Multi-view Emotion Recognition Using Deep Canonical Correlation Analysis

Jie-Lin Qiu[1], Wei Liu[1], and Bao-Liang Lu[1,2,3](✉)

[1] Center for Brain-Like Computing and Machine Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
{Qiu-Jielin,liuwei-albert,bllu}@sjtu.edu.cn

[2] Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognition Engineering, Shanghai Jiao Tong University, Shanghai, China

[3] Brain Science and Technology Research Center, Shanghai Jiao Tong University, Shanghai, China

Abstract. Emotion is a subjective, conscious experience when people face different kinds of stimuli. In this paper, we adopt Deep Canonical Correlation Analysis (DCCA) for high-level coordinated representation to make feature extraction from EEG and eye movement data. Parameters of the two views' nonlinear transformations are learned jointly to maximize the correlation. We propose a multi-view emotion recognition framework and evaluate its effectiveness on three real world datasets. We found that DCCA efficiently learned representations with high correlation, which contributed to higher emotion classification accuracy. Our experiment results indicate that DCCA model is superior to the state-of-the-art methods with mean accuracies of 94.58% on SEED dataset, 87.45% on SEED IV dataset, and 88.51% and 84.98% for four classification and two dichotomies on DEAP dataset, respectively.

Keywords: Emotion recognition · EEG · Eye movement
Deep Canonical Correlation Analysis · Coordinated representation
Multi-view deep networks

1 Introduction

Emotion recognition is important for communication, decision making, and human-machine interface. Since emotions are complex psycho-physiological phenomena associated with many nonverbal cues, it is difficult to build robust emotion recognition models using only one single modality. Signals from different modalities can represent different aspects of the emotions, and the complementary supplemental information from different modalities can be integrated to build a more robust emotional recognition model. Emotion recognition based on electroencephalography (EEG) and eye movement data have attracted increasing interest. Integrating different features with fusion technologies is important

L. Cheng et al. (Eds.): ICONIP 2018, LNCS 11305, pp. 221–231, 2018.
https://doi.org/10.1007/978-3-030-04221-9_20

to construct robust emotion recognition models [1]. The combination of signals from the central nervous system, EEG, and external behaviors, eye movement, has been a remarkable method for utilizing the complementarity of different modes of features [1–3].

In recent years, various deep neural networks have been introduced to affective computing and their attractive results showed the superior performance of such networks compared with the conventional shallow methods [13]. And various multimodal deep architectures have been proposed to leverage the advantages of two modalities, which can be concluded into two categories of representation: joint and coordinated [4]. The joint representation combines the unimodal signals into the same representation space, while the coordination representation processes the unimodal signals separately, enforces some similarity constraints on them, and brings them to the coordination space. Multimodal emotion recognition intends to distinguish emotions using different forms of physiological data collected at the same time, where complementary features of different modalities can be employed [2,3,12]. Deep neural networks have also been used for multimodal emotion recognition in an end-to-end method. Lu *et al.* used both EEG data and eye movement data to classify three kinds of emotions [3]. Liu *et al.* furthermore used Bimodal Deep AutoEncoder to extract high level representation features [5]. Tang *et al.* adopted the Bimodal-LSTM model to recognize multimodal emotions [6], and achieved better results than [5]. However, all the achievements above are based on joint representations and few coordinated based methods have been studied.

Coordinated representations first enforced similarity between representations. For example, the similarity models try to minimize distance between different modalities. With the rapid development of neural networks, they have shown the ability to reconstruct coordinated representations when learning jointly in an end-to-end manner [7]. What's more, structure coordinated space added more constrains between the modality representations [8]. Order-embedding is another example of a structured coordinated representation, which was proposed by Vendrov *et al.*, enforcing a dissimilarity metric and implementing the notion of partial order in the multimodal space [9]. Canonical correlation analysis (CCA) based structured coordinated is another case, where CCA computes the linear projection and maximizes the correlation between two modalities. CCA based models have been widely used for cross-modal retrieval and signal analysis. Kernel canonical correlation analysis (KCCA) uses reproducing kernel Hilbert spaces for projection but shows poor performance on large real-world datasets [10]. Deep canonical correlation analysis (DCCA) was introduced with deep network extension to optimize the correlation over the representations and showed better performance [11].

In this paper, we adopt DCCA to extract multimodal features for emotion recognition and achieved remarkable results. DCCA is a deep network based extension of canonical correlation analysis. It can learn separate representations nonlinearly for each modality, and coordinate them through a constraint. In this paper, we use deep networks to learn the nonlinear transformation of two views into a highly correlated space.

The main contributions of this paper are as follow:

(1) We first took coordinated representation of multimodal signals to recognize emotions, which means extracting more correlated high-level representations.
(2) We proposed a multi-view framework to deal with multimodal emotion recognition problem and achieved better classification accuracy than the state-of-the-art methods.

2 Deep Canonical Correlation Analysis

2.1 Background

Canonical correlation analysis can learn linear transformation of two vectors in order to maximize the correlation between them, which is widely used in economics, medical studies, and meteorology [21]. Lai *et al.* designed Colin's CCA, which performed Canonical Correlation Analysis with Artificial Neural Network [21]. With rapid development of deep learning, Andrew *et al.* proposed Deep Canonical Correlation Analysis (DCCA) with deep networks extension, which is a non-linear version of Canonical Correlation Analysis (CCA) that uses neural networks as the mapping functions [15]. DCCA calculates the representation of two views by multiplying them through stacked layers that are non-linearly transformed. Hossain *et al.* proposed a novel FS method based on Network of Canonical Correlation Analysis, NCCA, which is a robust method to acquisition noise and ignores mutual information computation based on Colin's CCA [21]. CCA is a standard statistical technique to find linear projections of two random vectors that are maximally correlated, while in Colin's CCA network [21], activation is fed forward from each input to the corresponding output through the respective weights to maximise the correlation. In Deep Canonical Correlation Analysis, deep networks are used for feature extraction with back propagation applied to maximise the correlation between two views.

2.2 Our Model

We use DCCA for feature transformation, fuse the features after extraction, and apply SVM as the classifier. The model is shown in Fig. 1, and the model contains three parts: non-linear feature transformation (L_2 and L_3 in Fig. 1), CCA calculation (CCA layer in Fig. 1), and feature fusion and classification. EEG features and eye movement features are separated into two views denoted as L_1 in Fig. 1, and we set two views' input features as X_1, and X_2.

Nonlinear Feature Transformation. In the deep networks, for simplicity, we assume that each intermediate layer in the network for the first view has c_1 units, and the output layer has o units, as shown in Fig. 1 as 'View 1'. Let $x_1 \in R^{n_1}$ be an instance of the first view, and the outputs of the first layer in the hidden layers for the instance x_1 are $h_1 = s(W_1^1 x_1 + b_1^1) \in R^{c_1}$, where $W_1^1 \in R^{c_1 \times n_1}$ is a matrix of weights, $b_1^1 \in R^{c_1}$ is a vector of biases, and $s(\cdot)$ is a non-linear function applied componentwise. The outputs h_1 can then be used to compute the outputs of the next one in hidden layers as $h_2 = s(W_2^1 x_1 + b_2^1) \in R^{c_1}$, and so on until the final representation in the hidden layers $f_1(x_1) = s(W_d^1 h_d - 1 + b_d^1) \in R^o$ is computed, for a network with d layers.

Given an instance x_2 of the second view, as shown in Fig. 1 as 'View 2', the representation $f_2(x_2)$ is computed the same way, with different parameters W_l^2 and b_l^2 (and potentially different architectural parameters c_2), here l is the number of layers in the View 2 network, and the total network function is defined as f_1 and f_2, from L_1 to L_2, for building two neural networks to transform features non-linearly, respectively. The layer sizes of both views are the same, including input layer L_1, hidden layers L_2, and output layer L_3 with each layer's nodes fully connected. The two views' output features are defined as H_1 and H_2, respectively. We use back propagation to update parameters of each view to acquire higher correlation in the CCA layer.

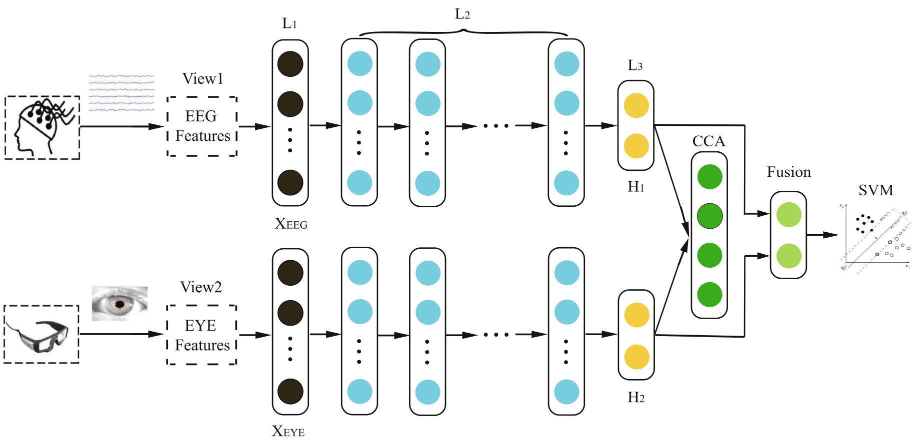


Fig. 1. Our Deep Canonical Correlation Analysis model, including deep networks (input layer, hidden layers and output layer), Canonical Correlation Analysis layer, and classifier SVM.

CCA Calculation. The goal is to jointly learn parameters for both views W_l^i and b_l^i, where $i = \{1, 2\}$, such that $corr(f_1(X_1), f_2(X_2))$ is as high as possible. Let θ_1 be the vector of all parameters W_l^1 and b_l^1 of the first view for $l = 1, \ldots, d$, where d is the number of hidden layers, and similarly for θ_2. The optimization function is:

$$(\theta_1^*, \theta_2^*) = \arg\max corr(H_1, H_2) = \underset{\theta_1, \theta_2}{\arg\max}\, corr(f_1(X_1; \theta_1), f_2(X_2; \theta_2)) \quad (1)$$

According to [15], the correlation of two views' transformed features (H_1 and H_2) can be calculated as follows:

$$corr(H_1, H_2) = corr(f_1(X_1), f_2(X_2)) = ||T||_{tr} = tr(T'T)^{1/2} \tag{2}$$

where

$$\begin{aligned} T &= \hat{\Sigma}_{11}^{-1/2}\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1/2} \\ \hat{\Sigma}_{11} &= \frac{1}{m-1}\overline{H}_1\overline{H}_1' + r_1 I, \\ \hat{\Sigma}_{22} &= \frac{1}{m-1}\overline{H}_2\overline{H}_2' + r_2 I, \\ \hat{\Sigma}_{12} &= \frac{1}{m-1}\overline{H}_1\overline{H}_2'. \end{aligned}$$

The $\overline{H}_1$ and $\overline{H}_2$ are the centered data matrixes:

$$\overline{H}_1 = H_1 - \frac{1}{m}H_1\mathbf{1}, \quad \overline{H}_2 = H_2 - \frac{1}{m}H_2\mathbf{1} \tag{3}$$

and r_1, r_2 are the regularization constants. To update the weighs of networks, we calculate the gradients. If the singular value decomposition of T is $T = UDV'$, then

$$\frac{\partial corr(H_1, H_2)}{\partial H_1} = \frac{1}{m-1}(2\nabla_{11}\overline{H}_1 + \nabla_{12}\overline{H}_2), \tag{4}$$

where

$$\nabla_{11} = -\frac{1}{2}\hat{\Sigma}_{11}^{-1/2}UDU'\hat{\Sigma}_{22}^{-1/2}, \; \nabla_{12} = \hat{\Sigma}_{11}^{-1/2}UV'\hat{\Sigma}_{22}^{-1/2}.$$

Feature Fusion and Classification. We take weighted average of two views' extracted features. DCCA is used for feature extraction, and linear SVM is used as classifier to recognize emotions. The fusion function is defined as follows:

$$F_{fusion} = \alpha H_1 + \beta H_2 \tag{5}$$

where F_{fusion} is fusion features, H_1 and H_2 are extracted features of EEG and eye movement, respectively, and α and β are the fusion weights. In our experiment, in order to balance the composition of features, we set $\alpha = \beta = 0.5$.

3 Experiment Settings

3.1 Dataset

We evaluate the performance of our approach on three real world datasets, the SEED[1] dataset, the SEED IV (See footnote 1) dataset, and the DEAP[2] dataset.

1 http://bcmi.sjtu.edu.cn/~seed/.
2 http://www.eecs.qmul.ac.uk/mmv/datasets/deap/.

- **SEED.** The SEED dataset contains EEG data with three emotions (happy, neutral, and sad) of 15 subjects, and all subjects' data were collected when they watching 15 four-minute-long emotional movie clips, where first 9 movie clips were used as training data and the rest were used as test data. The EEG signals were recorded with ESI NeuroScan System at a sampling rate of 1000 Hz with a 62-channel electrode cap. The eye movement signals were recorded with SMI ETG eye tracking glasses. To compare with the existing work, we used the same data, which contained 27 experiment results from 9 subjects.
- **SEED IV.** The SEED IV dataset contains EEG and eye movement features in total of four emotions (happy, sad, fear, and neutral) [16]. A total of 72 movie clips were used for the four emotions, and forty five experiments were taken by participants to evaluate their emotions while watching the movie clips with keywords of emotions and ratings out of ten points for two dimensions: valence and arousal. The valence scale ranges from sad to happy, and the arousal scale ranges from calm to excited.
- **DEAP.** The DEAP dataset contains EEG signals and other peripheral physiological signals from 32 subjects. These data were collected when participants were watching emotional music videos, which was one-minute-long each. We chose 5 as a threshold to divide the trials into two classes according to the rated levels of arousal and valence. We used 10-fold cross validation to compare our results with Liu *et al.* [5], Yin *et al.* [12], and Tang *et al.* [6] (Fig. 2).

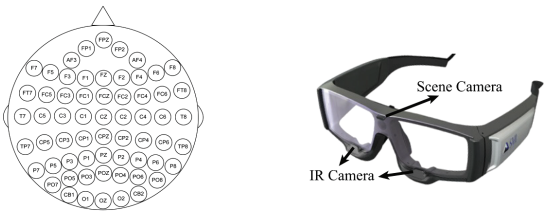


Fig. 2. The EEG electrode layout and SMI ETG eye tracking glasses.

3.2 Feature Extraction

For the SEED and SEED IV datasets, we extracted Differential Entropy (DE) features [19] from each EEG signal channel in five frequency bands: δ (1–4 Hz), θ (4–8 Hz), α (8–14 Hz), β (14–31 Hz), and γ (31–50 Hz). So at each time step, the dimension of EEG features is 310 (5 bands $\times$ 62 channels). As for eye movement features, we used the same features as Lu *et al.* 2015 [3], which were listed in Table 1. At each time step, there were 39 dimensions of pupil diameters in total, including both Power Spectral Density (PSD) and DE features. The extracted EEG features and eye movement features were scaled between 0 and 1.

For the DEAP dataset, because a 4–45 Hz bandpass frequency filter was applied during pre-processing, so we extracted DE features from EEG signals in four frequency bands: θ (4–8 Hz), α (8–14 Hz), β (14–31 Hz), and γ (31–45 Hz). Then in total, the dimension of extracted 32-channel EEG features is 128 (4 bands $\times$ 32 channels). As for peripheral physiological signals, six time-domain features were extracted to describe the signals in different perspective, including minimum value, maximum value, mean value, standard deviation, variance, and squared sum. So the dimension of peripheral physiological features is 48 (6 features $\times$ 8 channels).

Table 1. The details of the extracted eye movement features.

Eye movements parameters	Extracted features
Pupil diameter (X and Y)	Mean,standard deviation, DE in four bands (0–0.2 Hz, 0.2–0.4 Hz, 0.4–0.6 Hz, 0.6–1 Hz)
Disperson (X and Y)	Mean, standard deviation
Fixation duration (ms)	Mean, standard deviation
Blink duration (ms)	Mean, standard deviation
Saccade	Mean, standard deviation of saccade duration (ms) and saccade amplitude
Event statistics	Blink frequency, fixation frequency, fixation dispersion total, fixation duration maximum, fixation dispersion maximum, saccade frequency, saccade duration average, saccade latency average, saccade amplitude average

3.3 Parameter Details

In this paper, we build subject-specific models. We use grid search to find optimal hyperparameters, including learning rate, batch size, regulation parameters, and layer nodes. Taking several experiment results and time consuming into account, we choose learning rate as $1e^3$, batch size as 100, and regulation parameter as $1e^7$. The hidden units in our models are presented in Table 2.

Table 2. Layer's framework of different datasets in our experiments

Dataset	Layers
SEED	$400 \pm 40, 200 \pm 20, 150 \pm 20, 120 \pm 10, 60 \pm 10, 20 \pm 2$
SEEC IV	$400 \pm 40, 200 \pm 20, 150 \pm 20, 120 \pm 10, 90 \pm 10, 60 \pm 10, 20 \pm 2$
DEAP	$1500 \pm 50, 750 \pm 50, 500 \pm 25, 375 \pm 25, 130 \pm 20, 65 \pm 20, 30 \pm 20$

4 Experimental Results

4.1 Results on Different Datasets

Table 3 shows the comparison results of different approaches on the SEED dataset, different feature extraction methods are listed in the first line and SVM is used as classifier for all methods. From Table 3, DE Feature fusion tested on SVM achieved higher classification accuracy and less std than the CCA method, which directly used CCA on EEG and eye movement features. BDAE used RBM pre-training to build a multimodal autoencoder model performed a better result of 93.19% [5]. Tang *et al.* used Bimodal-LSTM to make fusion by considering timing and classification layer parameters and achieved the state-of-the-art performance [6]. In our DCCA model, we extracted highly correlated features, bringing closer these high-level representations, and achieved better results with test classification accuracy of 94.58% and std of 6.16.

Table 3. Average accuracies (%) and standard deviation of different approaches for three emotions classification on the SEED dataset

	CCA	DE features	BDAE [5]	Bimodal-LSTM [6]	DCCA
Accuracy(%)	40.35	81.21	93.19	93.97	**94.58**
Std	16.38	12.51	8.23	7.03	**6.16**

Comparison results on the SEED IV dataset is shown in Table 4. We regard Zheng *et al.*'s deep learning results as our baseline [16]. We compare our DCCA model with different existing feature extraction methods. Table 4 presents that BDAE achieved better results than DE features. Compared with CCA based approach and other methods, we conclude that DCCA model coordinating high-level features achieves the best results.

Table 4. Average accuracies (%) and standard deviation of different approaches for four emotions classification on the SEED IV dataset

	CCA	DE features	BDAE [16]	DCCA
Accuracy (%)	49.56	75.88	85.11	**87.45**
Std	19.24	16.14	11.79	**9.23**

Tables 5 and 6 demonstrate comparison results of different feature extraction methods on the DEAP dataset, which are for two dichotomous classification and four categories classification, respectively, while SVM is used as classifier. For two dichotomous classification, Liu *et al.*'s multimodal autoencoder model achieved 2% higher than AutoEncoder. Yin *et al.* used an ensemble of deep classifiers, making higher-level abstractions of physiological features [12]. Then Tang *et al.* used Bimodal-LSTM and achieved the state-of-the-art accuracy for two dichotomous classification [6]. For four categories classification, Tripathi achieved accuracy of 81.41% [18]. As for our DCCA method, we learned high-level correlated features and achieved better results than the state-of-the-art method with mean test accuracies of 84.33% and 85.62% for arousal and valence classification and 88.51% for four categories classification.

Table 5. Comparison of average accuracies (%) of different approaches on the DEAP dataset for two dichotomous

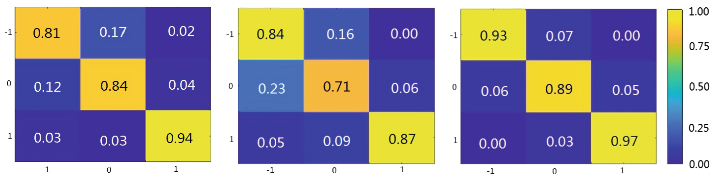
	CCA	AutoEncoder [3]	Liu *et al.* [5]	Yin *et al.* [12]	Tang *et al.* [6]	DCCA
Arousal (%)	61.25	74.49	80.5	84.18	83.23	**84.33**
Valence (%)	69.58	75.69	85.2	83.04	83.82	**85.62**

Table 6. Comparison of average accuracies (%) of different approaches on the DEAP dataset for four categories

Method	CCA	KNN+RF [17]	Tripathi *et al.* [18]	DCCA
Accuracy (%)	40.35	70.04	81.41	**88.51**

4.2 Discussion

The shortcoming of the existing feature-level fusion and multimodal deep learning methods is very difficult to relate the original features in one modality to features in other modality [14]. Moreover, the relations across various modalities are deep instead of shallow. In our DCCA model, we can learn coordinated representation from high-level features and make two views of features become more complementary, which in return improves the classification performance.



(a) (b) (c)

Fig. 3. Confusion matrices of DCCA outputs on the SEED dataset of single modality and feature fusion methods. Each row of the confusion matrices represents the target class and each column represents the predicted class. The element (i, j) is the percentage of samples in class i that is classified as class j. (a) EEG features; (b) Eye movement features; and (c) Fusion features.

Figure 3 shows the confusion matrices of SEED feature classification. The EEG features have classification accuracy of 0.86 while eye movement features' of 0.81, and the fusion feature has classification accuracy of 0.94. We can draw a conclusion from the confusion matrices that EEG features and eye movement features are complementary.

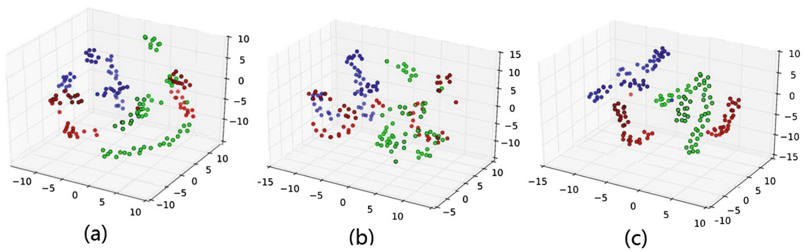


Fig. 4. t-SNE 3D visualization of extracted features on the SEED dataset, where blue for negative emotion, red for neutral emotion, and green for positive emotion. (a) EEG features; (b) Eye movement features; and (c) Fusion features. (Color figure online)

To find out the distribution of fusion features, we use t-SNE to make dimensionality reduction of the high-dimensional extracted features for visualization [20]. Figure 4 presents high-dimensional input features which are reduced to three dimensions for visualization. Comparing the EEG features, eye movement features, and fusion features, we can directly conclude that the fusion features are more reasonable and have better distribution than single-model of EEG and eye movement features, which are beneficial for classification.

5 Conclusion

In this paper, we have used Deep Canonical Correlation Analysis to extract highly correlated high-level features of two views on three real world datasets. The experimental results show that canonical correlation analysis with deep networks extension can achieve higher classification accuracy of emotion recognition when higher correlation is acquired. The deep networks with nodes' weights updated by back propagation can extract better features, which are more correlated of two views. Our work first put coordinated representation into multimodal emotion recognition and indicated a new way of multimodal representation in high-level fusion features.

Acknowledgments. This work was supported in part by the grants from the National Key Research and Development Program of China (Grant No. 2017YFB1002501), the National Natural Science Foundation of China (Grant No. 61673266), and the Fundamental Research Funds for the Central Universities.

References

1. Soleymani, M., Pantic, M., Pun, T.: Multimodal emotion recognition in response to videos. IEEE Trans. Affect. Comput. **3**, 211–223 (2012)
2. Zheng, W.L., Dong, B.N., Lu, B.L.: Multimodal emotion recognition using EEG and eye tracking data. In: EMBS 2014, pp. 5040–5043 (2014)
3. Lu, Y., Zheng, W.L., Li, B., Lu, B.L.: Combining eye movements and EEG to enhance emotion recognition. In: IJCAI 2015, pp. 1170–1176 (2015)
4. Baltrusaitis, T., Ahuja C., Morency, L.P.: Multimodal machine learning: a survey and taxonomy. IEEE Trans. Pattern Anal. 1–20 (2018)
5. Liu, W., Zheng, W.-L., Lu, B.-L.: Emotion recognition using multimodal deep learning. In: Hirose, A., Ozawa, S., Doya, K., Ikeda, K., Lee, M., Liu, D. (eds.) ICONIP 2016. LNCS, vol. 9948, pp. 521–529. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46672-9_58
6. Tang, H., Liu, W., Zheng, W.L., Lu, B.L.: Multimodal emotion recognition using deep neural networks. In: Liu, D., Xie, S., Li, Y., Zhao, D., El-Alfy, E.S. (eds.) ICONIP 2017. LNCS, vol. 10637, pp. 811–819. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70093-9_86
7. Frome, A., et al.: Devise: a deep visual-semantic embedding model. In: NIPS 2013, pp. 2121–2129 (2013)
8. Bronstein, M., Michel, F., Paragios, N.: Data fusion through cross-modality metric learning using similarity-sensitive hashing. In: CVPR 2010, pp. 3594–3601 (2010)
9. Zhang, H., Hu, Z., Deng, Y., Sachan, M., Yan, Z., Xing, E.P.: Learning concept taxonomies from multimodal data. In: ACL 2016, pp. 1791–1801 (2016)
10. Lai, P.L., Fyfe, C.: Kernel and nonlinear canonical correlation analysis. Int. J. Neural Syst. **10**, 365–377 (2000)
11. Sohn, K., Lee, H., Yan, X.: Learning structured output representation using deep conditional generative models. In: NIPS 2015, pp. 3483–3491 (2015)
12. Yin, Z., Zhao, M., Wang, Y., Yang, J., Zhang, J.: Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. Comput. Methods Progr. Biomed. **140**, 93–110 (2017)
13. Zheng, W.L., Zhu, J.Y., Peng, Y., Lu, B.L.: EEG-based emotion classification using deep belief networks. In: IEEE ICME 2014, pp. 1–6 (2014)
14. Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A.Y.: Multimodal deep learning. In: ICML 2011, pp. 689–696 (2011)
15. Andrew, G., Arora R., Bilmes, J.A., Livescu, K.: Deep canonical correlation analysis. In: ICML 2013, pp. 1247–1255 (2013)
16. Zheng, W.L., Liu, W., Lu, Y., Lu, B.L., Cichocki, A.: EmotionMeter: a multimodal framework for recognizing human emotions. IEEE Trans. Cybern. **99**, 1–13 (2018)
17. Chen, J., Hu, B., Wang, Y., Dai, Y., Ya, Y., Zhao, S.: A three-stage decision framework for multi-subject emotion recognition using physiological signals. In: IEEE BIBM 2016, pp. 470–474 (2016)
18. Tripathi, S., Acharya, S., Sharma, R.D., Mittal, S., Bhattacharya, S.: Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset. In: IAAI 2017 (2017)
19. Duan, R.N., Zhu, J.Y., Lu, B.L.: Differential entropy feature for EEG-based emotion classification. In: IEEE NER 2013, pp. 81–84 (2013)
20. Maaten, L., Hinton, G.E., Bengio, Y.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**, 2579–2605 (2008)
21. Hossain, M.Z., Kabir, M.M., Shahjahan, M.: A robust feature selection system with colins CCA network. Neurocomputing **173**, 855–863 (2016)