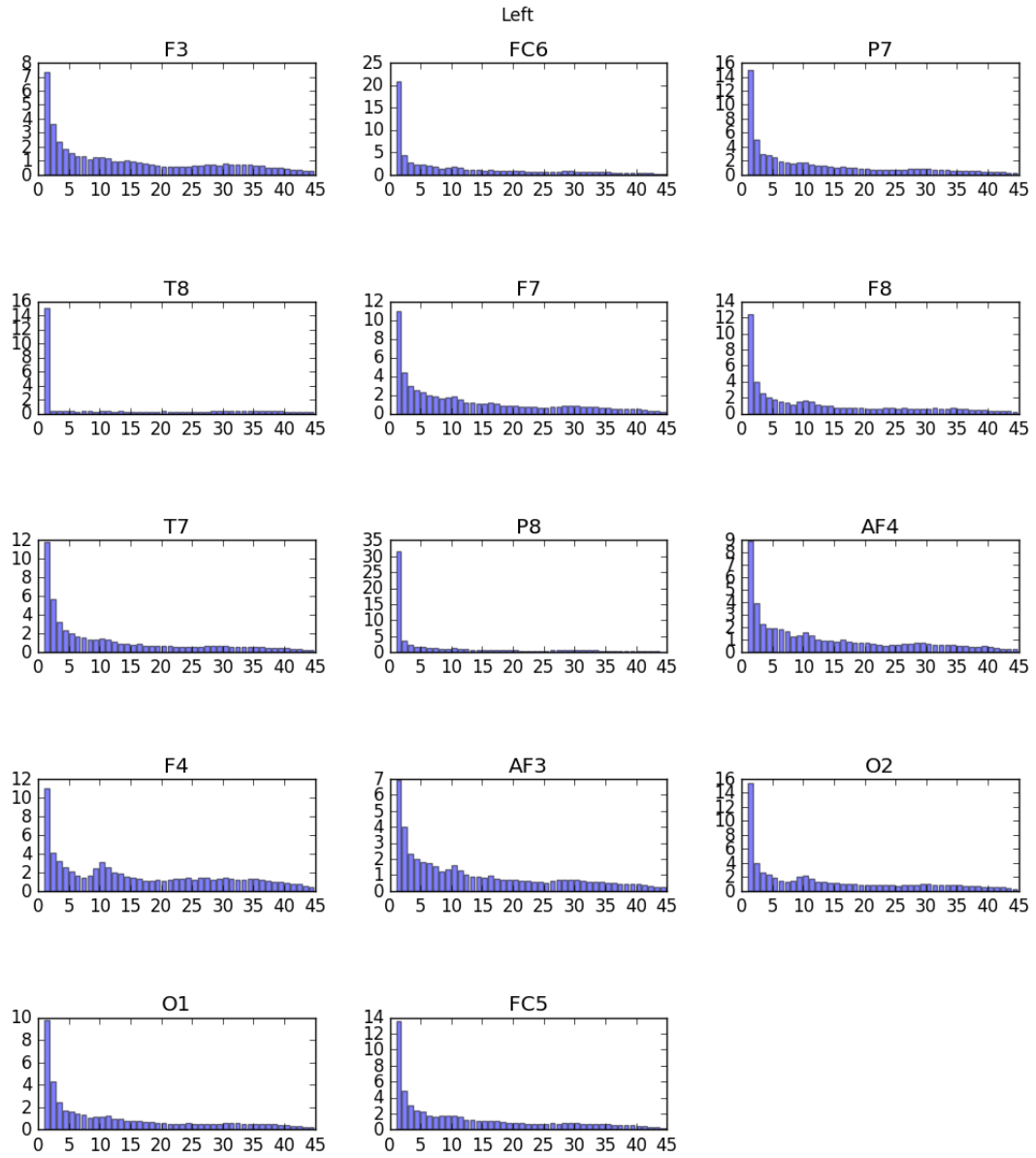[13] Numpy fundamental package for scientific computing. http://www.numpy.org/

Figure 7: A sample frequency domain plot while concentrating on "Left" target. Vertical axis is measured in Hz. Horizontal in μV.

Non-exclusive licence to reproduce thesis and make thesis public

I, Jevgeni Savostkin (date of birth: 13th of March 1990),

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

1.1 reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

1.2 make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

Improving accuracy of brain-computer interface with multiple trials

supervised by Ilya Kuzovkin and Raul Vicente

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, dd.mm.yyyy