Towards Learning Emotional Subspace

Tobey H. Ko¹, Zhonglei Gu², Tiantian He³, Yang Liu^{2,4}

¹Department of Industrial and Manufacturing Systems Engineering, University of Hong Kong, HKSAR, China

²Department of Computer Science, Hong Kong Baptist University, HKSAR, China

³Department of Computing, The Hong Kong Polytechnic University, HKSAR, China

⁴Institute of Research and Continuing Education, Hong Kong Baptist University, Shenzhen, China

ABSTRACT

We introduce a model designed to predict emotional impact of movies through affective video content analysis. Specifically, our approach utilizes a two-stage learning framework, which first conducts subspace learning using emotion preserving embedding (EPE) or biased discriminant embedding (BDE) to uncover the informative subspace from the original feature space according to the continuous or discrete emotional labels, respectively, and then carries out the prediction utilizing the support vector machine (SVM). Experimentation on a movie dataset validates the effectiveness of our learning framework.

1 INTRODUCTION

The Emotional Impact of Movies Task in MediaEval 2018 aimed at developing approaches which automatically and accurately predict the emotional impact of movie content, when the said movie content containing a certain stimulus, including either induced valence, induced arousal, or induced fear, is exposed to the general audience. Automatic video emotions discriminator capable of identifying movie content that is potentially inducing harmful emotions is expected to be developed through the successful implementation of this task. Approaches proposed for the task are trained and evaluated using the LIRIS-ACCEDE dataset (liris-accede.eclyon.fr) [1], which offers a collection of 160 professionally made and amateur movies shared under the Creative Commons license, out of which 44 of them are selected and annotated with their respective fear, valence, and arousal labels. More details of the task requirements and the data description can be found in the task paper [4].

In this paper, a two-stage learning framework is introduced for automatic prediction of the emotional impact of movie content. In order to learn an accurate feature representation of the induced emotions in movie content, the learning framework first projects the original data to a learned lowdimensional feature subspace using dimensionality reduction techniques, then conducts prediction on the learned subspace using classification techniques. Specifically, the dimensionality reduction process was completed using emotion preserving embedding (EPE) to learn the subspace for induced arousal and induced valence, whereas the biased discriminant embedding algorithm (BDE) [5] was implemented to learn the subspace for induced fear in movie content. On the learned low-dimensional feature subspace, we employ the classical support vector regression and classification techniques, as they are efficient and effective, to predict the induced affective emotion of the movie content in both a continuous and discrete manner.

2 LEARNING EMOTIONAL SUBSPACE

2.1 Emotion Preserving Embedding

EPE is proposed to learn the subspace for the continuous arousal and valence labels. Given the training set $\mathcal{X} = \{(\mathbf{x}_1, \mathbf{l}_1), (\mathbf{x}_2, \mathbf{l}_2), ..., (\mathbf{x}_n, \mathbf{l}_n)\}$, where $\mathbf{x}_i \in \mathbb{R}^D$ $(i = 1, \dots, n)$ is the feature vector of the *i*-th movie and $\mathbf{l}_i = [a_i, v_i]^T$ is the corresponding label vector containing the arousal label a_i and the valence label v_i . EPE aims to learn a $D \times d$ transformation matrix \mathbf{W} to map \mathbf{x}_i $(i = 1, \dots, n)$ to a low-dimensional subspace, where the emotion information and manifold structure of the dataset can be well preserved. To achieve this goal, EPE optimizes the following objective function:

$$\mathbf{W} = \underset{\mathbf{W}}{\operatorname{arg\,min}} \sum_{i,j=1}^{n} \|\mathbf{W}^{T}(\mathbf{x}_{i} - \mathbf{x}_{j})\|^{2} \cdot \left(\alpha S_{ij} + (1 - \alpha) N_{ij}\right), \quad (1)$$

where $S_{ij} = exp(-||\mathbf{l}_i - \mathbf{l}_j||^2/2\sigma^2)$ measures the label similarity of \mathbf{x}_i and \mathbf{x}_j , $N_{ij} = exp(-||\mathbf{x}_i - \mathbf{x}_j||^2/2\sigma^2)$ measures the closeness between \mathbf{x}_i and \mathbf{x}_j , and $\alpha \in [0, 1]$ is the parameter balancing the emotion information and the manifold structure. Eq. (1) could be equivalently rewritten as follows:

$$\mathbf{W} = \operatorname*{arg\,min}_{\mathbf{W}} tr(\mathbf{W}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{W}), \qquad (2)$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n] \in \mathbb{R}^{D \times n}$ is the data matrix, $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is the $n \times n$ Laplacian matrix [2], and \mathbf{D} is a diagonal matrix defined as $D_{ii} = \sum_{j=1}^{n} A_{ij}$ (i = 1, ..., n), where $A_{ij} = \alpha S_{ij} + (1 - \alpha)N_{ij}$. Then the optimal \mathbf{W} can be obtained by finding the eigenvectors corresponding to the smallest eigenvalues of the following eigen-decomposition problem:

$$\mathbf{X}\mathbf{L}\mathbf{X}^{T}\mathbf{w} = \lambda\mathbf{w}.$$
 (3)

After obtaining \mathbf{W} , we can obtain the low-dimensional representation of \mathbf{x}_i by $\mathbf{y}_i = \mathbf{W}^T \mathbf{x}_i$.

2.2 Biased Discriminant Embedding

BDE is a subspace learning algorithm we have proposed for the same task in the last year [5]. It aims to learn the

Copyright held by the owner/author(s).

MediaEval'18, 29-31 October 2018, Sophia Antipolis, France

MediaEval'18, 29-31 October 2018, Sophia Antipolis, France

subspace for the binary fear labels. In this scenario, each data sample \mathbf{x}_i is associated with a binary label $l_i \in \{0, 1\}$, with 1 for fear and 0 otherwise. BDE aims to maximize the *biased* discriminant information in the learned subspace. As mentioned in [5], the so-called *biased* discrimination is designed to emphasize the importance of the *fear* class. The objective function of BDE is given as follows:

$$\mathbf{W} = \operatorname*{arg\,max}_{\mathbf{W}} tr\left(\frac{\mathbf{W}^T \mathbf{S}^b \mathbf{W}}{\mathbf{W}^T \mathbf{S}^w \mathbf{W}}\right),\tag{4}$$

where $\mathbf{S}^w = \sum_{i,j=1}^n (N_{ij} \times l_i \times l_j) (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^T$ and $\mathbf{S}^b = \sum_{i,j=1}^n (N_{ij} \times |l_i - l_j|) (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^T$ denote the biased within-class and between-class scatters, respectively.

The optimal \mathbf{W} then can be obtained by finding the eigenvectors corresponding to the largest eigenvalues of the following generalized eigen-decomposition problem:

$$\mathbf{S}^{b}\mathbf{w} = \lambda \mathbf{S}^{w}\mathbf{w}.$$
 (5)

3 RESULTS AND ANALYSIS

In this section, we evaluate the performance of our approach on the MediaEval 2018 Emotional Impact of Movies Task. There are 93337 and 26600 frames in the development set and the test set, respectively. We use 11 types of features to construct the original feature vector for each frame, i.e., 1583-D Auto Color Correlogram (ACC), 256-D Color and Edge Directivity Descriptor (CEDD), 144-D Color Layout (CL), 33-D Edge Histogram (EH), 80-D Fuzzy Color and Texture Histogram (FCTH), 192-D Gabor, 60-D Joint descriptor joining CEDD and FCTH in one histogram (JCD), 168-D Scalable Color (SC), 256-D Tamura, 64-D Local Binary Patterns (LBP), and 18-D VGG16 fc6 layer (FC6). The total dimension of the original feature space is therefore 2854.

For valence/arousal prediction, we use EPE to learn the transformation matrix \mathbf{W} from the development set and use \mathbf{W} to project the *D*-dimensional development and test data (D = 2854) to the *d*-dimensional subspace. We set d = 4, 5, 9, 10 for Runs 1, 2, 3, 4, respectively. We set $\alpha = 0.5$ in our experiment to equally consider the emotion information and the manifold structure. Then we train the ν -SVR [6] on the *d*-dimensional development set and apply the trained model for prediction on the *d*-dimensional test set. For SVR, we use RBF kernel and the default settings recommended by libsvm [3]: $\nu = 0.5$ and $\gamma = 1/d$.

For fear prediction, we use BDE to learn **W**. Similar to the previous experiment, we set d = 4, 5, 9, 10 for Runs 1, 2, 3, 4, respectively. Then we train the ν -SVC [6] on the *d*-dimensional development set and apply the trained model for classification on the *d*-dimensional test set. Similarly, We use RBF kernel and the default settings recommended by libsvm [3]: $\nu = 0.5$ and $\gamma = 1/d$.

Tables 1-3 present the results of our approach in which several observation can be derived. First, in Table 1, Run 1 (d = 4) performs the best while Run 4 (d = 10) performs the worst. Moreover the performance drops when dimensionality of subspace increases. This indicates that the arousal information may embed in a very low-dimensional subspace, and thus

Table 1: Results of arousal prediction on the Media-Eval 2018 Emotional Impact of Movies Task.

	Run 1	$\operatorname{Run}2$	Run 3	Run 4
MSE	0.1493	0.1574	0.1608	0.1623
PCC	0.0828	0.0650	0.0487	0.0255

Table 2: Results of valence prediction on the Media-Eval 2018 Emotional Impact of Movies Task.

	Run 1	$\operatorname{Run}2$	Run 3	Run 4
MSE	0.1016	0.1089	0.1089	0.1076
\mathbf{PCC}	0.0499	0.0164	0.0872	0.1142

Table 3: Results of fear prediction on the MediaEval2018 Emotional Impact of Movies Task.

	$\operatorname{Run}1$	$\operatorname{Run}2$	Run 3	$\operatorname{Run}4$
Intersection Union	0.1052	0.0612	0.0360	0.0196

increasing further the dimensionality of subspace may have an adverse effect in terms of prediction of induced arousal. However, results in Table 2 do not yield clear implication in the optimality of valence prediction with respect to the dimensionality of the learned subspace. The reason might be that we have not yet discovered the optimal dimension of the subspace for valence. Further investigation is needed if we intend to uncover the key to obtaining an optimal dimension for learned subspace. From Table 3, we can see that the performance of our method on fear prediction is unsatisfactory. A possible reason is the high imbalance between fear class and non-fear class, which makes the traditional learning mechanism inefficient, even though we have made some effort in modeling the class imbalance during subspace learning.

4 CONCLUSION

The paper describes our approach designed for predicting emotional impact of movies and validate the approach on the MediaEval 2018 Emotional Impact of Movies Task. The future work will be conducted from the following two aspects. First, we are interested in exploring how to build a joint learning mechanism for both arousal and valence, as these two emotional dimensions are related to each other. Second, we will investigate more effective ways to model the class imbalance in subspace learning and the subsequent classification, especially for the extremely imbalanced cases.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61503317, in part by the General Research Fund (GRF) from the Research Grant Council (RGC) of Hong Kong SAR under Project HKBU12202417, and in part by the SZSTI Grant with the Project Code JCYJ20170307161544087. Towards Learning Emotional Subspace

REFERENCES

- 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] M. Belkin and P. Niyogi. 2001. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. In Advances in Neural Information Processing Systems 14 (NIPS). 585–591.
- [3] 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.
- [4] E. Dellandrea, M. Huigsloot, L. Chen, Y. Baveye, and M. Sjoberg. 2018. The MediaEval 2018 Emotional Impact of Movies Task. In *Mediaeval 2018 Workshop*.
- [5] Y. Liu, Z. Gu, and T. Ko. 2017. HKBU at MediaEval 2017 Emotional Impact of Movies Task. In *Mediaeval 2017 Work-shop*.
- [6] Bernhard Schölkopf, Alex J. Smola, Robert C. Williamson, and Peter L. Bartlett. 2000. New Support Vector Algorithms. *Neural Comput.* 12, 5 (2000), 1207–1245.