Affective States Classification using EEG and Semi-supervised Deep Learning Approaches

Haiyan Xu, Konstantinos N. Plataniotis
The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto,
10 King's College Road, Toronto, ON, M5S3G4, CANADA,
Email: xuhaiyan@comm.utoronto.ca, kostas@comm.utoronto.ca

***Abstract*—Affective states of a user provide important information for many applications such as, personalized information (e.g., multimedia content) retrieval/delivery or intelligent human-computer interface design. In recently years, physiological signals, Electroencephalogram (EEG) in particular, have been shown to be very effective in estimating a user's affective states during social interaction or under video or audio stimuli. However, due to the large number of parameters associated with the neural expression of emotion, there is still a lot of unknowns on the specific spatial and spectral correlation of the EEG signal and the affective states expression. To investigate on such correlation, two types of semi-supervised deep learning approaches, stacked denoising autoencoder (SDAE) and deep belief networks (DBN), were applied as application specific feature extractors for the affective states classification problem using EEG signals. To evaluate the efficacy of the proposed semi-supervised approaches, a subject-specific affective states classification experiment were carried out on the DEAP database to classify 2-dimensional affect states. The DBN based model achieved averaged F1 scores of 86.67%, 86.60% and 86.69% for arousal, valence and liking states classification respectively, which has significantly improved the state-of-art classification performance. By examining the weight vectors at each layer, we were also able to gain insights on the spatial or spectral locations of the most discriminating features. Another main advantage of applying the semi-supervised learning methods is that only a small fraction of labeled data, e.g., 1/6 of the training samples, were used in this study.**

I. Introduction

A picture is worth a thousand words, and a video of a moment can capture even more dynamics than a picture can. With the evolution of social media, sharing and trading multimedia contents (pictures or videos) are becoming the new way of communication between friends and family. As a result, with the increasing amount of multimedia content that we are dealing with each day, from the video viewing/posting (e.g., Youtube, Instagram) to on-demand delivery of TV/movie content, the conventional name-based indexing approach can no longer meet the needs of current multimedia storage and retrieval systems. One possible solution to such increasingly personalized information retrieval problem is to apply content-based multimedia indexing approach where a user's response or estimated response to a multimedia content (preference: like/dislike, feedback: felt relaxed/engaged) as an explicit tagging method. Thus, such context awareness where integrated with a user's emotional states is critical for future adaptive, intelligent multimedia sharing system design.

To estimate a person's emotional response to an image or video clip is not an easy task as the responses to a given stimulus can vary greatly from subject to subject. However, it has been shown that the expression of emotion (i.e., affect), such as intensity, can be quantized and used to estimate the true emotional states of a user. Many modalities have been proposed for affective states classification in the literature with promising results, such as facial expression, speech or body gesture. In recent years, several studies on extra-cranial EEG signals [1], [2], [3], [4] have shown that EEG signals are strongly correlated with the expression of emotion and can be used as an alternative and effective way to estimate a user's affective states. In these studies, features were first extracted from each EEG channel (e.g., 32 channels) and then concatenated into one feature vector (often high in dimension) and used with a supervised learning approach, such as support vector machines [2], [5] to estimate a user's affective states. Although promising results have been shown with these supervised machine learning approaches, due to the high-dimensional inputs, they are constrained by the number of labeled samples available, which are manually labeled by either experts or participants, and often lead into small sample problem in practice. On the other hand, with the advancements of sensor design and the digital communication technology, new and innovative applications of EEG signals driven by real-life problems are being proposed and becoming very useful in improving the quality of our lives. However, due to practical limitations, such as cost and intrusiveness, only a small number of sensors are feasible in constructing the commercial-grade EEG headsets. Therefore, for a desired application, the spatial location of the sensors and the analysis on the characteristics (e.g., spectral) of the captured signal determine the successful of such applications.

Motivated by the above challenges, the main objective of this study is to develop an EEG-based affective signal processing framework that can learn both from labeled and unlabeled data to alleviate the practical constraint on the number of labeled data. More specifically, our aim was to further examine the efficacy of the semi-supervised learning approaches, such as stacked denoising autoencoders (SDAE) [6] and deep belief networks (DBN) [7], as an application specific feature extractor/selector on analyzing EEG signals. The extracted features should ultimately provide insights on the spatial or spectral characteristics of the EEG signals that

are most suited for the chosen application, e.g., affective states. To evaluate the efficacy of the proposed semi-supervised learning framework, we have carried out a subject-dependent classification experiment on the DEAP dataset [5], the most comprehensive EEG-emotion dataset that is publicly available. The semi-supervised learning methods, DBN in particular, were able to significantly outperform the baseline SVM-based classifier with radial basis kernel, and achieved the state-of-the-art classification performance on the dataset. More importantly, the proposed system only requires a much smaller number of labeled data ($< 15\%$ of the total samples) in comparison to the supervised learning methods, which make the proposed system much more practical.

II. Approaches

A. Problem Definition

In a semi-supervised learning process, we are given both labeled and unlabeled data. Typically we assume that both the labeled and unlabeled data are from the same distribution D. To make use of the unlabeled data, X_u, unsupervised learning approaches, e.g., Restricted Boltzmann Machines (RBMs), can be applied to first learn a model of $P(x)$ of the input data, then develop a (non-linear) mapping $F(x;W)$ that maps the input vectors in X into a feature space Z [8]. The high-level features learned through this mapping should allow for accurate reconstruction of input data and also preserve the higher-order structure of the input data. To fine-tune the model parameters for a classification task, the obtained discriminative model can be fitted to a labeled dataset (preferable to constitute the same dynamic structure) using a chosen backpropagation method.

B. Deep Belief Networks

Deep Belief Networks (DBNs) are probabilistic generative models that contain an input (visible) layer and many hidden layers where each hidden layer captures high-order correlations between the hidden units in the layer below.

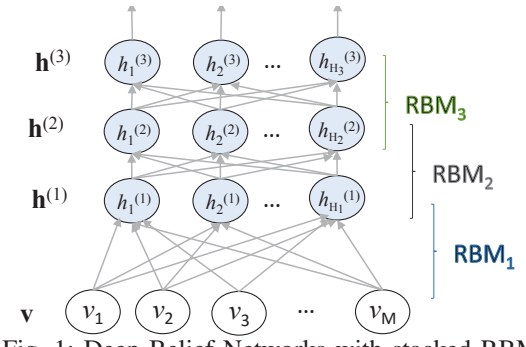

Fig. 1: Deep Belief Networks with stacked RBMs

As shown in Fig. 1, the main building block of a DBN is a bipartite undirected graphical model called the Restricted Boltzmann Machine (RBM), and when stacked together, the top two layers of the DBN form an undirected bipartite graph with the lower layers forming a directed sigmoid belief network [7]. A key feature of this RBM-based algorithm is its unsupervised greedy layer-by-layer training to effectively form a deep, hierarchical model and its ability to work with labeled data to generate both the label and the data.

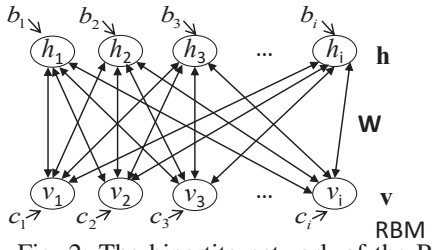

Fig. 2: The bipartite network of the RBMs

As shown in Fig. 2, RBM is a generative stochastic neural network that contains one single layer of hidden units **h** that are not connected to each other and one layer of visible units **v** representing the input variables. The visible and hidden layers are connected through the symmetric weights **W**, biases **b** and **c** to form a bipartite network. The parameters are trained to minimize the reconstruction error or maximize the log probability of the input distribution. In details, a RBM has a joint probability distribution over the visible units **v** and hidden units **h** defined as:

$$P(\mathbf{v},\mathbf{h};\theta) = \frac{1}{Z(\theta)} e^{-\mathbf{h^T W v - b^T h - c^T v}} \tag{1}$$

where $\theta = \{\mathbf{b}, \mathbf{c}, \mathbf{W}\}$, $Z(\theta)$ is a normalization constant defined as $Z(\theta) = \sum_{\mathbf{v}} \sum_{\mathbf{h}} \exp(-E(\mathbf{v},\mathbf{h};\theta))$ and the energy function can be defined as:

$$E(\mathbf{v},\mathbf{h};\theta) = \mathbf{h^T W v - b^T h - c^T v} \tag{2}$$

The binary units of the hidden layer are Bernoulli random variables where a hidden unit h_i is activated with an activation function, for example, logistic Sigmoid, based on each visible unit v_i with probability

$$P(h_i = 1|v;\theta) = \sigma(b_i + \sum_j v_j W_{ij}) \tag{3}$$

$$P(v_j = 1|h;\theta) = \sigma(c_j + \sum_i h_i W_{ij}) \tag{4}$$

where $\sigma(x) = 1/(1+e^{-x})$, and W_{ij} is the symmetric interaction term between the visible unit v_i and hidden unit h_j. The marginal probability that the model assigns to a visible unit is thus defined as:

$$P(\mathbf{v},\theta) = \frac{\sum_{\mathbf{h}} \exp(-E(\mathbf{v},\mathbf{h};\theta))}{Z(\theta)} \tag{5}$$

Taking the gradient of the log likelihood of **v**, i.e., $\log P(\mathbf{v};\theta)$, we can derive the weight updating rule as:

$$\Delta W_{ij} = \alpha(E_{P_{data}}(v_j h_i) - E_{P_{model}}(v_j h_i)) \tag{6}$$

where α is the learning rate, $E_{P_{data}}[\cdot]$ denotes the data distribution observed in the training set and $E_{P_{model}}[\cdot]$ is the same expectation but defined by the model. Since $E_{P_{model}}(v_j h_i)$ is

intractable in practice, thus contrastive divergence approximation to the gradient is used instead where $E_{P_{model}}(v_j h_i)$ is replaced by running Gibbs sampling initialized at the data.

C. Stacked Denoising AutoEncoder

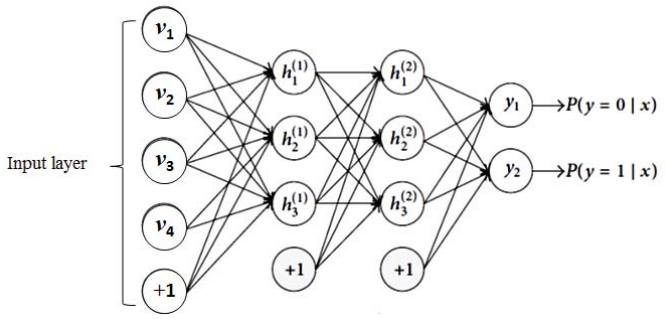

Fig. 3: Stacked AutoEncoder with Softmax Classifier

Stacked denoising autoencoder (SDAE) [6] is another type of semi-supervised deep learning algorithm that is adapted from the conventional stacked encoder architecture shown in Fig. 3. The basic building block of the SDAE is a denoising autoencoder as shown in Fig. 4, which has a similar structure as the one used by RBM shown in Fig. 2, where a hidden unit h_i is activated from the visible unit as:

$$h_i = f_\theta(b_i + \sum_j v_j W_{ij}) \tag{7}$$

where $f_\theta(\cdot)$ is the activation function for the encoder. Similarly to the setup for DBN, we have used $f_\theta(x) = 1/(1+e^{-x})$, as the activation function in this study. In the decoder phase, the last hidden layer is decoded back to reconstructions of previous input layers. One pass of an decoder with tied weights is calculated as:

$$\hat{v_j} = g_{\theta'}(c_j + \sum_i h_i W_{ij}) \tag{8}$$

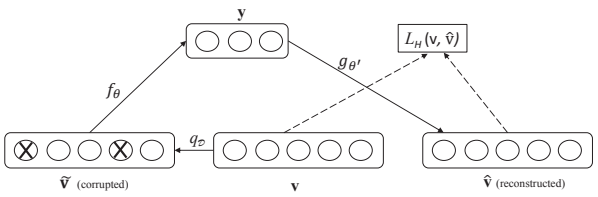

Fig. 4: The Denoising Autoencoder Structure

The essence of the denoising autoencoder is to recover the original inputs even when the inputs are corrupted by noise.This is done by first corrupting the input $\mathbf{x}$ into $\tilde{\mathbf{x}}$ by means of a stochastic mapping $\tilde{\mathbf{x}} \sim q_D(\tilde{\mathbf{x}}|\mathbf{x})$. The cost function is defined as the negative log likelihood of the probability distribution of the visible layer given the hidden layer. If we assume the inputs have a Gaussian distribution and share the same mean as the reconstructed visible layer, $\hat{v}$, and identity covariance, i.e., $P(v|h) = \mathcal{N}(\hat{v}, \mathbf{I})$, then the cost function is defined as:

$$L = -\log P(v|h) = \mathcal{N}(\hat{v}, \mathbf{I}) = \frac{1}{2N}\sum_{n=1}^{N}\sum_i (v_i^n - \hat{v}_i^n)^2 \tag{9}$$

which is essentially is to minimize the averaged reconstruction error of a training batch of N training samples [6].

For the supervised fine-tuning process when used in classification tasks, a logistic regression layer is added on top of the network (more precisely on the outputs of the output layer). This setup also used for the DBN approach. The main difference between SDAE and DBN is that DBN uses undirected Gibbs sampling process to estimate the probability distribution of the visible unit $P(\mathbf{v})$, whereas the SDAE seeks the most robust and efficient (i.e., less hidden units) way to reconstruct the input vector, take into consideration of potential noises in the inputs by injecting artificial noises (zero-masking) on the inputs. With the EEG classification problem, which are often very noisy with interferences from other neural non-emotional noises as well as ambient interferences, these two semi-supervised approaches will provide us ways to either better reconstruct or extract the most emotion-related characteristics of EEG signal.

III. Experimental Setup

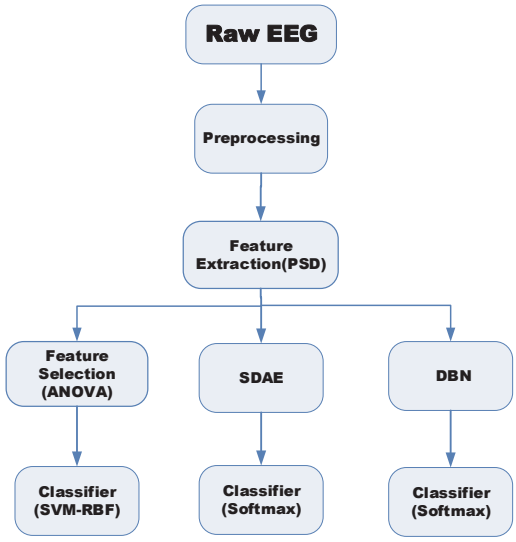

Fig. 5: Signal Processing pipeline for the semi-supervised approaches

A. DEAP Database

The Database for Emotion Analysis using Physiological Signals (DEAP) [5] is the largest multi-modality database for emotion analysis to date. The dataset contains 40 trials of multi-modality emotion-related recordings of thirty-two participants while they were watching one-minute long excerpts of music videos. For EEG modality, in each trial, 32 channels of EEG signal were recorded at 512Hz, the recording includes

a 5 sec. long baseline before the display of video clips as well as during the music video. The annotation of the data was obtained by self-assessment, after the display of each music video clip, each participant reported his/her felt emotional state using the 2-D model of emotion [9], which included scores for arousal, valence, liking, dominance and familiarity. In this study, these self-assessment inputs were used as ground truth and were further categorized into two classes along each affective axis, i.e., high/low arousal or activated /deactivated valence, and were used as ground truth for the learning process followed. On the scale from 1-9, the high or activated states are within [1 5], and low or deactivated states are from (5 9].

To take into account of the response time after the stimuli exposure, in this study, only the last 40s of each recording was used for further analysis. To populate the training and testing samples for each subject, each recording was further segmented into 4 sec. long non-overlapping epochs and used as input samples. The epoch length of 4 seconds were chosen for better spectral resolutions, which was shown empirically more suitable for this study. In total, there were 400 samples (40 trials x 10 epochs) for each subject. For preprocessing, each EEG segment was first referenced to the common average due to the use of active electrodes, then down-sampled to 256Hz and finally, applied a bandpass filter of [4 45] Hz to reduce noises and interferences.

B. Narrow-band PSD Features

The recorded extra-cranial EEG signal is very high in dimension (due to high spatial and sampling rate) and noisy with multiple non-emotional noise sources, therefore, the first step is to reduce such interferences. Due to the high dimensional characteristic of the input signal, e.g., 32 channels at 256Hz sampling rate (after down-sampling) and the available number of samples. i.e., 400 for each subject, a feature extraction step is thus advisable to reduce the input dimensionality as well as any potentiality redundant information.

The spectral power variation in the EEG signal has been shown to be highly correlated to the neural expression of emotions [10], [2], [3]. To extract the narrow-band features, power spectral density (PSD) was first estimated from each 4s epoch using short time Fourier transform (STFT) with 256-point hamming window and no overlapping. The narrow-band Feature set was directly calculated from the estimated PSDs, where spectral components within [4 45]Hz were divided non-uniformly into 15 narrow-bands. More specifically, each conventional wide-bands, Theta (4-7Hz), Alpha (8-12 Hz), Lower Beta (13-21Hz), Upper Beta (22-30 Hz) and Gamma (30-45Hz) were further divided into three sub-bands, and energy within each sub-band was summarized and used as features. The narrow-band psd features from each channel were then concatenated into one feature vector, denoted as PSD32, with the dimension of 32x15=480. To further reduce the non-emotional interferences of other neural activities, the same PSD features were also computed on the baseline recordings and subtracted from the each PSD features of that recording. Finally, the extracted features were standardized by subtracting the mean and divided by the standard deviation, then added 0.5 to set the features in the range of [0 1].

C. Baseline SVM with ANOVA Feature Selection

To construct the baseline model as proposed in [2], each feature of the concatenated feature vector was first ranked using one-way ANOVA (ANalysis of VAriance) analysis. The one-way ANOVA analysis is very closely related to the Fisher score in this case, in the sense that they both calculate the ratio of between-class variance over within class variance. However, the difference is that, Fisher score only ranks the features without providing a relative significance (discriminating) measure on the fisher score obtained. However, often we do not known implicitly the number of features that is optimal, one-way ANOVA analysis on the other hand, will allow us to select the number of features based on the statistical significance of each features. In this study, the significance level was set to 0.85, i.e., P=0.15 in the ANOVA feature selection process.

Support vector machines (SVMs) have been successfully applied in the literature on EEG signal classification tasks [5]. The SVMs first maps the input data to a higher dimensional space and then through the use of support vectors, it maximize the separation margin between the input classes [11]. For the baseline SVM classifier, RBF kernel was used to generate the simulation results here. The SVM implementation was carried out using LIBSVM [12].

D. Performance Metric

Influenced by many personal factors, such as one's past experiences or culture background, there is often a large variation amongst the responses of the participants to a set of experimental stimuli, thus, resulting an imbalanced dataset. To accurately evaluate the performance of the classifiers, the averaged F1 scores were chosen as the performance metric for all the developed models [13].

The averaged classification accuracy was also reported to provide some perspectives on how well our proposed system work in comparison to the previous works reported in the literature where F1 scores were not presented, although, it's worth noting that accuracy alone can provide misleading results on the performance of the system when the classes are imbalanced.

IV. Simulation Results and Discussions

For the proposed subject-dependent EEG-based affective states classification system, a five-fold cross validation process was carried out to evaluate the performance of each model. More specifically, samples of one subject were partitioned into 5 disjoint subsets and at each fold, one of the subsets was used as test samples, and the rest of the subsets were used as training samples. To generate the validation set, the training samples were further divided into training and validation sets with 5:1 ratio, which gave us the train-validation-test ratio of 267-53-80. This repeated 5 times till all the subsets were used as testing sets once. Averaged F1 scores for each subject were

TABLE I: Summary of Major Parameters and Search Ranges.

Classifier	Parameters
SVM	RBF kernel, the regularization parameter C and kernel width γ was chosen in the range of $2^{[-3:10]}$ and $2^{[-5:5]}$
SDAE	2 hidden layers, for parameter update, RMSE error used during non-supervised pre-training and averaged F1 value for supervised fine-tuning. Layer 1 & 2 sizes $\in$ [100 400] with 50 step-size, sparsity set to 0.05, learning rate $\in$ [0.1 0.01 0.001], zero-masking was 0.5, unrolled to feed-forward neural nets for classification
DBN	2 hidden layers and with one regression layer added as layer 3, for parameter update, cross entropy error used during non-supervised pre-training and averaged F1 value for supervised fine-tuning. Layer 1 &2 sizes $\in$ [50 200] with 50 as step-size, layer 3 was set to 500. learning rate $\in$ [0.01 0.001 0.0001],

TABLE II: The averaged F1 scores using the narrow-band PSD features.

	Arousal (%)	Valence (%)	Liking (%)
SVM-RBF	81.01 $\pm$2.54	82.48 $\pm$ 3.29	80.90 $\pm$ 4.85
SDAE	81.50$\pm$3.69	82.43 $\pm$3.64	81.06$\pm$4.80
DBN	86.67*$\pm$3.93	87.08* $\pm$4.69	86.69*$\pm$4.40

*The significance test using one-way ANOVA between the averaged F1 scores obtained using SVM and the DBN approach provide a $p < 0.001$

computed at each fold and were averaged at the end of the experiment.

From Table II, we see that the DBN-based approach outperformed both the SVM-RBF model with ANOVA feature selection as well as the SDAE based model, which indicates that the features extracted by the DBN approach were more discriminative. By comparing the feature learning processes between the DBN and SDAE approaches, where DBN-based approach aims to estimate the statistical property of the input samples, and through repeated Gibbs sampling, we believe the higher-level abstract features learned by the DBN model were able to better represent the class distributions. Such learning ability is even more important when the input signals are complex (e.g., EEG) and the sample size is small w.r.t the signal dimension. SDAE did not produce significantly better results than the SVM-based model, this could due to the limitation of the sample size and the complexity of the signal. Theoretically, the emphasis of SDAE approach is to better reconstruct the input signal (e.g., small cross-entropy) with lower output (last layer before the regression layer) dimension, but this does not grant better classification results if the input samples are small. The SDAE-based model achieved comparable classification performance when compared to the SVM-based supervised learning approach with only 1/6 of the training samples were used with labels.

In summary, the semi-supervised approaches used in this study provide the framework for future commercial BCI applications, where it is more practical to obtain a larger amount of unlabeled data and learn the best data representation on the more abstract level and fine-tune system parameters for the application with a smaller amount of labeled data.

A. Comparison to Related Works

For affective states classification results reported on the DEAP dataset, in [5], an averaged classification accuracy

TABLE III: Averaged correct classification rates and F1 scores.

Classifier	Arousal		Valence		Liking	
	ACC	F1	ACC	F1	ACC	F1
SVM-RBF	82.93	0.8101	83.19	0.8248	83.19	0.8090
SDAE	85.86	0.8150	84.77	0.8243	87.11	0.8106
DBN	88.33	0. 8667	88.59	0.8708	89.20	0.8669

of 62% for arousal and 57.6% for valence were reported using SVM when only EEG modality was used. In [3], an averaged classification accuracies of 68.4% for arousal, 76.9% for valence, 73.9% for dominance and 75.3% for liking were reported when segment-level fusion were used with K-PCA. In [14], an averaged F1 score of 0.549 and 0.607 were reported on a subset of the database using fuzzy adaptive resonance theory mapping (ARTMAP) as the classifier. In [15], DBN were used to classify affective (Liking) states using EEG and EOG inputs, and an averaged AUC score of 75.5% were reported.

For related works reported on other dataset using semi-supervised learning approaches. In [16], DBN-HMM based model was applied to classify two affective states over six subjects, and an accuracy of 87.62% were reported. In [17], DBNs were used to classify arousal, valence and liking states using two EOG and two EMG channels, a classification rate 60.9%, 51.2%, and 68.4% were achieved on leave-one-video out single-subject classification setup.

Overall, the DBN-based affective signal processing framework proposed in this study have significantly improved the state-of-the-art performance on DEAP dataset. Since our model were evaluated on a dataset of 32 subjects with music-video stimuli, e.g., DEAP dataset, we are confident that the results obtained in this study should be considered as state-of-art more generally.

B. Features Selected Through DBN Method

By analyzing the weight vectors between the visible and hidden units, we can obtain important insights on the emotion-specific spatial and spectral locations of the most discriminate features. The shape and magnitude of the weight filters learned at each layer provides us insights on how DBNs have picked out strong features from the input data.

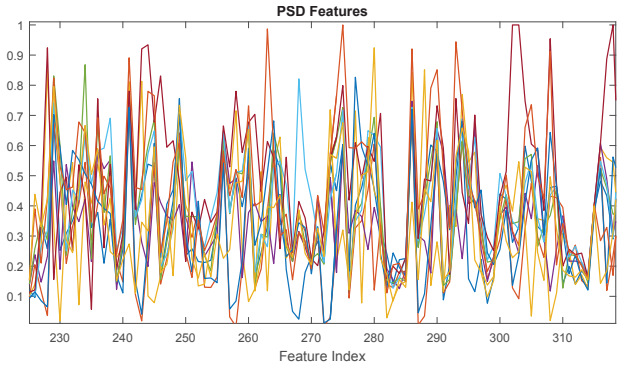

Fig. 6: Samples of the input PSD features extracted from 4s segments of each channel and concatenated into a feature vector (shown here from subject 1, trial 1).

Samples of the input PSD features are shown in Fig. 6, where the EEG inputs were first preprocessed as discussed in

Sec. III-A, then the power spectral features of each channel (in 4s segments) were extracted and concatenated into one feature vector (32*15=480), only the first 5 samples were plotted here to provide better readable details on the input structure.

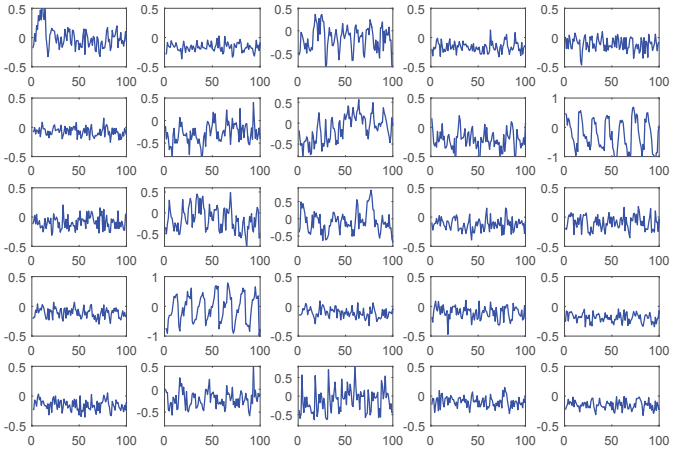

Fig. 7: First 25 Learned filters using DBN and the PSD features of the EEG inputs. Each filter is referenced to according to its sub. figure location, i.e., (3,2) for the sub. figure on the third row and second column.

Since the larger weights of each filer corresponds to the feature components that contain more discriminative information for the tasks learned by the neural networks, Fig. 7 shows the weights of the trained DBNs at the first layer, which is learned using both the unlabeled data (generative model) and labeled data (fine-tuning for classification). From Fig. 7, we can see that the learned filters represent the following key characteristics: ***a)*** they capture the general shape (i.e., the energy variation) among the 32 channels, as shown in sub. figure (2,5) and (3,2), where the start and end of the features of each channels are clearly captured. More importantly, it also captures the rising or decreasing of PSD power within each channel. ***b)*** Finer power variations within each channel are emphasized, as shown in sub. Fig. (1,3), (2,2-4), (3, 2) and (5,2-3). ***c)*** The spectral and spatial information of the critical channels are selected during the learning process , e.g., in (1,1), (2,3) and (3,3), features from channel 1 (Fp1) and 6 (FC1) are given more weight, indicating more discriminating power, this feature selection process becomes more obvious when all 480 features are plotted (only the first 100 features shown here).

V. Conclusion and Future Work

In this paper, we have carried out an EEG-based affective states classification study using semi-supervised deep learning approaches. The semi-supervised deep learning approaches not only significantly reduced the number of labeled data required for a learning system, but also significantly improved the current state-of-the-art classification performance when tested on the DEAP database (most comprehensive to date). The DBN based model achieved averaged F1 scores of 86.67%, 86.60% and 86.69% in classifying the high and low arousal, valence and liking states respectively, when only less than 15% of the input data were labeled. Overall, the use of a semi-supervised learning algorithm has not only alleviated the critical challenges faced by the current EEG-based brain-computer interface systems, such as the requirement of a large number of manually labeled data, but also provides an application-driven feature selector that takes into account of both the spatial and spectral locations of each feature, thus provides much needed insights on the correlation between a desired application and the underlying neural networks.

References

[1] Xiao-Wei Wang, Dan Nie, and Bao-Liang Lu, "Emotional state classification from EEG data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, 2014.

[2] Yuan-Pin Lin, Chi-Hong Wang, Tzyy-Ping Jung, Tien-Lin Wu, Shyh-Kang Jeng, Jeng-Ren Duann, and Jyh-Horng Chen, "EEG-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, July 2010.

[3] V. Rozgic, S.N. Vitaladevuni, and R. Prasad, "Robust EEG emotion classification using segment level decision fusion," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 1286–1290.

[4] Sander Koelstra and Ioannis Patras, "Fusion of facial expressions and EEG for implicit affective tagging," *Image and Vision Computing*, vol. 31, no. 2, pp. 164–174, 2013.

[5] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[6] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *The Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.

[7] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[8] Ruslan Salakhutdinov, *Learning deep generative models*, Ph.D. thesis, University of Toronto, 2009.

[9] James A. Russell, "A circumplex model of affect," *Journal of personality and social psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[10] E Başar, C Başar-Eroğlu, S Karakaş, and M Schürmann, "Brain oscillations in perception and memory," *International Journal of Psychophysiology*, vol. 35, no. 2, pp. 95–124, 2000.

[11] Vladimir N. Vapnik, *Statistical learning theory*, Wiley, 1 edition, September 1998.

[12] Chih-Chung Chang and Chih-Jen Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 27, 2011.

[13] Haibo He, Edwardo Garcia, et al., "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

[14] Wei Shiung Liew, Manjeevan Seera, Chu Kiong Loo, and Einly Lim, "Affect classification using genetic-optimized ensembles of fuzzy artmaps," *Applied Soft Computing*, vol. 27, pp. 53–63, 2015.

[15] Kang Li, Xiaoyi Li, Yuan Zhang, and Aidong Zhang, "Affective state recognition from EEG with deep belief networks," in *2013 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Dec 2013, pp. 305–310.

[16] Wei-Long Zheng, Jia-Yi Zhu, Yong Peng, and Bao-Liang Lu, "Eeg-based emotion classification using deep belief networks," in *2014 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2014, pp. 1–6.

[17] Dan Wang and Yi Shang, "Modeling physiological data with deep belief networks," *International journal of information and education technology (IJIET)*, vol. 3, no. 5, pp. 505, 2013.