



# Using Support Vector Machine on EEG for Advertisement Impact Assessment

Zhen Wei<sup>1\*</sup>, Chao Wu<sup>1,2</sup>, Xiaoyi Wang<sup>3</sup>, Akara Supratak<sup>1</sup>, Pan Wang<sup>1</sup> and Yike Guo<sup>1\*</sup>

<sup>1</sup> Data Science Institute, Imperial College, London, United Kingdom, <sup>2</sup> School of Public Affairs, Zhejiang University, Hangzhou, China, <sup>3</sup> School of Management, Zhejiang University, Hangzhou, China

The advertising industry depends on an effective assessment of the impact of advertising as a key performance metric for their products. However, current assessment methods have relied on either indirect inference from observing changes in consumer behavior after the launch of an advertising campaign, which has long cycle times and requires an ad campaign to have already been launched (often meaning costs having been sunk). Or through surveys or focus groups, which have a potential for experimental biases, peer pressure, and other psychological and sociological phenomena that can reduce the effectiveness of the study. In this paper, we investigate a new approach to assess the impact of advertisement by utilizing low-cost EEG headbands to record and assess the measurable impact of advertising on the brain. Our evaluation shows the desired performance of our method based on user experiment with 30 recruited subjects after watching 220 different advertisements. We believe the proposed SVM method can be further developed to a general and scalable methodology that can enable advertising agencies to assess impact rapidly, quantitatively, and without bias.

**Keywords:** EEG, SVM, advertisement impact assessment, neuromarketing, machine learning

## OPEN ACCESS

### Edited by:

Peter Lewinski,  
University of Oxford, United Kingdom

### Reviewed by:

Xiaoli Li,  
Beijing Normal University, China  
Dominika Basaj,  
Warsaw University of Technology,  
Poland

### \*Correspondence:

Zhen Wei  
zw708@ic.ac.uk  
Yike Guo  
y.guo@imperial.ac.uk

### Specialty section:

This article was submitted to  
Neural Technology,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 31 May 2017

**Accepted:** 30 January 2018

**Published:** 12 March 2018

### Citation:

Wei Z, Wu C, Wang X, Supratak A,  
Wang P and Guo Y (2018) Using  
Support Vector Machine on EEG for  
Advertisement Impact Assessment.  
*Front. Neurosci.* 12:76.  
doi: 10.3389/fnins.2018.00076

## 1. INTRODUCTION

Advertising plays a critical role in marketing. Every year, companies allocate a significant proportion of marketing budget to attempt to quantify the impact of their advertising, particularly video advertising on TV and on the web which receives wide viewership (Brady, 2014; Bradley, 2015). However, current methodologies, including both direct observation (questionnaires and focus groups prior to starting of the advertising campaign), and indirect (trends in sales or consumer interest during and after a campaign), tend to have practical or experimental challenges that reduce the effectiveness of the assessment (Goldberg, 1990; Ducoffe, 1996; Elliott and Speck, 1998; Lewis and Reiley, 2009; Ostrovsky and Schwarz, 2011).

Direct observation methods include questionnaires, and focus groups, where yet to be released advertisements (or multiple versions of an advertisement) are shown to a select group of viewers selected to be representative of the advertisement's intended audience. The viewers answer questions and provide feedback during a survey or engage in discussion with the organizers; then the results or discussion are analyzed by the advertising team to try to assess how well the advertisement fulfils the criteria of their campaign (Gaines et al., 2007).

However, direct approach methods are subject to the same kinds of challenges as other experimental psychology approaches: experimental biases are introduced due to the experimental environment (typically an office room with multiple participants) being different to how a viewer

would normally view the advert (often in the comfort of one's own home), leading to a different state of mind of the viewer. There is a tendency for respondents to feel obliged to give more favorable reviews under experimental conditions of being under observation than they would typically do. The way the experiment is carried out in groups can also result in peer-pressure and group-dynamics altering the responses of individuals, leading to participants reporting attitudes or preferences that may not truly represent their own when they are in private (De Pelsmacker et al., 2002; Maison et al., 2004; Shen and Li, 2009). The analysis of survey feedback or discussion too can introduce experimenter-bias where the opinions of the experimenters impact the evaluation of the results; and the size and cost of conducting the study and analyzing the results is often cost-prohibitive past certain scales (and when considering TV advertising for national release, the size of these focus studies are often a tiny proportion of eventual viewership, resulting in uncertainty of the statistical relevance of the results).

Indirect observation methods of assessing advertising impact involve inferring advertising impact based on the result of an advertising campaign (Sharma et al., 2011). These methods avoid the inaccuracies imposed by artificial experimental conditions of direct observation methods but suffer from their own challenges. By looking only at the results and effects of the advertisement campaign, and because it is only possible to look at aggregated effects (such as the impact on sales or customer interest in a product), it is often difficult to identify if, why, and how a particular aspect of the advertisement causes an impact. Furthermore due to the life-cycle of an advertising campaign, it is only possible to infer impact during or after the launch of an advertising campaign, meaning in many cases, much or all of the cost of the campaign having been sunk, limiting adaptability in the event of lackluster response to the campaign (Kanetkar et al., 1992; Grewal et al., 1998; Sundar and Kalyanaraman, 2004).

In recent decades, research in neuroscience has brought new understanding and tools that can change how the impact of advertising can be assessed and has formed the new field of neuromarketing, in which recent neuroscience and experimental psychology tools and understanding are being applied to marketing. One key hypothesis in neuromarketing is that a consumers decisions can be driven more by emotion than by a careful comparison of product benefits or differentiators. Therefore, measuring an advertisement's emotional impact on an individual could correlate well with the impact of the advertisement.

In neuromarketing-based advertisement impact assessment, biometrics are gathered from individuals participation in the study, and these biometrics are used to assess impact, rather than voluntarily self-reported information from surveys or discussion. The recorded data includes biometrics such as eye-tracking, facial coding, Galvanic skin response and electrodermal activity, and EEG. EEG is a noninvasive electrophysiological recording of brain activity, using electrodes placed along the scalp. EEG has multiple advantages over other methods of measuring brain activity in that it has a high temporal resolution, is non-invasive, quick to instrument and tolerant to subject movement, and low cost with the use of single electrode equipment. In 2010,

Murugappan found in a study of human-computer interaction, a correlation between a user's emotion and EEG, providing useful information in understanding a user's reaction to advertisements (Murugappan et al., 2010). Therefore, an increasing amount of research into neuromarketing has turned to EEG as a key sensor in measuring emotion. Lucchiari and Pravettoni observed that EEG signals with a frequency of 16–31 Hz (i.e., Beta wave) could be modulated by the experience of pleasure when a consumer was presented with a favorite brand (Lucchiari and Pravettoni, 2012). In 2016, Wang, Chang and Chuang found that a narratives structure in video commercials induced higher EEG signals with a frequency band of 4–7 Hz (i.e., higher Theta) power of the left frontal region resulting in higher preference for branded products (Wang et al., 2016).

In psychology research, a person's emotion can be quantified through self-reported measures such as liking (valence) and excitability (awaken) (Poels and Dewitte, 2006; Smit et al., 2006). Questionnaires can be used to gather this type of information. According to the AIDA model, four quantified metrics are used to characterize the experience for a consumer watching an advertisement: attention, interest, desire, and action (Strong, 1925).

In the literature, Support Vector Machine has been widely used on EEG data; research on EEG based emotion recognition using frequency domain features and Support Vector Machine (SVM) was done by Wang et al. (2011) Research on EEG-based emotion recognition in music listening using Support Vector Machine (SVM) was done by Lin et al. (2009) Though other regression in binary results can also be used to build this model, in the literature, Support Vector Machine is the most widely used method in this field. Also, Support Vector Machine is suitable because of the sparse dataset the experiment uses. The technique used will build a prediction model based on several different brainwaves, which include frequency band less than 4Hz (i.e. Delta), frequency band between 4 and 7 Hz (i.e., Theta), frequency band 8–15 Hz (i.e., Alpha), frequency band 16–31 Hz (i.e., Beta) and frequency bigger than 32Hz (i.e., Gamma).

In this paper, we test the hypothesis that it is possible to use low-cost EEG equipment to collect brainwaves of subjects viewing advertisement, and to apply the latest methods from neuromarketing, and machine learning as a more accurate method of assessing advertisement impact and the likelihood of a person purchasing the advertised product than the current state-of-the-art.

## 2. METHODOLOGY

This section of the paper will describe the method used to collect data from a single-electrode wearable EEG device, self-reported measures for the impact of an advertisement, and train a predictive model against the data using SVM (Poels and Dewitte, 2006; Smit et al., 2006; Chen et al., 2014).

### 2.1. Data Collection

In the experiment, thirty right-handed male participants aged 20–35 from the University participated in this experiment as paid volunteers. Thirty participants were in the experiment, each of

them is given 4–5 advertisements, to create a sample size of 450, big enough to carry out statistical analysis. The experiment was carried in China, and the participants are bi-lingual in Chinese and English speaker. The participants had normal or corrected vision without any histories of neurological/mental diseases. The purpose of choosing bi-lingual participants is to avoid language-barrier caused by the advertisement's language content being in Chinese or English. We selected male participants to avoid biases from gender-specificity of the advertised products. Many products are specifically targeted at a particular gender, for example, male clothing and female clothing are targeted differently; as are hair and cosmetic products. We constructed a database of 220 TV advertisements from four gender-neutral or male-targeted products: cars (55 ads), digital products (55 ads), clothing (55 ads) and food (55 ads), which were randomly selected from Youku.com (one of the largest online media websites). Each advertisement was 15–20 s in length. The video resolution and audio volume of each video were normalised to the same level using professional video and audio editing software. In each test, 4–5 advertisements were randomly selected from the 220 TV advertisements database. The shortcomings and future research because of this design will be discussed in section 4.

Before the experiment, the volunteers received detailed instructions on all the tasks they would perform. Each participant was fitted with a single-electrode EEG headset by an experimenter, was seated comfortably in a lab room at 1.20 meters from a 19-inch PC monitor, and shown five advertisements randomly selected from every genre for a total of 20 advertisements. An E-prime system was used to control the presentation of the stimuli.

The experiment consisted of 4 blocks, each containing 20 trials. During every trial, the volunteers were presented with the advertising for about 15–20 s. The advertisement was followed by an evaluation questionnaire, including the willingness of the participant to purchase the advertised product (yes or no), and liking the advertisement (7-point Like scale).

Each volunteer performed two practice trials before the start of the formal experiment. The frontal EEG was recorded with the single-channel dry electrode-device and system (NeuroCAR1.0, Neuromanagement Lab, Zhejiang University, China). The integrated chip of the device was the ThinkGear (NeuroSky, Wuxi, China). The sampling rate of the device for gathering EEG signals was 512 HZ and the data saved into a computer. Four trials data were rejected due to voltage abnormality, and in total, 450 sets of brainwave data were used in our study.

## 2.2. Questionnaire Collection

The participants were asked to fill out a questionnaire with 22 questions asking about different aspects (objective and subjective) of the advertisement, or their experience. The questionnaire is listed in **Table 1**. The questionnaires are chosen to record different aspects of the advertising can impact the viewer's response to the advertising, which may indicate the likelihood of the viewer wanting to purchase the product. The aspects are chosen from several studies in the literature that each focus on one of two aspects of an advertisement's impact. M.

**TABLE 1** | The output dataset list.

	Output feature
$Y_1(1 - 7)$	Content Quality
$Y_2(1 - 7)$	Image Quality
$Y_3(1 - 7)$	Excitement
$Y_4(1 - 7)$	Attractiveness
$Y_5(1 - 7)$	Easiness for understanding
$Y_6(1 - 7)$	Clearness of the brand
$Y_7(1 - 7)$	Brand awareness
$Y_8(1 - 7)$	Familiarness of the brand
$Y_9(1 - 7)$	Willingness to buy
$Y_{10}(1 - 7)$	Intention to further learn the product
$Y_{11}(1 - 7)$	likeliness of memorize the advertisement content the next day morning
$Y_{12}(0/1)$	If the brand of the product can be memorized the next day
$Y_{13}(0/1)$	Chinese language or not
$Y_{14}(0/1)$	If it is moving
$Y_{15}(0/1)$	If it has significant vision and sound impact
$Y_{16}(0/1)$	If it is interesting
$Y_{17}(0/1)$	If it surprises you
$Y_{18}(0/1)$	Whether it has celebrities
$Y_{19}(0/1)$	If it is sexy
$Y_{20}(0/1)$	If it has children
$Y_{21}(0/1)$	If it has cartoon
$Y_{22}(0/1)$	If it is a story telling ads

Vaismoradi, Kimberly, and Klaus show that the audio and visual fidelity of the advertisement ( $Y_2$  and  $Y_{15}$  of the questionnaire) had an impact (Vaismoradi et al., 2013). The content impact are covered by questions  $Y_1$ ,  $Y_5$ , and  $Y_{14}$ . The brand quality is covered in ( $Y_6$ ,  $Y_7$ , and  $Y_8$ ). The overall feeling of the advertisement ( $Y_3$  and  $Y_4$ ) was found to make an impact on a customer's decision to purchase or not (Dahlén, 2002; Niazi et al., 2012). The conscious decision reported by customers on whether they would make the purchase or not ( $Y_9$  and  $Y_{10}$ ). Padgett and Douglas Allen showed that advertising memorability also plays a role in purchase power ( $Y_{11}$  and  $Y_{12}$ ). (Padgett and Allen, 1997) Other impacts include: Celebrity endorsement ( $Y_{18}$ ) (Bocheer and Nanjegowda, 2013; Srikanth et al., 2013); the feature of children or cartoons ( $Y_{20}$  and  $Y_{21}$ ) (Blatt et al., 1972; Fischer et al., 1991); the language ( $Y_{13}$ ) (Noriega and Blair, 2008); the narrative style and story-telling ( $Y_{16}$ ,  $Y_{17}$ ,  $Y_{22}$ ) (McQuarrie, 2002; Phillips and McQuarrie, 2002); and sexual-appeal ( $Y_{19}$ ) is widely established to have an impact on advertisement (Severn et al., 1990; Weller et al., 2015).

Eleven of the questions had binary answers of 0 or 1; another eleven were ranked answers from 1 to 7. The questions are listed in each column as output data (see **Table 1**).

## 2.3. Modeling

Raw EEG data is first augmented into frequency domain EEG signal before creating a larger dataset necessary for analysis. In this paper, the frequency domain EEG signal is listed in **Table 2**. Further details of input data  $X$  is listed in section 2.3.1

**TABLE 2** | The EEG band dataset list.

	Input feature	Meaning
$X_1$	Time Stamp	A sequence of numbers indicate the time index of raw signal
$X_2$	Signal Quality	Value ranges from 0 to 255
$X_3$	Raw	Voltage
$X_4$	Attention	Intensity of a user's level of mental focus/attention, value ranges from 0 to 100
$X_5$	Meditation	Level of a user's mental calmness/relaxation, value ranges from 0 to 100
$X_6$	Delta	>4 Hz
$X_7$	Theta	$\geq 4$ Hz and <8 Hz
$X_8$	Low Alpha	7.5–9 Hz
$X_9$	High Alpha	9.5–12.5 Hz
$X_{10}$	Low Beta	12–15 Hz
$X_{11}$	High Beta	15–18 Hz
$X_{12}$	Low Gamma	30–80 Hz
$X_{13}$	High Gamma	>80 Hz

The output data is labeled as  $Y$ . Each row of output data  $Y$  has 22 dimensions, and each dimension represents a feature, and all features are independent.  $Y_{450 \times 22}$  can be written as  $Y_{450 \times 22} = [Y_1, Y_2, \dots, Y_{22}]$  where  $Y_i$  represents each column vector of  $Y$ , and  $i = 1, 2, \dots, 22$ , see **Table 1**.

The purpose of this research is to use SVM to train a machine learning model that can find the map  $f$ , where  $Y = f(X)$ , the reason to use Support Vector Machine (SVM) is explained in section 2.3.4. Once found, any new input EEG signals can be translated into a prediction of an advertisement's impact on a consumer, and their likelihood of either a positive impression or willingness to buy the product.

Due to the format of the raw EEG signal data, data extraction is applied to amend the input data into the applicable format. Feature selection is conducted on input data, and label selection is conducted on output data. Among the thirty participants, data corresponding to twenty nine participants are used as in sample data. The sample workflow of this research is shown in **Figure 1**. 10% of the sample data is then separated as a testing dataset, with the remaining 90% used for training dataset. Bootstrapping is used to bootstrap the training dataset. The data corresponding to the remaining participant is used as out of sample test. The out of sample test is carried out by selecting each last person in the thirty participants and then averaged the thirty out of sample as judgment.

Support Vector Machine is applied as the machine learning method to predict the label given output data. Cross-validation and learning curve is eventually used to check over-fitting/under-fitting. We will elaborate each component in the workflow in the following sections.

### 2.3.1. Feature Extraction

Raw EEG signal is a voltage over time signal. Hence, a crucial process is needed to break the raw EEG signal into constituent frequencies (High alpha/beta/gamma/delta waves), etc. It is crucial because the majority of the informational content in EEG

signals are in the frequency domain, so breaking up the signal into constituent frequencies is most appropriate to apply SVM to classify EEG frequency domain data. Besides, according to the research that has been done by Dong et al. (2017), frequency data is better for feature extraction than the raw time domain data. Feature extraction, specifically Fast Fourier transform (FFT) is used to transform the raw EEG signal from time domain into frequencies domain (Heideman et al., 1985; Van Loan, 1992; Pritchard et al., 1994). We computed the EEG frequency band power using traditional EEG frequency band definitions (Delta: 1–3 Hz; theta: 4 Hz, alpha I: 89Hz, alpha II: 1,012 Hz, beta I: 1,317 Hz, beta II: 1,830 Hz, gamma I: 3,140 Hz, gamma II: 4,150 Hz), and these are time signals for each frequency band. Also, because the band EEG signal is collected from the four different product types, and with duration of each advertisement being different, data is also needed to be normalized. Timestamps are used to represent the duration, and advertisement lengths are normalized by cutting advertisements to the same length as the shortest advertisement in its category (it is 24 in cars, 13 in clothing, 28 in digital, and 16 in food).

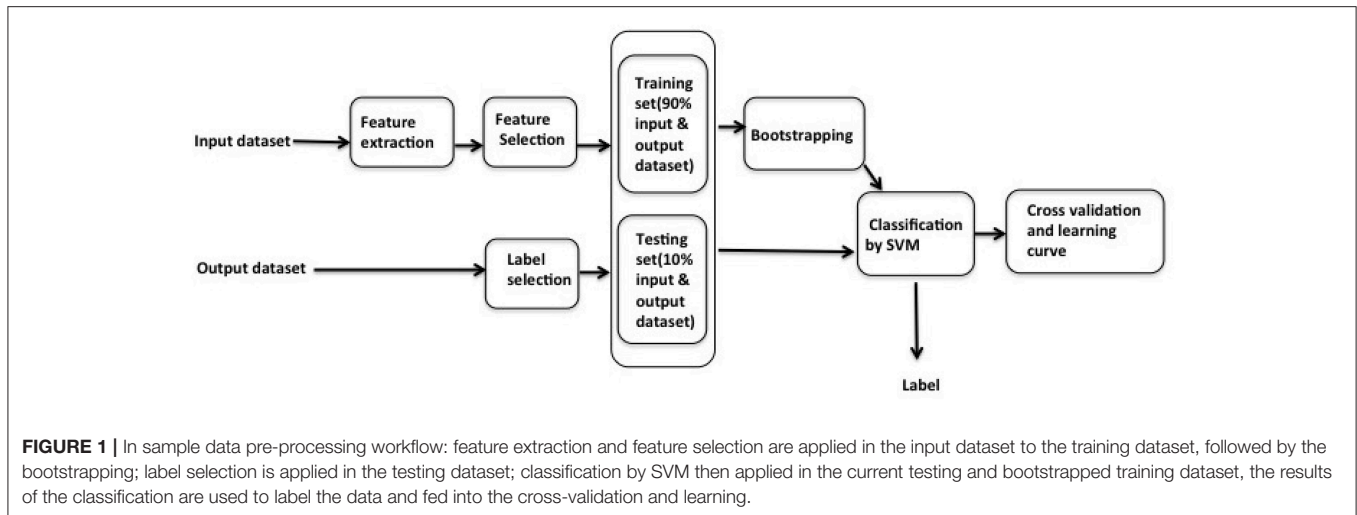
After normalization and extraction, the raw EEG signal is transformed into a  $450 \times 13$  dimension vector, where each of the elements themselves contain a vector of data corresponding to the advertisement length. The 13 features include wavelength, time, and signal quality and etc.  $X$  is a 13 columns, 450 rows vector,  $X_{450 \times 13}$  can be represented by 13 columns, i.e.,  $X_{450 \times 13} = [X_1, X_2, \dots, X_{13}]$ , each vector is  $X_i$  where  $i = 1, 2, \dots, 12, 13$ , then each column vector of  $X$  is listed in **Table 2**. However, in the experiment, selected columns  $X$  are used in the analysis, and they are Delta, Theta, Low Alpha, HighAlpha, LowBeta, HighBeta, LowGamma and HighGamma, Raw, Mediation and Attention, further selection details are in section 2.3.2.

### 2.3.2. Feature Selection and Label Selection

The map  $Y = f(X)$  can be expressed as  $label = f(features)$ . It is implied by the equation that a different  $X$  and  $Y$  will result in different  $f$ , and different certainty profiles. Feature extraction and selection is, therefore, a very important process in machine learning

Feature selection can be classified into three types: flat features, stream features, and structured features (Tomasi and Kanade, 1991; John et al., 1994; Chowdhury and Lavelli, 2012). In the experiment, each feature of the input data is independent, hence features in the experiment are classified as flat.

The 13 features are grouped into different combinations to test their ability to predict purchasing chance. In the end, 11 features give the best prediction in the model. These grouped-features also corresponding to the literature that is related to emotion: Delta, Theta, LowAlpha, HighAlpha, LowBeta, HighBeta, LowGamma, and HighGamma (Klimesch, 1999; Teplan, 2002). Attention, Meditation and Raw values (see **Table 2** are also considered in the model because their physical meaning is related to emotion. Output data  $Y$  is the information of questionnaire answers. In this research, we are using the customer's questionnaire score to determine whether participants will strike an emotion to purchase the product. Each label  $Y$  represents different response/measure to the advertisement, if it is 1, it meant positive



feedback is linked to emotion. Hence, we label the overall emotion as 1 (positive), and each label in the output data  $Y$  is independent. Different people can be influenced by different emotion, so we built a model that considers all emotions equally and then make the prediction. Questionnaire answers can be classified into two types: binary answers (yes or no answers) and ranked answers (values ranked between one and seven). All labels are selected and tested independently to see their significant impact on emotion. Due to the questionnaire answer types, labels are then grouped as binary answers, ranked answers and a combination of binary and ranked answers, each group's signification is also tested to the purchase possibility through emotion.

### 2.3.3. Data Augmentation via Bootstrapping

The original dataset contains 450 samples, and is a sparse dataset, and therefore bootstrapping is needed to increase the size of the dataset (Efron and Tibshirani, 1994; Efron et al., 2003). In the experiment, a dataset of twenty-nine people are used as in sample size of 435; while the other participant data of a sample size of 15 is used for out of sample testing. The 435 in sample size is too sparse to divide into a 90% training data (i.e., 393 samples) and a 10% testing data (i.e., 42 samples).

The in-sample data 391 samples are first divided into ten-folds under the condition that each fold has the same ratio of all type of answers in output data  $Y$ . Nine of the folds are used for training, while the tenth is used for testing, the ten folds algorithm is shown in Algorithm 2. Bootstrapping is used on each in sample training dataset. In bootstrapping process, Gaussian distribution bootstrapping method has been applied to each column in each training data  $X$ , shown in Algorithm 1. Gaussian process is a method to construct a parametric bootstrap approach from Bayesian non-parametric statistics, which implicitly considers the time-dependence of the data. (Efron, 1998) Many features are different frequency bands. Therefore, Gaussian is suitable also because it can generate continuous new data. However, Gaussian is a probability distribution between 0 and 1, while the data can be numbers even bigger than 1,000 and any number between 0 and 1 is too small to be the new bootstrapped number to

be used in the experiment, hence in the bootstrapping process, a scalar has been calculated to multiply the probability, and the scalar is calculated as  $scale = 0.5 \times (min + max)$ , where the min represents the minimum number of that column, and the max represents the maximum number of that column. The new bootstrapped number is the original value plus the scaled Gaussian probability. In this way, the original data spreading will not be affected. Mathematical combination is then applied in the ninefold training data, which is shown in Algorithm 3. The corresponding output data  $Y$  is duplicated to match the size of bootstrapped training data  $X$  in each of the bootstrapping processes.

#### Algorithm 1 Bootstrapping1

```

1: procedure GAUSSIAN PROCESS BOOTSTRAP TO EACH COLUMN
2:   Input : Gaussian model, input data  $X_{train}$ 
3:   Output : edited input data  $X_p$ 
4:   ▷ Use Gaussian process regression bootstrap the whole  $X_{train}$  dataset, name as bootstrapping1
5:   for Each column in  $X_{train}$  as  $i$  do
6:     ▷ Edit  $X_{train}$  as  $X_p$ 
7:      $X_p = X_{train}$ 
8:     ▷ select the maximum element
9:      $max = max(X_{p_i})$ 
10:    ▷ select the minimum element
11:     $min = min(X_{p_i})$ 
12:     $scale = 0.5 \times (min + max)$ 
13:    ▷ bootstrapping using Gaussian Distribution in  $(-1, 1)$ 
14:     $bootc = scale \times Gaussian\_bootstrap(X_p) + X_p$ 
15:    ▷ update  $X_p$  with all new columns
16:     $X_p = bootc$ 
  
```

### 2.3.4. Support Vector Machine

In literature, there are many different classification selection processes in machine learning. For example, Navie Bayes classifier, Random Forest, K-Nearest Neighbors, and Decision

Tree (Quinlan, 1987; Altman, 1992; Ho, 1995; Rennie et al., 2003). Support vector machine is a machine learning model used for classification and regression analysis (Aizerman et al., 1964; Boser et al., 1992; Jin and Wang, 2012). When SVM is used for classification, they separate a given set of binary labeled training data and a hyperplane that is maximally distance from them. Assume the input data is  $\mathbf{x}^i = (x_1^i \dots x_n^i)$  be the realization of the random vector  $\mathbf{x}^i$  While  $\phi$  is the map mapping the feature space to a label space  $\mathbf{y}$ , where label space contains many vectors, mathematically label as  $\{(\mathbf{x}^1, y^1), \dots, (\mathbf{x}^m, y^m)\}$ . The SVM learning algorithm finds a hyperplane  $(\mathbf{w}, b)$  such that the quantity

$$\gamma = \min_i y^i \{ \langle \mathbf{w}, \phi(\mathbf{x}^i) \rangle - b \} \quad (1)$$

is maximized. In this equation, the dimension of  $\phi$  is the same as the dimension of the label  $y$  and  $\langle \mathbf{w}, f(\mathbf{x}^i) \rangle - b$  corresponds to the distance between point  $\mathbf{x}^i$  and the decision boundary.  $\gamma$  is the margin and  $b$  is a real number. The kernel of this function is  $K_{i,j} = \langle \phi(\mathbf{x}^i), \phi(\mathbf{x}^j) \rangle$ . Given a new data  $x$  to classify, a label is assigned according to its relationship to the decision boundary, and the corresponding decision function is written as  $f(x) = \text{sign}(\langle \mathbf{w}, \phi(x) \rangle - b)$ .

In the experiment, as each label in the output is independent, and are binary (ranked answers are binarized via threshold as described below), therefore, regression that can end up with binary results are tested, i.e. non-linear regression, logistic regression and Support Vector Regression. All three methods have been utilized and proven appropriate in the previous research in EEG emotion detection (Schröder et al., 2001; Wang et al., 2011; Bejaei et al., 2015). However, Support Vector Machine is more appropriate on this type of sparse data (Wang et al., 2011). Besides, the experiment SVM has been shown to give the highest prediction accuracy result. The number of samples is larger than the number of features in our experiment, therefore the variable is chosen to reflect so; the loss function is set to 'Squared\_hinge' because it is commonly used in classification and 'Squared\_hinge' is convex and smooth and matches the function  $0 - 1$ , which is suitable for our experiment. 'Squared\_hinge' is mathematically written as  $(1 - yf(x))^2$ . The map  $f$  is linear. Therefore the kernel is zero. Regularization is  $L2$  as output data is no longer sparse, nor feature selection as features are targeted in output data (Ng, 2004).

In order to combine both binary and ranked answers, the rank answers are binarized using a thresholds according to the following rules:

- When the selected labels involve only the ranked answers, the threshold is set to be 3.5, 4 (i.e.  $= (7 + 1)/2$ ) and 4.5.
- When the selected labels involve only binary answers, the threshold is set to be 0, 0.5 (i.e.  $= (0 + 1)/2$ ) and 1.
- When the Threshold is a combination both types of answers, then the threshold can be written as equation:  $Thre = \alpha * Thre_7 + \beta * Thre_2$ , where  $\alpha$  and  $\beta$  are the weights

### 2.3.5. Cross-Validation and Learning Curve

When the SVM model is built up, cross-validation and learning curve is used to assess whether the model is over-fitted or under-fitted (Geisser, 1993; Kohavi et al., 1995; Babyak, 2004; Frost, 2015).

### Algorithm 2 Tenfolds

```

1: procedure TEN-FOLDS: IN SAMPLE TRAINING AND
   TESTING DATA SELECTION THROUGH INDEXES
2:   ▷ Ten-fold the whole dataset into ten sub dataset: same
   ratio of ones and zeros in nine subsets, combine the rest
   of unselected zeros and ones into the tenth subset
3:   Input : input data X, output data Y
4:   Output : nine different training input/output dataset
   Xtraini
5:   ▷ Find the index of ones in the Y
6:   index_ones = index_of_ones(Y)
7:   ▷ Find the index of zeros in the Y
8:   index_zeroes = index_of_zeroes(Y)
9:   ▷ Calculate the ratio of ones in the length of Y
10:  ratio = length(index_ones)/length(Y)
11:  ▷ After divide the Y into ten folds, find the number of
   ones in each fold
12:  ones = round(ratio * length(Y)/10)
13:  ▷ After divide the Y into ten folds, find the number of
   zeros in each fold
14:  zeros = round((1 - ratio) * length/10)
15:  ▷ Find the left over ones after first nine folds
16:  ones_left = length(index_ones) - 9 * ones
17:  ▷ Find the left over zeros after first nine folds
18:  zeros_left = length(index_zeroes) - 9 * zeros
19:  ▷ In the tenth fold, append the selected index of zeros and
   ones
20:  Xfold[10] = append(X[ones_left], X[zeros_left])
21:  Yfold[10] = append(Y[ones_left], Y[zeros_left])
22:  for i in 0:9 do
23:    ▷ Find the Y index of ones in each fold
24:    selection_ones = index_of_ones[i*ones:(i+1) * ones]
25:    ▷ Find the Y index of zero in each fold
26:    selection_zeroes = index_of_zeroes[i * zeros:(i + 1) *
   zeros]
27:    ▷ Find the Y index in each fold
28:    selection = append(selection_ones, selection_zeroes)
29:    ▷ Find the Yfold that is corresponding to the selected
   column index in Y
30:    Yfold[i] = Y[selection]
31:    ▷ Find the Xfold that is corresponding to the selected
   column index in each X
32:    Xfold[i] = X[selection]
33:  ▷ Xtrain and Ytrain is achieved by appending any nine
   of ten subsets in Xfold and Yfold
34:  Xtrain = append(Xfold[C(10, 9)])
35:  Ytrain = append(Yfold[C(10, 9)])

```

## 3. RESULT

### 3.1. Behavior Result

In this model, we use labels and EEG signals as variables, and emotion is the hidden variable bridge labels and EEG signals, they can be written as  $Y = f(X|E)$ , where  $E$  is emotion. The result of this experiment shows that if we have an individual's EEG

**Algorithm 3** Bootstrapping2

```

1: procedure BOOTSTRAPPING THE TRAINING X DATASET
   FROM TENFOLDS USING MATHEMATICAL COMBINATION
2:   Input : input data Xtrain
3:   Output : nine different training input dataset X2j
4:   ▷ Start with the original dataset Xtrain, but edit
   the selected column by the corresponding column in
   algorithm1
5:   X2 = Xtrain
6:   for Each column in X2 as j do
7:     ▷ make a copy of X2j for editing
8:     column = X2j
9:     for r in range(1, 12) do
10:      ▷ generate all unique iterations of column using
      nCr method, and then replace the corresponding
      index in the Xtrain while the rest of the columns
      remain the same
11:      newdata = nCriterations(12, r, row)
12:      X2j = replace_bootstrapping1(Xtrain[j]_replace
      Xp[newdata])

```

signal, which has been collected from the single electron device after watching the advertising, the accuracy of predicting if this individual would or not make a purchase of the corresponding product in the advertising is around 75% based on the SVM model. In the experiment, each participant has been selected as the out sample data, and then the rest of the participants are in sample data (90% of in sample data is the training dataset, while the 10% of the in-sample data is the testing dataset, bootstrapping has been applied to the 90% in sample training dataset.

The accuracy of prediction using SVM over the ranked answers is 77.28%. In this setting, the threshold of ranked answers results is 4. The recall score of this model is 72% and the F score for this model is 75%. In the same threshold of ranked answers result is 4, each category prediction is: the likelihood of purchasing the car is 63.5%, the likelihood of purchasing the cloth is 92.3%, the likelihood of purchasing the digital is 68.5% and the likelihood of purchasing the food is 82.76% (Table 3).

The accuracy of prediction using SVM model over combined ranked and binary answers is 75.4% under the conditions that ranked and binary answers have equal weighting (of 0.5), the threshold of ranked answers is 4, the threshold of a binary answer is 0.4, the threshold of the whole dataset is 0.4 (Table 8). The recall score of this model is 69% and the F score for this model is 71%. In the same experiment setting, each category prediction is: the likelihood of purchasing the car is 59.6%, the likelihood of purchasing the clothes is 92.3%, the likelihood of purchasing the digital is 64.8% and the likelihood of purchasing the food is 81.03% (Tables 4–7).

This indicates that EEG collected using single-electrode wearable devices is above 70% accuracy for prediction whether customers would purchase the product after watching the advertisement, and it can achieve higher accuracy prediction and reach about 75%.

The out of sample predictions are tested in two cases, and in each case, the results are the average/mean of picking up different

**TABLE 3** | Accuracy prediction of the ranked answer to different type of product at different thresholds.

Product type	Threshold 3.5	Threshold 4	Threshold 4.5
Car	0.365384615385	0.634615384615	0.826923076923
Food	0.603448275862	0.827586206897	0.931034482759
Digital	0.388888888889	0.685185185185	0.87037037037
Clothes	0.673076923077	0.923076923077	0.961538461538
All dataset	0.595090082962	0.772837217714	0.898908153808

**TABLE 4** | Likelihood of purchasing the car in the combined answer model at different thresholds.

Threshold to each type answer in car	Threshold of the car dataset		
	Thre = 0.4	Thre = 0.5	Thre = 0.6
<i>Thres<sub>7</sub></i> = 0.4, <i>Thres<sub>2</sub></i> = 3.5	0.365384615385	0.365384615385	0.826923076923
<i>Thres<sub>7</sub></i> = 0.4, <i>Thres<sub>2</sub></i> = 4	0.596153846154	0.596153846154	0.865384615385
<i>Thres<sub>7</sub></i> = 0.4, <i>Thres<sub>2</sub></i> = 4.5	0.75	0.75	0.903846153846
<i>Thres<sub>7</sub></i> = 0.5, <i>Thres<sub>2</sub></i> = 3.5	0.365384615385	0.365384615385	0.903846153846
<i>Thres<sub>7</sub></i> = 0.5, <i>Thres<sub>2</sub></i> = 4	0.634615384615	0.634615384615	0.903846153846
<i>Thres<sub>7</sub></i> = 0.5, <i>Thres<sub>2</sub></i> = 4.5	0.807692307692	0.807692307692	0.923076923077
<i>Thres<sub>7</sub></i> = 0.6, <i>Thres<sub>2</sub></i> = 3.5	0.365384615385	0.365384615385	0.980769230769
<i>Thres<sub>7</sub></i> = 0.6, <i>Thres<sub>2</sub></i> = 4	0.634615384615	0.634615384615	0.980769230769
<i>Thres<sub>7</sub></i> = 0.6, <i>Thres<sub>2</sub></i> = 4.5	0.807692307692	0.807692307692	1

thirty individual participants as out sample test. When it is only ranked answers, if the samples from the remaining participant from all four categories are combined into one dataset, then the prediction of purchasing power can reach 72.4%; if each product category is tested, then the car purchasing power can reach up to 50.09%, the clothing purchasing power can reach 78.01%, the digital products purchase power can reach 56.56% and the food purchase power can reach 69.34%. When it is the combined result of ranked and binary answers, if the samples from the remaining participant from all four categories are combined into one dataset, then the prediction of purchasing power can reach 71.2%; if each product category is tested, then the car purchasing power can reach up to 51.8%, the clothing purchasing power can reach 65.9%, the digital products purchase power can reach 57.2% and the food purchase power can reach 60.62%.

Also, in the research, each label Y has been tested, but it is a difficult test and the result is not balanced, the recall and F scores are both low, that is why those predictions are not mentioned in the results. Different thresholds for each model have also been tested during the experiment, however, they are not good enough to accurately explain the model.

Seventy-five percentage is a relatively high result to judge the purchasing power based on the EEG signal after watching the advertising, though there is no research on the direct link of EEG

**TABLE 5 |** Likelihood of purchasing the food in the combined answer model at different thresholds.

Threshold to each type answer in food	Threshold of the food dataset		
	Thre = 0.4	Thre = 0.5	Thre = 0.6
Thres <sub>7</sub> = 0.4, Thres <sub>2</sub> = 3.5	0.603448275862	0.603448275862	0.965517241379
Thres <sub>7</sub> = 0.4, Thres <sub>2</sub> = 4	0.810344827586	0.793103448276	0.965517241379
Thres <sub>7</sub> = 0.4, Thres <sub>2</sub> = 4.5	0.913793103448	0.913793103448	0.98275862069
Thres <sub>7</sub> = 0.5, Thres <sub>2</sub> = 3.5	0.603448275862	0.603448275862	0.98275862069
Thres <sub>7</sub> = 0.5, Thres <sub>2</sub> = 4	0.827586206897	0.827586206897	0.98275862069
Thres <sub>7</sub> = 0.5, Thres <sub>2</sub> = 4.5	0.931034482759	0.931034482759	1
Thres <sub>7</sub> = 0.6, Thres <sub>2</sub> = 3.5	0.603448275862	0.603448275862	0.948275862069
Thres <sub>7</sub> = 0.6, Thres <sub>2</sub> = 4	0.827586206897	0.827586206897	0.948275862069
Thres <sub>7</sub> = 0.6, Thres <sub>2</sub> = 4.5	0.931034482759	0.931034482759	0.948275862069

**TABLE 6 |** Likelihood of purchasing the digital in the combined answer model at different thresholds.

Threshold to each type answer in digital	Threshold of the digital dataset		
	Thre = 0.4	Thre = 0.5	Thre = 0.6
Thres <sub>7</sub> = 0.4, Thres <sub>2</sub> = 3.5	0.407407407407	0.407407407407	0.944444444444
Thres <sub>7</sub> = 0.4, Thres <sub>2</sub> = 4	0.648148148148	0.648148148148	0.962962962963
Thres <sub>7</sub> = 0.4, Thres <sub>2</sub> = 4.5	0.814814814815	0.814814814815	0.981481481481
Thres <sub>7</sub> = 0.5, Thres <sub>2</sub> = 3.5	0.388888888889	0.388888888889	0.981481481481
Thres <sub>7</sub> = 0.5, Thres <sub>2</sub> = 4	0.685185185185	0.685185185185	0.981481481481
Thres <sub>7</sub> = 0.5, Thres <sub>2</sub> = 4.5	0.87037037037	0.87037037037	0.981481481481
Thres <sub>7</sub> = 0.6, Thres <sub>2</sub> = 3.5	0.388888888889	0.388888888889	0.981481481481
Thres <sub>7</sub> = 0.6, Thres <sub>2</sub> = 4	0.685185185185	0.685185185185	0.981481481481
Thres <sub>7</sub> = 0.6, Thres <sub>2</sub> = 4.5	0.87037037037	0.87037037037	NA

signal and purchase intention. However, research has been done on using emotions to quantify the purchasing power. In 2006, John Pawle and Peter Cooper showed that emotional factors to brand decision making range from 63 to 85% (Tsai, 2005; Pawle and Cooper, 2006).

### 3.2. Cross-Validation and Learning Curve

The two models with the whole dataset are selected to draw the cross-validation and learning curve. The reasons for only

**TABLE 7 |** Likelihood of purchasing the clothes in the combined answer model at different thresholds.

Threshold to each type answer in clothes	Threshold of the clothes dataset		
	Thre = 0.4	Thre = 0.5	Thre = 0.6
Thres <sub>7</sub> = 0.4, Thres <sub>2</sub> = 3.5	0.653846153846	0.653846153846	1
Thres <sub>7</sub> = 0.4, Thres <sub>2</sub> = 4	0.923076923077	0.923076923077	1
Thres <sub>7</sub> = 0.4, Thres <sub>2</sub> = 4.5	0.961538461538	0.961538461538	1
Thres <sub>7</sub> = 0.5, Thres <sub>2</sub> = 3.5	0.692307692308	0.692307692308	1
Thres <sub>7</sub> = 0.5, Thres <sub>2</sub> = 4	0.923076923077	0.923076923077	1
Thres <sub>7</sub> = 0.5, Thres <sub>2</sub> = 4.5	0.961538461538	0.961538461538	1
Thres <sub>7</sub> = 0.6, Thres <sub>2</sub> = 3.5	0.692307692308	0.692307692308	0.980769230769
Thres <sub>7</sub> = 0.6, Thres <sub>2</sub> = 4	0.923076923077	0.923076923077	0.980769230769
Thres <sub>7</sub> = 0.6, Thres <sub>2</sub> = 4.5	0.961538461538	0.961538461538	0.980769230769

**TABLE 8 |** Likelihood of purchasing in the combined answer model at different thresholds.

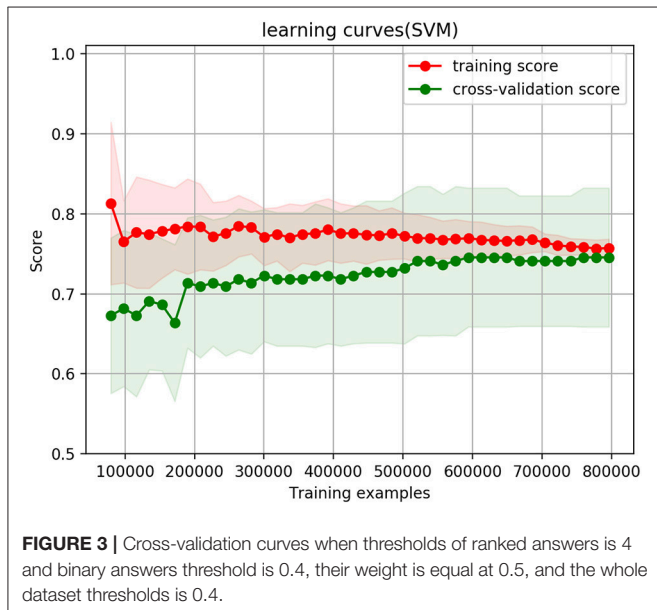
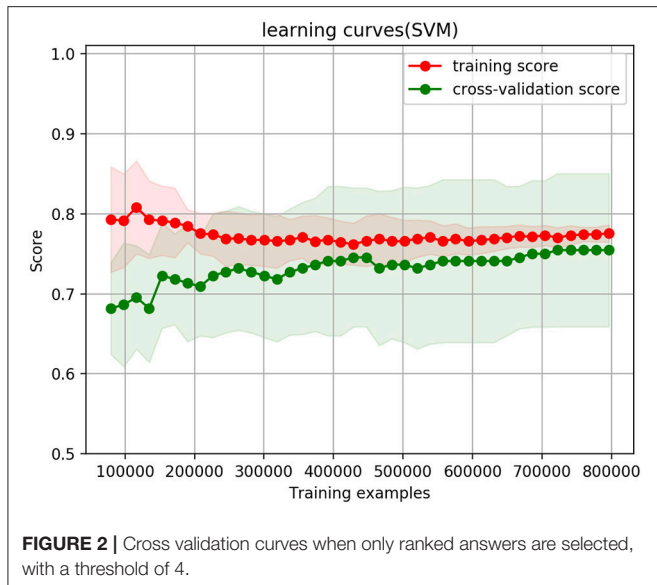
Threshold to each type answer in whole dataset	Threshold of the whole dataset		
	Thre = 0.4	Thre = 0.5	Thre = 0.6
Thres <sub>7</sub> = 0.4, Thres <sub>2</sub> = 3.5	0.6044	0.60022153401	0.940276239347
Thres <sub>7</sub> = 0.4, Thres <sub>2</sub> = 4	0.754142459932	0.758595745643	0.950177453263
Thres <sub>7</sub> = 0.4, Thres <sub>2</sub> = 4.5	0.866265795601	0.866265795601	0.968578339399
Thres <sub>7</sub> = 0.5, Thres <sub>2</sub> = 3.5	0.595090082962	0.595090082962	0.977326672243
Thres <sub>7</sub> = 0.5, Thres <sub>2</sub> = 4	0.772837217714	0.772837217714	0.977326672243
Thres <sub>7</sub> = 0.5, Thres <sub>2</sub> = 4.5	0.894454868097	0.894454868097	0.977326672243
Thres <sub>7</sub> = 0.6, Thres <sub>2</sub> = 3.5	0.595090082962	0.595090082962	0.977326672243
Thres <sub>7</sub> = 0.6, Thres <sub>2</sub> = 4	0.772837217714	0.772837217714	0.977326672243
Thres <sub>7</sub> = 0.6, Thres <sub>2</sub> = 4.5	0.894454868097	0.894454868097	0.977326672243

selecting these two are they give the best prediction results, and they are directly related to the paper hypothesis: predicting the likelihood of customers purchasing the products.

The cross-validation and learning curve is shown in **Figures 2, 3**.

**Figure 2** shows the cross-validation and learning curve when only the ranked answers are selected. **Figure 3** shows the cross-validation and learning curve when the thresholds of ranked





answers is 4, and binary answers threshold is 0.4, their weight is equal at 0.5, and the whole dataset thresholds is 0.4. In both figures, the X-axis represents the size of training examples; Y-axis represents the accuracy prediction score. The red lines represent the training score, and green lines represent cross-validation score/testing score.

In the ranked answers only case in **Figure 2**, training and testing curves converge to 77.28%; when the combination of ranked answers and binary answers are selected (ranked answers threshold is 4, and binary answers threshold is 0.4), as shown in **Figure 3**, the training and testing curves converge to 75.4%. The shaded red and shaded green are standard deviations at confidence interval 10% to its corresponding training score and corresponding testing score. Mathematically, they are calculated

as  $(training_{mean} - training_{std}, training_{mean} + training_{std})$  and  $(testing_{mean} - testing_{std}, testing_{mean} + testing_{std})$ . Both figures show that the training curves are slightly over-fitting at the beginning, but it gradually decrease and reach a stable point with the decreasing variance. In testing curves, the variance remains approximately the same from the beginning till the end throughout the training example size increases; it remains the same because the testing data size does not change regardless of how training example size changes. Convergence of training and testing curves means there is no over-fitting and no under-fitting issues of the model; the SVM model we set up and use therefore performs well.

## 4. DISCUSSION

The approximately 75% accuracy prediction and the converging cross-validation curve result together show that the EEG signal and edited AIDA metrics model we found is a suitable model. This result is sufficient to declare EEG a useful dataset to collect in the advertising industry.

Further improvements can be made to the accuracy of prediction through improving the chosen thresholds and increasing the study sample size rather than relying on bootstrapping. The accuracy of prediction may be further improved through selecting both male and female participants and gender-neutral products and their advertisement.

The statistical results show that the ranked answers with a threshold of 4 has a better result than the combined rank and binary answers, with about 2% more accuracy. It also shows that each category prediction is lower than the whole dataset prediction in both models. It shows products like cars have the lowest accuracy of prediction and clothes have the highest accuracy of prediction among the four categories. This may be caused by cars being a costly product, and the decision to purchase one depends more on lifestyle and personal circumstance than advertising. In this case, the prediction accuracy can be improved by pre-selecting participants to control for these factors, or by adjusting predictions based on the personal circumstances. In the experiment, the F scores and the recall scores are also good in the model; it happens to be consistent that the whole dataset scores are better than each category dataset. The out of sample test result is not as good as the sample test, but it is acceptable, which further indicates the model is good.

In the current research, the emotions and opinions of the user have been used as a hidden variable to bridge EEG signal and self-reported metrics to evaluate the impact of advertisement and its power to influence purchasing. In the literature, research has been done on the relationship of Theta wavelength with emotion and the relationship of Beta Wavelength with emotion (Lucchiari and Pravettoni, 2012; Wang et al., 2016). Considering that EEG signals contain components at many other frequencies, it is worth further investigation of the relationship between EEG signal and emotion, and how they impact a decision to make a purchase.

In past attempts, an AIDA model has been used to quantify emotion metrics in terms of scores for attention, interest, desire

and action aspects after watching the advertisement (Strong, 1925). In our research, we follow the idea of the AIDA model and add additional dimensions to quantify these emotions to achieve a more precise measure. The ranked and binary answers have equal weighting in accordance with the AIDA model and literature mentioned in section 2.1 which considers all labels with equal impact. It is also interesting to further execute a non-parametric machine learning model to assess the importance of each label or a grouping of similar labels for purchasing power. More details on the emotion model our experiment uses are described in section 2.1.

This is a new method in Neuromarketing. The advantage of this method is to be able to assess impact fast. The model, once found, can work in real-time on EEG signals, in a highly automated way. Something not possible in traditional methods that involve data collection via surveys or discussion, and then analysis, which requires both time and man-power to accomplish; or assessing the impact through sales figures afterwards, which can be easier to automate, but has much longer cycle times.

Our proposed evaluation method also shows less bias due to EEG data being involuntary and therefore not subject to conscious and experimental bias. Unlike in a focus group study conducted by an advertisement company, the study in an academic setting removes the potential for study biases in which participants tend to give kinder answers than they may think, out of goodwill or some sense of social obligation (De Pelsmacker et al., 2002; Maison et al., 2004; Shen and Li, 2009). The questionnaire questions are written neutrally, and asks similar questions from slightly different angles to validate consistency. Because these questions yield consistent results, it indicates a high likelihood that they represent the true thought of participants. The questions in the questionnaire do not include identifiable or confidential information of any individual; therefore removing potential reasons for participants to hide their true opinion.

Furthermore, the device used in this research is a single-electrode wearable device that is low cost, portable, and easy to use, allowing this kind of data collection to be scaled much more rapidly than machines traditionally associated with functional brain imaging in neuroscience, such as traditional clinical-grade multi-channel EEG setups, and fMRI (Signal, 2015). Multiple types of EEG devices have been studied in the literature and applied to neuromarketing studies; the results do not show an appreciable improvement in accuracy of multi-channel EEG systems vs. single electrode EEG systems (Hamzy and Dutta, 2000; Liu et al., 2013). Therefore, single electrode EEG devices is sufficient for the experiment, and the ease of use of a simple, self-contained and battery-operated wearable device opens up new kind of customer engagement opportunities where their EEG can be recorded throughout the day in settings that are much more normal than would be the case in a lab setting, removing any influence that may come from the experimental setup.

The most significant part of this research is the extension of modeling emotions with a single-wavelength collected from EEG to multidimensional wavelengths, allowing the extraction of more informational content from EEG than previously attempted.

The result of this research can be further applied to individual consumers behavior; to allow the advertising industry to tailor their advertisement for maximum impact; and adapted to work for TV and movie studios to predict viewership rates of movies and TV shows from trailers.

## 5. CONCLUSION

We proposed a model that can use EEG signals measured using low-cost consumer-grade EEG headsets taken while a consumer watches an advertisement, to rapidly predict the consumer's likelihood of purchasing the product. While further research can be made on the selection of thresholds, and the quality of the result can be improved with the collection of larger datasets, the method as shown is nevertheless easy to deploy, yields rapid results, scales better than any existing method, and introduces less experimental and environmental bias. If employed in place of existing focus-group studies, any company involved in mass-media advertising stands to improve the effectiveness of their advertising, improve estimates of the impact on sales of their advertising, and be more informed when building their advertising strategy, leading to increased ROI.

## ETHICS STATEMENT

The experiment was carried out at Zhejiang University and the approval was obtained from the Ethics committee at the Zhejiang University, China. In addition, written informed consent forms were obtained from all volunteers before the experiment started. The ethics approval number from Zhejiang University is NSL20160012.

## AUTHOR CONTRIBUTIONS

ZW carried out the experiment and wrote the paper. CW and AS edited the paper. PW prepared the experiment. XW prepared the experiment data and provided advice with some of the editing. YG supervised the project.

## FUNDING

This work was supported by Grant No. 14YJC630129 from the Humanities and Social Sciences Foundation of the Ministry of Education of China and Grant No. 71572176 from the National Natural Science Foundation of China.

## ACKNOWLEDGMENTS

This is a collaboration work between School of Management Zhejiang University China and Data Science Institute, Imperial College London. We would also like to show our gratitude to the Jaywing company for sharing their pearls of wisdom and their insights with us during this research.

## REFERENCES

- Aizerman, A., Braverman, E. M., and Rozner, L. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Automat. Remote Control* 25, 821–837.
- Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Statist.* 46, 175–185. doi: 10.1080/00031305.1992.10475879
- Babiyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom. Med.* 66, 411–421. doi: 10.1097/01.psy.0000127692.23278.a9
- Bejaei, M., Wiseman, K., and Cheng, K. (2015). Developing logistic regression models using purchase attributes and demographics to predict the probability of purchases of regular and specialty eggs. *Br. Poul. Sci.* 56, 425–435. doi: 10.1080/00071668.2015.1058917
- Blatt, J., Spencer, L., and Ward, S. (1972). A cognitive development study of children's reactions to television advertising. *Telev. Soc. Behav.* 4, 452–467.
- Bocheer, K., and Nanjegowda, H. (2013). The impact of celebrity advertisement on Indian customers. *CLEAR Int. J. Res. Comm. Manag.* 4, 59–65.
- Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory* (Pittsburgh, PA: ACM), 144–152.
- Bradley, J. (2015). *Big Spenders on a Budget: What the Top 200 US Advertisers are Doing to Spend Smarter*. Advertising Research Foundation.
- Brady, S. (2014). What percent of revenue do publicly traded companies spend on marketing and sales. *Vital*.
- Chen, Y. H., de Beecq, M. O., Vanderheyden, L., Carrette, E., Mihajlović, V., Vanstreels, K., et al. (2014). Soft, comfortable polymer dry electrodes for high quality ECG and EEG recording. *Sensors* 14, 23758–23780. doi: 10.3390/s141223758
- Chowdhury, M. F. M., and Lavelli, A. (2012). "Combining tree structures, flat features and patterns for biomedical relation extraction," in *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (Avignon: Association for Computational Linguistics), 420–429.
- Dahlén, M. (2002). Thinking and feeling on the world wide web: the impact of product type and time on world wide web advertising effectiveness. *J. Market. Commun.* 8, 115–125. doi: 10.1080/13527260210142347
- De Pelsmacker, P., Geuens, M., and Anckaert, P. (2002). Media context and advertising effectiveness: the role of context appreciation and context/ad similarity. *J. Advert.* 31, 49–61. doi: 10.1080/00913367.2002.10673666
- Dong, H., Supratak, A., Pan, W., Wu, C., Matthews, P. M., and Guo, Y. (2017). Mixed neural network approach for temporal sleep stage classification. *IEEE Trans Neural Syst Rehabil Eng.* 26, 324–333. doi: 10.1109/TNSRE.2017.2733220
- Ducoffe, R. H. (1996). Advertising value and advertising on the web. *J. Adver. Res.* 36, 21–21.
- Efron, B. (1998). *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia, PA: SIAM
- Efron, B., and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. Boca Raton, FL: CRC Press.
- Efron, B. (2003). Second thoughts on the bootstrap. *Statist. Sci.* 18, 135–140. doi: 10.1214/ss/1063994968
- Elliott, M. T., and Speck, P. S. (1998). Consumer perceptions of advertising clutter and its impact across various media. *J. Adver. Res.* 38, 29–30.
- Fischer, P. M., Schwartz, M. P., Richards, J. W., Goldstein, A. O., and Rojas, T. H. (1991). Brand logo recognition by children aged 3 to 6 years: Mickey mouse and old joe the camel. *JAMA* 266, 3145–3148.
- Frost, J. (2015). The danger of overfitting regression models. *Minitab Blog*.
- Gaines, B. J., Kuklinski, J. H., and Quirk, P. J. (2007). The logic of the survey experiment reexamined. *Polit. Anal.* 15, 1–20. doi: 10.1093/pan/15/1/008
- Geisser, S. (1993). *Predictive Inference, Vol. 55*. Boca Raton, FL: CRC Press.
- Goldberg, M. E. (1990). A quasi-experiment assessing the effectiveness of tv advertising directed to children. *J. Market. Res.* 27, 445–454. doi: 10.2307/3172629
- Grewal, D., Monroe, K. B., and Krishnan, R. (1998). The effects of price-comparison advertising on buyers' perceptions of acquisition value, transaction value, and behavioral intentions. *J. Market.* 62, 46–59. doi: 10.2307/1252160
- Hamzy, M. and Dutta, R. (2000). *Visual and Audible Consumer Reaction Collection*. US Patent App. 09/731,870.
- Heideman, M. T., Johnson, D. H., and Burrus, C. S. (1985). Gauss and the history of the fast fourier transform. *Arch. Hist. Exact Sci.* 34, 265–277. doi: 10.1007/BF00348431
- Ho, T. K. (1995). "Random decision forests," in *Proceedings of the Third International Conference on Document Analysis and Recognition, 1995*, Vol. 1 (Montreal, QC: IEEE), 278–282.
- Jin, C., and Wang, L. (2012). "Dimensionality dependent pac-bayes margin bound," in *Advances in Neural Information Processing Systems* (Lake Tahoe, NV), 1034–1042.
- John, G. H., Kohavi, R., and Pfleger, K. (1994). "Irrelevant features and the subset selection problem," in *Proceedings of the Eleventh International Conference on Machine Learning* (New Brunswick, NJ), 121–129.
- Kanetkar, V., Weinberg, C. B., and Weiss, D. L. (1992). Price sensitivity and television advertising exposures: some empirical findings. *Market. Sci.* 11, 359–371. doi: 10.1287/mksc.11.4.359
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev.* 29, 169–195. doi: 10.1016/S0165-0173(98)00056-3
- Kohavi, R. (1995). "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, Vol. 14 (Stanford, CA), 1137–1145.
- Lewis, R., and Reiley, D. (2009). "Retail advertising works! measuring the effects of advertising on sales via a controlled experiment on yahoo!," in *The FTC Microeconomics Conference, and Economic Science Association Meetings* (Tucson, AZ; Pasadena, CA; Lyon).
- Lin, Y.-P., Wang, C.-H., Wu, T.-L., Jeng, S.-K., and Chen, J.-H. (2009). "EEG-based emotion recognition in music listening: a comparison of schemes for multiclass support vector machine," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009* (IEEE), 489–492.
- Liu, Y., Sourina, O., and Hafiyandi, M. R. (2013). "EEG-based emotion-adaptive advertising," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII)* (IEEE), 843–848.
- Lucchiari, C., and Pravettoni, G. (2012). The effect of brand on EEG modulation. *Swiss J. Psychol.* 71, 199–204. doi: 10.1024/1421-0185/a000088
- Maison, D., Greenwald, A. G., and Bruin, R. H. (2004). Predictive validity of the implicit association test in studies of brands, consumer attitudes, and behavior. *J. Cons. Psychol.* 14, 405–415. doi: 10.1207/s15327663jcp1404\_9
- McQuarrie, E. F. (2002). The development, change and transformation of rhetorical style in magazine advertisements 1954–1999. *J. Adver.* 31, 1–13.
- Murugappan, M., Ramachandran, N., and Sazali, Y., (2010). Classification of human emotion from EEG using discrete wavelet transform. *J. Biomed. Sci. Eng.* 3, 390–396. doi: 10.4236/jbise.2010.34054
- Ng, A. Y. (2004). "Feature selection, L1 vs. L2 regularization, and rotational invariance," in *Proceedings of the Twenty-First International Conference on Machine Learning* (Banff, AB: ACM), 78.
- Niazi, G. S. K., Siddiqui, J., Alishah, B., and Hunjra, A. I. (2012). Effective advertising and its influence on consumer buying behavior. *Inform. Manag. Bus. Rev.* 4, 114–119.
- Noriega, J., and Blair, E. (2008). Advertising to bilinguals: does the language of advertising influence the nature of thoughts? *J. Market.* 72, 69–83. doi: 10.1509/jmkg.72.5.69
- Ostrovsky, M., and Schwarz, M. (2011). "Reserve prices in internet advertising auctions: a field experiment," in *Proceedings of the 12th ACM Conference on Electronic Commerce* (San Jose, CA: ACM), 59–60.
- Padgett, D., and Allen, D. (1997). Communicating experiences: a narrative approach to creating service brand image. *J. Advert.* 26, 49–62.
- Pawle, J., and Cooper, P. (2006). Measuring emotion—lovemarks, the future beyond brands. *J. Adver. Res.* 46, 38–48. doi: 10.2501/S00218499060060053
- Phillips, B. J., and McQuarrie, E. F. (2002). The development, change, and transformation of rhetorical style in magazine advertisements 1954–1999. *J. Adver.* 31, 1–13. doi: 10.1353/asr.2006.0010
- Poels, K., and Dewitte, S. (2006). How to capture the heart? Reviewing 20 years of emotion measurement in advertising. *J. Adver. Res.* 46, 18–37. doi: 10.2501/S00218499060060041
- Pritchard, W. S., Duke, D. W., Coburn, K. L., Moore, N. C., Tucker, K. A., Jann, M. W., et al. (1994). EEG-based, neural-net predictive classification

- of alzheimer's disease versus control subjects is augmented by non-linear EEG measures. *Electroencephalogr. Clin. Neurophysiol.* 91, 118–130. doi: 10.1016/0013-4694(94)90033-7
- Quinlan, J. R. (1987). Simplifying decision trees. *Int. J. Man Mach. Stud.* 27, 221–234. doi: 10.1016/S0020-7373(87)80053-6
- Rennie, J. D., Shih, L., Teevan, J., and Karger, D. R. (2003). "Tackling the poor assumptions of naive bayes text classifiers," in *ICML Vol. 3* (Washington, DC), 616–623.
- Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M., and Gielen, S. (2001). "Acoustic correlates of emotion dimensions in view of speech synthesis," in *Seventh European Conference on Speech Communication and Technology* (Aalborg).
- Severn, J., Belch, G. E., and Belch, M. A. (1990). The effects of sexual and non-sexual advertising appeals and information level on cognitive processing and communication effectiveness. *J. Adver.* 19, 14–22.
- Sharma, G. D., Mahendru, M., and Singh, S. (2011). Advertisement cause sales or sales cause advertisement: a case of Indian manufacturing companies. *SSRN*.
- Shen, G. B., and Li, S. (2009). *Dynamic Earch with Implicit User Intention Mining*. U.S. Patent 7,599,918.
- Signal, N. B. (2015). *Of Neurosky, Inc. Neorosky*. Available online at: <https://www.ant-neuro.com/>
- Smit, E. G., Van Meurs, L., and Neijens, P. C. (2006). Effects of advertising likeability: a 10-year perspective. *J. Adver. Res.* 46, 73–83. doi: 10.2501/S0021849906060089
- Srikanth, J., Saravanakumar, M., and Srividhya, S. (2013). The impact of celebrity advertisement on Indian customers. *Life Sci. J.* 10, 59–65.
- Strong, E. K. (1925). Theories of selling. *J. Appl. Psychol.* 9:75. doi: 10.1037/h0070123
- Sundar, S. S., and Kalyanaraman, S. (2004). Arousal, memory, and impression-formation effects of animation speed in web advertising. *J. Adver.* 33, 7–17. doi: 10.1080/00913367.2004.10639152
- Teplan, M. (2002). Fundamentals of EEG measurement. *Measur. Sci. Rev.* 2, 1–11.
- Tomasi, C., and Kanade, T. (1991). *Detection and Tracking of Point Features*. Pittsburgh, PA: School of Computer Science, Carnegie Mellon University.
- Tsai, S.-P. (2005). Utility, cultural symbolism and emotion: a comprehensive model of brand purchase. value. *Int. J. Res. Market.* 22, 277–291. doi: 10.1016/j.ijresmar.2004.11.002
- Vaismoradi, M., Turunen, H., and Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nurs. Health Sci.* 15, 398–405. doi: 10.1111/nhs.12048
- Van Loan, C. (1992). *Computational Frameworks for the Fast Fourier Transform*. Philadelphia, PA: SIAM.
- Wang, X.-W., Nie, D., and Lu, B.-L. (2011). "EEG-based emotion recognition using frequency domain features and support vector machines," in *International Conference on Neural Information Processing* (Berlin; Heidelberg: Springer), 734–743.
- Wang, R. W. Y., Chang, Y. C., and Chuang, S. W. (2016). EEG spectral dynamics of video commercials: impact of the narrative on the branding product preference. *Sci. Reports* 6:36487. doi: 10.1038/srep36487
- Weller, R. B., Sibley, S. D., and Neuhaus, C. (2015). "Experimental results concerning the affect of the female model in television commercials on product and brand recall," in *Proceedings of the 1982 Academy of Marketing Science (AMS) Annual Conference* (Las Vegas, NV: Springer), 468–472.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Wei, Wu, Wang, Supratak, Wang and Guo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.