International Journal of Computer and Communication Technology

Volume 3 | Issue 3 Article 3

July 2012

Semantic Learning and Web Image Mining with Image Recognition and Classification

Lambodar Jena
Dept. of CSE,C.V.Raman College of Engineering, Bhubaneswar, India, lmjena@yahoo.com

Ramakrushna Swain
Department of CSE and IT, , C.V.Raman College of Engineering, Bhubaneswar, India;, rkswain1@gmail.com

N.K. kamila
C.V.Raman College of Engineering and Technology, nkamila@yahoo.com

Semantic Learning and Web Image Mining with Image Recognition and Classification

Lambodar Jena[1], Ramakrushna Swain[2], Narendra K. Kamila[3]
Department of CSE C.V.Raman College of Engineering, Bhubaneswar, India
[1]lmjena@yahoo.com, [2]rkswain1@gmail.com, [3]nkamila@yahoo.com

Abstract- **Image mining is more than just an extension of data mining to image domain. Web Image mining is a technique commonly used to extract knowledge directly from images on WWW. Since main targets of conventional Web mining are numerical and textual data, Web mining for image data is on demand. There are huge image data as well as text data on the Web. However, mining image data from the Web is paid less attention than mining text data, since treating semantics of images are much more difficult. This paper proposes a novel image recognition and image classification technique using a large number of images automatically gathered from the Web as learning images. For classification the system uses image-feature-based search exploited in content-based image retrieval(CBIR), which do not restrict target images unlike conventional image recognition methods and support vector machine(SVM), which is one of the most efficient & widely used statistical method for generic image classification that fit to the learning tasks. By the experiments it is observed that the proposed system outperforms some existing search systems.**

Keywords- ***Web image mining, image-gathering, image classification, SVM***

I. INTRODUCTION

Due to wide spread of digital imaging devices, digital images of various kinds of real world scenes can be obtained easily, so that demand for image recognition of various kinds of real world images becomes more essential. It is, however, hard to apply conventional image recognition methods to such generic recognition, because most of their applicable targets are restricted[1]. Hence, semantic processing of images such as automatic attaching keywords to images, classification and search in terms of semantic contents of images are desired.

Web images are as diverse as real world scenes, since Web images are taken by a large number of people for various kinds of purpose[2]. This property is completely different from commercial or personal photo collections built by one or a few persons. It can be expected that such diverse images on the Web enable us to measure general visualness[3] of a concept by analyzing Web images associated with the word concept. We can easily extract keywords related to an image on the Web (Web image) from the HTML file linking to it, so that we can regard a Web image as an image with related keywords[4]. The system is constructed as an assembly of three modules[1], depicted in Fig.- 1.

The processing in the system consists of three steps. In the gathering stage, the system gathers images related to given class keywords from the Web automatically. In the learning stage, it extracts image features from gathered images and associates them with each class. In the classification stage, the system classifies an unknown image into one of the classes corresponding to the class keywords by using the association between the image features and the classes.

In order to process the modules and search the semantic concepts[10] in image databases, the Web images are engaged for learning the semantic concepts searching since the Web images associated with textual descriptions can serve as an important knowledge base. Here the strategy is to search the semantic concepts by words from the Web and learn the returned Web images associated with the words. The Web images after filtering out the noisy images serve as the training set for learning in the image databases.

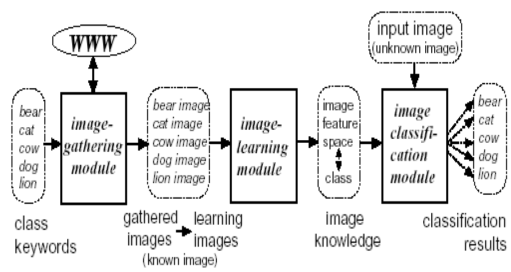

Fig. 1: The Three Modules.

A. BACKGROUND OF SVMS

Support Vector Machine (SVM) techniques [5] are employed for attacking the learning tasks and image classification. The support vector machine (SVM) is a promising classification technique. It can separate the classes with a particular hyperplane which maximizes a quantity called the margin. The margin is the distance from a hyperplane separating the classes to the nearest point in the dataset. The advantage of maximum margin criterion is its robust characteristic against noise in data, and making a solution unique for linearly separable problems. In addition, it is important that the SVM with a theoretically strong support is based on the statistical learning theory framework. An important finding of the statistical learning theory is that the generalization error can be bound by the sum of the empirical error and term, which depends on the VC dimension that characterizes the complexity of the approximating function class . SVM has been extensively used as a classification tool with a great deal of success in a variety of area from object recognition to classification of cancer morphologies . It has also been successfully applied to a number of real-world problems such as handwritten characters and image recognition, face detection and speaker identification.

B. A GEOMETRICAL INTERPRETATION OF SVMS

A geometric interpretation of the SVM illustrates how this idea of smoothness or stability gives rise to a geometric quantity called the margin which is a measure of how well separated the two classes can be. We start by assuming that the classification function is linear[19].

$$f(x) = w \cdot x = \sum_{i=1}^{n} w_i x_i$$

where x_i and w_i are the ith elements of the vectors x and w, respectively. The operation w · x is called a dot product. The label of a new point xnew is the sign of the above function, ynew = sign [f(xnew)]. The classification boundary, all values of x for which f(x) = 0, is a hyperplane1 defined by its normal vector w. see figure (2)

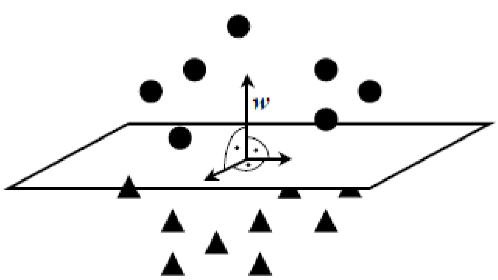

Fig. 2

Assume we have points from two classes that can be separated by a hyperplane and x is the closest data point to the hyperplane, define x0 to be the closest point on the hyperplane to x. This is the closest point to x that satisfies

w · x = 0 (see Figure 3).

We then have the following two equations:

w · x = k for some k, and

w · x = 0.

Subtracting these two equations,

we obtain w · (x − x0) = k.

Dividing by the norm of w (the norm of w is the length of the vector w), we obtain:

$$\frac{w}{\|w\|} \cdot (x - x_0) = \frac{k}{\|w\|}$$

Where $\|w\| = \sqrt{\sum_{i=1}^{n} w_i^2}$.

Noting that w / ||w|| is a unit vector (a vector of length 1), and the vector x − x0 is parallel to w, we conclude that

$$\|x - x_0\| = \frac{|k|}{\|w\|} .$$

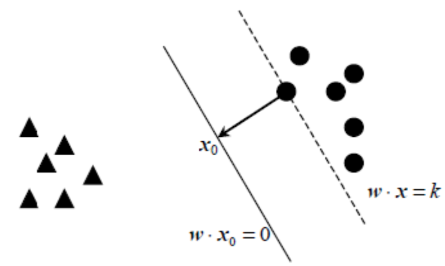

Fig. 3

Our objective is to maximize the distance between the hyperplane and the closest point, with the constraint that the points from the two classes fall on opposite sides of the hyperplane. The following optimization problem satisfies the objective:

$$\max_{w} \min_{x_i} \frac{y_i (w \cdot x_i)}{\|w\|} \quad \text{subject to} \quad y_i (w \cdot x) > 0 \text{ for all } x_i$$

II. IMAGE GATHERING

At first, it needs to decide some class keywords, which represent classes into which unknown images are classified. For example, ``bear'', ``dog'' and ``lion''. For each class keyword, it gathers related images from the Web. To gather images from the Web, it uses the Image Collector and the module is called as an image-gathering module[2,7].

An image-gathering module gathers images from the Web related to the class keywords. Due to the recent explosive progress of WWW , a large number of images can be easily accessed on WWW. There are, however, no established methods to make use of WWW as a large image database[6]. Ther image-gathering module does not need to make a huge index for a great number of images on the whole WWW because of taking advantage of commercial keyword-based text-search engines. It can gather a lot of images related to given keywords full-automatically without a user's intervention during the processing (Fig. 4). The system has been implemented on a scheme of cluster, which enables to gather more than one hundred images from WWW in about one minute.

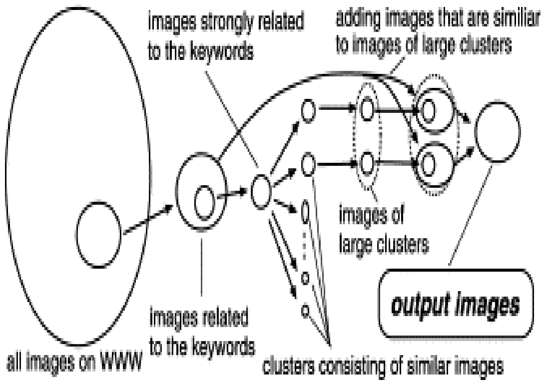

Fig. 4. Processing flow of image-gathering.

The image-gathering process consists of two stages, which are a collection stage and a selection stage. In the collection stage, it gathers many images and HTML documents related to the given keywords using Web search engines. Then it performs evaluation of the relevancy of images by analyzing associated HTML documents. According to the relevancy of images to the given keywords, images are divided into two groups: images in group A are highly relevant to the keywords, and others are classified into group B. The possibility that images in group A are relevant is high, so that we use them as training data of a SVM classifier in the next stage, although they includes a small number of irrelevant ones.

In the selection stage, it selects relevant images from all the downloaded images by employing image analysis. This paper uses the bag-of-key points model as an image representation and an SVM classifier as a classification method. In general, to use machine learning methods like an SVM to select true images, we need labeled training images. However, we do not want to pick up good images by hand. Instead, we regard images classified into group A as training images, although they always include some irrelevant images. It provides a classifier with all group-A images as relevant training images.

In the selection stage, first it convert all the downloaded images into feature vectors based on the bag-of-keypoints representation, and then train an SVM classifier with all the vectors in the group A as training data. Next, it classifies all the vectors in the group A and B as relevant or irrelevant with the trained SVM. Finally, we can get only images classified as relevant to the provided keywords as a result. The detail of this processing is as follows:

1. Sample many image patches from each image

2. Extract patch feature vectors from all the points by SIFT descriptor [2]

3. Generate codebooks with k-means clustering over extracted patch feature vectors

4. Assign all patch feature vectors to the nearest codebooks, and convert a set of patch feature vectors for each image into one histogram vector of assigned codebooks.

5. Train an SVM classifier with all the histogram vectors in the group A as training data.

6. Classify all the histogram vectors of downloaded images as relevant or irrelevant with applying the trained SVM.

The main idea of the bag-of-keypoints model is representing images as collections of independent local patches, and vector-quantizing them as histogram vectors [15].

III. SEARCHING OF SEMANTIC CONCEPTS BY A LEARNING SCHEME

The overview of the scheme for learning Web images to search the semantic concepts in image databases. We illustrate each step of the system as follows. The description is described in Fig. 5.

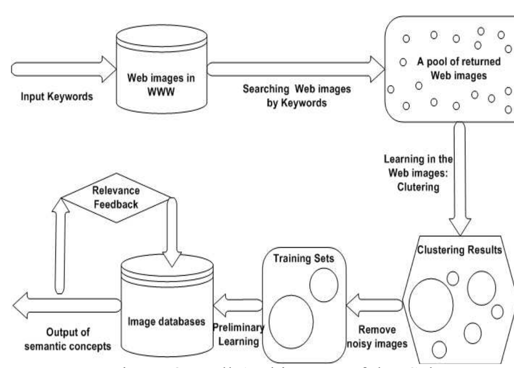

Fig. 5: Overall Architecture of the Scheme

A. A. Searching and clustering Web images

In the proposed system, a user first keys in words to represent their desired semantic concepts. Then, it searches the images on the Web which are associated with the related words. From the Web, it collects a pool of images which have textual descriptions related the semantic concepts. However, the image pool may contain many noisy images which are not relevant. Thus, the clustering techniques are employed to remove the noisy images. The strategy is to cluster the images into 'k' clusters. Then, the top 'p' clusters with the most images will be selected, and other clusters will be regarded as noises.The engaged clustering technique is based on the k-means algorithm.

B. B. Learning semantic concepts by SVMs

After removing the noisy images, we can obtain a set of training images which roughly represent the semantic concepts. Then, we employ the SVM techniques to learn the semantic concepts in the image databases since SVMs provide good generalization performance and can achieve excellent results on pattern classifications problems [5]. In the preliminary searching round, we employ the One-class SVMs (1-SVM) to learn the training set of images in the database. 1- SVM is derived from classical SVMs for solving density estimation problems. After learning by 1-SVMs, we can obtain the preliminary searching results. Then, we employ the relevance feedback with two-class SVMs to improve the retrieval performance. Details for relevance feedback by SVMs can be found in [8].

IV. LEARNING AND CLASSIFICATION

Image classification by Web images is performed by combination of an image gathering system and an image classification system which is depicted in (Fig.6).

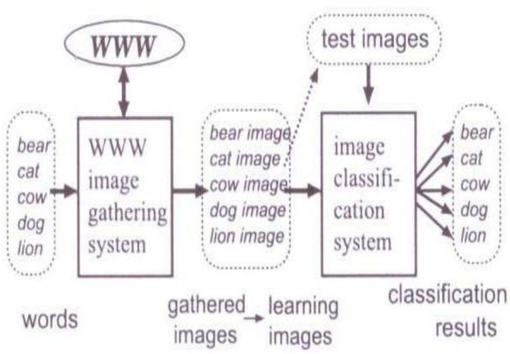

Fig. 6: Image classification by Web images.

First, images are gathered related to some kinds of words from the Web by utilizing the Image Collector. Next, image features are extracted from gathered images and associate image features with words for image classification. Finally, we classify an image into one of classes corresponding to class keywords by comparing its image features with ones of images gathered from the Web in advance. In this paper, we describe image gathering from the Web, learning and classification. In the system, image classification is performed by image-feature-based search. First, in the learning stage, an image-learning module extracts image features from gathered images and associates image features with the classes represented by the class keywords. Next, in the classification stage, we classify an unknown image into one of the classes by comparing image features.

V. RESULTS AND DISCUSSION

In our image database, we collect 10,000 images from the Web by using Image Collector which include semantic categories, such as cat, car , butterfly and sunset, etc. To evaluate the performance of the proposed scheme in a large image database, we choose 10-semantic concepts, including cat, autumn, butterfly, car, elephant, firework, iceberg, sunset, surfing and waterfall. To search Web images, we choose the Google Image Search Engine . For each query semantic concept, top 50 returned image from Google were collected. For the clustering algorithm in our proposed scheme, we choose the parameters k =12 and p=4 in the k-mean algorithm. The kernel function used in SVMs is based on the Radial Basis Function [5]. Fig. 4 shows the experimental results. We observe that the average retrieval precision on Top 20, Top 50, and Top 100 results is over 14%, 8%, and 5%, respectively. The preliminary searching results are further improved by relevance feedbacks using SVMs. In each feedback round, 50 images are presented to users for judging their relevance. Table 1 shows the retrieval performance improved by 3-round relevance feedbacks. We can see that the average precision in Top 20 , Top 50 and Top 100 after 3-round feedbacks can achieve 61 % ,34% and 20% respectively.

Table 1: Average Retrieval Precision by Relevance Feedbacks

Feedback Round	**Top 20**	**Top 50**	**Top 100**
No Feedback	15.5%	9.9%	6.7%
1 Feedback	30.0%	16.2%	16.4%
2 Feedback	49.0%	28.4%	18.1%
3 Feedback	61.5%	34.2%	20.3%

In the experiment, we gathered images from the Web for 10 kinds of class keywords . The total number of gathered image is 10,000, we choose 4582 images on 10-semantic concepts, including cat, autumn, butterfly, car, elephant, firework, iceberg, sunset, surfing and waterfall, and the precision(pri.) by subjective evaluation was 66.2%(table-2), which is defined to be NOK/(NOK+NNG), where NOK, NNG are the number of relevant images and the number of irrelevant images to their keywords.

Table 2: Results of image-gathering and classification in experiments .

	Img.	
class	***Num.***	***Pre.***
cat	419	56.4
autumn	354	62.0
butterfly	575	75.7
car	506	65.5
elephant	275	89.9
firework	504	77.0
iceberg	576	57.0
sunset	347	64.0
surfing	405	68.7
waterfall	595	72.4
Total/avg.	**4590**	**66.2**

VI. CONCLUSION

In this paper, we described design, implementation, and experiments of an automatic image-gathering system from WWW. The only input we have to give to the system is a list of query keywords, and then the system carries out collection of Web images by on-demand crawling over WWW and analyzing HTML files and selection by image-feature-based clustering and picking up larger clusters without a user's intervention during the processing. Here, we integrate the color, texture, and shape as image features, which focuses on a generic image recognition and classification by using gathered images from WWW as training images. In this paper, we propose a scheme to learn Web images for searching semantic concepts in image databases. We suggest to implement the SVMs techniques for learning tasks and classification tasks to obtain more classification rate. For future works, we plan to make much improvement in classification methods and extraction of image features to obtain high precision of fast-image gathering from WWW and more improved classification rate.

REFERENCES

1. K. Yanai. Web image mining toward generic image recognition. In Poster Proc. 12th International World Wide Web Conference, 2003.
2. Keiji Yanai, Web image gathering with region-based bag-of-features and multiple instance learning, Proceedings of the 2009 IEEE international conference on Multimedia and Expo, p.450-453, June 28-July 03, 2009, New York, NY, USA
3. Keiji Yanai, Generic image classification using visual knowledge on the web, Proceedings of the eleventh ACM international conference on Multimedia, November 02-08, 2003, Berkeley, CA, USA
4. Hideki Nakayama , Tatsuya Harada , Yasuo Kuniyoshi, Canonical contextual distance for large-scale image annotation and retrieval, Proceedings of the First ACM workshop on Large-scale multimedia retrieval and mining, October 23-23, 2009, Beijing, China
5. C. Burges. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2(2):121.167, 1998
6. Keiji Yanai , Kobus Barnard, Probabilistic web image gathering, Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval, November 10-11, 2005, Hilton, Singapore
7. K.. Yanai. Image collector II: A system for gathering more than one thousand images from the web for one keyword. In Proc. of IEEE International Conference on Multimedia and Expo, volume I, pages 785--788, 2003.
8. S. Tong and E. Chang. Support vector machine active learning for image retrieval. In Proc. ACM Multimedia, pages 107.118, 2001
9. A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Analysis and Machine Intelligence, 22(12):1349.1380, 2000.
10. K. Barnard and D. A. Forsyth. Learning the semantics of words and pictures. In Proc. of IEEE International Conference on Computer Vision, volume II, pages 408--415, 2001.
11. Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision, 40(2):99--121, 2000.
12. J.R. Smith and S.F. Chang, Visually searching the Web for content. IEEE Multimedia 4 3 (1997), pp. 12–20.
13. S. Sclaroff, M. LaCascia, S. Sethi and L. Taycher, Unifying textual and visual cues for content-based image retrieval on the World Wide Web. Computer Vision and Image Understanding 75 1/2 (1999), pp. 86–98.
14. N. Rowe and B. Frew, Automatic caption localization for photographs on World-Wide Web pages. Information Processing and Management 34 1 (1998), pp. 95–107.
15. M.J. Swain and D.H. Ballard, Color indexing. International Journal of Computer Vision 7 1 (1991), pp. 11–32
16. U. Gargi and R. Kasturi. An evaluation of color histogram based methods in video indexingin: Proceedings of International Workshop on Image Databases and Multimedia Search (1996) pp. 75–82
17. J. Hafner, H.S. Sawhney, W. Equitz, M. Flickner and W. Niblack, Efficient color histogram indexing for quadratic form distance functions. IEEE Transactions on Pattern Analysis and Machine Intelligence 17 7 (1995), pp. 729–736.
18. R.O. Duda, P.E. Hart and D.G. Stork. Pattern Classification (2nd ed.), Wiley (2001).
19. S. Mukherje ,Classifying Microarray Data Using Support Vector Machine: MIT/Whitehead Institute for Genome Research and Center forBiological and Computational Learning at MIT.