Classifying the Stickiness of Mind-Wandering

Bachelor's Project Thesis

Stijn de Vries, s.de.vries.47@student.rug.nl,
Supervisors: C.Y. Jin & M. van Vugt

Abstract: The experience of your mind wandering away from the task at hand is familiar to many. Sometimes disengaging from mind-wandering and returning to the task at hand can be very difficult. This can be disruptive and potentially dangerous if caution is required. Detection of this "sticky" mind-wandering can be beneficial for safety and productivity reasons. The occurrence of sticky mind-wandering was predicted on the basis of electroencephalography (EEG) data recorded during a sustained attention to response task and a visual search task. The types of mind-wandering were assessed by means of questions inserted in the tasks that asked participants about their mental state, on or off task, as well as the content of their thoughts. A Logistic Regression (LR), Random Forest Classifier (RFC) and Support Vector Machine (SVM) were then trained on feature vectors obtained from temporal and spatial sampling points where a significant difference was observed between sticky and non-sticky mind-wandering. It was found that SVM (accuracy 62.2%) outperformed the LR (57%) and the RFC (58.1%), potentially because it is better able to deal with the high complexity of EEG data. The research furthermore suggests that activation of the visual brain areas just after the stimuli is lower in sticky mind wandering than in non-sticky mind wandering.

Keywords Mind-wandering, Stickiness, EEG, Classifier.

1 Introduction

Most people will know the feeling of their minds wandering away from the task at hand, often while doing long and monotonous activities such as driving. This feeling is called task unrelated thought or mind-wandering, a term popularized by Smallwood and Schooler (2006). When mind-wandering, returning to the task at hand can sometimes be difficult as people get stuck in their thoughts. The difficulty of disengagement from mind-wandering has been called the stickiness of mind-wandering (Joormann et al., 2011; van Vugt and Broers, 2016). Because mind-wandering is very common, it can be beneficial to detect and interrupt it in order to return to the task at hand. This could be especially useful if sticky mind-wandering occurs in people working in positions requiring attention, such as truck drivers or operators of heavy machinery.

In order to detect sticky mind-wandering it is required to know where and how it occurs. However, little research has been done on the subject of stickiness in mind-wandering. In contrast, a closely related phenomenon, rumination, did receive a lot of attention and can provide valuable insights into stickiness. Sticky mind-wandering and rumination are both related to thinking that is difficult to disengage from (van Vugt and Broers, 2016). Rumination occurs when people actively think about the negative aspects or problems in their lives but refrain from finding a solution to their problems, and is a symptom of depression (Davis and Nolen-hoeksema, 2000). Because rumination and mind-wandering occur in the brain, one method of learning more about these processes and their similarities is to measure brain activity when mind-wandering or rumination occurs in the brain.

Rumination and mind-wandering have received a lot of attention in literature (Nolen-Hoeksema et al., 2008; Smallwood and Schooler, 2006). Research using brain imaging techniques have provided insights into brain activity in rumination and mind-wandering. For example, decreased activation of the prefrontal brain areas has been linked to depression (Davidson, 1994), and decreased activation in the dorsolateral prefrontal cortex was found in both rumination (Ferdek et al., 2016) and during sadness (Liotti et al., 2000). Furthermore, it has been suggested by McLaughlin et al. (2007) that rumination elicits enhanced processing of negative material, even in nondepressed individuals (Lewis et al., 2015). One research studying sticky mind-wandering (Huijser et al., 2020) suggested that participants allocate less attention to the task during sticky thoughts when compared with neutral or non-sticky thought.

Since sticky mind-wandering and rumination are closely related, it is probable that there are similar brain areas involved in both processes. Thus, it is expected that there is decreased activation in the dorsolateral prefrontal cortex in sticky mind-wandering when compared to non-sticky mind-wandering. Furthermore it is expected to find decreased attention towards the task at hand during sticky mind-wandering when compared to non-sticky mind-wandering, as this was found in the study (Huijser et al., 2020) concerning sticky thought.

To find the differences between sticky and non-sticky mind-wandering in the brain and verify whether these predictions hold, this study will use electroencephalogram (EEG) recordings. Information about participants' mental state will be obtained using self-reports, also known as thought probes. Thought probes or self-reports are a method where questions are interspersed throughout the experiment to obtain information about a participants state of mind at the time of the thought probe. For example, the thought probe might ask the participant whether they were mind-wandering or not.

The thought probe method is often used in mind-wandering research (Weinstein, 2018). However, these thought probes disturb the experiment and might influence the performance of the participants and thus the results of the experiment. A solution for this problem could be automated detection of mind wandering. In recent years, different approaches have been taken to detect mind wandering without thought probes (Hutt et al., 2019; Beninger et al., 2020).

Jin et al. (2019) also tried to solve this problem using machine learning in combination with EEG recordings. Using thought probes, EEG trials were either labeled on or off-task depending on whether the participant reported mind-wandering or not. A Support Vector Machine (SVM) was trained to classify feature vectors consisting of EEG markers, and it predicted whether participants were on or off-task. On average, an across-task prediction accuracy of 60% was obtained. Furthermore the EEG marker most predictive of mind-wandering was found to be the alpha power. Not only does this provide insight into the various important brain areas that are involved in mind-wandering, furthermore such a classifier could also be useful in mind-wandering research because it removes the need for thought probes.

Another contribution from the research of Jin et al. (2019) is opportunity for further research. The thought probes in their study contained four additional questions about the content of the participants' thoughts. One question about the stickiness of their thoughts asked participants to give a stickiness 'rating' to their thoughts prior to the thought probe.

The dimension of stickiness was not explored in their research but will be the focus of this study. A similar approach to the research of Jin et al. (2019) will be taken. The goal of this research is to investigate where stickiness can be found in the brain and when it occurs. This will be done by training efficient classifiers to classify stickiness in mind-wandering.

The EEG and behavioural data from Jin et al. (2019) will be used to create feature vectors that highlight the differences between sticky and non-sticky mind-wandering. The feature vectors will be used to train and evaluate the classifiers.

2 Methods

2.1 Tasks

Stickiness in mind wandering will be investigated with data from Jin et al. (2019). Participants were required to complete two tasks, a Sustained Attention to Response Task (SART) and a Visual Search (VS). The SART is a simple and boring task that is commonly used in mind-wandering research. When participants perform such simple and boring tasks they are prone to mind-wandering, which allows for consistent collection of mind-wandering data. The

VS task provides a different context that can be used to check whether inferences made from the SART can be generalized to other tasks that have a different focus. Training the classifiers with data from both tasks ensures that the classifiers will not be dependent on a specific experimental paradigm.

The experiment consisted of two sessions, each session lasting approximately 2.5 h. Participants performed three blocks of each task per session. Task sequence was counterbalanced between blocks and sessions. In the SART, participants focused on a screen that displayed words. Words were displayed in lower case, which were frequent, nontarget stimuli, or in upper case which were infrequent, target stimuli. Frequent stimuli appeared 89% of the time and infrequent stimuli appeared 11% of the time. In response to an lower case word, participants were required to press the button 'm', and in response to a upper case word, participants were required to withhold the response. A SART trial started with a fixation cross that lasted for a uniformly sampled period between 1.5 $\sim$ 2.1 s. The stimulus appeared for 300 ms and was followed by a 900 ms mask. The intertrial interval was 3 s.

In the VS, blue stimuli were shown on the screen. Before each trial, participants were told what stimuli they were required to search. If the target was present, participants were required to press the left arrow button. If the target was not present, participants were required to press the right arrow button. The target differed from the non-targets in shape. The probability of the target appearing was equal to the probability of the target not appearing. A VS trial started with a fixation cross that lasted for 1.5 $\sim$ 2.1 s. Each search panel appeared for 3 s.

2.1.1 Thought probes

54 thought probes were included throughout both the SART and VS tasks. Each thought probe consisted of 4 questions asking about the thoughts the participant had just before the thought probe. The four question were about the content, the temporal orientation, the emotional valence and the stickiness of the thought. The question regarding stickiness was 'How hard was it to let the idea go?' with the responses (1) Very difficult (2) Difficult (3) Not easy or Difficult (4) Easy (5) Very easy. Each response corresponds to a stickiness rating, 1 for very sticky and 5 for not sticky at all. The degree of stickiness was judged by the participants themselves.

The 3 trials prior to a thought probe were labeled with the stickiness rating that was giving by the participant in that thought probe. Bastian and Sackur (2013) suggested that the average length of a mind wandering episode is 11.1 seconds. The average length of a trial in this research was 6 s. Because the focus of this research is mind wandering and its stickiness, it would be expected that the 2 trials preceding a thought probe are mind wandering trials if the participant indicated mind wandering. However, the research was only based on a SART task, whereas the current research also has a VS task. Furthermore, using 3 trials instead of 2 allows for a bigger sample size which is desirable because data from only 30 participants was obtained, resulting in a small sample. Thus, this research will use the 3 trials preceding a thought probe where the participant indicated mind-wandering.

Subject information and a detailed description of the task procedure for both tasks are described in the original research paper (Jin et al., 2019).

2.1.2 EEG data

The continuous EEG data was obtained using a Biosemi 128-channel system with a sampling rate of 512 Hz. Six more electrodes were used to measure mastoid signals and eye movements. Mastoid signals serve as a non-brain reference and tracking eye movements allows for easier artifact identification and rejection.

The continuous EEG data was processed offline. The data was referenced to the averaged mastoid signals, band-pass filtered (0.5 - 40 Hz) and downsampled to 256 Hz. Lastly, the trials were segmented in epochs of 1600 ms, of which 400 ms were before and 1200 ms after the stimulus onset.

Channels that had excessive spikes or were noisy compared to surrounding channels were identified by visual inspection and replaced through spherical interpolation. Infomax independent component analysis (ICA) and visual inspection were then used for ocular artifact detection and removal. Details of the online recording parameters and the offline processing procedures can be found in the original study (Jin et al., 2019).

The amount of channels used for this research was reduced to 32 channels using the 10-20 system.

Since EEG channels record highly correlated signals, reducing the amount of channels reduces the complexity of the data without a large loss of information.

2.1.3 Trials

The data used for training and evaluating the classifiers consists of both session one and session two. Depending on whether participants' mind-wandering was sticky or not during trials, the trials were divided into two groups. The 'sticky' group, containing stickiness ratings 1 and 2, and the 'non-sticky' group, containing stickiness ratings 4 and 5. Trials with a stickiness rating of 3 were excluded because they are neither sticky nor non-sticky, and thus they contain little information that is useful for the classification of stickiness.

The data was divided in two categories because it simplifies the training process for both Logistic regression and SVM as they are binary classifiers. It furthermore ensures there is enough data in each class and helps with the classification of stickiness because this research focuses on the differences in sticky and non-sticky thought and less so on the degree of stickiness of that thought.

From all the trials across all participants, 8 trials were identified to have a stickiness rating other than 1 through 5, and were not included in the data. The deviant stickiness ratings were due to a wrong input from participants. Participants 16 and 25 did not contribute any trials from the VS task as they were all excluded. Participant 19 did not contribute any trials from the SART task as they were all excluded. Across all participants, 1861 trials were obtained in total, of which 1042 trials from the SART and 819 trials from the VS task. 1097 trials were non-sticky trials and 764 trials were sticky trials.

2.2 Data Analysis

An initial analysis was performed on the data to find what features might be important in the classification of stickiness in mind wandering. This was done for two reasons. First, EEG data is of high dimensionality, and even when only 32 channels were used, every trial is a large data object in the form of a matrix. This complicates the training process and extends the run time of the classifiers notably.

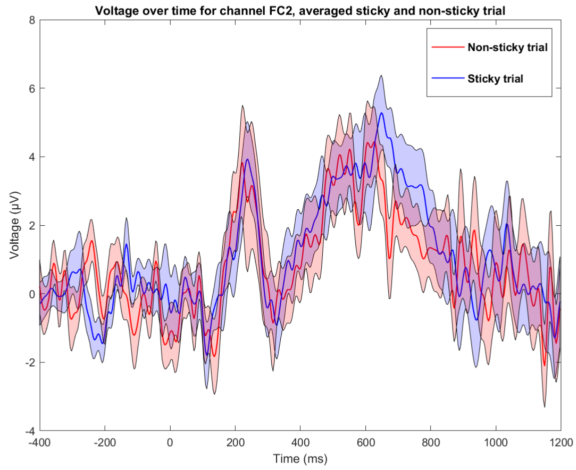

Figure 2.1: Average SART trial over all participants, in the sticky and non-sticky condition over one channel. The shaded error shows the SEM of each time point. Time points where the shaded areas do not overlap are significantly different.

Furthermore, both RFC and Logistic regression do not deal well with data that has a low signal-to-noise ratio. In order to reduce the complexity of the data, a filter was created using the differences between the sticky and non-sticky data.

2.2.1 Feature computation

All trials per participant were averaged to obtain an averaged EEG recording of the sticky condition and an averaged EEG recording of the non-sticky condition for each participant. For each channel, the standard error of the mean (SEM) was calculated across participants.

Thus for each channel a vector with the average values and the SEM was obtained for the sticky and the non-sticky condition. The time points that did not have an overlapping error were treated as significantly different between conditions. Figure 2.1 shows an example of the average sticky and non-sticky trial with their SEMs for the channel FC2.

To obtain the filter, the vectors for every channel were concatenated in one matrix. Significantly different time points were marked with a '1' and insignificantly different time points were marked with a '0'. Two feature matrices were obtained, one for the SART and one for the VS. These two matrices were combined using their dot product to obtain a 'task-general' feature matrix that contains time points where both tasks have a significant difference between the two conditions. Applying this filter to samples results in a feature vector that was used for training and testing the classifiers.

Each sample was filtered as follows: if there were some amount of consecutive number of 1's in the feature matrix, the values at those time points were taken from the sample and averaged. The average value was treated as a feature and was stored in the feature vector. A parameter was used to discard irrelevant information from the feature vector. If the amount of consecutive 1's was lower than the parameter, the average value would not be included in the feature vector. This parameter will be referred to as the minimum time window in the remainder of this paper.

For example, in figure 2.1 there is no overlap from -267 ms to -227 ms. The average of the values from these time points will form one feature. All the features are put in a vector to obtain the feature vector.

10 fold cross validation was performed in order to prevent over fitting of the data. The trials were normalized and the data set was balanced during cross validation to ensure that the classifiers did not have a bias towards the majority class.

2.2.2 Parameter selection

Two parameters were identified that could influence the performance of the classifiers. The first parameter was the amount of trials that every participant contributed. Some participants contributed a small amount of trials (e.g. participant 1 contributed only one SART trial and 2 VS trials) whereas other participants contributed a large amount of trials (e.g. participant 12 contributed 79 SART trials and 63 VS trials). The feature matrix was constructed using the average trial data across participants and not across trials. Thus, the average trial of participant 1 would have the same weight as the average trial of participant 12, even though participant 12's data contains a lot more information.

The second parameter was the minimum time window that was used in creating the feature vector. If it was set too high, too much information would be lost because only a small amount of time points could be used. However, if it was set at 1 then the filter might take in irrelevant information because only one significant time point is needed to contribute a value to the feature vector, even though that single time point could reflect just a false positive.

To examine the influence of the amount of trials in the dataset of each participant and the minimum time window, an ANOVA (Analysis of Variance) was performed on three variables, the minimum amount of trials a participant had to have in order to contribute any data to the classification, the minimum time window that was needed to contribute a value to the feature vector, and the type of classifier.

For the minimum amount of trials that a participant had to contribute, the difference between 1 trial and 30 trials will be investigated. For the minimum time window parameter the values 1 through 4 will be used.

2.3 Classifiers

This study compared 3 different algorithms for classification of EEG signals. The first algorithm is logistic regression (LR). LR was used as a baseline to compare the performance of the other two classifiers to. Even though LR is not able to model complex data such as EEG data since it is of high dimensionality, as a baseline it is useful because it is easy to train and it shows what a simple model can achieve. Furthermore, a LR is easily interpretable because feature importance can be inferred from the coefficients obtained in the LR, and the features can be linked back to the original data with ease.

The second algorithm that was used is the Support Vector Machine (SVM). SVM has been applied to EEG classification problems widely and are good at generalizing, are robust to over fitting and perform well with a low amount of data (Lotte et al., 2007), which is often the situation with EEG data. Furthermore, SVM is able to use a kernel function that allows a non-linear hyperplane, thus being able to find non-linear interactions in high dimensional data sets.

The third algorithm that was used is Random Forest Classification (RFC). RFC is known to rarely over fit, and makes use of an ensemble technique, voting, which reduces the variance of the model and thus reduces classification errors (Hastie et al., 2009, p. 588; Lotte et al., 2007). Furthermore, RFC was used in an aEEG (Amplitude integrated EEG) classification study, and outperformed widely-used classifiers including SVM and LR (Wang et al, 2014).

2.3.1 Logistic Regression

LR is a supervised classification method. It is a binomial classifier because the target variable can be either '0' or '1', which represent the categories 'non-sticky' (0) and 'sticky' (1) that are to be classified. LR creates a model that is based on the sigmoid function (1).

$$g(z) = \frac{1}{1 + e^{-z}} \tag{2.1}$$

This model ensures that the target variable is always between zero and one instead of a continuous value. By comparing the predicted probability and the decision boundary, x_i can be assigned a target variable.

2.3.2 Support Vector Machine

SVM is a supervised classification algorithm that uses a hyperplane to distinguish between classes. The hyperplane chosen by the algorithm optimizes the distance to the nearest data point to obtain the optimal hyperplane that separates the data. The SVM can use a kernel to improve classification. The kernel function used in the SVM of the current study is the Guassian kernel (3).

$$K(x, y) = exp(\frac{-||x - y||^2}{2\sigma^2}) \tag{2.2}$$

The Gaussian kernel allows the SVM to consider non-linear hyperplanes, which is most likely better able to model than a linear kernel because EEG data is of high dimensionality.

2.3.3 Random Forest classification

RFC is an ensemble classification technique. RFC builds a population of randomly grown decision trees. A decision tree is a tree-like structure that has nodes which split into two branches. This node represents a test on a certain property. A sample will be passed on from node to node depending on the outcome of the test. All leaf nodes contain a label, and if the sample is passed to a leaf node it is classified as the label that that leaf node contains. When a new observation is presented, every tree in the population casts a vote on the predicted class and the majority vote decides what the observation will be classified as. Decision trees are able to capture complex structures in data. However, this also allows them to capture noise more easily, which leads to an unstable tree. Using the average of a collection of trees reduces the large variance that a single tree has to a small variance of the overall ensemble and so RFC takes the strength of decision trees and reduces its weaknesses.

2.4 Classification

The three algorithms were trained and evaluated using the classification app learner in MATLAB (version R2020b), using the feature vectors as described in the 'Analysis of Data' section. The classifiers used in MATLAB for LR is 'Logistic Regression', for SVM is 'Medium Gaussian SVM' and for the RFC is 'Bagged trees'. The code for each classifier was extracted and altered to include k-fold cross-validation and allow for extraction of data from each fold.

All three classifiers were validated using 10-fold cross validation. The performance of the model was measured using prediction accuracy, the precision and the recall.

The performance of the classifiers will be compared using a one way ANOVA and the classifiers will be evaluated using the prediction accuracy, the precision and the recall.

3 Results

Classification was performed on the feature vectors obtained from the participants' trials. Classification accuracy in parameter selection ranged from 51% to 59%. Classification accuracy with optimal parameters was above chance level for all classifiers.

Furthermore the average sticky and non-sticky trials were visualized. Figure 3.2 illustrates the temporal and spatial locations of differences between the sticky and non-sticky condition.

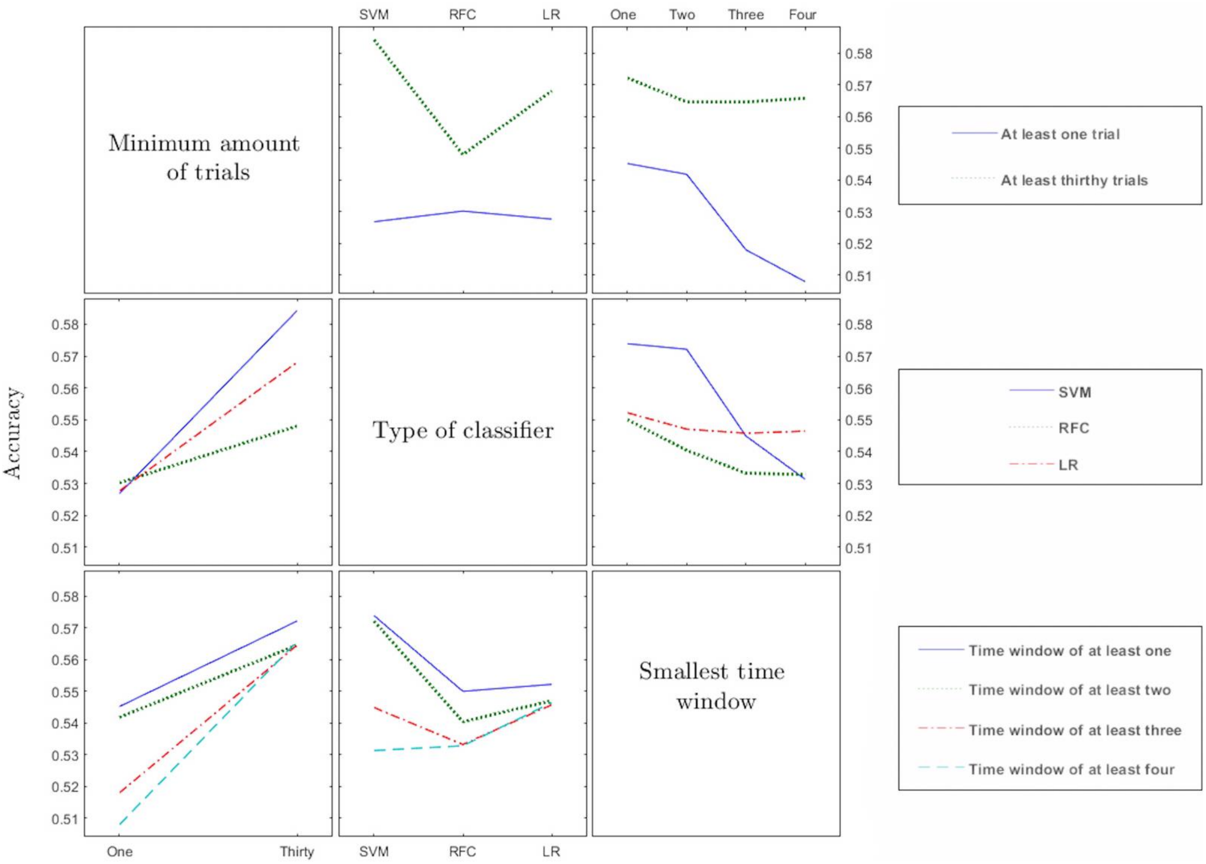

Figure 3.1: Achieved accuracies of the SVM, LR and RFC with minimum trial amount per dataset (one vs. thirty) and the smallest time window (one, two, three or four) as two interacting factors.

3.1 Parameters

A three-way ANOVA was conducted to examine the effect of three factors (type of classifier, minimum amount of trials, smallest time window) on the accuracy of classification.

Figure 3.1 shows an overview of these effects. Classifier type includes three levels (Logistic regression, RFC, SVM), minimum amount of trials includes two levels (one and thirty) and smallest time window includes four levels (one, two, three and four). All main effects were statistically significant (figure 3.1).

The main effect for classifier type ($F(2,216) = 3.86$, $p < 0.05$) indicates a significant difference between the logistic regression (M = 0.548, SD = 0.0435), the RFC (M = 0.539, SD = 0.035) and the SVM (M = 0.556, SD = 0.052). The main effect for minimum amount of trials ($F(1,216) = 63.62$, $p < 0.001$) indicates significant difference between having at least one trial (M = 0.529, SD = 0.038) and having at least thirty trials (M = 0.567, SD = 0.042). The main effect for smallest time window ($F(3,216) = 4.42$, $p < 0.005$) indicates significant difference between using a minimum time window of one (M = 0.559, SD = 0.043), two (M = 0.553, SD = 0.043), three (M = 0.541, SD = 0.042) and four (M = 0.537, SD = 0.047).

The interaction effect between minimum amount of trials and classifier was significant $F(2,216) = 5.64$, $p < 0.005$. The interaction effect between minimum amount of trials and smallest time window was also significant, $F(3,216) = 2.92$, $p < 0.05$. No significant effect was found for the interaction between classifier and smallest time window, ($F(6,216) = 1.43$, $p > 0.05$), and no significant interaction was found for the interaction of all three

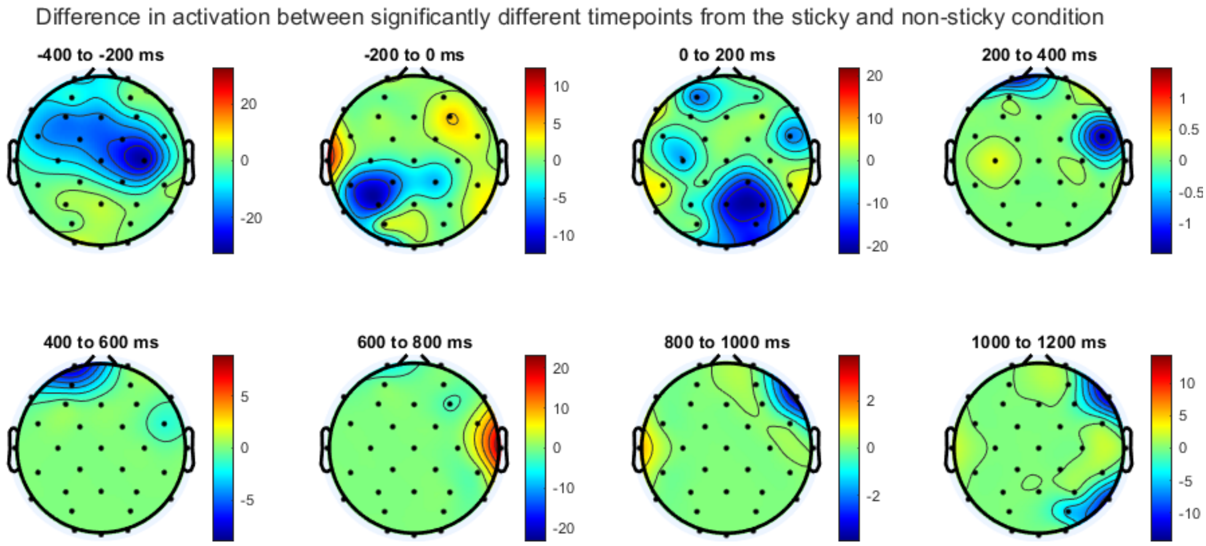

Figure 3.2: Visualization of differences between sticky and non-sticky condition in intervals of 200 ms. The stimulus appears at 0 ms. Data plotted is the average sticky trial subtracted by the average non-sticky trial. Blue color indicates lower activation in sticky condition and red color indicates lower activation in non-sticky condition.

independent variables, F(6,216) = 1.03, $p > 0.05$.

The ANOVA showed that there was a significant interaction for the minimum amount of trials. Figure 3.1 shows that using at least thirty trials results in a higher accuracy in all cases. Thus, when evaluating the classifiers, participants will need to have at least thirty trials in order for their data to be used in training the classifiers.

Regarding the smallest time window, a significant interaction was found between the minimum amount of trials and the smallest time window. Post hoc comparisons using the Tukey HSD test indicated that the mean score for a time window of one (M = 0.559, SD = 0.043) was significantly different than the time window of 4 (M = 0.537, SD = 0.047). The time window of 2 (M = 0.553, SD = 0.043) and the time window of 3 (M = 0.541, SD = 0.042) did not significantly differ with any conditions. Because there was no difference between a time window of 1, 2 or 3, a time window of 1 will be used as it will retain the most information.

The optimal parameters for each classifier are the same and thus the same parameters will be used for all three classifiers.

3.2 Classifiers

Three one-way ANOVAs were performed on the influence of the classifier type (SVM, RFC, LR) on the accuracy, the precision and the recall. The ANOVAs were performed with the parameters obtained in section 3.1.

First, a one-way ANOVA was performed on the influence of the classifier type (SVM, RFC, LR) on the accuracy. The interaction effect was significant, F(2,27) = 4.52, $p < 0.05$. This indicates a significant difference between the influence of SVM (M = 0.6227, SD = 0.0358), RFC (M = 0.5813, SD = 0.0350) and LR (M = 0.5708, SD = 0.05) on the accuracy.

Secondly, a one-way ANOVA was performed on the influence of the classifier type on the precision. The interaction effect was not significant, F(2,27) = 0.74, $p > 0.05$. This indicates no significant difference between the influence of SVM (M = 0.6491, SD = 0.0454), RFC (M = 0.6249, SD = 0.0372) and LR (M = 0.6361, SD = 0.0501) on the precision.

Lastly, a one-way ANOVA was performed on the influence of the classifier type on the recall. The interaction effect was significant, F(2,27) = 31.01, $p < 0.001$. This indicates a significant difference between the influence of SVM (M = 0.7225, SD =

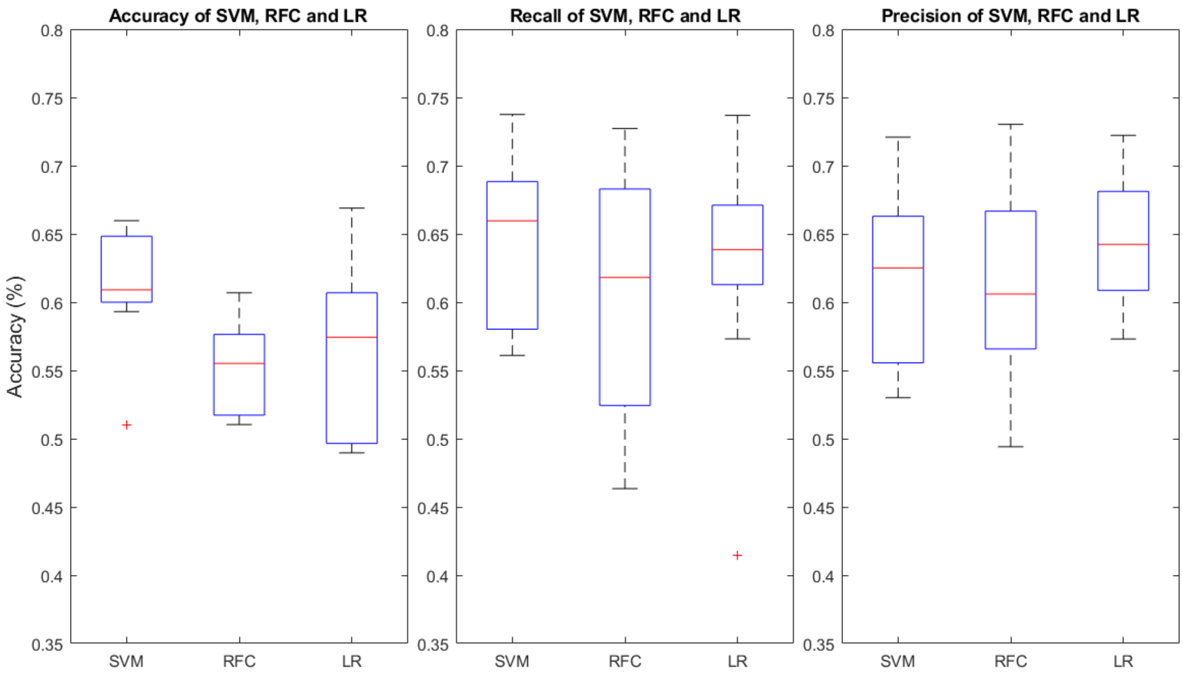

Figure 3.3: Three box plots showing the accuracy, precision and recall of the three classifiers from left to right. The values were obtained from training the classifiers using 10-fold cross validation.

0.0352), RFC (M = 0.6483, SD = 0.0458) and LR (M = 0.5702, SD = 0.0477) on the recall.

Figure 3.3 shows box plots of the three classifiers for their accuracy, recall and precision.

4 Discussion

The goal of this research was to understand where stickiness occurs in the brain and to train efficient classifiers to detect stickiness. This was done by training a Logistic Regression, a Random Forest Classifier and a Support Vector Machine to classify EEG trials of participants performing a Visual Search task or a Sustained Attention to Response task.

Prediction accuracy across classifiers ranged from 57% to 62%. The overall accuracy was higher than chance, which suggests that stickiness is not something that is a subjective experience but can also be found in EEG data. However, overall accuracy was not high enough to reliably predict stickiness in mind-wandering.

There are several reasons for the low overall accuracy. First of all, the data consists of self-reports, which rely on the subjective experience of participants and thus always introduce some noise. The same mental states might be experienced differently per participant and this blurs the line between sticky and non-sticky thought.

Secondly the data was very limited. Data from only 30 subjects was collected in the original study of Jin et al. (2019). After filtering the data and discarding data from participants with low amounts of data points as explained in the methods section, data from only 21 participants remained.

Lastly, data of high dimension often introduces the curse-of-dimensionality, which states that properly describing different classes requires exponentially more data as the dimensionality of feature vectors grow (Lotte et al., 2007). Because limited data was used the curse-of-dimensionality could be a major reason why the overall accuracy was low.

No optimization of the parameters of the algorithms was performed because it was outside the scope of this study. However, it is likely that optimization would have increased the prediction accuracy and larger differences could have been found between the algorithms which would aid in a more informative conclusion regarding choice of classification algorithm.

Furthermore it was found that the accuracy and the recall of SVM were significantly higher than that of the RFC and LR, but that there was no significant difference between the RFC and the LR in both the accuracy and recall. There was no difference between the three classifiers in terms op precision. Lotte et al. (2007) states that the SVM is known to be insensitive to the curse-of-dimensionality and to have good generalization properties. Better performance of the SVM can be due to the aforementioned reasons.

First of all, limited data introduced the curse-of-dimensionality. However, the reason that SVM outperformed the LR and RFC could be because the SVM is insensitive to the curse-of-dimensionality but the LR and RFC are not.

Secondly, data was obtained from both a SART and a VS task. Because the SVM has good generalization properties it might have performed better overall.

Visualization of the average sticky and non-sticky trial show differences in temporal and spatial locations throughout the brain.

Figure 3.2 shows that the overall activation in the brain in the 400ms before the stimulus appears is lower in sticky thought when compared to non-sticky thought. This could indicate that attenuation of brain activity occurs in sticky mind-wandering when compared to non-sticky mind-wandering.

Furthermore figure 3.2 shows a lower activation in the visual brain areas in the 200 ms after the stimulus in sticky mind-wandering as compared to non-sticky mind-wandering. Amano et al. (2006) suggests that it takes participants 150 $\sim$ 200 ms to react to visual stimuli, which means that visual processing of the stimulus occurs within 200 ms after the stimulus appears. Attenuation in this time frame of the visual brain areas could suggest decreased attention towards the stimulus in sticky mind-wandering when compared to non-sticky mind-wandering. This finding falls in line with the expectation based on the results of Huijser et al. (2020) that participants pay less attention to the stimulus during sticky mind-wandering.

There is no clear decrease in activation in the dorsolateral prefrontal cortex during sticky mind-wandering which was expected. The -400 to -200 ms time frame in 3.2 shows decreased activation in the frontal and part of the parietal lobe, however because the spatial resolution of EEG data is poor these results cannot be linked to the dorsolateral prefrontal cortex.

Another reason that there was no clear decrease in activation in the dorsolateral prefrontal cortex during sticky thought could be that this is an aspect of rumination that is not present in stickiness. Decreased activation in the dorsolateral prefrontal cortex was found in both rumination (Ferdek et al., 2016) and sadness (Liotti et al., 2000), which could suggest that it is related to processing of negative thoughts and not necessarily thoughts that are difficult to disengage from.

Future research using brain imaging techniques that have a higher spatial resolution than EEG could investigate whether stickiness is related to the dorsolateral prefrontal cortex. This would provide insights in both stickiness and the difference between stickiness and rumination.

A more obvious direction is the collection of more EEG data and classifying stickiness in mind-wandering using a SVM with optimized parameters. If a higher classification rate can be obtained, the classifier could be used in real-life situations requiring close attention, although some difficulties regarding convenience and the feasibility of using EEG data in real time need to be overcome.

This study showed that above chance classification of stickiness is possible using several classification algorithms, and that SVM outperforms both LR and RFC in the classification of stickiness in mind-wandering. Furthermore decreased brain activation was found in the 400 ms before a stimulus en decreased activation of the visual brain areas was found in the 200 ms after a stimulus, indicating attenuation during sticky mind-wandering when compared to non-sticky mind-wandering.

References

K. Amano, N. Goda, S. Nishida, Y. Ejima, T. Takeda, and Y. Ohtani. Estimation of the timing of human visual perception from magnetoencephalography. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(15):3981–91, 2006.

M Bastian and J Sackur. Mind wandering at the fingertips: automatic parsing of subjective states based on response time variability. *Frontiers in psychology*, 4:573–573, 2013.

J. Beninger, A. Hamilton-Wright, H.E.K. Walker, and L.M. Trick. Machine learning techniques to identify mind-wandering and predict hazard response time in fully immersive driving simulation. *Soft Computing*, 2020.

R.J. Davidson. Asymmetric brain function, affective style, and psychopathology: The role of early experience and plasticity. *Development and Psychopathology*, 6:741–741, 1994.

R.N. Davis and S. Nolen-hoeksema. Cognitive inflexibility among ruminators and nonruminators. *Cognitive Therapy and Research*, 24(6):699–711, 2000.

M.A. Ferdek, C.M. van Rijn, and M. Wyczesany. Depressive rumination and the emotional control circuit: an eeg localization and effective connectivity study. *Cognitive, Affective, Behavioral Neuroscience*, 16(6):1099–1113, 2016.

S. Huijser, M. Verkaik, M.K. van Vugt, N.A. Taatgen, and M. Faber. Captivated by thought: "sticky" thinking leaves traces of perceptual decoupling in task-evoked pupil size. *PLOS ONE*, 15(12):0243532, 2020.

S. Hutt, K. Krasich, C. Mills, N. Bosch, S. White, J.R. Brockmole, and S.K. D'Mello. Automated gaze-based mind wandering detection during computerized learning in classrooms. *User Modeling and User-Adapted Interaction : The Journal of Personalization Research*, 29(4):821–867, 2019.

C.Y. Jin, J.P. Borst, and M.K van Vugt. Predicting task-general mind-wandering with eeg. *Cognitive, Affective, & Behavioral Neuroscience*, 19: 1059–1073, 2019.

J. Joormann, S.M. Levens, and I.H. Gotlib. Sticky thoughts: Depression and rumination are associated with difficulties manipulating emotional material in working memory. *Psychological Science*, 22(8):979–983, 2011.

K.L. Lewis, L.E. Taubitz, E.L. Duke, M.W. Steuer, and C.L. Larson. State rumination enhances elaborative processing of negative material as evidenced by the late positive potential. *Emotion*, 15(6):687–693, 2015.

M. Liotti, H.S. Mayberg, S.K Brannan, S. McGinnis, P. Jerabek, and P.T. Fox. Differential limbic-cortical correlates of sadness and anxiety in healthy subjects: implications for affective disorders. *Biological Psychiatry*, 48(1):30–42, 2000.

F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for eeg-based brain–computer interfaces. *Journal of Neural Engineering*, 4(2):R1–R13, 2007.

K.A. McLaughlin, T.D. Borkovec, and N.J. Sibrava. The effects of worry and rumination on affect states and cognitive activity. *Behavior Therapy*, 38(1):23–38, 2007.

S. Nolen-Hoeksema, B.E. Wisco, and S. Lyubomirsky. Rethinking rumination. *Perspectives on Psychological Science*, 3(5): 400–424, 2008.

J. Smallwood and J.W. Schooler. The restless mind. *Psychological Bulletin*, 132(6):946–958, 2006.

M.K. van Vugt and N. Broers. Self-reported stickiness of mind-wandering affects task performance. *Frontiers in Psychology*, 7(732), 2016.

Y. Weinstein. Mind-wandering, how do i measure thee with probes? let me count the ways. *Behavior Research Methods*, 50(2):642–661, 2018.