

HKBU at MediaEval 2017 Emotional Impact of Movies Task

Yang Liu^{1,2}, Zhonglei Gu¹, Tobey H. Ko³

¹Department of Computer Science, Hong Kong Baptist University, HKSAR, China

²Institute of Research and Continuing Education, Hong Kong Baptist University, Shenzhen, China

³Department of Industrial and Manufacturing Systems Engineering, University of Hong Kong, HKSAR, China
csygliu@comp.hkbu.edu.hk, cszlg@comp.hkbu.edu.hk, tobeyko@hku.hk

ABSTRACT

In this paper, we describe our model designed for automatic prediction of emotional impact of movies. Specifically, a two-stage learning framework is proposed. First, the dimensionality reduction techniques are employed to discover the key emotion information embedded in the original feature space. Specifically, we use a classical method principal component analysis (PCA) and a new algorithm biased discriminant embedding (BDE) to learn the subspace. After dimensionality reduction, SVM is utilized for prediction. Experimental results validate the effectiveness of our approaches.

1 INTRODUCTION

In this 2017 Emotional Impact of Movies Task, the participants are asked to predict the expected emotional impact of movie content. Specifically, to predict the response of a general audience to a given stimulus [4], either induced valence, induced arousal, or induced fear, from movie clip segments. The dataset used in this task is the LIRIS-ACCDEE dataset (liris-accdeec.ec-lyon.fr), which contains videos from a set of 160 professionally made and amateur movies, shared under Creative Commons licenses that allow redistribution [1]. More details of the task requirements as well as the dataset description can be found in [3].

In this paper, we propose a two-stage learning framework to predict the emotional impact of movies. First, we use dimensionality reduction to project the original data to the low-dimensional subspace, with the key emotional information being well preserved. Specifically, we use principal component analysis (PCA) to extract features for arousal and valence prediction and propose a new algorithm called biased discriminant embedding (BDE) to extract features for fear prediction. After dimensionality reduction, we use support vector regression and classification [2, 5] for continuous and discrete predictions, respectively.

2 OUR MODEL

2.1 Feature Extraction

2.1.1 Principal Component Analysis. Given the data matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, where $\mathbf{x}_i \in \mathbb{R}^D$ denotes the feature vector of the i -th data point, principal component analysis

(PCA) aims to learn a $D \times d$ transformation matrix \mathbf{W} , which maximizes the following objective:

$$\mathbf{W} = \arg \max_{\mathbf{W}} \text{tr} \left(\mathbf{W}^T (\mathbf{X} - \bar{\mathbf{x}} \mathbf{1}_n^T) (\mathbf{X} - \bar{\mathbf{x}} \mathbf{1}_n^T)^T \mathbf{W} \right), \quad (1)$$

where $\bar{\mathbf{x}} = \sum_{i=1}^n \mathbf{x}_i / n$ and $\mathbf{1}_n$ denotes the $n \times 1$ vector with all entries being 1. The optimization problem in Eq. (1) could be solved by the standard eigen-decomposition.

2.1.2 Biased Discriminant Embedding. Given the data matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ and the label vector $\mathbf{l} = [l_1, l_2, \dots, l_n]$, where $l_i \in \{0, 1\}$ denotes the corresponding label of \mathbf{x}_i , 1 for fear and 0 otherwise, biased discriminant embedding (BDE) aims to maximize the *biased* discriminant information in the reduced subspace. The motivation for proposing the *biased* discrimination is that in fear prediction, one might be more interested in the *fear* class than the *non-fear* one.

The objective function of BDE is given as follows:

$$\mathbf{W} = \arg \max_{\mathbf{W}} \text{tr} \left(\frac{\mathbf{W}^T \mathbf{S}^b \mathbf{W}}{\mathbf{W}^T \mathbf{S}^w \mathbf{W}} \right), \quad (2)$$

where $\mathbf{S}^w = \sum_{i,j=1}^n (N_{ij} \times l_i \times l_j) (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^T$ denotes the biased within-class scatter, $\mathbf{S}^b = \sum_{i,j=1}^n (N_{ij} \times |l_i - l_j|) (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^T$ denotes the biased between-class scatter, and $N_{ij} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma)$ measures the closeness between two data samples \mathbf{x}_i and \mathbf{x}_j . The optimization problem could be solved by generalized eigen-decomposition.

2.2 Emotion Prediction

2.2.1 Support Vector Regression. For predicting the arousal and valence values, we use ν -SVR [5] to train two regressors separately. The dual problem that ν -SVR aims to solve is:

$$\begin{aligned} \min_{\alpha, \alpha^*} \quad & \frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^T \mathbf{K} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) + (\mathbf{1}^{(m)})^T (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) \\ \text{s.t.} \quad & \mathbf{e}^T (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) = 0, \quad \mathbf{e}^T (\boldsymbol{\alpha} + \boldsymbol{\alpha}^*) \leq C\nu \\ & 0 \leq \alpha_i, \alpha_i^* \leq C/n, \quad i = 1, \dots, n. \end{aligned} \quad (3)$$

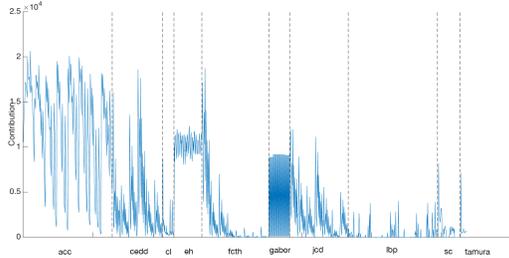
The prediction label of a new coming vector \mathbf{y} is:

$$l = \sum_{i=1}^n (\alpha_i^* - \alpha_i) K(\mathbf{y}_i, \mathbf{y}) + b. \quad (4)$$

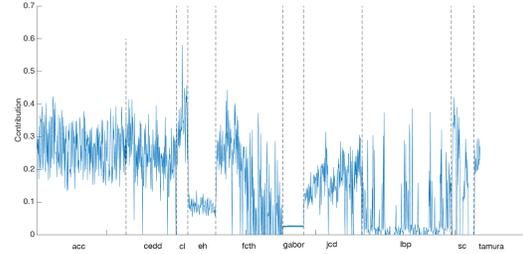
2.2.2 Support Vector Classification. To predict the binary fear labels, we use ν -SVC [5]. The dual problem that ν -SVC

Table 1: Performance of the proposed model (provided by the organizers).

	arousal_mse	arousal_r	valence_mse	valence_r	fear_accuracy	fear_precision	fear_recall	fear_f1
Run 1	0.126	0.045	0.191	-0.151	0.831	0.226	0.020	0.036
Run 2	0.138	-0.075	0.256	-0.253	0.762	0.168	0.065	0.078
Run 3	0.144	-0.081	0.257	-0.239	0.763	0.168	0.065	0.078
Run 4	0.140	-0.081	0.259	-0.254	0.763	0.168	0.065	0.078
Run 5	0.137	-0.079	0.249	-0.267	0.792	0.222	0.050	0.058



(a) PCA for arousal-valence prediction



(b) BDE for fear prediction

Figure 1: Contribution of each individual feature in arousal-valence and fear prediction tasks.

aims to solve is:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \alpha^T \mathbf{Q} \alpha \\ \text{s.t.} \quad & \mathbf{1}^T \alpha = 0, \quad \mathbf{e}^T \alpha \geq \nu, \quad 0 \leq \alpha_i \leq 1/n, \quad i = 1, \dots, n. \end{aligned} \quad (5)$$

The prediction label of a new coming vector \mathbf{y} is given by:

$$l = \text{sgn} \left(\sum_{i=1}^n l_i \alpha_i K(\mathbf{y}_i, \mathbf{y}) + b \right). \quad (6)$$

3 RESULTS

In this section, we report our experimental settings and the evaluation results. For each 1-second segment, we construct a 1,271-D feature set, including 256-D Auto Color Correlogram (acc) features, 144-D Color and Edge Directivity Descriptor (ceddd) features, 33-D Color Layout (cl) features, 80-D Edge Histogram (eh) features, 192-D Fuzzy Color and Texture Histogram (fctth) features, 60-D Gabor (gabor) features, 168-D Joint descriptor joining CEDD and FCTH in one histogram (jcd) features, 256-D Local Binary Patterns (lbp) features, 64-D Scalable Color (sc) features, and 18-D Tamura (tamura) features.

Subtask 1: For Run 1, we use the original 12,710-D feature set (10 seconds) as the input. For Runs 2-5, we use PCA to reduce the original feature set to the 50-D, 80-D, 57-D, and 40-D subspaces, respectively. ν -SVR with RBF kernel is then used for prediction. We set $\nu = 0.5$ and $\gamma = 1/D$.

Subtask 2: For Run 1, we use the original 12,710-D feature set (10 seconds) as the input. For Runs 2-5, we use BDE to reduce the original feature set to the 2-D subspaces. ν -SVC with RBF kernel is then used for prediction. We set $\nu = 0.1$ and $\gamma = 1/D$.

Table 1 reports the evaluation results of our model provided by the task organizers. We can see that the performance in

the low-dimensional subspace is generally worse than that in the original feature space. The reason might be that the dimension of the original feature space is quite high so that such a low-dimensional subspace cannot fully capture the discriminant information embedded in the original data.

In addition to the overall performance, we analyze the contribution of each dimension in the original feature space. The contribution of the i -th dimension is defined as $\text{Contribution}_i = \sum_j \lambda_j |w_{ij}|$, where λ_j denotes the j -th eigenvalue, w_{ij} denotes the (i, j) -th element of \mathbf{W} , and $|\cdot|$ denotes the absolute value operator. From Figure 1 we can see that the acc feature makes important contributions in both arousal/valence and fear prediction tasks, which indicates its importance in emotional discriminant analysis.

4 DISCUSSION AND OUTLOOK

This paper introduces our model designed for predicting emotional impact of movies. To extract the compact representation of original feature set, dimensionality reduction is utilized. We then use SVM for prediction. For the future work, we are interested in analyzing the relation between arousal/valence and fear, which could help understanding the emotional impacts deeply. Moreover, as the ground truth (labels) of emotions are provided by human beings, they generally vary with each individual and are somewhat subjective. We are therefore particularly interested in refining the human labeled ground truth via machine learning technologies.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant 61503317, and in part by the Faculty Research Grant of Hong Kong Baptist University (HKBU) under Project FRG2/16-17/032.

REFERENCES

- [1] Y. Baveye, E. Dellandréa, C. Chamaret, and L. Chen. 2015. LIRIS-ACCEDE: A Video Database for Affective Content Analysis. *IEEE Transactions on Affective Computing* 6, 1 (Jan 2015), 43–55.
- [2] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2 (2011), 27:1–27:27. Issue 3.
- [3] Emmanuel Dellandrea, Martijn Huigsloot, Liming Chen, Yoann Baveye, and Mats Sjöberg. 2017. The MediaEval 2017 Emotional Impact of Movies Task. In *Mediaeval 2017 Workshop*.
- [4] A. Hanjalic. 2006. Extracting moods from pictures and sounds: towards truly personalized TV. *IEEE Signal Processing Magazine* 23, 2 (March 2006), 90–100.
- [5] Bernhard Schölkopf, Alex J. Smola, Robert C. Williamson, and Peter L. Bartlett. 2000. New Support Vector Algorithms. *Neural Comput.* 12, 5 (2000), 1207–1245.