Knowledge-Based Systems 84 (2015) 214–223

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Pornographic images recognition based on spatial pyramid partition and multi-instance ensemble learning

Daxiang Li [a,c,*], Na Li [a,c], Jing Wang [b], Tingge Zhu [a,c]

[a] School of Telecommunication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China
[b] Computer Graphics, Imaging and Vision (CGIV) Research Group, School of Computing and Engineering University of Huddersfield, United Kingdom
[c] Key Laboratory of Electronic Information Application Technology for Scene Investigation, Ministry of Public Security, Xi'an 710121, China

ARTICLE INFO

Article history:
Received 21 September 2013
Received in revised form 13 April 2015
Accepted 18 April 2015
Available online 28 April 2015

Keywords:
Multi-instance learning
Pornographic images recognition
Extreme learning machine

ABSTRACT

For tackling the problem of pornographic image recognition, a novel multi-instance learning (MIL) algorithm is proposed by using extreme learning machine (ELM) and classifiers ensemble. Firstly, a spatial pyramid partition-based (SPP) multi-instance modeling technique has been deployed to transform the pornographic images recognition problem into a typical MIL problem. The method has deployed a bag corresponding to an image and an instance corresponding to each partitioned sub-block described by low-level visual features (i.e. color, texture and shape). Secondly, a collection of visual word (VW) has been generated by using hierarchical k-mean clustering method, and then based on the fuzzy membership function between instance and VW, a fuzzy histogram fusion-based metadata calculation method has been proposed to convert each bag to a single sample, which allows the MIL problem to be solved directly by a standard single instance learning (SIL) machine. Finally, by using ELM, a group of base classifiers with different number of hidden nodes have been constructed, and their weights bas been dynamically determined by using performance weighting rule. Therefore, the strategy of classifiers ensemble is used to improve the overall adaptability of proposed ELMCE-MIL algorithm. Experimental results have shown that the method is robust, and its performance is superior to other similar algorithms.

1. Introduction

With the rapid development of computer and communication technologies, the Internet has become one of the primary resources to access variable information and knowledge. However, the ocean of information from the Internet not only contains useful knowledge that people need, but also rapidly increasing objectionable messages, such as violence, pornography, and rumors. Because image contains much more semantic information than texts does, the pornographic images have the most straightforward impact to the teenager's physical and mental health. Thus, purifying the network environment and keeping teenagers away from pornographic web pages becomes a serious challenge to their family and our society. Therefore, an automated, effective and accurate pornographic images recognition method, which can prevent teenagers from browsing online pornographic images, becomes very important and necessary [1,2].

Blacklist and keyword-based approaches are two kinds of common pornographic image recognition methods [3–6]. The blacklist-based methods block all the accesses to the websites in a list of URLs where pornographic contents have been reported. However, since the highly dynamic feature of the model network system, the methods cannot filter all the URLs containing pornographic contents effectively. On the other hand, keyword-based methods attempt to filter images by analyzing the sensitive texts from webpages. But many words belonging to the pornographer's lexicon can be also appeared in webpages for other purpose (i.e. the educational sites about breast cancer). In addition, some pornographic websites embedding their text into images which make the keyword-based analysis impossible to be applied. Therefore, in recent years, content-based pornographic image recognition technology becomes a hot research topic for pornographic images recognition, which has better adaptability when identifying pornographic content from webpages.

In recent years, many content-based pornographic image recognition methods have been proposed and they can be roughly categorized into three groups [7–9], which are model-based, feature-based and region-based methods. (1) Model-based method was first proposed by Fleck and Forsyth in 1996 [10,11]. The

* Corresponding author at: School of Telecommunication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China.
E-mail address: 389097028@qq.com (D. Li).

http://dx.doi.org/10.1016/j.knosys.2015.04.014
0950-7051/

authors firstly detect the skin color pixels in an image using color and texture information and then find all connected columned skin regions that are candidates for trunks and limbs. These skin columns are used to combine constrained subjects derived from a geometrical model of the human body. If the combination can form a reasonable shape of a human body, the image is then treated as pornographic. This method is quite straightforward and easy to apply. However, as human bodies are non-rigid objects and have a wide variety of postures, it is very difficult to build a comprehensive body grouper effectively. Thus, the recognition accuracy is usually not very high [7,8]. (2) Feature-based methods emphasize the extraction of image features [7,8]. For example, Wang et al. [12] uses wavelet transformation to represent images, and then color histogram and normalized central moments are used to form semantic-matching vectors to classify those images. However, it generates high missing and false detection since local details of the image are ignored. (3) Rule-based methods [9] decide pornographic image according to some predefined rules such as the ratio of the skin areas against the image area, the number of connected skin regions, the ratio of the largest skin area to all skin regions etc. As rule-based methods are relatively simple, the recognition results are unsatisfactory. Recently, a bag-of-feature approach based on SIFT has been proposed [13] to classify images into nude and non-nude while the reported average recognition rate is no more than 70%. Yan et al. [14] also proposed a new approach of extracting salient region from pornographic image regions, which remarkably improves the detection rate.

Above mentioned content-based pornographic image recognition methods can be regarded as binary classification problems that the image should be classified into either pornographic or non-pornographic. Because of the users often distinguish the pornographic images by their "semantic features" rather than "visual features", in order to bridge the "semantic gap", the machine learning method, such as support vector machine (SVM) or neural network, can be employed to train the "visual feature" vectors for generating the classification model. But the first challenging task of such methods is to collect a large number of training samples for the learning. Unfortunately, manually collecting (and possibly further annotating, aligning, cropping, etc.) during the supervised learning processes is an expensive endeavor, which is both time consuming and error-prone [15].

To solve this problem, in this paper, the pornographic image recognition problem has been transformed into a multi-instance learning (MIL) problem by regarding each image as a bag and their low-level visual features of divided sub-blocks as instances [16]. An image can be labeled as a positive bag if it is a pornographic image (otherwise it should be labeled as a negative bag). One advantage of using the MIL is its coarse labeling processes, rather than fine labeling at region level, so it can significantly improve the efficiency when applying supervised learning.

There are three major contributions in this paper:

- A SPP-based multi-instance modeling method has been proposed. We applied SPP method to divide each image into a sequence of increasingly finer sub-blocks, and then based on the color, texture and shape features extracting from all the sub-blocks, an image are modeled as three different multi-instance bags. Compared with other method, the SPP method is not only simple and efficient, but also more adaptable and robust.
- A brand new ELM and classifier ensemble based MIL algorithm has been proposed. Firstly, ELM is used to train a group of base classifiers, which the ELM with different number of hidden nodes has been deployed. Secondly, the performance weighting rule is used to determine the weights of the base classifiers dynamically. To the best of our knowledge, this is the first inductive ELM method for the MIL problem with superb efficiency and performance.
- For demonstrating the promising performance of our method, an extensive comparison work was carried out for evaluating our method with many state-of-the-art MIL approaches in a real-world setting when solving pornographic image recognition problem.

The remaining paper is organized as follows: In Section 2, a brief review of recent development of MIL is introduced. Section 3 provides the details of our proposed MIL method, including the multi-instance modeling, metadata extraction, extreme learning machine and classifier ensemble. The experimental results are presented in Section 4 and the Section 5 concludes the paper.

2. Related work

MIL has become an active research area in machine learning since it was first introduced for solving drug activity prediction problems [17]. In the MIL applications, the training samples are usually regarded as bags, where each bag consists of multiple instances. Training labels are associated with bags rather than their instances. According to the original MIL definition, a bag is labeled as positive if at least one of its instances is positive, and it is labeled as negative if all of its instances are negative. The goal of a MIL algorithm is to generate a classifier that can classify unseen bags correctly.

During the last decade, many MIL algorithms have been presented, including axis parallel hyper-rectangles [17], Citation-kNN, Diverse Density (DD), DD with Expectation Maximization (EM-DD) and Neural Network et al. [18]. In those variety of solutions, support vector machines (SVM)-based methods have been highlighted, which have been successfully deployed for solving many machine-learning problems. For example, Andrews et al. [19] modified the SVM formulation, and presented mi-SVM and MI-SVM algorithms in his work. However, unlike the standard SVM, those methods usually suffer from local minima which lead to non-convex optimization problems. Gehler et al. [20] applied deterministic annealing to solve the problem by proposing so-called AL-SVM method, which could find better local minima of the objective function. Gartner et al. [21] designed kernel functions (i.e. set kernel and statistics kernel) directly from the bags, and used a standard SVM to solve the MIL problem. Due to instance labels were unavailable, these multi-instance kernels implicitly simply assumed that all instances in a bag have equally important, which is a very rough assumption. Therefore, Kwok and Cheung [22] designed marginalized kernel and Zhou et al. [23] designed graph kernel for the problem by considering the different contributions from the instances. Recently, some meta-data based methods have also been proposed, which convert each bag in the MIL problem into a single meta-data so that standard single instance learning (SIL) methods (i.e. SVM method) can be used. Name a few, DD-SVM [24], Multi-Instance Learning via Embedded Instance Selection (MILES) [25], MILD [26], MILIS [27], EC-SVM [28], etc. It has been proved that converting each bag of the MIL into a single representation vector, and then using a SIL method to solve the problem is a very effective solution. However, few existed feature representation methods are effective enough to describe the MIL bags, which lead difficulties to adapt some well-known SIL methods when solving the MIL problems. For example, the DD-SVM [24] method has to learn a collection of instance prototypes to construct a new feature space using Diverse Density (DD) function. Therefore, its representational features are very sensitive to noise and suffer from very high computation cost.