Modeling Physiological Data with Deep Belief Networks

Dan Wang and Yi Shang

***Abstract*—Feature extraction is key in understanding and modeling of physiological data. Traditionally hand-crafted features are chosen based on expert knowledge and then used for classification or regression. To determine important features and pick the effective ones to handle a new task may be labor-intensive and time-consuming. Moreover, the manual process does not scale well with new or large-size tasks. In this work, we present a system based on Deep Belief Networks (DBNs) that can automatically extract features from raw physiological data of 4 channels in an unsupervised fashion and then build 3 classifiers to predict the levels of arousal, valance, and liking based on the learned features. The classification accuracies are 60.9%, 51.2%, and 68.4%, respectively, which are comparable with the results obtained by Gaussian Naïve Bayes classifier on the state-of-the-art expert designed features. These results suggest that DBNs can be applied to raw physiological data to effectively learn relevant features and predict emotions.**

***Index Terms*—Deep belief networks, emotion classification, feature learning, physiological data.**

I. Introduction

Inspired by the relationship between emotional states and physiological signals [1], [2], researchers have developed many methods to predict emotions based on physiological data [3]-[11]. Emotions could be classified into a few classes as follows: the negative emotions include anger, anxiety, disgust, embarrassment, fear, and sadness, whereas the positive emotions have affection, amusement, contentment, happiness, joy, pleasure, pride, and relief [7]. Arousal-valence space [12] is an alternative way to define emotions by continuous values. The dimension of arousal represents calmness or excitement, whereas the dimension of valence ranges from highly positive to highly negative.

Empirically, physiological data can be obtained from biosensors in various ways: Electrodermal Activity (EDA) measures electrical conductivity or skin conductance; Electrocardiogram (ECG) measures heart beats; Electroencephalography (EEG) measures brain activities; Electrooculogram (EOG) measurs eye movements [13], and, in a broader sense, related data such as accelerometer data, voice, GPS, etc.

Traditionally, to map biological signals to emotions, the first step is to extract features from the raw data. For example, the R-R intervals extracted from ECG represent a person's heart beat periods, which is related to emotional changes between calmness and excitement. These features are usually hand-engineered using task dependent techniques developed by domain experts [14]-[16] and selected by experts or feature selection algorithms like principal components analysis (PCA). The process is labor-intensive and time-consuming. When physiological data types and/or prediction tasks change, the whole process needs to be repeated. A system that could extract important features automatically from raw physiological data with minimal human input is more scalable.

Deep belief networks (DBNs) [17], as a semi-supervised learning algorithm, is promising for this problem. It consists of a multilayer neural network with each layer a restricted Boltzmann machine (RBM) [18]. The network can be trained efficiently using both labeled and unlabeled data. The trained weights and biases in each layer correspond to features at different levels.

DBNs have been applied in handwriting recognition [17], [19], image recognition [20], [21], and modeling human motions [22]. For physiological data, [23] used DBNs to classify EEG signals to five clinically significant waves and [24] used DBNs classifiers to determine sleep stages from EEG, EOG, and EMG. To the best of our knowledge, there is no existing work to predict emotional states from physiological data using DBNs, which is the focus of this paper.

The paper is organized as follows. Section II gives a brief introduction to DBNs, followed by a description of the physiological data set DEAP [25] used in the experiment in Section III. Section IV presents the DBN structures and experimental settings in predicting emotions from raw physiological data. Section V presents the learned features and experiment result. Finally, Section VI summarizes the paper.

II. Basics of Deep Belief Networks

Shallow neural networks with an input layer, a single hidden layer, and an output layer require more computational elements or are hard to model complex concepts and multi-level abstractions. In contrast, multi-layer neural networks provide better representational power and could derive more descriptive multi-level models due to their hierarchical structures, with each higher layer representing higher-level abstraction of the input data. Unfortunately, it is difficult to train all layers of a deep neutral network at once [26]. With random initial weights, the learning is likely to get stuck in local minima.

Hinton et al. proposed deep belief networks (DBNs) to overcome the difficulties by constructing multilayer restricted Boltzmann machines (RBMs) [18] and training them layer-by-layer in a greedy fashion [17], [27]. The training process has two stages. The first stage is unsupervised pre-training, in which data without labels are used and the training is done in an unsupervised way. The

Manuscript received April 10, 2013; revised July 1, 2013.

Dan Wang and Yi Shang are with the Department of Computer Science, University of Missouri, Columbia, MO 65211 USA (e-mail: dwdy8@mail.missouri.edu, shangy@missouri.edu).

training begins at the bottom of the network, at layer 1, to obtain features in the first layer of hidden nodes from the raw data input. Then the training moves up a layer, between hidden node layer 1 and layer 2, treating hidden node layer 1 as the new input to get features in hidden node layer 2. The greedy layer-wise training is conducted until reaching the highest hidden node layer. The first stage trains a generative model with weights between layers to capture the raw input's features, resulting in better starting point for the second stage to learn than randomly assigned initial weights.

The second training stage consists of fine-tuning the weights and supervised learning at the top layer. In this stage, a new layer is added on top of the stacked RBMs learned in the first stage to construct a discriminative model. Labeled data are used to train the new layer, which acts as a classifier. The overall structure of a DBN with three layers (three layers of hidden nodes) is shown in Fig. 1.

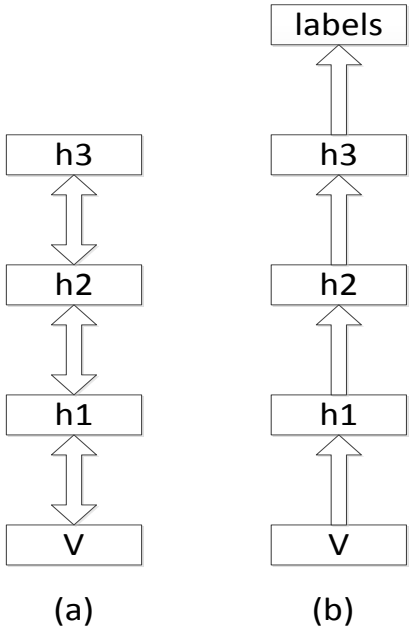


Fig. 1. The structure of a DBN with three hidden layers: (a) The pre-training stage using un-labeled data and (b) The fine-tuning stage with a new top layer added and trained using labeled data.

A. Restricted Boltzmann Machines–the Building Blocks

As the building blocks of DBNs, a restricted Boltzmann machine has a visible layer consisting of stochastic, binary nodes as the input and a hidden output layer consisting of stochastic, binary feature detectors as the output, connected by symmetrical weights between nodes in different layers. There is no connection between the nodes in the same layer. A graphical depiction of an RBM is shown in Fig. 2.

A RBM is a generative stochastic neural network that can learn a probability distribution over its set of inputs. A joint configuration (v, h) of the visible nodes v and hidden nodes h can be represented by the following energy function

$$E(v,h) = -\sum_{i,j} v_i h_j w_{ij} - \sum_i b_i v_i - \sum_j b_j h_j \tag{1}$$

where v_i is the binary state of visible node i, h_j is the binary state of hidden node j, w_{ij} is the weight between node i and j, and b_i is the bias term of visible node i and b_j the bias term of hidden node j.

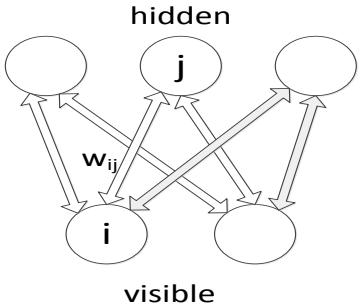


Fig. 2. A graphical depiction of an RBM.

The probability of a given configuration is the normalized energy function:

$$p(v,h) = \frac{e^{-E(v,h)}}{\sum_{u,g} e^{-E(u,g)}} \tag{2}$$

The binary nodes of the hidden layer are Bernoulli random variables. The probability that node h_j is activated, given visible layer v, is

$$P(h_j = 1 \mid v) = \sigma(b_j + \sum_i w_{ij} v_i) \tag{3}$$

where

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{4}$$

The probability that node v_i is activated, given hidden layer h, can be calculated in a similar way as follows

$$P(v_i = 1 \mid h) = \sigma(b_i + \sum_j w_{ij} h_j) \tag{5}$$

Restricted Boltzmann machines are trained to maximize the product of probabilities of a set of training examples X:

$$\underset{W}{\operatorname{argmax}} \prod_{x \in X} P(x)$$

or equivalently to maximize the log likelihood

$$\underset{W}{\operatorname{argmax}} \sum_{x \in X} \log P(x)$$

It is intractable to compute the gradient of the log likelihood. Therefore, [27] proposed contrastive divergence by doing k iterations of Gibbs sampling to approximate it.

$$\Delta w_{ij} = \varepsilon(< v_i h_j >^0 - < v_i h_j >^k) \tag{6}$$

where $<\cdot>^m$ is the average in a contrastive divergence iteration m and ε is the learning rate.

B. Unsupervised Learning Stage of DBN

A single RBM is not good enough to extract high-level features from physiological data due to its complexity. In a DBN of stacked RBMs, the bottom RBM is first trained. Then the outputs of the bottom RBM given the original training examples becomes the inputs of the next layer RBM, i.e., the training examples of the next level RBM. Continuing this process, iteratively RBMs at upper layers are trained

using outputs of lower layers, in a way that the learned feature detectors in a lower layer become the visible input layer for an upper-layer RBM. This greedy layer-by-layer learning method computes weights and biases in different layers that represent different levels of abstractions.

As an example, the features learned by a 784-500-500-2000-10 DBN with 3 hidden layers on MNIST handwritten digits dataset [28] are shown in Fig. 3-Fig. 5. The experiment settings are the same as in [17]. Many features found by the algorithms in the first layer roughly represent dots in different positions. The features in the second layer look like strokes built upon the first layer features. The third layer's features are parts of the whole digits, corresponding to higher-level components.

C. Supervised Learning Stage of DBN

The supervised learning stage adds a label layer on the top on DBN and removes the links in the top to down direction. Now the neural network becomes a feed-forward neural network, as shown in Fig. 1 b), and the standard backpropogation algorithm is used to learn the network weights and biases based on labeled training examples. The learning goal is to minimize the classification error given labeled examples. The weights and biases are initialized as the values learned in the unsupervised learning stage, except for those in the top layer, which are randomly initialized.

III. DEAP DATASET

DBNs are applied to the DEAP dataset [25] to study features learned from raw physiological signals and their effectiveness in predicting emotion states.

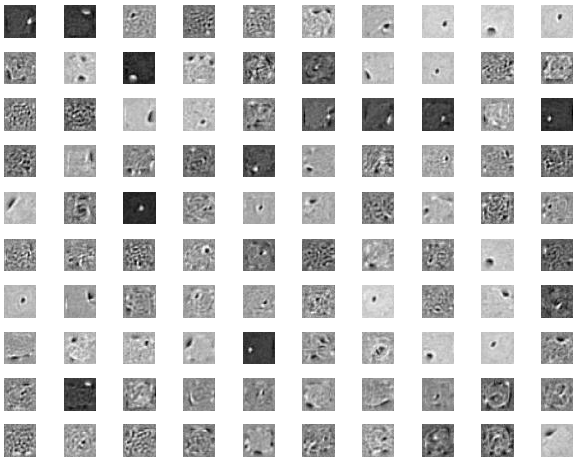


Fig. 3. The features learned in the 1st hidden layer of a 3-layer DBN on MNIST handwritten digits dataset.

DEAP is a multimodal dataset for the analysis of human affective states. The EEG and peripheral physiological signals (down sampled to 128Hz) of 32 subjects were recorded as each subject watched 40 one-minute long videos. Subjects rated the levels as continuous values of arousal, valence, liking, dominance, and familiarity.

DBN is applied to predict the levels of arousal, valence, and liking, respectively, given raw physiological data. The arousal scale ranges from calm (1) to excited (9). The valence scale ranges from sad (1) to joyful (9). Liking also has values from 1 to 9.

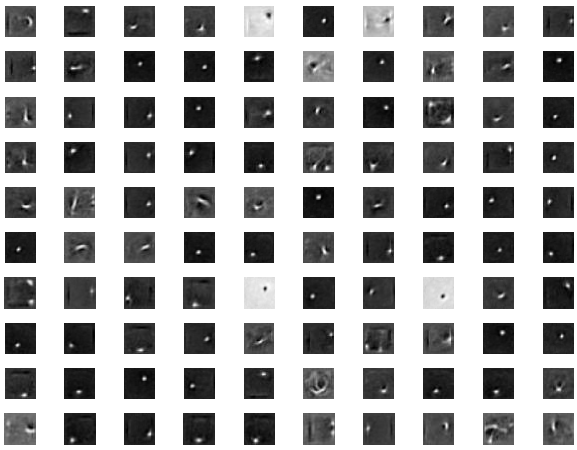


Fig. 4. The features learned in the 2nd hidden layer of a 3-layer DBN on MNIST handwritten digits dataset.

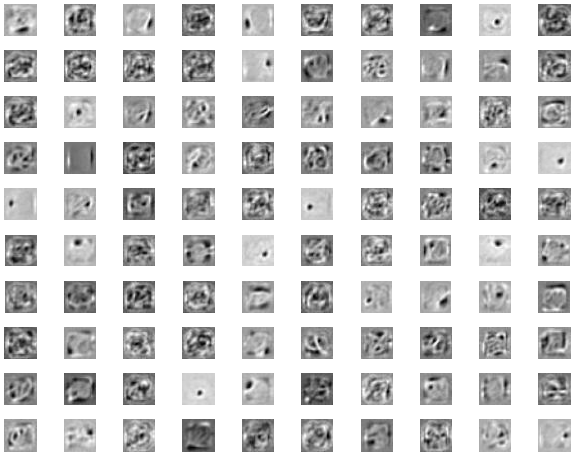


Fig. 5. The features learned in the 3rd hidden layer of a 3-layer DBN on MNIST handwritten digits dataset.

32 EEG channels and 8 peripheral nervous system channels were recorded, including hEOG (horizontal EOG), vEOG (vertical EOG), zEMG (Zygomaticus Major EMG), tEMG (Trapezius Major EMG), GSR (values from Twente converted to Geneva format in Ohm), respiration belt, plethysmograph, and body temperature.

In our experiment, 4 peripheral channels, two EOG and two EMG channels, are used to do predication. The EOG channels record eye movements. The activity of the Zygomaticus major is monitored in zEMG to capture subject's laughs or smiles, whereas the Trapezius muscle (neck) is recorded by tEMG to reflect possible head movements.

As a reference, [25] trained a Gaussian Naïve Bayes classifier for each subject due to the high inter-subject variability. For each subject, three different binary classifiers were trained to map the 8 peripheral channels to low (1-5) / high (5-9) arousal, valence, and liking, respectively. A leave-one-video-out cross validation scheme was conducted, i.e., in each trial a video was taken out for testing and the remaining 39 videos were used for training. As was done in most machine-learning work on physiological data, engineered features such as eye blinking rate, energy of the signal, and mean and variance of EMG and EOG were extracted. All the extracted features were fed into the classifier to train the model. Then the model was used to predict the test cases. The average accuracies over all

subjects were 0.570, 0.627, and 0.591 for arousal, valence, and liking, respectively.

IV. Training a DBN Classifier on Physiological Data

The goal is to train a single DBN classifier for all subjects on the raw data from two EEG and two EOG channels to predict arousal, valence, and liking in two classes (low or high). Performance is evaluated based on the classification accuracy defined as follows:

$$accuracy = \frac{number\ \ of\ correctly\ \ classified\ \ examples}{total\ \ number\ \ of\ examples} \quad (7)$$

There are five steps in the experiment: raw data pre-procession, raw data selection and division, normalization, randomization, and DBN training and classification.

1) In the first step, all signals were pre-processed by notch filtering at 50 Hz in order to remove power line disturbance. Bandpass filters of 0.3 to 32 Hz and 10 to 32 Hz were applied to EOG and EMG, respectively, as suggested by [24].
2) Part of the raw data is discarded and the remaining data is divided into training set and test set. Since a subject's physiological signals are more likely to be elicited by a video at the end of the one-minute watching period as the plot develops, it is reasonable to discard the first 50 second and use only the last 10 seconds' data in each one-minute record. Then the 10 seconds' data were divided into 10 one-second segments. Therefore, the learning process uses one-second examples, each in 512 dimensions (128 Hz * 1 sec * 4 channel). In the total, there are 12800 examples (32 subject * 40 video * 10 example). In each trial, the 10 samples of one randomly chosen video from each subject were left out for testing, resulting in the size of test set being 320 (32 subject * 1 video * 10 example). The remaining 12480 (32 subject * 39 video * 10 example) examples formed the training set.
3) A channel-wise normalization is performed to scale all values to [0 1] according to the following formula:

$$ch_{ij} = \frac{ch_{ij} - \min(ch_i)}{\max(ch_i) - \min(ch_i)} \quad (8)$$

where ch_i represents all the data in the channel i and ch_{ij} is a data element in the channel i. The reason to normalize data this way is two-folded: a) the ranges of different channels may vary, so normalization can make them comparable when concatenating all channels as input, and b) a DBN's node in the input layer has to have values between 0 and 1, to be treated as probabilities of activation of this node.

4) Randomizing training samples is necessary because the mini-batch technique in training DBNs requires samples of each class are evenly distributed.
5) In the last step, a DBN is first pre-trained using unlabeled data (without using the labels), which means the same features learned in the unsupervised learning stage can be used for the three different classification problems in the supervised learning stage. Since the features obtained in the unsupervised learning stage captures the properties of the data, they can be saved for future classification problems. After supervised learning based on labeled data and backpropagation to train a DBN for arousal classification, it is used to predict arousal test set. The same process is applied to the valence and liking classification problems.

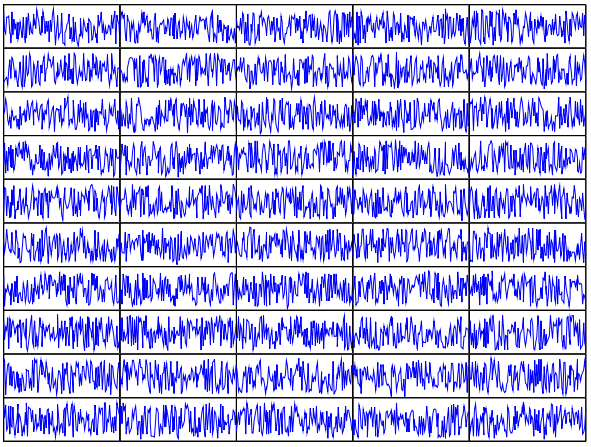
For each classification problem, 10 trials were run to get the mean accuracies, as well as the standard deviations. The DBNToolbox matlab code [24] was used in the experiment. A DBN with two hidden layers, each layer with 50 nodes was constructed. Therefore the DBN structure for the unsupervised learning stage is 512-50-50 and for the supervised learning stage is 512-50-50-2. DBN parameters are listed in Table I.

TABLE I: DBN Parameters Used in the Experiments

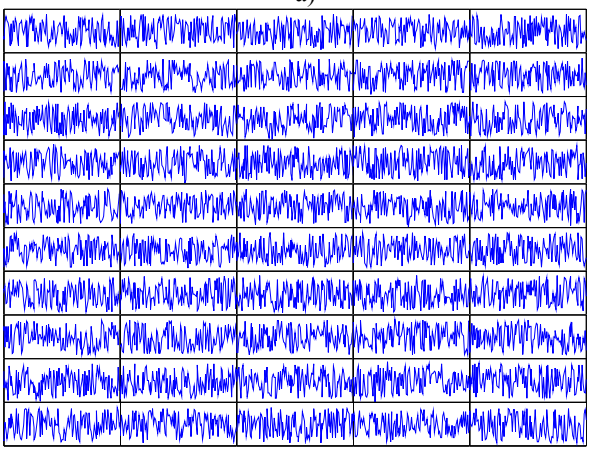
Unsupervised learning rate	0.05
Supervised learning rate	0.05
Number of epochs in unsupervised learning (pre-training)	50
Number of epochs in supervised learning (fine-tuning)	20
Mini-batch size in both stages	100

V. Experimental Results

A. Learned Features



a)



b)

Fig. 6. EOG features at the first hidden layers with 50 nodes learned using a) hEOG data and b) vEOG data.

It is interesting to see what features DBNs can learn from the raw physiological data. Because the data of all channels were concatenated as the input, the weights between two layers need to be separated to show those only relevant to one channel at a time. Although [29] proposed two techniques

called activation maximization and sampling from a unit to show clear patterns in higher layers, they need to clamp input or somehow use the training set's information, which is not convenient. Instead, to show the features learned in the first hidden layer, we take w_{ij}, where i is a node corresponding to one channel in the visible layer and j is a node in the hidden layer, and draw a one-dimensional signal for node j on one channel. For the DBN described in Table I, Fig. 6 shows the features learned in the 1st hidden layer of 50 nodes using EOG data, a) hEOG and b) vEOG. Fig. 7 shows the features learned at the 1st layer nodes using EMG data, a) zEMG and b) tEMG.

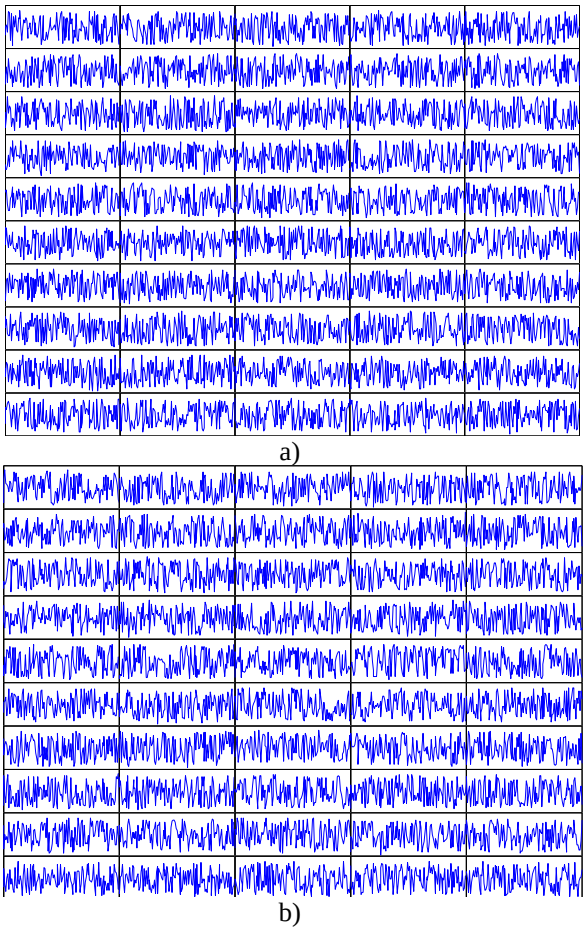


Fig. 7. EMG features at the first hidden layers with 50 nodes learned using (a) zEMG data and (b) tEMG data.

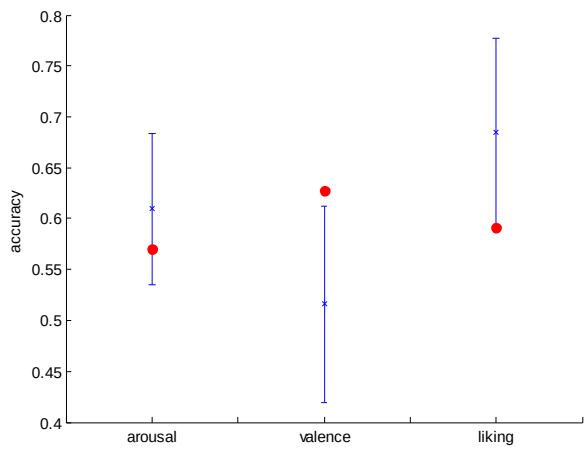


Fig. 8. Performance comparison of DBN and Gaussian Naïve Bayes classifier on three classification problems. The 'x' and error bars are for DBN classification accuracies on raw data, while the red dots are for the Naïve Bayes classifier.

The result doesn't show obvious patterns, which is different from the nicely structured features learned from the handwritten digits shown in Figs. 3-Fig. 5.

B. Experiment Results

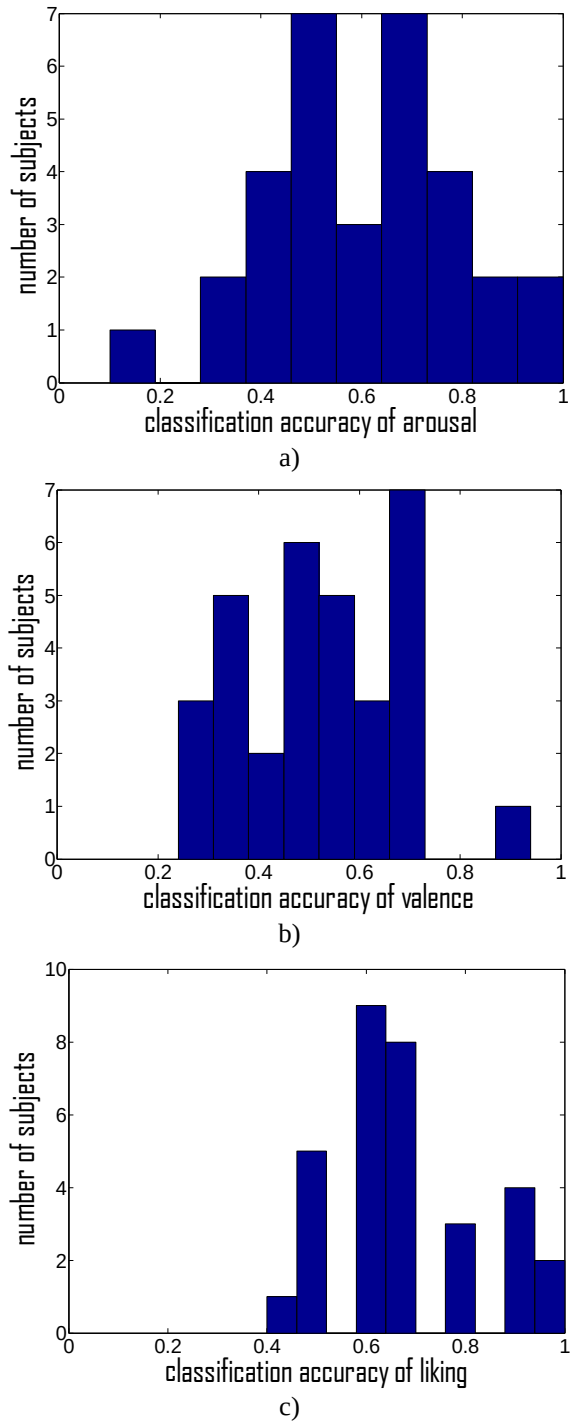


Fig. 9. The histogram of prediction accuracies of the DBN-based method on 32 subjects.

The experiments of the DBN described in Table 1 were run on a Windows 7 machine with 3.0 G dual-core CPU and 4G memory. The whole experiment with 10 trials took about 1 hour.

The means and standard deviations of accuracies of arousal, valence, and liking are 0.609/0.074, 0.512/0.097, and 0.684/0.093, respectively, which are drawn in Fig. 8. The classification accuracies obtained by the Naïve Bayesian classifier in [25] are also shown in the figure for comparison.

There are three major differences in these two results. First, the previous result of [25] is per-subject, i.e., a different classifier is trained for each subject, whereas one DBN classifier is trained for all subjects in our work. Secondly, the previous work used all 8 peripheral channels, compared to only 4 channels used in this work. Finally, the previous work used hand-crafted features, whereas this work simply fed the raw data into the DBN. The proposed method is more general, automated, and obtained comparable solutions.

The experimental results show the DBN method obtained higher liking accuracy, slightly higher arousal accuracy, but lower valence accuracy than the previous work. The reasons of lower valence classification performance could be a) the 4 channels used in this work do not contain sufficient information as in the 8 channels used in the previous work, b) the features of valence are too complex for the DBN in our settings to learn, c) the inter-subject variability is too large for a single model to capture, as stated in [25].

Fig. 9 shows the histograms of the DBN-based method's classification prediction accuracies on 32 subjects.

VI. Conclusion and Future Work

In this paper, DBNs are applied to a physiological problem to learn features from raw physiological data and predict emotions. The trained generic model of DBNs for all subjects in the DEAP dataset achieved good performance compared to previous results. The result shows that DBNs are capable to learn useful features in an unsupervised fashion and obtain classification performance comparable with Gaussian Naïve Bayes classifiers on hand-crafted features. The proposed method is general, fully automated, and can be easily adapted to other physiological signal analysis and prediction problems. It is especially useful for new problems where important features are unknown. The main drawback of the method is the relatively long time for training, e.g., a few hours for the dataset in the experiments.

In future work, more channels will be used in experiments.

The performances of DBNs on the raw data from more than 4 channels in the dataset, up to all the 40 channels, should be investigated. Since the dimensionality of the data is high, an effective channel selection algorithm is necessary. Secondly, the relationship between a DBN' structure and its performance is of interest for finding appropriate tradeoffs between solution quality and training time. Thirdly, in the fine-tuning (supervised learning) stage, there exists an active learning problem for selecting appropriate examples for labeling in some applications. For example, human experts are needed to evaluate the true conditions of subjects as the labels of training data, which is time consuming and expensive. When only a limited number of examples can be labeled, an interesting question is which examples should be labeled, an interesting decision problem. In addition, since this work shows the DBN-based method is poor in predicting valence, but better in predicting arousal and liking, the DBN learned features and hand-engineered features could be combined to obtain better performance.

Acknowledgement

This work is supported by National Institutes of HealthGrants R21/R33-GM078601 and R01-GM100701.

References

[1] R. R. Cornelius, *The science of emotion: Research and tradition in the psychology of emotions*, vol. 133001539, NJ: Prentice Hall Upper Saddle River, 1996.

[2] D. Sander, D. Grandjean, and K. R. Scherer, "2005 Special Issue: A systems approach to appraisal mechanisms in emotion," *Neural networks*, vol. 18, pp. 317-352, 2005.

[3] D. Novak, M. Mihelj, and M. Munih, "A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing," *Interacting with Computers*, vol. 24, pp. 154-172, 2012.

[4] D. McDuff, A. Karlson, A. Kapoor, A. Roseway, and M. Czerwinski, "AffectAura: an intelligent system for emotional memory," in *Proc. the ACM SIGCHI Conference on Human Factors in Computing Systems*, Austin, TX, 2012, pp. 849-858.

[5] E. W. Boyer, R. Fletcher, R. J. Fay, D. Smelson, D. Ziedonis, and R. W. Picard, "Preliminary Efforts Directed Toward the Detection of Craving of Illicit Substances: The iHeal Project," *Journal of Medical Toxicology*, pp. 1-5, 2012.

[6] K. Plarre *et al.*, "Continuous inference of psychological stress from sensory measurements collected in the natural environment," in *Proc. 2011 10th International Conference on Information Processing in Sensor Networks (IPSN)*, Chicago, IL, USA, 2011, pp. 97-108.

[7] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biological Psychology*, vol. 84, pp. 394-421, 2010.

[8] J. Wagner, E. Andre, and F. Jung, "Smart sensor integration: A framework for multimodal emotion recognition in real-time," in *Proc. 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 2009, pp. 1-8.

[9] M. E. Miiller, "Why Some Emotional States Are Easier to be Recognized Than Others: A thorough data analysis and a very accurate rough set classifier," in *Proc. International Conference on Systems, Man and Cybernetics*, 2006, pp. 1624-1629.

[10] C. Peter, E. Ebert, and H. Beikirch, "A wearable multi-sensor system for mobile acquisition of emotion-related physiological data," *Affective Computing and Intelligent Interaction*, pp. 691-698, 2005.

[11] R. W. Picard, "Affective computing: challenges," *International Journal of Human-Computer Studies*, vol. 59, pp. 55-64, 2003.

[12] M. M. Bradley and P. J. Lang, "Measuring emotion: Behavior, feeling, and physiology," *Cognitive neuroscience of emotion*, vol. 25, pp. 49-59, 2000.

[13] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Transactions on Affective Computing*, vol. 1, pp. 18-37, 2010.

[14] G. Chanel, J. J. Kierkels, M. Soleymani, and T. Pun, "Short-term emotion assessment in a recall paradigm," *International Journal of Human-Computer Studies*, vol. 67, pp. 607-627, 2009.

[15] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 2067-2083, 2008.

[16] P. Rainville, A. Bechara, N. Naqvi, and A. R. Damasio, "Basic emotions are associated with distinct patterns of cardiorespiratory activity," *International journal of psychophysiology*, vol. 61, pp. 5-18, 2006.

[17] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, pp. 504-507, 2006.

[18] P. Smolensky, "Information processing in dynamical systems: Foundations of harmony theory," *Parallel distributed processing: explorations in the microstructure of cognition*, Cambridge, MA, USA: MIT Press, 1986, pp. 194-281.

[19] H. Lee, C. Ekanadham, and A. Ng, "Sparse deep belief net model for visual area V2," *Advances in neural information processing systems*, vol. 20, pp. 873-880, 2008.

[20] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Master's thesis, Department of Computer Science, University of Toronto, 2009.

[21] P. Vincent, H. Larochelle, Y. Bengio, and P. A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. the 25th international conference on Machine learning*, 2008, pp. 1096-1103.

[22] G. W. Taylor, G. E. Hinton, and S. T. Roweis, "Modeling human motion using binary latent variables," *Advances in neural information processing systems*, vol. 19, pp. 1345, 2007.

[23] D. F. Wulsin, J. R. Gupta, R. Mani, J. A. Blanco, and B. Litt, "Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement," *Journal of Neural Engineering*, vol. 8, pp. 036015, 2011.

[24] M. Längkvist, L. Karlsson, and A. Loutfi, "Sleep Stage Classification Using Unsupervised Feature Learning," *Advances in Artificial Neural Systems*, vol. 2012, 2012.

[25] S. Koelstra *et al.*, "Deap: A database for emotion analysis using physiological signals," *IEEE Transactions on Affective Computing*, pp. 1-1, 2011.
[26] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," *Advances in neural information processing systems*, vol. 19, pp. 153, 2007.
[27] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural computation*, vol. 14, pp. 1771-1800, 2002.
[28] Y. LeCun and C. Cortes. The MNIST database of handwritten digits [Online]. Available: http://yann.lecun.com/exdb/mnist/
[29] D. Erhan, Y. Bengio, A. Courville, and P. Vincent, "Visualizing higher-layer features of a deep network," Technical report, University of Montreal2009.

Dan Wang was born in Chengdu, China. He is currently a PhD student in Computer Science Department at the University of Missouri, Columbia, MO 65211, USA. He got MS in Communication and Information System and BS in Communications Engineering from Southwest Jiaotong University, Chengdu, China, in 2007 and 2003, respectively.

He has published 4 papers. His research interests include mobile computing, machine learning, and wireless sensor networks.

Yi Shang is a professor and director of Graduate Studies in the Computer Science Department, University of Missouri, Columbia, Missouri. He received his Ph.D. degree in computer science from University of Illinois at Urbana-Champaign in 1997, M.S. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, in 1991, and B.S. degree from the University of Science and Technology of China, Hefei, in 1988. He has published over 140 refereed papers in the areas of nonlinear optimization, wireless sensor networks, mobile computing, intelligent systems, and bioinformatics, and received funding from NSF, NIH, Army, DARPA, Microsoft, and Raytheon. He is a lifetime member of ACM and senior member of IEEE.