CONCEPT-ORIENTED SAMPLE IMAGES SELECTION

D. Djordjevic, A. Dorado, and E. Izquierdo

Electronic Engineering Department
Queen Mary, University of London, UK
E-mail: divna.djordjevic@elec.qmul.ac.uk

W. Pedrycz

Electrical and Computer Engineering Department
University of Alberta, Edmonton, Canada
E-mail: pedrycz@ece.ualberta.ca

ABSTRACT

In semantic-based image classification, learning concepts for adding knowledge to the image descriptions is an issue of special interest. This learning increases the capabilities for more "intelligent" image processing. The classifier learns by generalizing specific facts present in a number of design samples. Due to the fact that the learning and classification processes run over image descriptions containing part of the image content, selection of training patterns should take into account relationships among those descriptions. Proposed framework uses unsupervised clustering to support the selection of design samples and user feedback to refine the classifier model.

1. INTRODUCTION

Learning concepts from features is an ongoing challenge for researchers and practitioners in different communities such as pattern recognition, machine learning and image analysis, among others.

Although the problem of learning concepts has been studied during decades, it is still an open issue. Saitta and Bergadano present an interesting comparative analysis of results from pattern recognition and from theoretical machine learning on this problem [1].

Learning concepts is addressed in the context of semantic-based image classification herein. Concepts are used to add knowledge to the image descriptions linking human interpretations of the image content. Augmented descriptions are useful to perform more "intelligent" processing on large-scale image databases.

Bhanu and Dong present a framework for learning concepts combining partially supervised clustering and relevance feedback [2]. Training strategies have been presented in [3][4]. In contrast, the introduced approach is addressed to assist the selection of design samples without overloading the role of the professional annotator. It exploits the capabilities of support vector classifiers of learning from relatively few examples.

The semantic component places the classifier into the supervised learning scope. Using inductive learning the classifier can learn by generalizing specific facts present in a number of design samples (or training patterns).

Taking into account that the learning and classification processes run over image descriptions containing part of the image content, there is a clear drawback of selecting design samples looking only at randomly selected images.

The proposed framework combines unsupervised clustering and user hints to assist the learning and classification processes in order to refine the classifier model. Conversely, the teaching assistance is given by selecting data-driven (clustering results) and user-driven (relevance feedback) samples. A two-class support vector is used to classify new patterns [5].

Next section introduces the problem of learning concepts and gives some definitions and notations. Section 3 describes the proposed framework. Selected experimental results are presented in Section 4. Concluding remarks appears in Section 5.

2. THE PROBLEM OF LEARNING CONCEPTS

Among the different approaches for learning concepts are concept learning from examples and concept learning by observation. In the former, the machine learns using independent instances representing certain class.

The inductive learning process is carried out presenting declarative knowledge through a number of labeled samples. It arises a problem that can be stated as follows:

How to assist the learning process in the selection of samples for a given concept?

In the case of image classification, normally, those samples are picked from a database. Time, relaxation on selection of "good" samples, and database size are some shortcomings of randomly selection. Accordingly, the framework is proposed.

3. A FRAMEWORK TO ASSIST CONCEPT LEARNING FROM EXAMPLES

3.1. General Overview

A support vector classifier performs the task of using content-based descriptions (feature vectors) to assign certain images to a given concept (semantic class or category).

Concept-wise human understanding is introduced by labeling images as either positive or negative samples of a concept depending on human perception of their content.

An image is considered positive sample of a given concept if it satisfies a criterion defined by a professional annotator. For instance, a picture is a positive sample of a “building image” if it depicts a visible building object.

There are several identified problems in the selection of sample images such as subjectivity of the beholder, quality of the picture (occlusion, shadows, rotation), and quantity of available examples, among others.

Another problem is the selection of samples based on human perception missing the fact that the classifier will work on partial descriptions with limited domain knowledge.

Unsupervised clustering is introduced in order to facilitate the selection of design samples. In addition, incremental domain knowledge is also used to tune the classifier model. The system flowchart is illustrated in Figure 1.

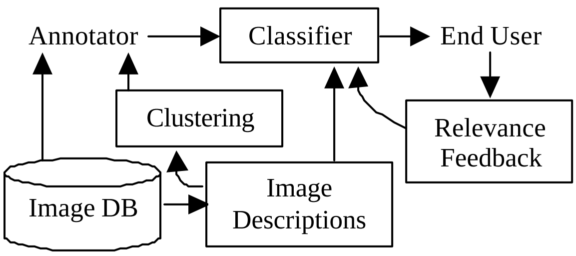

Figure 1. Data Flowchart.

3.2. Support Vector Classifier

There has been an increasing interest in Support Vector Machines (SVM) over the past few years due to their good generalization performances over various pattern recognition problems [6][7]. SVM are often used as classifiers or learning methods for relevance feedback problems in image retrieval systems.

The idea of this supervised leaning approach is not to estimate distributions of the known/unknown patterns but to learn the *support vectors*. These vectors support the optimal nonlinear decision hyperplane and are determined from a known training set.

SVM perform well even with very small training sets [8]. Based on those training patterns and assuming that both training and testing patterns are generated by an unknown probability distribution, it is possible to estimate the separating hyperplane:

$$(\mathbf{w} \cdot \mathbf{x}) + b = 0, \quad w \in \mathbf{R}^p, \quad b \in \mathbf{R} \tag{1}$$

where $\mathbf{x} = (x_1, x_2, \ldots x_p)$ is a p-dimensional *pattern* (or vector, or instance) describing some features of the image content. A feature $x_i, i = 1 \ldots p$, is an individual scalar component of $\mathbf{x}$, which normally corresponds to a measurement of a given visual primitive.

The support vectors are used to classify unknown data samples.This classification is based on an optimal hyperplane, being the solution of the following optimization problem:

$$\arg\min \left(\frac{1}{2} \|\mathbf{w}\|^2 \right) \tag{2}$$

Such that

$$y_i((\mathbf{w} \cdot \mathbf{x}_i) + b) \geq 1, \quad i = 1, \ldots, N \tag{3}$$

where $< \mathbf{x}_i, y_i >$ is a sample involving pairing information. $y_i = \Omega(\mathbf{x}_i) = 1$ if $\mathbf{x}_i$ satisfies the user-defined criterion regarding to a given concept and $y_i = \Omega(\mathbf{x}_i) = 0$ otherwise. $\Omega(\cdot)$ denotes the classifier expressed in the simplest case as a function

$$\Omega : R^p \to \{0,1\} \tag{4}$$

This solution maximizes a margin that is the minimal distance from the closest data samples to the decision surface. The optimization problem (2) and (3) is a quadratic problem often solved by conversion to Wolfe dual [9].

Furthermore SVM are extended for solving classification in nonlinear feature spaces by introducing kernels. The SVM performance is highly dependant on the chosen kernel and the tuning of its parameters to a particular problem in question. The proposed approach uses a Gaussian kernel with a fix kernel width s :

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\|\mathbf{x} - \mathbf{x}'\|^2 \Big/ 2s^2 \right) \tag{5}$$

3.2. Assisting the Learning Process through Unsupervised Clustering

As defined in [10], *cluster analysis* is the organization of a collection of patterns into clusters based on similarity. Such a similarity between patterns is quantified or measured using a proximity metric (e.g. Euclidean, Mahalanobis). Clustering techniques are used to uncover group of patterns, if there is a cluster tendency. Otherwise, the clusters do not reflect any meaningful grouping.

Perceptually separable visual primitives are clustered using the standard Fuzzy C-Means algorithm [11], which is based on minimization of the criterion function:

$$J_m = \sum_{k=1}^{N}\sum_{i=1}^{C} u_{ik}^{m} \left\| \mathbf{x}_k - \mathbf{v}_i \right\|^2, \quad 1 < m < \infty \qquad (6)$$

where u_{ik} is the degree of membership of $\mathbf{x}_k$ in the cluster i, $\mathbf{v}_i$ is the p-dimension prototype (center) of the cluster, and $\|\cdot\|$ is any norm expressing the proximity between a given pattern and the corresponding cluster prototype.

The nearest patterns to the cluster prototypes are used as candidates of sample instances (Figure 2). The source images for those patterns are presented to a professional annotator who will identify positive and negative examples of the concept.

Sample selection based on clustering seeks to exploit any underneath structure in the feature space and minimizes the drawbacks of manual (or randomly) selection of sample images.

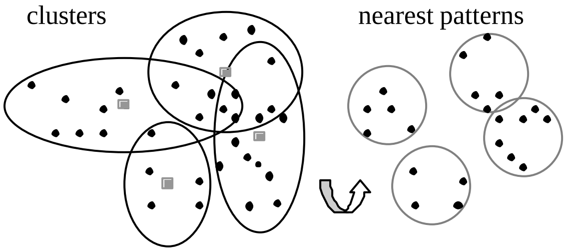

Figure 2. Selecting Candidates from the Nearest Patterns to the Cluster Prototypes.

3.3. Assisting the Classification Process through Relevance Feedback

User hints are captured during the learning and classification phases. During training, a professional annotator provides labels during the selection of positive and negative samples. Afterwards, the classifier model can be affined accumulating domain knowledge through relevance feedback.

As is illustrated in Figure 3, the proposed framework integrates relevance feedback to increase domain knowledge and reinforce the boundaries between patterns containing (or not) the concept. These boundaries are defined by a hyperplane based on the support vectors.

4. EXPERIMENTAL RESULTS

Experiments were conveyed with imagery selected from Corel stock gallery (corel.com). More than 1000 pictures were collected from the Corel categories containing indoor images from collections as office, interior, bathroom-kitchen and outdoor images from building, rome, ny_city, etc.

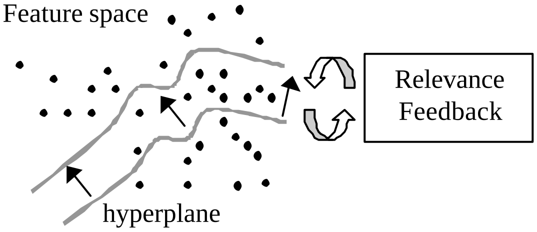

Figure 3. Refining the hyperplane.

The feature space consists of vectors containing color layout descriptions (cf. [12]).The MPEG-7 color layout is a histogram-based descriptor, and it was set up to 58 coefficients (28 luminance and 15 chrominance coefficients, each). The matching procedures in our test use the L2 norm.

While aware that based only on low lever color content hints we would not be able to infer semantics, we have chosen only one feature to show the possible advantage of our approach not in correlation with feature combination.

The test conditions are based on using over 435 indoor and 600 outdoor images. With equal number of positive and negative samples chosen for most tests except when relevance feedback is based on random display of images. Some samples of positive and negative images used to obtain the support vectors are depicted in Figures 4 and 5.

Figure 4. Samples of indoor images.

Figure 5. Samples of outdoor images.

Figure 6 presents accuracies achieved by the indoor/outdoor classification problem. There is a close trend between approaches (1) and (2). Whereas the amount of supervision in (1) requires knowledge of the database in advance, in (2) the supervision is governed by the number of images shown to the user. Also a shortcoming is the total subjectivity due to the selection of sample relies completely on the images ignoring any relationship (low-level similarity) among the image descriptions.

Accuracy in the approach (3) decreases notoriously though this is expected because of the sensible reduction on the required supervision. The professional annotator needs only to provide hints to indicate the class label of each cluster. This lightens the burden of annotation while as introducing some noise.

Approach (4), which corresponds to the proposed one, reported the best performance with the aforementioned advantages of taking into account the underlying structures (clusters), minimizing the required supervision, and reducing the randomness of the results.

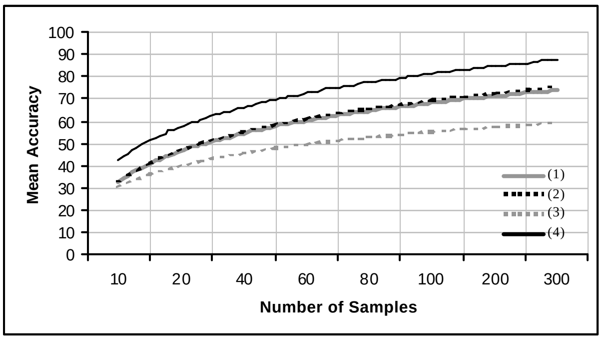

Figure 6. Comparative Results.

(1) Support vector (SVM) classifier using randomly selection of sample images from the whole feature space.
(2) SVM using only with relevance feedback (RF). The classifier is not trained with the assistance of a professional annotator.
(3) SVM assisted by hints provided by a professional annotator and clustering results during the learning phase. Samples are selected from the nearest patterns (see Figure 2 and 7).
(4) SVM assisted by annotator's hints, clustering results and relevance feedback.

Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5

Figure 7. Samples of nearest patterns to prototypes.

5. CONCLUSIONS

A framework to assist learning concepts from image descriptions in the context of semantic image classification is presented. The approach seeks to overcome the drawback of randomly selection of design samples. It combines data-driven and user-driven information through clustering and relevance feedback, respectively. Such a combination acts likewise a negotiation between what is observed by the user (images) and by the machine (image descriptions).

6. ACKNOWLEDMENTS

Support from the Natural Sciences and Engineering Research Council (NSERC) and Canada Research Chair (W. Pedrycz) is gratefully acknowledged.

Part of the research leading to this paper was also done within the framework of the project aceMedia – Integrating knowledge, semantics and content for user-centred intelligent media services (www.acemedia.org).

7. REFERENCES

[1] L. Saitta and F. Bergadano, "Pattern recognition and Valiant's learning framework," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 2, pp. 145-155, Feb. 1993.

[2] B. Bhanu and A. Dong, "Concepts learning with fuzzy clustering and relevance feedback," *Engineering Applications of Artificial Intelligence*, Vol. 15, pp. 123-138, 2002.

[3] M.R. Boutell, J. Luo, X. Shen, and C. Brown, "Learning multi-label scene classification," *Pattern Recognition*, Vol. 37, pp. 1757-1771, 2004.

[4] A. Dorado, D. Djordjevic, E. Izquierdo, and W. Pedrycz, "Supervised semantic scene classification based on low-level clustering and relevance feedback," *in Proc. of the European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology*, EWIMT, pp. 181-188, 2004.

[5] A.K. Jain, P.W. Duin, and J. Mao, "Statistical Pattern Recognition: A Review," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 1, pp. 4-37, Jan. 2000.

[6] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Neural Networks*, Vol. 12, No. 2, pp. 181-201, May 2001.

[7] P.-H. Chen, C.-J. Lin, and B. Schölkopf, "A tutorial on nu-support vector machines," May 2003. To appear in *Applied Stochastic Models in Business and Industry*, 2005.

[8] R. Duin, "Classifiers in almost empty spaces," in *Proc. of 15th Int'l Conf. on Pattern Recognition*, Vol. 2, pp. 1-7, Barcelona, Spain, 2000.

[9] S.S. Keerthi, S.K. Shevade, C. Bhattacharyya, and K.R.K. Murthy, "Improvements to Platt's SMO algorithm for SVM classifier design," *Technical report, Dept of CSA*, IISc, Bangalore, India, 1999.

[10] A.K. Jain, M.N. Murty, and P.J. Flynn, "Data Clustering: A Review," *ACM Computing Surveys*, Vol. 31, No. 3, pp. 264-323, Sep. 1999.

[11] J. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*, Plenum Press, New York, NY, 1981.

[12] B.S. Manjunath, J.-R. Ohm, V.V. Vasudevan, and A. Yamada, "Color and Texture Descriptors," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 11, No. 6, pp. 703-715, Jun. 2001.