Selection of Stable Features for Modeling 4-D Affective Space from EEG Recording

Rakib Al-Fahad, Mohammed Yeasin, ASM Iftekhar Anam, Bahareh Elahian
Department of Electrical and Computer Engineering
The University of Memphis, Memphis, TN 38152, USA
Email: {ralfahad, myeasin, aanam, belahian}@memphis.edu

***Abstract*—Recent advances in neuroscience made it possible to understand how the human brain processes emotions and affective states. However, the modeling of emotion remains elusive due to inherent ambiguity and complexity related to the perception of emotions, interpersonal variabilities, and context-specific interpretations. Here, we present a robust method of modeling 4-D continuous affective space (Valence, Arousal, Like, and Dominance). First, we determined the functional areas and frequency bands related to 4-D affective space. Second, we extracted and selected a set of stable features. For both steps, we used two different feature selection methods namely: Recursive Feature Elimination (RFE) and stability selection method. Moreover, compare their performances. For the RFE, we used Random Forest (RF), Support Vector Regression (SVR), Tree-based bagging, and for the stability selection, we used Randomized Lasso as an estimator. Empirical analysis on the DEAP data set shows that the stability selection method consistently provides relevant set of bands, electrode location and features over a range of model parameters. We also observed that only a small number of locations (40%-63%) and certain frequency bands specifically, gamma band frequency over Superior Temporal Gyrus, Supramarginal Gyrus, and Somatosensory Association Cortex were the highest ranked features across the affective dimensions. The selected features using the stability criteria were used to model 4-D affective space using SVR. Empirical analyses shows that the Root Mean Square Error (RMSE) for Valence, Arousal, Dominance, and Like are 2.13, 2.00, 2.07, and 2.11 respectively. In addition, we also compare the performance of this method with feature fusion and ensemble classification. It was observed that the SVR with selected features outperformed all other approaches. The predicted Valence-Arousal-Dominance were converted to categorical emotions for seamless interpretation.**

I. Introduction

The inherent ambiguity and complexity involved in the perception of emotion and its intensity, frequency, and duration introduce significant variability that complicates modeling affects. Appraisal, Categorical, and Dimensional are commonly used approaches for modeling affect [1]. The appraisal-based approach is feed-forward only, begins from a stimulus and ends with a response. It does not account for body loop. This model needs complex measurements of changes in all steps involved in the process [2]. In the categorical approach, affects are expressed using a small number of basic emotions (for example, happy, disgust, sad, surprise). This approach can depict some complex, non-basic, and delicate affects (e.g., depression or embarrassment) but fails to illustrate the affective states [3]. To overcome such limitation, the dimensional theory is used to define affect according to one or more continuous dimensions. Those dimensions are considered independent of each other. Valence-Arousal [4], Pleasure-Arousal-Dominance (PAD) model [5] are among popular dimensional model. Despite the controversy, the 4-D Valence-Arousal-Dominance-Like is widely used dimensions to describe emotions and affective states. (e.g. [1], [3], [5]. Key challenges in the modeling of continuous affective space are (but are not limited to): 1) a database with ground truth information, 2) framework to extract meaningful features, 3) find the relevant cortical area and EEG frequency bands, 4) select stable and relevant features for better generalization and interpretation, and 5) sophisticated methods that can model affect in presence of technical noise and inter-subject variability.

The DEAP (A Database for Emotion Analysis using Physiological signals) dataset [6] partially addresses the database issue. However, methodologies used in different studies (e.g. [6]–[9]) to analyze DEAP dataset are elemental in nature. A plethora of literature also reported modeling emotions from EEG data (e.g. [6], [10]–[13]). These studies are limited to modeling categorical emotions and usually use a naïve approach to analyze each predictor individually and then compensate for possible error arising from multiple comparisons (e.g., family-wise error or false discovery rate). These studies mostly yield unstable set of features.

Variable selection attempts to identify the most salient subset of variables from a larger set of features mixed with irrelevant variables. This problem is especially challenging when the number of available data samples are smaller compared to the number of possible predictors. Using generic subsampling and high-dimensional selection algorithms, stability selection not only yields a stable set of features but also provides the family-wise error control. It has previously been used in diverse fields of science including gene selection and neuroimaging [14], [15]. One of the downsides of the multivariate approaches is that outcomes often depend on model parameters (e.g., regularization factor). Compared to conventional multivariate approaches, stability selection produces more reliable estimations because of its internal randomization implemented as bootstrap based subsampling [16], [17].

We propose a systematic approach to determine the role of relevant functional areas of brain and frequency bands and

978-1-5090-6182-2/17/$31.00 ©2017 IEEE

select features that are consistent across model parameters. In particular, we compare the performances of Recursive Feature Elimination (RFE) and stability selection method in our framework to select relevant and stable features. For the RFE, we used Random Forest (RF), Support Vector Regression (SVR), Tree-based bagging, and for the stability selection we used Randomized Lasso as an estimator. We also developed a robust model to predict the 4-D affective space and converted them to categorical emotions for seamless interpretation. Empirical analyses were performed to determine the functional areas of brain and frequency bands related to emotions and affective states, selecting a stable set of features, and modeling 4-D continuous affective space.

The rest of the paper is organized as follows: In Section II, we discuss the context of the research. Subsequently, dataset and methods used in this work are presented in Section III. Following this, Section IV presents empirical analyses on DEAP dataset explaining the role of functional areas of the brain and frequency bands in modeling of continuous affective space. We also present the performance comparison of various models. Finally, Section V concludes the paper with lessons learned and a few remarks on future direction.

II. Research context

Major brain waves (θ, α, β, and γ band) are correlated with emotions and affective states. For example, the δ wave represents deep unconsciousness, intuition, and insight; the θ wave represents creativity and deep relaxation; the α wave represents pacey and dreamy state of mind; the β wave represents conscious thought and external focus; the γ wave is linked to perception, alertness or anxiety [18]. Different frequency bands and functional areas of the brain a have significant role in modeling perception, action, and cognition. For example, Valence has a high correlation with all frequency bands [6], an increase in θ and α band power correlates with positive Valence in the occipital regions (over visual cortices) [19]. Onton et al. [20] suggested a positive correlation between Valence and high-frequency power (mainly γ bands) that emerges from anterior temporal cerebral sources. Cole et al. [21] described that positive emotional self-induction and external stimulation is related to an increase of beta power over the right temporal sites. Arousal involves activation mainly of the reticular activating system, autonomic nervous system, as well as endocrine system. [21]. The central α power decreases with the higher levels of Arousal. It shows negative correlations in the θ, α, and γ band [6]. Dominance ranges from a helpless or submissive or weak feeling (without control) to an empowered feeling (in control of everything). The lower value of Dominance separates anger, hostility, and contempt from negative emotions like shyness, sadness, shame, etc. On the other hand, it also distinguishes self-confidence from arrogance, cockiness or awe [22]. Liking increases θ and α band power activities over left frontocentral cortices [6]. We have adopted RFE and stability selection method to find out certain functional areas of brain and frequency bands that carry most discriminative information and build a robust model.

Feature selection is an age-old problem in statistics. It improves generalization, reduces computational cost and, defying the curse of dimensionality, and enhance system interpretability and efficiency [23]–[25]. Much of the available literature on feature selection from EEG data are commonly based on correlation, F-test, chi-square test, ANOVA analysis, SVM ranking, Wrapper based methods, GA Wrapper, Principal Component Analysis, Linear Discriminant Analysis, and Forward-Backward based methods [6], [10]–[13]. For example, Shan et al. [26] investigated the use of feature ranking criteria to select the key EEG features from a broad pool of available features. They combined Random Forest with the RFE approach for better performance. Similarly, Rakib et al. [27] used RFE to select relevant frequency band and electrode location. Conventional feature selection methods are not stable and may select irrelevant or noise variables degrading the modeling performance. To overcome such limitation, we adopted Randomized Lasso stability selection method [16]. Meinshausen et al. [16] proved that, even if the necessary conditions needed for consistency of the original Lasso models are violated, stability selection with Randomized Lasso will be consistent in feature selection.

III. Method

A. Dataset

In DEAP dataset [6], thirty-two healthy participants participated in experiment where male-female ratio was 50% and aged between 19 and 37 (mean = 26.9). Forty selected music video clips were used as the visual stimuli to elicit participants' emotions. Their EEG and peripheral physiological signals were recorded using Biosemi Active System. They rated video stimuli on a discrete 9-point likert scale (Valence, Arousal, Dominance, and Like rating) using Self-Assessment Manikins during the experiment. The data was recorded in two separate geographical locations. Twenty-two participants were recorded in Twente, Netherlands and the rest of the participants in Geneva, Switzerland.

B. Data visualization

The self-annotation of Valence-Arousal-Dominance and Like in DEAP dataset are biased and strongly correlated with each other. Histogram visualization and density plot indicate that the class distribution of dimensions is imbalanced and shows multimodal and skewed Gaussian-like distribution. Figure 1 shows a strong correlation between Valence and Like (0.62). Dominance has a significant correlation with Valence (0.52). Other correlations between dimensions are less than 0.3. The number of examples, ranges from 3-7 is higher than ranges 1-3 and 7-9 in all dimension. Self-annotations of trails are not even consistent for every participant. For example, Valence of trial #1 ranges from 2.51-9 has high variance (Min: 2.51, Median: 7, Mean: 6.76, Max: 9). Figure 2 shows the variability of annotation for each trial. Most of the classifiers are sensitive to unbalanced dataset. Some machine learning packages use optimization algorithms (e.g. LIBSVM) that can overcome such problems [28]. Ensemble methods are often used to get better predictive performance, generalization, or robustness for unbalanced data.

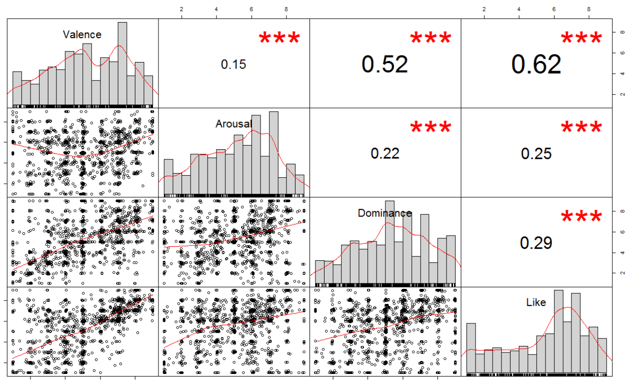

Fig. 1. Statistical summary of DEAP dataset. The diagonal plots with red line shows the distribution of the data. The upper triangle shows absolute value of the Pearson Correlation and the result of the correlation test represented by stars. The lower triangle shows the bivariate scatter plots, with a fitted line (linear model fitting). It is easy to note that the number of samples is unbalanced, self-annotations are highly correlated across all dimensions, and the distribution of data is skewed.

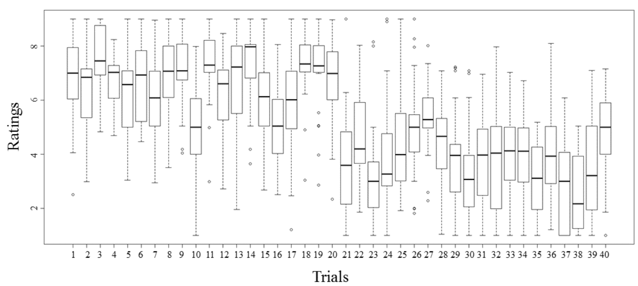

Fig. 2. Box plot view of participants' (32 participants) self-annotation variability for every trial (40 trials) in Valence dimension. Each boxplot shows the distribution based on summary of minimum, first quartile, median, third quartile, and maximum value for one trial. We observed the similar trend in all other dimensions.

C. Preprocessing

Raw EEG was were down-sampled to 128 Hz for computational efficiency and further processing. Bandpass frequency filter of 4-45 Hz was applied for filtering. Signals were then re-referenced on electrode Cz. EOG artifacts were removed by the Blind Source Separation (BSS) technique using Principal Component Analysis (PCA) with analysis window of 512. The shift between correlated windows was applied. Then EEG data was segmented into 60-second trials and a 3-second pre-trial for baseline removal. The baseline removal process provided 63 second EEG data with 8064 samples for each channel. Segmented trials were then reordered from presentation order to an experiment ID order. Overall we have 1280 trials. The final output is a matrix of $40 \times 32 \times 8064$ (trail number $\times$ number of channel $\times$ samples for each channel) dimension with a 1280×4 class label.

D. Feature extraction

Different types of features from the EEG recording that are relevant to emotions and affective state can be distinguished by time, frequency, and time-frequency domain features [29]. Some popular feature extraction methods based on state of the art are:

1) Time domain Statistical Features: Avg. Band Power, Mean, Standard deviation, 1st difference, Normalized 1st difference, 2nd difference, Normalized 2nd difference, Skewness, Kurtosis of the raw EEG signal.
2) Hijorth Features [30]: Mobility, Complexity
3) Fractal Dimension Feature [31]
4) Frequency Domain Features: Average band power
   a) Wavelet Entropy [32],
   b) Recursive Energy Efficiency [33].

Some neuroscientific findings suggest the relation between hemispherical asymmetry and affective state [34]. The Magnitude Squared Coherence Estimation (MSCE) is a widely used technique in asymmetry feature extraction method. Based on the literature review [29], the asymmetric relationship between FP1−FP2, F3−F4, F7−F8, C3−C4, P3−P4 electrode pairs may be considered as dominant features in affect detection. It was observed that 100 components of PCA analysis explained 96% variance of MSCE features. Instead of using higher dimensional feature vector, It is convenient to use 100 components.

E. Feature selection methods

Feature selection methods are used to reduce the dimensionality on sample sets that improve the estimator's accuracy on high-dimensional datasets. They enhanced generalization by reducing overfitting and enhanced generalization [35]. There are two basic feature selection strategies: Filter and Wrapper method.

1) Filter method: The filter method finds out a consistent set of variables outside of the predictive model based on some filtering criteria, e.g. the variables are individually evaluated to check the probable relationship between classes. The sets of variables in this technique are selected based on a threshold of importance. Commonly used filter based methods include correlation coefficient, Chi-squared test. Filter methods calculate each feature separately and consequently. The highly-correlated or redundant features may be selected and significant interactions and relation between variables may not be able to be quantified. The selection criterion of filter based methods is not directly related to the effectiveness of the model.
2) Wrapper method: The wrapper based method evaluates multiple models and adds or removes predictors to find optimal combinations that can maximize model performance. Widely used methods of wrapper-based include RFE, GA (Genetic Algorithm), SA (Simulated Annealing). The wrapper based methods need massive computational resources and time, and there is a risk of over-fitting.

RFE is based on the idea to repeatedly construct models (e.g. SVR, RF, tree-based bagging) and choose the best performing feature based on the weight of coefficients. RFE selects features by gradually reducing the candidate feature set. First, the estimator (e.g. SVR, RF, tree-based bagging) is trained on the initial set of features, and absolute weights of the coefficients are calculated. Features with the smallest weights are pruned from the current set of features. This pruning procedure is recursively repeated on data set until the desired number of features remains in the set. Some RFE algorithms are based on chunking the whole dataset into small subsets and consider their profiles. Those algorithms pick a small subset of representative features sacrificing a small amount of performance (e.g. 10% RMSE tolerance). Here, stability of variable selection depends on the type of model used for feature ranking at each iteration. If the necessary conditions needed for consistency of model used in RFE vary, the number of selected variables and their ranks will vary significantly.

Feature selection methods often yield unstable features where the number of features $>>$ the number of samples. Performance and stability of RFE based feature selection are very limited in this case. To deal with such problem, Meinshausen et al. [16] introduce stability selection based on subsampling in combination with selection algorithms. It was reported that, even if the necessary conditions needed for consistency of the original Lasso (L1 norm penalized linear models) method are violated, stability selection with Randomized Lasso will be consistent in variable selection. The main advantages of this algorithm are [16]:

1) It works efficiently with high-dimensional data
2) It enables finite sample control for predetermined error rates. Therefore, it provides a transparent principle to choose a proper amount of regularization for estimators.
3) The method is extremely general i.e. has a very wide range of applicability.

We adopted the Randomized Lasso (generalization of the Lasso). It is consistent for variable selection even though the "irrepresentable condition" [36] is violated. Lasso penalizes the absolute value of coefficients $|\beta|_k$ of every component with a penalty term proportional to the regularization parameter $\lambda \in \mathbb{R}$. On the other hand, Randomized Lasso penalizes using randomly chosen values in a range [λ, λ/α]. Where, $\alpha \in (0, 1)$ is the weakness parameter. The concept of weakness parameter is closely related to weak greedy algorithms [37]. Let W_k be an i.i.d. random variable in a range from [α,1] for $k = 1....p$. The Randomized Lasso estimator can be written as [16]:

$$\hat{\beta}^{\lambda,W} = argmin_{\beta \in R^p} ||Y - X\beta||_2^2 + \lambda \sum_{k=1}^{p} \frac{|\beta_k|}{W_k} \tag{1}$$

Here, Y and X is the class label and feature matrix respectively. Implementation of equation 1 is a straightforward two-stage process: 1) Re-scaling of the feature variables (with scale factor W_k for the k-th variable), and then 2) LARS algorithm is applied on re-scaled variables [38]. Here, the re-weighting is simply chosen at random. It is not sensible to expect improvement from randomization with one random perturbation. However, applying Randomized Lasso with many iterations (e.g. 1000 times) and looking for variables that are chosen frequently is a useful tool to find out stable feature [16].

From the stability selection method, we can expect 1) Highly important, stable, and relevant features are always selected. Hence their scores are close to 100%, 2) Moderately important, less relevant features are selected when stronger features are not present in the currently selected subset. Hence they have non-zero score, 3) Irrelevant features would have zero scores or close to zero since they would never be among selected features.

IV. Empirical analysis

Our main goals were to: 1) identify frequency bands and functional areas of brain that are related to emotions and affective states, 2) select a set of stable variables in predicting 4-D continuous affective space. First, we performed an empirical analysis to find frequency bands and functional areas of brain that are related to emotions. Following this, we compared the performance of different feature selection methods. Then, we used the selected features to build model of 4-D continuous affective state using SVM. We also compare the performance with feature fusion and ensemble methods. Finally, we converted the 4-D affective space to categorical emotions for human interpretation. Subsequent subsections discuss our objectives in chronological order.

A. Selection of frequency bands and functional areas of brain

We calculated average band power feature from EEG signal using Power spectral density (PSD). PSD was calculated using Hamming window of 1-s length with no overlapping. The average power for θ(4 - 7Hz), α(8 - 15 Hz), β(16 - 31 Hz), and γ($>$ 32Hz) band for each channel over all trials were calculated. RFE was applied to those average band power features using outer re-sampling with 10-fold cross validation and 10 times repetitions. Error tolerance of 10% was used to eliminate non-informative variables. Then RFE method was applied to these features. A number of methods, i.e. RFE and stability selection method were applied to select features and compare performances. The RFE is a resource greedy algorithm because it needs parameter tuning in every iteration. Hence, it is prudent to select classifiers that do not require (or less) parameter tuning for model fitting. We evaluated the performance of various classifiers that were available in the RFE suit. Figure 3 shows feature selection performance comparison of different estimators on the average band power feature. It was observed that tree-based bagging method provides better performance compared to the SVR and RF.

It was observed that the optimal numbers of selected electrodes are 12, 11, 11, and 9 to predict Valence, Arousal, Dominance, and Like respectively. RFE implies that the γ band activities are key to reduce prediction errors in Valence and Like. On the other hand, all fours bands contribute roughly

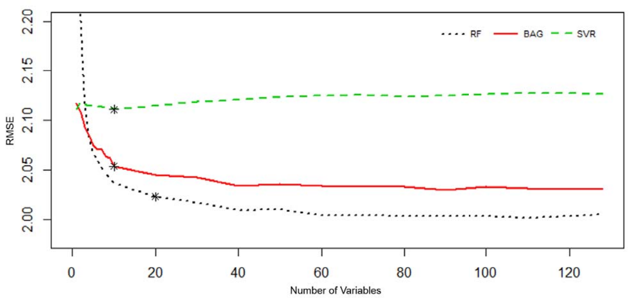

Fig. 3. Performance comparison of different RFE estimator on average band power features. It is observed that tree based bagging (BAG) estimator in RFE is faster (less parameter to tune) and effective than SVR (Support Vector Regression) or RF (Random Fores). Star (*) indicates the optimal point of variable selection.

equally to reduce prediction errors in Arousal and Dominance. The γ and the β band are better predictors for Valence and Like. It was also observed that lower frequency bands (θ and α) have less influence on Like and Valence compared to higher frequency band (β and γ). Penalty parameter C is a hyper-parameter for SVR. We observed that if we vary C of SVR (as an estimator of RFE), the number of selected features changes significantly.

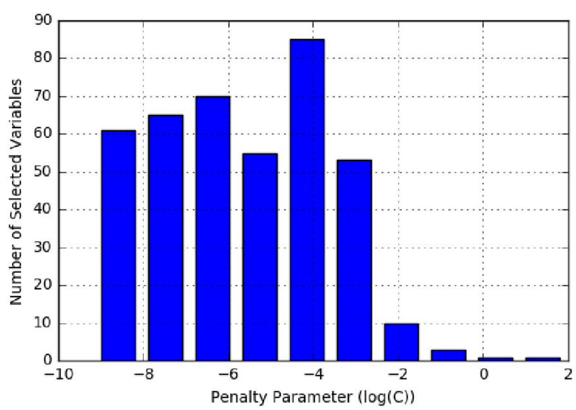

Fig. 4. The effects of penalty parameter (of SVR estimator) variation in RFE feature selection. Bar plot indicates that the number of selected variables changes significantly with hyper-parameter of an estimator.

Figure 4 shows the effects of penalty parameter (of SVR estimator) variation in RFE feature selection. It was observed that the number of selected feature changes significantly (from 1 to 85) with the variation of parameters of estimator. The condition number (C) of a matrix X is defined as the norm (e.g. L2-norm) of X times, the norm of the inverse of X. A system of equations is well- conditioned if a small change in the coefficient matrix results in a small change in the solution vector. For a well-conditioned matrix, mutual coherence is small and L1-recovery with the Lasso performs very well. On the other hand, ill-conditioned matrix has very high condition number (roughly $\log(C) \geq$ the precision of matrix entries). It has high mutual incoherence, and the regressors are highly correlated [39]. In section III-B, we observed that intersubjective variability is very high. On the other hand, the average band power feature matrix is ill-conditioned (C = 3067.75). The stability of RFE is highly dependent on the type of model that is used for feature ranking at each iteration. Therefore, the feature set chosen by methods like RFE are mostly unstable. Randomized Lasso can recover the ground truth from an ill-conditioned (feature) matrix [40].

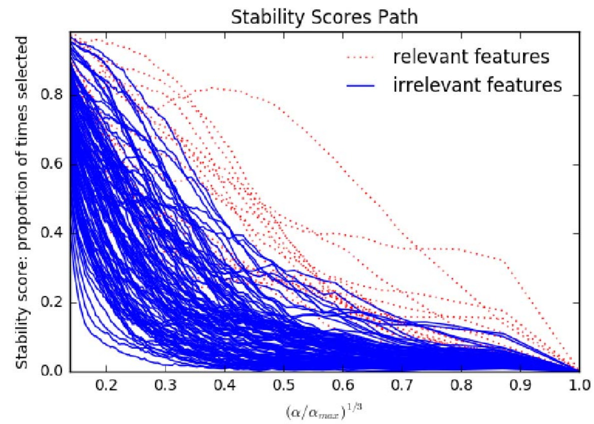

Fig. 5. The stability paths for L1-penalized Logistic Regression with stability selection for Valence.

We applied Randomized Lasso stability selection method to predict stable and relevant EEG channel location and frequency band. Figure 5 shows the stability paths for Lasso with stability selection for Valence. We plot the path as a function of $(\alpha/\alpha_{max})^{1/3}$. The power $1/3$ scales the path less brutally than the log and enables to see the progression along the path. Each of the lines shows the proportion of times selected vs. $(\alpha/\alpha_m ax)^{1/3}$ out of 1000 randomized models. We varied the regularization parameter using a grid base approach and used sample fraction = 0.75, number of resampling = 1000, with tolerance = 0.001 and selection threshold = 0.25. Table I shows ranks of selected channels, frequency bands and stability score of features for Valence, Arousal, Dominance, and Like.

Music videos were used as stimuli in DEAP dataset. Listening to music videos involved audio processing, rhythm detection, and visual related cognitive task [41]. Stability selection ranked very high (100 %) on γ frequency band across dimensions. Table I shows γ band activity over Superior Temporal Gyrus (T8), Supramarginal Gyrus (Cp6) and Somatosensory Association Cortex (Cp2) have a score of 100% for Valence, Arousal, Dominance, and Like. Somatosensory Association Cortex are involved in visuospatial processing, rhythm detection, and temporal context recognition. Supramarginal Gyrus are involved with language perception processing. Superior Temporal Gyrus plays a major role in nonverbal sound processing (e.g. word, music, speech) [41]. It was observed that activation of θ, α, β bands over Premotor Cortex (FC1), and θ, α, β bands over Somatosensory Association Cortex (CP1) are very important in Valence and Arousal modeling. The θ, β and γ band activity over Dorsolateral Prefrontal Cortex

TABLE I
LIST OF SELECTED FEATURE FROM STABILITY SELECTION METHOD.

	Valence			Arousal			Dominance			Like		
Rank	Channel	Band	Score	Channel	Band	Score	Channel	Band	Score	Channel	Band	Score
1	T8	γ	1	CP6	γ	1	AF3	θ	1	CP2	γ	1
2	F3	θ	0.64	AF4	γ	0.86	CP2	γ	0.99	FC1	β	0.96
3	FC1	θ	0.54	CP1	γ	0.78	FC1	β	0.81	F3	θ	0.89
4	P4	γ	0.53	AF3	α	0.64	CP1	γ	0.79	T8	γ	0.81
5	CP6	β	0.52	F3	γ	0.61	Fp2	θ	0.69	AF3	θ	0.71
6	T8	α	0.49	P3	α	0.57	C3	β	0.65	P4	γ	0.63
7	FC1	β	0.46	P8	α	0.5	O1	β	0.65	CP6	γ	0.62
8	F7	γ	0.38	F3	β	0.45	CP6	γ	0.63	FC6	α	0.56
9	FC5	α	0.37	FC1	β	0.45	Fz	γ	0.62	T8	α	0.51
10	CP2	γ	0.35	C3	γ	0.45	FC6	α	0.58	Fz	γ	0.51
11	AF4	α	0.35	C3	θ	0.44	C4	θ	0.58	F3	γ	0.51
12	PO4	γ	0.34	FC1	θ	0.43	F3	α	0.57	C4	θ	0.5
13	F3	α	0.33	O1	α	0.4	P8	β	0.49	P7	β	0.47
14	Fp2	θ	0.32	C4	θ	0.4	Pz	β	0.48	P8	θ	0.47
15	FC1	α	0.32	Fp1	γ	0.39	Fp1	θ	0.45	Fp2	γ	0.46
16	CP6	γ	0.31	CP1	α	0.37	T7	β	0.44	Cz	α	0.45
17	C4	θ	0.29	C4	γ	0.36	F3	γ	0.44	P3	θ	0.41
18	Fz	γ	0.27	P7	γ	0.33	Fp1	γ	0.43	O1	β	0.36
19				CP1	β	0.33	AF3	γ	0.42	AF3	γ	0.36
20				P8	β	0.32	FC6	β	0.42	FC1	θ	0.35
21				T8	γ	0.29	PO4	θ	0.39	AF3	α	0.33
22				CP2	γ	0.29	F4	γ	0.39	PO4	γ	0.31
23				AF3	β	0.29	FC5	θ	0.36			
24				O1	β	0.26	AF3	β	0.32			
25				P3	θ	0.26	PO3	γ	0.32			
26				PO3	γ	0.25						

(AF3) played a vital role in Dominance modeling. Similarly, θ, γ, α band activity over Dorsolateral Prefrontal Cortex (AF3) is important in Like modeling. The Dominance scale represents the control or dominance versus submissiveness. For example, both fear and anger are unpleasant emotions. However, anger is a dominant emotion, while fear is a submissive emotion [5]. Dorsolateral prefrontal cortex is related to social judgment, executive memory, intentionality and abstract thinking [41]. Therefore, θ band (associated with creativity and deep relaxation) activity over Dorsolateral Prefrontal Cortex is closely related to music listening and judgment. Therefore, It was observed that θ band activity over the Dorsolateral Prefrontal Cortex is the highest ranked feature (score$\sim$100%) for Dominance in stability selection algorithm.

B. Performance of Stability Selection Vs. RFE

We applied SVR on selected average band power features and evaluated model performance based on RMSE using 10-fold cross-validation. Table II summarizes the results of comparison between RFE and stability selection method. It was observed that RMSE of SVR significantly improved (22%, 13%, 40%, and 19% for Valence, Arousal, Dominance, and Like respectively) with stability selection method. Hence stability selection is preferable in selecting stable and relevent features.

TABLE II
PERFORMANCE (RMSE) COMPARISON BETWEEN RFE AND STABILITY SELECTION

Method	Valence	Arousal	Dominance	Like	Number of Selected Channels
RFE	2.73	2.28	3.43	2.6	12, 11, 11, 9
Stability Selection	2.13	2	2.07	2.11	13, 16, 20, 17

C. Robust model

Our next step was to perform a systematic analysis to build robust models that can predict affect in continuous 4-D affective space. In this study, we only considered the functional areas of brain, selected in the previous experiment. Besides average band power feature, we extracted different types of features from EEG data as listed in Section III-D. We used SVR, RF, and Tree-based bagging regression algorithm for modeling. Steps involved in the model fitting are: 1) Z-score normalization, 2) hyper-parameter tuning (e.g. C and gamma for SVR) before model fitting, 3) model fitting on data, and 4) RMSE reporting using 10-fold cross-validation.

Table III shows the performance comparison of models, buildup using average band power feature (α, β, γ and θ band) and their fusion with other types of features discussed in section III-D. It was observed that performance of feature fusion is comparable to average band power feature. On the other hand, SVR performs better than RF and tree based bagging regression.

TABLE III
PERFORMANCE (RMSE) COMPARISON OF ROBUST MODELS

Affect	SVM		RF		Bagging	
	BPF	FF	BFF	FF	BFF	FF
Valence	2.13	2.13	2.20	2.15	2.27	2.20
Arousal	2.00	2.03	2.14	2.09	2.12	2.09
Dominance	2.07	2.11	2.19	2.25	2.24	2.23
Like	2.11	2.10	2.17	2.13	2.32	2.16

BPF: Band Power Feature, FF: Feature Fusion

D. 4-D affective space to categorical emotion

There are eight basic categorical emotion within the PAD model [5]. The possible combinations are: high versus low pleasure (+P and -P), high versus low arousal (+A and -A) and high versus low dominance (+D and -D). The labels, that can be used to describes the resulting octants are listed in the table IV. Mehrabian et al [5] stated "Hostile (-P+A+D)" states include feeling angry, defiant, insolent, and nasty, catty; "Anxious (-P+A-D)" states include feeling aghast, bewildered, in pain, insecure, or upset, distressed; and "Exuberant (+P+A+D)" states include feeling admired, excited, mighty, and triumphant, bold, carefree.

TABLE IV
PAD OCTANT

+P+A+D	Exuberant	-P-A-D	Bored
+P+A-D	Dependent	-P-A+D	Disdainful
+P-A+D	Relaxed	-P+A-D	Anxious
+P-A-D	Docile	-P+A+D	Hostile

Affect can be visualized as points in a 3-D, PAD emotion space. When the PAD (Pleasure-Arousal-Dominance) scale scores are standardized (e.g. between -1 to +1), each affect can be described regarding its values on the pleasure-displeasure, arousal-nonarousal, and dominance-submissiveness axes. Some examples of affects in PAD scale (range from -1 to +1) are dignified $(.55, .22, .61)$, elated $(.50, .42, .23)$, hungry $(-.44, .14, -.21)$, inhibited $(-.54, -.04, -.41)$, angry $(-.51, .59, .25)$, sleepy $(.20, -.70, -.44)$, unconcerned $(-.13, -.41, .08)$, violent $(-.50, .62, .38)$, bored $(-.65, -.62, -.33)$, curious $(.22, .62, -.01)$, loved $(.87, .54, -.18)$, puzzled $(-.41, .48, -.33)$ [5]. According to these ratings "loved" $(P = .87, A = .54, D = -.18)$ is a highly pleasant, moderately aroused, and lower submissive affective state. On the other hand, "hungry $(P = -.44, A = .14, D = -.21)$" is a moderately unpleasant, lower aroused, and moderately submissive affective state. We predicted Valence-Arousal-Dominance measure for trial 1-5 using our robust model and projected them in 3-D pad model (standardized range of dimensions are between -1 to +1) like in figure 6. The green marked circle has Valence: 8.58, Arousal: 7.54, and Dominance: 8.5. This point is very close to Exuberant. Therefore, we can infer that emotion is very close to feeling admired, bold, carefree, excited, mighty, and triumphant.

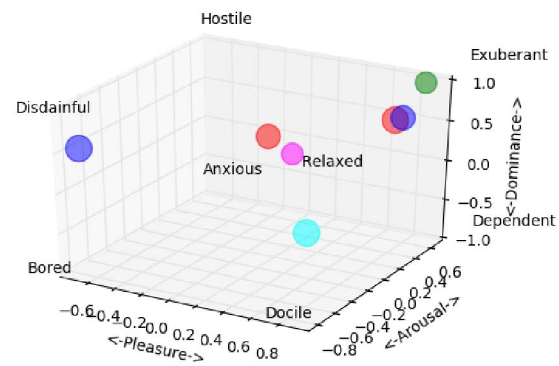

Fig. 6. Relationship between affective space and categorical emotions.

V. CONCLUSIONS

We developed an efficient computational framework to study the role of functional areas of the brain and frequency bands to model continuous affective space. In particular, we examined the activities in EEG frequency bands and electrode locations. We compared the performances of a number of wrapper-based feature selection methods. Based on empirical analyses, we found that stability selection method yields stable set of features. We also showed that stability selection method produces reliable set of features even if the number of samples is small and condition number of the data matrix is very large. In addition, the proposed approach provides robust estimate of continuous 4-D affective space that was converted to categorical emotions using method proposed in [5] for seamless interpretation. We plan to improve the model by including physiological signals along with the EEG recordings.

REFERENCES

[1] D. Grandjean, D. Sander, and K. R. Scherer, "Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization," *Consciousness and cognition*, vol. 17, no. 2, pp. 484–495, 2008.

[2] H. Gunes, M. A. Nicolaou, and M. Pantic, "Continuous analysis of affect from voice and face," in *Computer Analysis of Human Behavior*. Springer, 2011, pp. 255–291.

[3] J. A. Russell, "Core affect and the psychological construction of emotion." *Psychological review*, vol. 110, no. 1, p. 145, 2003.

[4] ——, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[5] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychology*, vol. 14, no. 4, pp. 261–292, 1996.

[6] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *Affective Computing, IEEE Transactions on*, vol. 3, no. 1, pp. 18–31, 2012.

[7] H. Hoffmann, A. Scheck, T. Schuster, S. Walter, K. Limbrecht, H. C. Traue, and H. Kessler, "Mapping discrete emotions into the dimensional space: An empirical approach," in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2012, pp. 3316–3320.

[8] G. K. Verma and U. S. Tiwary, “Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals,” *NeuroImage*, vol. 102, pp. 162–172, 2014.
[9] A. E. Vijayan, D. Sen, and A. Sudheer, “Eeg-based emotion recognition using statistical measures and auto-regressive modeling,” in *Computational Intelligence & Communication Technology (CICT), 2015 IEEE International Conference on*. IEEE, 2015, pp. 587–591.
[10] M. Li and B.-L. Lu, “Emotion classification based on gamma-band eeg,” in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 1223–1226.
[11] S. A. Hosseini, M. A. Khalilzadeh, M. B. Naghibi-Sistani, and V. Niazmand, “Higher order spectra analysis of eeg signals in emotional stress states,” in *Information Technology and Computer Science (ITCS), 2010 Second International Conference on*. IEEE, 2010, pp. 60–63.
[12] S. Valenzi, T. Islam, P. Jurica, and A. Cichocki, “Individual classification of emotions using eeg,” *Journal of Biomedical Science and Engineering*, vol. 7, no. 8, p. 604, 2014.
[13] C. Guerrero-Mosquera, M. Verleysen, and A. N. Vazquez, “Eeg feature selection using mutual information and support vector machine: A comparative analysis,” in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, 2010, pp. 4946–4949.
[14] A.-C. Haury, F. Mordelet, P. Vera-Licona, and J.-P. Vert, “Tigress: trustful inference of gene regulation using stability selection,” *BMC systems biology*, vol. 6, no. 1, p. 1, 2012.
[15] S. Ryali, T. Chen, K. Supekar, and V. Menon, “Estimation of functional connectivity in fmri data using stability selection-based sparse partial correlation with elastic net penalty,” *Neuroimage*, vol. 59, no. 4, pp. 3852–3861, 2012.
[16] N. Meinshausen and P. Bühlmann, “Stability selection,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, no. 4, pp. 417–473, 2010.
[17] R. D. Shah and R. J. Samworth, “Variable selection with error control: another look at stability selection,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 75, no. 1, pp. 55–80, 2013.
[18] N. Doidge, *The brain that changes itself: Stories of personal triumph from the frontiers of brain science*. Penguin, 2007.
[19] W. Klimesch, P. Sauseng, and S. Hanslmayr, “Eeg alpha oscillations: the inhibition–timing hypothesis,” *Brain research reviews*, vol. 53, no. 1, pp. 63–88, 2007.
[20] J. A. Onton and S. Makeig, “High-frequency broadband modulation of electroencephalographic spectra,” *Frontiers in human neuroscience*, vol. 3, p. 61, 2009.
[21] H. W. Cole and W. J. Ray, “Eeg correlates of emotional tasks related to attentional demands,” *International Journal of Psychophysiology*, vol. 3, no. 1, pp. 33–41, 1985.
[22] H. A. Demaree, D. E. Everhart, E. A. Youngstrom, and D. W. Harrison, “Brain lateralization of emotional processing: historical roots and a future incorporating dominance,” *Behavioral and cognitive neuroscience reviews*, vol. 4, no. 1, pp. 3–20, 2005.
[23] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A training algorithm for optimal margin classifiers,” in *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 1992, pp. 144–152.
[24] V. N. Vapnik and V. Vapnik, *Statistical learning theory*. Wiley New York, 1998, vol. 1.
[25] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.
[26] K.-Q. Shen, C.-J. Ong, X.-P. Li, Z. Hui, and E. P. Wilder-Smith, “A feature selection method for multilevel mental fatigue eeg classification,” *IEEE transactions on biomedical engineering*, vol. 54, no. 7, pp. 1231–1237, 2007.
[27] R. Al-Fahad and M. Yeasin, “Robust modeling of continuous 4-d affective space from eeg recording,” in *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on*. IEEE, 2016, pp. 1040–1045.
[28] R. Akbani, S. Kwek, and N. Japkowicz, “Applying support vector machines to imbalanced datasets,” in *Machine Learning: ECML 2004*. Springer, 2004, pp. 39–50.
[29] R. Jenke, A. Peer, and M. Buss, “Feature extraction and selection for emotion recognition from eeg,” *Affective Computing, IEEE Transactions on*, vol. 5, no. 3, pp. 327–339, 2014.
[30] B. Hjorth, “Eeg analysis based on time domain properties,” *Electroencephalography and clinical neurophysiology*, vol. 29, no. 3, pp. 306–310, 1970.
[31] Y. Liu and O. Sourina, “Real-time fractal-based valence level recognition from eeg,” in *Transactions on Computational Science XVIII*. Springer, 2013, pp. 101–120.
[32] M. Murugappan, M. Rizon, R. Nagarajan, S. Yaacob, I. Zunaidi, and D. Hazry, “Eeg feature extraction for classifying emotions using fcm and fkm,” *International Journal of Computers and Communications*, vol. 1, no. 2, pp. 21–25, 2007.
[33] M. Murugappan, N. Ramachandran, Y. Sazali *et al.*, “Classification of human emotion from eeg using discrete wavelet transform,” *Journal of Biomedical Science and Engineering*, vol. 3, no. 04, p. 390, 2010.
[34] L. A. Schmidt and L. J. Trainor, “Frontal brain electrical activity (eeg) distinguishes valence and intensity of musical emotions,” *Cognition & Emotion*, vol. 15, no. 4, pp. 487–500, 2001.
[35] M. L. Bermingham, R. Pong-Wong, A. Spiliopoulou, C. Hayward, I. Rudan, H. Campbell, A. F. Wright, J. F. Wilson, F. Agakov, P. Navarro *et al.*, “Application of high-dimensional feature selection: evaluation for genomic prediction in man,” *Scientific reports*, vol. 5, p. 10312, 2015.
[36] P. Zhao and B. Yu, “On model selection consistency of lasso,” *Journal of Machine Learning Research*, vol. 7, no. Nov, pp. 2541–2563, 2006.
[37] V. N. Temlyakov, “Weak greedy algorithms [*] this research was supported by national science foundation grant dms 9970326 and by onr grant n00014-96-1-1003.” *Advances in Computational Mathematics*, vol. 12, no. 2-3, pp. 213–227, 2000.
[38] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani *et al.*, “Least angle regression,” *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.
[39] G. Strang, *Linear Algebra and Its Applications*. Orlando, FL, Academic Press, Inc., 1988.
[40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
[41] “Cortical,” https://thebrainstimulator.net/pdf/Trans_Cranial_Technologies-cortical_functions_ref_v1_0_pdf.pdf, accessed:10-17-2016.