6th Annual International IEEE EMBS Conference on Neural Engineering
San Diego, California, 6 - 8 November, 2013

Differential Entropy Feature for EEG-based Emotion Classification

Ruo-Nan Duan, Jia-Yi Zhu and Bao-Liang Lu* *Senior Member*, *IEEE*

***Abstract*— EEG-based emotion recognition has been studied for a long time. In this paper, a new effective EEG feature named differential entropy is proposed to represent the characteristics associated with emotional states. Differential entropy (DE) and its combination on symmetrical electrodes (Differential asymmetry, DASM; and rational asymmetry, RASM) are compared with traditional frequency domain feature (energy spectrum, ES). The average classification accuracies using features DE, DASM, RASM, and ES on EEG data collected in our experiment are 84.22%, 80.96%, 83.28%, and 76.56%, respectively. This result indicates that DE is more suited for emotion recognition than traditional feature, ES. It is also confirmed that EEG signals on frequency band Gamma relates to emotional states more closely than other frequency bands. Feature smoothing method - linear dynamical system (LDS), and feature selection algorithm - minimal-redundancy-maximal-relevance (MRMR) algorithm also help to increase the accuracies and efficiencies of EEG-based emotion classifiers.**

I. INTRODUCTION

Emotion is an overall performance of human's consciousness and behavior. It largely reflects human's perception and attitudes. Emotion always plays a critical role in our daily life, especially in human-human interaction. Furthermore, in the area of human-machine interaction, the emotion recognition which is based on the computer system has also become a key part of the advanced brain-machine interaction system [1]. Thus, as its great significance and wide usage, the emotion recognition has been a popular focus in the field of modern neuroscience, psychology, neural engineering, and computer science as well.

Emotion recognition can be divided into two categories: one is based on non-physiological signals while the other is based on physiological signals. Many previous studies of emotion recognition are based on non-physiological signals, such as facial-expression-based [2] and voice-based [3] emotion recognition. However, facial expressions and tone of voice can be deliberately hidden so that the method based on them is obviously not reliable. In contrast, the method based on physiological signals, which refer to the electroencephalography (EEG), electromyogram (EMG), electrocardiogram (ECG), skin resistance (SC), pulse rate and respiration signals, are seem to be more effective and reliable because humans cannot control them intentionally. Among these methods, the EEG-based emotion recognition has become quite common nowadays. There are also many research projects focusing on EEG-based emotion recognition recently. Li *et al.* indicated that gamma band EEG signals are suitable to classify happiness and sadness with high time resolution [4]. Nie *et al.* found that the emotion associated EEG is mainly produced in the right occipital lobe and parietal lobe for the alpha band, the central site for beta band, and the left frontal lobe and right temporal lobe for gamma band [5]. Wang *et al.* used Isomap to estimate the emotion state, and the trajectory of emotion obtained by Isomap is almost consistent with the change of emotional states [6].

In this paper, we focused on the different performance of subjects' EEG when they were watching movies that were designed for eliciting positive or negative emotional states. After collecting EEG data, we extracted energy spectrum (ES), differential entropy (DE), rational asymmetry (RASM), and differential asymmetry (DASM) as features and compared their classification accuracy in five frequency bands. We also chose the linear dynamical system (LDS) to smooth the features before classifying and adopted the linear SVM as the classifier. Finally, we applied the principal component analysis (PCA) and the minimal redundancy maximal relevance (MRMR) algorithm to reduce the feature dimension to save the storage space and speed up the classification procedure.

II. EXPERIMENTS

A. Stimuli

In the experiment, movie clips were chosen to elicit the positive or negative emotional states. There were totally twelve clips in one experiment, six of which were designed for eliciting the positive emotion and six were for negative emotion. All the movie clips used were in English, and each of them lasted for about 4 minutes long. Only classical and popular movies that were regarded impressive were used as stimuli, such as Schindler's List, High School Musical, and The Day After Tomorrow.

A pressure-sensing steering wheel was employed to record the response from the subjects who were asked to squeeze the steering wheel tightly when he/she felt intense emotion.

B. Subjects

Three men and three women participated in the experiments, who were aged between 22 and 24. They were all healthy and right-handed, with adequate sleep the day before experiment. Each of them participated the experiment twice at intervals of one week or longer. All of them were informed of the harmlessness of the equipment.

R. N. Duan, J. Y. Zhu and B. L. Lu are with the Center for Brain-Like Computing and Machine Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University and MOE-Microsoft Key Laboratory for Intelligent Computing and Intelligent Systems, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China.
* Corresponding author (bllu@sjtu.edu.cn)

978-1-4673-1969-0/13/$31.00 ©2013 IEEE

C. Procedure

The experiments were conducted in the morning or early in the afternoon. A 62-channel electrode cap was used to collect the EEG signals of the subject during the experiment, and ESI NeuroScan System was applied to record the data with the sample rate 1000Hz synchronously. The movie clips were played in random order, and there was a 10s hint before each clip to ask the subject to focus on the following movie clip and a 20s rest after each clip. Figure 1 shows the procedure of the stimuli playing.

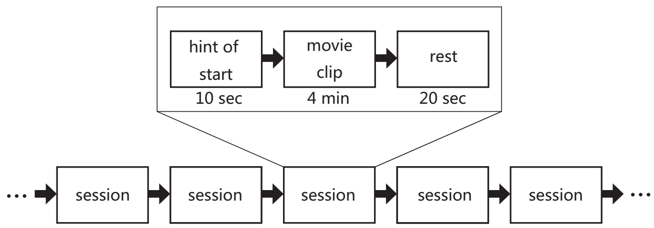

Fig. 1. Procedure of the stimuli playing

III. METHODS

A. Preprocessing

According to the pressure response from the subjects, only the data collected when the target emotion was elicited were used. EEG data were down-sampled with sampling frequency 200Hz in order to speed up the computation, and artifacts were removed manually.

B. Feature Extraction

Frequency domain features and their combinations were employed in this study. A 512-point short-time Fourier transform with a non-overlapped Hanning window of 1s was used to calculate the frequency domain features. Four kinds of features were compared, which were ES, DE, DASM, and RASM.

ES was the average energy of EEG signals in five frequency bands (delta: 1-3Hz, theta: 4-7Hz, alpha: 8-13Hz, beta: 14-30Hz, gamma: 31-50Hz).

DE was defined as

$$\begin{aligned} h(X) &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) dx \\ &= \frac{1}{2}\log(2\pi e\sigma^2), \end{aligned} \tag{1}$$

where the time series X obeys the Gauss distribution $N(\mu, \sigma^2)$. It has been proven that, for a fixed length EEG sequence, DE is equivalent to the logarithm ES in a certain frequency band [7]. DE was employed to construct features in five frequency bands mentioned above.

DASM and RASM were the differences and ratios between DE of 27 pairs of hemispheric asymmetry electrodes. The detail of the 27 pairs of electrodes is shown in Table I. DASM and RASM can be, respectively, expressed as

$$DASM = h(X_i^{left}) - h(X_i^{right}), \text{and} \tag{2}$$

$$RASM = h(X_i^{left})/h(X_i^{right}), \tag{3}$$

where $h(X)$ is defined in Eq. 1, and i is the pair number.

TABLE I

27 PAIRS OF HEMISPHERIC ASYMMETRY ELECTRODES

Pair No.	1	2	3	4	5	6	7	8	9
Left	Fp1	F7	F3	FT7	FC3	T7	P7	C3	TP7
Right	Fp2	F8	F4	FT8	FC4	T8	P8	C4	TP8
Pair No.	10	11	12	13	14	15	16	17	18
Left	CP3	P3	O1	AF3	F5	F7	FC5	FC1	C5
Right	CP4	P4	O2	AF4	F6	F8	FC6	FC2	C6
Pair No.	19	20	21	22	23	24	25	26	27
Left	C1	CP5	CP1	P5	P1	PO7	PO5	PO3	CB1
Right	C2	CP6	CP2	P6	P2	PO8	PO6	PO4	CB2

C. Feature Smoothing

In order to remove the component which has nothing to do with the emotional states, the moving average filter and linear dynamic system (LDS) approach [8] with window length 20s were applied to smooth the feature sequence, respectively.

D. Classification and Classifiers Combination

In this study, two different training and test data partition strategies are used. 1) data from the first eight sessions of one experiment were used to train the model, and data from the rest four sessions in the same experiment were used to test it; 2) data from one experiment of a subject were used to train the model, and data from the other one experiment of the same subject were used to test it.

Two kinds of classifiers were employed, which were linear-kernal support vector machine (SVM) and k-nearest neighbors (kNN) algorithm.

In order to make full use of the advantages of different features, combination of classifiers was used to improve the performance of classifiers. Simple arithmetic combination methods were applied to combine the three classifiers trained by DE, DASM, and RASM in gamma frequency bands, respectively. Maximum, sum, and product rules are simplest rules for combination, which classify the instance into the class with the maximal probability that is the maximum, sum, or product of probabilities from different classifiers in this class.

E. Dimensionality Reduction

Dimensionality reduction could help to increase the speed and stability of the computation. Principal component analysis (PCA) algorithm and minimal-redundancy-maximal-relevance (MRMR) [9] algorithm helped to reduce the dimensionality in this study.

IV. RESULTS AND DISCUSSION

A. Classification

The performance of different kinds of features on Delta, Theta, Alpha, Beta, and Gamma frequency bands is shown in Table II. ASM stands for asymmetry feature that is the combination of DASM and RASM. Total frequency band means that features in all the five frequency bands are used to train and test the model. The training data and the test data are from the different sessions of the same experiment. Experiments 1 - 6 refer to the first experiment of subject 1 -

6, and experiments 7 - 12 refer to the second experiment of subject 1 - 6, respectively (the same below). All the features used here had been smoothed by LDS, and SVM models were applied as classifiers. As we can see from the table, Gamma frequency bands perform better than other frequency bands. And it is obvious that the accuracies of classifiers trained with features calculated using DE (DE, DASM, RASM, and ASM) are higher than those trained with traditional ES features.

This result confirms that the emotional states related to EEG in Gamma frequency band more closely than other frequency bands. It can be implied from this result that DE is more suited for EEG-based emotion classification than traditional ES features.

We also used the data from one's first/second experiment to train a model, and the data from his/her second/first experiment to test it. The features employed here were the DE of all the five frequency bands after LDS smoothing. Performance of SVM and kNN models is also compared in this task. The results are described in Table III, which shows that SVM models increase the accuracies on this data set. The average accuracy of the SVM classifiers trained and tested using data from different experiments reaches 74.10% and 71.79%. This result implies that the relation between the variation of emotional states and the EEG signal is stable for one person with the passage of time. It may be noticed that the accuracy of subject 1's model using training data and test data from different experiments is much lower than the average level, which is caused by the difference of stimuli set between the two experiments subject 1 participated.

The results of combination of classifiers are depicted in Fig. 2. For the reason that EEG signals on Gamma frequency band correlate to emotional states closely, SVM classifiers trained with features DE, DASM, and RASM on Gamma frequency band are adopted to show the performance of maximum, sum, and product rules. From this figure, we can tell that the combination of three classifiers will increase the performance for the classifiers that are with terribly low accuracies when working alone, such as classifiers in experiment 3. And for classifiers that perform relatively well, the combination of classifiers helps to increase the stability of the classifier, and to raise the accuracies slightly.

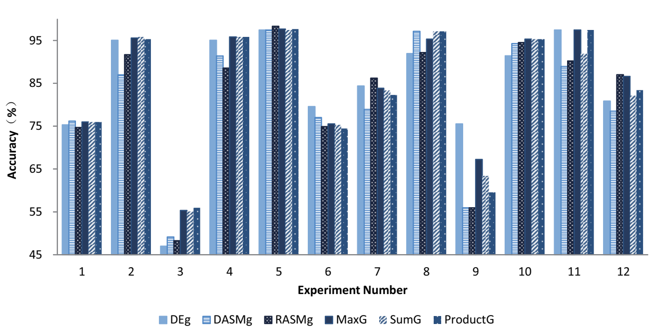

Fig. 2. Performance of combination of classifiers

TABLE II
CLASSIFICATION ACCURACIES USING DIFFERENT KINDS OF FEATURES

Exp.	Feature	δ	θ	α	β	γ	Total
1	ES	38.94	50.42	44.26	58.72	**79.57**	64.47
	DE	45.32	44.26	45.74	60.85	75.32	**77.66**
	DASM	45.11	44.26	44.26	69.36	**76.17**	**76.17**
	RASM	44.26	44.26	44.26	44.26	74.68	**79.15**
	ASM	43.40	33.83	44.26	68.51	75.32	**76.38**
2	ES	55.47	46.61	72.92	**96.35**	95.83	82.55
	DE	54.69	42.45	88.02	**96.61**	95.05	88.02
	DASM	61.97	55.73	72.40	**96.61**	86.98	89.06
	RASM	58.33	58.33	68.75	**96.61**	91.67	92.71
	ASM	61.97	55.73	71.88	**95.05**	90.10	89.06
3	ES	53.62	**63.83**	56.17	60.64	42.76	54.47
	DE	53.83	65.96	60.21	**67.87**	47.02	65.53
	DASM	61.28	57.02	47.66	62.98	49.15	**67.02**
	RASM	54.89	54.90	47.23	58.94	48.30	**62.98**
	ASM	66.17	54.04	47.23	**67.23**	50.00	66.81
4	ES	55.47	59.11	64.32	**82.03**	59.64	75.00
	DE	41.93	48.44	48.44	89.84	**95.05**	77.60
	DASM	50.00	57.55	45.57	86.19	**91.40**	68.75
	RASM	54.47	59.38	45.31	85.94	**88.54**	67.45
	ASM	55.47	59.90	53.91	87.50	**91.40**	67.19
5	ES	76.17	70.43	89.79	90.21	95.32	**97.45**
	DE	62.55	68.72	93.62	**97.45**	**97.45**	**97.45**
	DASM	41.91	63.83	**97.45**	97.23	**97.45**	**97.45**
	RASM	44.26	68.30	95.75	97.87	**98.30**	97.45
	ASM	42.34	65.53	**97.45**	**97.45**	**97.45**	**97.45**
6	ES	72.13	72.55	75.53	69.15	71.91	**81.06**
	DE	64.89	60.85	63.83	79.15	79.57	**84.26**
	DASM	58.09	57.87	53.62	65.75	77.01	**81.70**
	RASM	68.09	60.43	55.75	65.96	74.89	**88.72**
	ASM	60.43	56.60	54.26	65.96	77.02	**81.70**
7	ES	54.17	55.47	56.77	62.76	**81.77**	64.06
	DE	47.92	53.39	55.47	63.02	**84.38**	77.08
	DASM	58.33	57.81	54.43	56.51	**78.91**	73.44
	RASM	60.15	50.26	55.47	58.85	**86.20**	74.22
	ASM	59.64	58.08	55.47	57.03	**92.19**	73.70
8	ES	67.17	63.80	**94.27**	91.41	91.41	87.50
	DE	54.17	77.34	91.93	**96.61**	91.93	90.89
	DASM	60.42	55.47	85.16	91.67	**97.14**	91.41
	RASM	61.46	55.21	77.86	91.67	92.18	**96.09**
	ASM	60.42	55.47	85.42	**95.31**	93.23	91.41
9	ES	47.45	75.74	61.91	58.30	66.60	**76.81**
	DE	47.87	66.17	54.89	65.11	75.53	**78.30**
	DASM	47.23	60.00	55.74	73.20	55.96	**78.94**
	RASM	44.89	57.87	55.74	**78.94**	55.96	72.77
	ASM	52.34	58.51	55.74	76.17	56.17	**78.72**
10	ES	66.93	67.19	57.55	80.99	74.74	**81.25**
	DE	63.54	66.93	60.68	90.88	91.41	**94.27**
	DASM	45.05	55.73	77.60	83.85	**94.27**	77.60
	RASM	55.47	55.73	67.97	80.21	**94.53**	90.10
	ASM	55.47	55.47	81.77	82.55	**94.27**	78.39
11	ES	76.17	70.43	89.78	95.75	95.31	**97.02**
	DE	62.55	68.72	93.62	**97.45**	**97.45**	**97.45**
	DASM	63.19	64.89	65.96	91.06	88.94	**91.70**
	RASM	59.79	70.64	65.32	93.19	90.21	**94.47**
	ASM	63.19	66.81	65.53	91.49	91.49	**94.26**
12	ES	60.00	50.00	50.00	67.66	**69.36**	57.02
	DE	56.38	54.68	51.06	72.98	80.85	**82.13**
	DASM	50.85	51.70	49.36	74.89	**78.51**	78.30
	RASM	56.81	58.30	44.26	71.49	**87.02**	83.20
	ASM	52.77	57.23	50.43	75.32	**85.96**	78.30
Ave.	ES	60.31	62.13	67.77	76.16	**77.02**	76.56
	DE	54.64	59.83	67.29	81.49	**84.25**	84.22
	DASM	53.62	56.82	62.43	79.11	**80.99**	80.96
	RASM	55.32	57.80	60.31	76.99	81.87	**83.28**
	ASM	56.13	56.43	63.61	79.96	**82.88**	81.11

B. Feature Smoothing

The rapid fluctuations in the feature sequence should be removed, because emotional state varies smoothly. The accu-

TABLE III

CLASSIFICATION ACCURACIES WITH DATA FROM DIFFERENT EXPERIMENTS USING TWO KINDS OF CLASSIFIERS

Data Train	Data Test	Clsfr.	Sub.1	Sub.2	Sub.3	Sub.4	Sub.5	Sub.6	Ave.
First	First	SVM	77.66	88.02	65.53	77.60	97.45	84.26	81.75
		KNN	48.94	72.40	48.30	62.50	95.32	74.68	67.02
Sec.	Sec.	SVM	77.08	90.89	78.30	94.27	97.45	82.13	86.69
		KNN	38.80	90.89	55.74	83.07	95.32	60.00	70.64
First	Sec.	SVM	55.06	81.20	69.24	77.97	99.39	61.76	74.10
		KNN	48.44	73.28	61.45	66.82	99.24	55.11	67.39
Sec.	First	SVM	58.07	78.47	57.40	76.32	99.39	61.07	71.79
		KNN	57.44	73.01	61.83	67.63	99.24	56.30	69.24

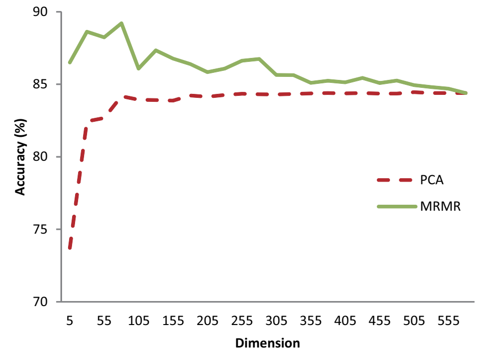

Fig. 3. Average performance of different dimensionality reduction methods

racies of SVM classifiers trained by features with no smoothing method, moving average method, and linear dynamical system smoothing method are compaired in Table IV. The training data and test data are from the different sessions of the same experiment, and the table shows the average results. The table shows that for subject 1, 3, 4, and 6, the smoothing methods help to improve the performance of classifiers in the emotion classification task. And the LDS method is more effective when smoothing the feature sequence. But for subject 2 and 5, the smoothed feature sequence decreases the accuracies of classifiers for the reason that the smoothing causes information lost in some degree.

TABLE IV

PERFORMANCE OF DIFFERENT FEATURE SMOOTHING METHODS

Smooth	Sub.1	Sub.2	Sub.3	Sub.4	Sub.5	Sub.6	Ave.
No	62.02	92.06	63.83	83.72	98.62	78.41	79.78
Mv. Ave.	73.72	90.24	67.56	87.11	98.19	83.94	83.46
LDS	77.37	89.46	71.92	85.94	97.45	83.20	84.22

C. Dimensionality Reduction

Figure 3 demonstrates the performance of PCA and MRMR methods in dimensionality reduction. The training data and test data are from different experiments on the same subject, and the SVM classifiers trained by LDS smoothed features are employed here. The original features we use are the combination of DE, DASM, and RASM features in all the five frequency bands, which means that the original feature dimensionality is as high as 580. The figure reveals that the accuracy changes slightly after dimensionality reduction using PCA when the dimensionality is higher than 100. And the MRMR algorithm helps to increase the accuracy of the classifier when dimensionality of the feature decreases.

It can be indicated that most of the feature values we employ are not necessary for the task. Some of them are irrelevant to emotion recognition, and some are redundant in this task. This discovery helps us to reduce the computations of features and the complexity of the classification models.

V. CONCLUSION

In this paper, a series of experiments were conducted to collect the EEG signals during the subject was watching movie clips. A new EEG feature named differential entropy (DE) was proposed and compared with the traditional frequency domain feature named energy spectrum (ES). According to the result, DE and its combination on symmetrical electrodes perform better than ES feature. And the combination of classifiers increases the stability and the accuracy of classifier. At the same time, it is confirmed that the relation between the variation of emotional states and the EEG signals does not change for one person during a period of one week.

ACKNOWLEDGMENT

This work was partially supported by the National Natural Science Foundation of China (Grant No. 61272248), the National Basic Research Program of China (Grant No. 2013CB329401), and the Science and Technology Commission of Shanghai Municipality (Grant No. 13511500200).

REFERENCES

[1] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, "Emotion recognition in human-computer interaction," *Signal Processing Magazine*, vol. 18, no. 1, pp. 32–80, 2001.

[2] K. Anderson and P. W. McOwan, "A real-time automated system for the recognition of human facial expressions," *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 36, no. 1, pp. 96–105, 2006.

[3] J. Ang, R. Dhillon, A. Krupski, E. Shriberg, and A. Stolcke, "Prosody-based automatic detection of annoyance and frustration in human-computer dialog," in *Proc. ICSLP*, vol. 3. Citeseer, 2002, pp. 2037–2040.

[4] M. Li and B.-L. Lu, "Emotion classification based on gamma-band EEG," in *Proc. 31nd Annu. Int. Conf. IEEE EMBS*. IEEE, 2009, pp. 1223–1226.

[5] D. Nie, X.-W. Wang, L.-C. Shi, and B.-L. Lu, "EEG-based emotion recognition during watching movies," in *5th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2011, pp. 667–670.

[6] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from EEG data using machine learning approach," accepted by *Neurocomputing*.

[7] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, "Differential entropy feature for EEG-based vigilance estimation," in *Proc. 35th Annu. Int. Conf. IEEE EMBS*. IEEE, 2013, pp. 6627–6630.

[8] L.-C. Shi and B.-L. Lu, "Off-line and on-line vigilance estimation based on linear dynamical system and manifold learning," in *Proc. 32nd Annu. Int. Conf. IEEE EMBS*. IEEE, 2010, pp. 6587–6590.

[9] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.