Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Understanding emotional impact of images using Bayesian multiple kernel learning

He Zhang *, Mehmet Gönen, Zhirong Yang, Erkki Oja

Department of Information and Computer Science, Aalto University School of Science, FI-00076 Aalto, Espoo, Finland

ARTICLE INFO

Article history:
Received 23 February 2014
Received in revised form
15 August 2014
Accepted 9 October 2014

Keywords:
Image emotions
Multiple kernel learning
Multiview learning
Variational approximation
Low-level image features

ABSTRACT

Affective classification and retrieval of multimedia such as audio, image, and video have become emerging research areas in recent years. The previous research focused on designing features and developing feature extraction methods. Generally, a multimedia content can be represented with different feature representations (i.e., views). However, the most suitable feature representation related to people's emotions is usually not known a priori. We propose here a novel Bayesian multiple kernel learning algorithm for affective classification and retrieval tasks. The proposed method can make use of different representations simultaneously (i.e., multiview learning) to obtain a better prediction performance than using a single feature representation (i.e., single-view learning) or a subset of features, with the advantage of automatic feature selections. In particular, our algorithm has been implemented within a multilabel setup to capture the correlation between emotions, and the Bayesian formulation enables our method to produce probabilistic outputs for measuring a set of emotions triggered by a single image. As a case study, we perform classification and retrieval experiments with our algorithm for predicting people's emotional states evoked by images, using generic low-level image features. The empirical results with our approach on the widely-used `International Affective Picture System (IAPS)` data set outperform several existing methods in terms of classification performance and results interpretability.

1. Introduction

Affective computing [1] aims to help people communicate, understand, and respond better to affective information such as audio, image, and video in a way that takes into account the user's emotional states. Among the emotional stimuli, affective image classification and retrieval have attracted increasing research attention in recent years, due to the rapid expansion of the digital visual libraries on the Web. While most of the current Content-Based Image Retrieval (CBIR) systems [2] are designed for recognizing objects and scenes such as plants, animals and outdoor places, an Emotional Semantic Image Retrieval (ESIR) system [3] aims at incorporating the user's affective states to enable queries like "beautiful flowers", "cute dogs", "exciting games", etc.

Though emotions are highly subjective human factors, still they have stability and generality across different people and cultures [5]. As an example, Fig. 1 shows two pictures taken from a photo sharing site (`ArtPhoto` [4]). The class labels of "Amusement" and "Fear" are determined by the emotion that has received the most votes from people. Intuitively, an "Amusement" picture usually makes people feel pleasant or induces high valence, whereas a "Fear" picture may induce low valence but high arousal to the viewer.

In analogy to the concept of "semantic gap" that implies the limitations of image recognition techniques, the "affective gap" can be defined as "the lack of coincidence between the measurable signal properties, commonly referred to as features, and the expected affective state in which the user is brought by perceiving the signal" [6]. Concerning the studies related to image affect recognition, three major challenges can be identified: (a) the modeling of affect, (b) the extraction of image features to reflect affective states, and (c) the building of classifiers to bridge the "affective gap".

Most of the current works (e.g., [5,7,4,8]) use descriptive words (e.g., the scenario in Fig. 1) to represent affective space. To obtain the ground truth label for learning, each image is assigned with a single emotional label among various emotional categories based on the maximum votes from the viewers. However, an image can usually evoke a mixture of affective feelings in people rather than a single one. Furthermore, the emotions often conceptually correlate with each other in the affective space. For example, the two paintings shown in Fig. 2 are labeled as "Excitement" and "Sad (ness)" according to the maximum votes (from the web survey in [4]). Nevertheless, by examining the votes from the viewers, each image actually has evoked a distribution of emotions rather than a

* Corresponding author.
E-mail addresses: he.zhang@aalto.fi (H. Zhang), mehmet.gonen@aalto.fi (M. Gönen), zhirong.yang@aalto.fi (Z. Yang), erkki.oja@aalto.fi (E. Oja).

http://dx.doi.org/10.1016/j.neucom.2014.10.093
0925-2312/

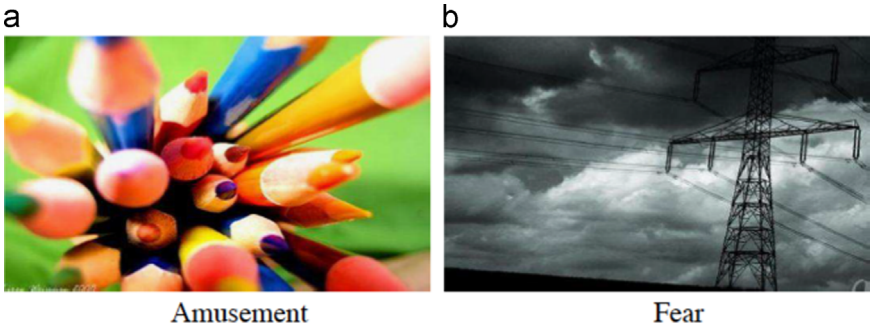

Fig. 1. Example of images from a photo sharing site (`ArtPhoto` [4]) with the ground truth labels of Amusement and Fear.

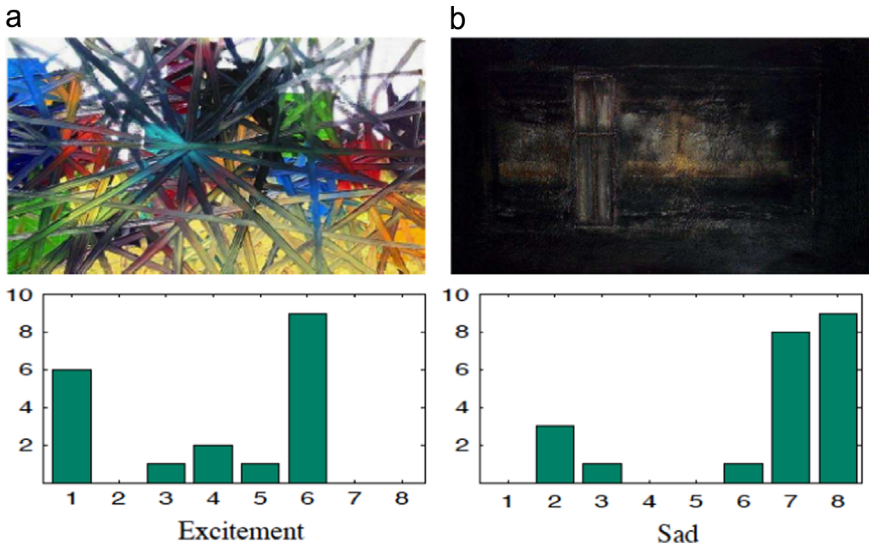

Fig. 2. Example of images from an online user survey showing that images can evoke mixed feelings in people instead of a single one [4]. The *x*-axis shows emotions (from left to right): Amusement, Anger, Awe, Contentment, Disgust, Excitement, Fear, Sad. The *y*-axis shows the number of votes.

single one. Moreover, the correlations can be observed between certain emotions. For example, "Amusement" is closely associated with "Excitement", and "Fear" often comes with "Sadness".

Feature extraction is a prerequisite step for image classification and retrieval tasks [2], especially for the recognition of emotions induced by pictures or artworks. In the literature, much effort has been spent on designing features specific to image affect recognition (e.g., [9,7,4,10,11]). Other works (e.g., [12–14,8]) used the generic low-level color, shape, and texture features for detecting the image emotions. Concerning the inference, supervised learning has been used more often than unsupervised learning for inferring the image emotions. Among the classifiers, Support Vector Machines (SVMs) have been adopted by most of the works (e.g., [13,15,7,16,8]). Since the most suitable feature representation or subset related to people's emotions is not known a priori, feature selection has to be done for better prediction performance prior to the final prediction, which increases the computational complexity. Instead of using a single representation or view, we can also make use of different representations or views at the same time. This implies that multiview learning [17] is preferred to single-view learning. Multiview learning with kernel-based methods belongs to the framework of Multiple Kernel Learning (MKL), which is a principled way of combining kernels calculated on different views to obtain a better prediction performance than single-view learning methods (see [18] for a recent survey).

In this paper, we propose a novel Bayesian multiple kernel learning algorithm for affective classification and retrieval tasks with multiple outputs and feature representations. Thanks to the MKL framework, our method can learn the feature representation weights by itself according to the data and task at hand without an explicit feature selection step, which makes the interpretation easy and straightforward. Our method has been implemented within a multilabel setup in order to capture the correlations between emotions. Due to its probabilistic nature, our method is able to produce probabilistic values for measuring the intensities of a set of emotions triggered by a single image. As a case study, we conduct classification and retrieval experiments with our proposed approach for predicting people's emotional states evoked by images, using conventional low-level color, shape, and texture image features. The experimental results on the widely-used `International Affective Picture System (IAPS)` data set show that our proposed Bayesian MKL approach outperforms other existing methods in terms of classification performance, feature selection capacity, and results interpretability.

Our contributions are thus two-fold:

1. Instead of single view representation, a multiview learning with kernel-based method has been applied to emotional image recognition, with the advantages of better prediction performance, automatic feature selection, and interpretation of image emotional impact.
2. A novel Bayesian multiple kernel learning algorithm with multiple outputs and feature representations has been proposed for affective classification and retrieval tasks. Our method is able to

Please cite this article as: H. Zhang, et al., Understanding emotional impact of images using Bayesian multiple kernel learning, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.10.093