Learning Image to Symbol Conversion*

Malini K. Bhandaru **Bruce A. Draper** **Victor R. Lesser**

Dept of Computer Science
Univ. of Mass at Amherst
Amherst, MA 01003

Abstract

A common paradigm in object recognition is to extract symbolic and/or numeric features from an image as a preprocessing step for classification. The machine learning and pattern recognition communities have produced many techniques for classifying instances given such features. In contrast, learning to extract a distinguishing set of features that will lead to unambiguous instance classification has received comparatively little attention. We propose a learning paradigm that integrates feature extraction and classifier induction, exploiting their close interrelationship to give improved classification performance.

Introduction

Object recognition systems can conceptually be divided into two phases: feature extraction and recognition. In the feature extraction phase, feature vectors are extracted for each instance. In general, the features are hand selected, as are their parameters, for example the cut-off frequencies of a bandpass filter or the window size of a convolution operator. In the recognition phase, the extracted feature vectors are classified by comparing them to implicit or explicit object representations.

In other words, the conventional approach to image to symbol conversion is to use a fixed set of *hand selected* features, where a feature is a specific feature extraction procedure (FEP) with preset parameter values (see Figure 1). The learning algorithm induces a classifier by partitioning the feature space into distinct regions corresponding to different classes of objects. Classification success is directly dependent on whether the selected features are sufficient to distinguish among the classes. To increase the likelihood of including a distinguishing set of features, some systems use a large set of features. The drawbacks of such an approach are: a need for greater computational resources (to compute the additional features), redundant, possibly confusing features for the learning element, and a need for more training examples to cover the expanded feature space. Other classification systems restrict their domains of applicability to tasks for which the hand-selected features are known to be sufficient.

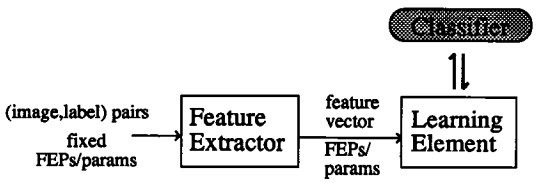

Figure 1: Conventional approach to learning image to symbol conversion.

For complex scenarios that are characterized by varying image to noise ratio, unpredictable activity and possible object occlusion, the optimal features are neither easily identifiable nor stationary. Adaptive feature extraction (adaptive front-end) capabilities are necessary to cope with such perceptual tasks. The Schema Learning System [Draper93], Goldie [Kohl87], and TRIPLES [Ming90] concentrate on selecting a subset from a large but finite set of features to efficiently disambiguate classes. Flexible software architectures [Bhanu90, Lesser93] have been developed for adapting FEP and parameters to the environment being monitored. The above systems rely, however, on either implicit or explicit hand-crafted object descriptions. The issue of automating the acquisition of object descriptions for such adaptive front-end systems has not been addressed.

The space of potential features is infinite in adaptive front-end systems because the parameters of feature extraction procedures may take on an infinite number

*This work was supported by the Rome Air Development Center of the Air Force Systems Command under contract F30602-91-C-0037 and F30602-91-C-0038, by the Office of Naval Research under contract N00014-92-J-1450 and by the Advanced Research Projects Agency (via Tacom) under contract DAAE07-91-C-R035. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

of values. Each parameter instantiation gives rise to a distinct feature. The features are considered distinct in that when different instantiation of a FEP are applied to an image, the feature vector or "signatures" obtained may be different. For example, consider an edge detection algorithm that first smoothes the image before looking for discontinuities. When such a FEP is applied to an image with two nearby parallel discontinuities, it will yield zero, one or two edges depending on the amount of smoothing. However, the signatures are consistent in that the loss of detail inherent in smoothing can be predicted given knowledge about the FEP and the image.

The edge extraction example serves to illustrate that features are a function of feature extraction procedures and their associated parameter values, that is, instantiated feature extraction procedures (IFEPs). In the absence of domain knowledge, the signatures obtained from two distinct parameterizations of a single FEP are not directly comparable. Additionally, the degree of detail that must be extracted cannot be predetermined, but is a function of the recognition task at hand. For example, the more similar the object classes, greater the detail that may be necessary to disambiguate among them.

Since feature extraction, object representation, and classification are so closely tied in perceptual domains, we propose a learning paradigm that integrates them. Learning image to symbol conversion is envisaged as a search process in the infinite space of features towards selecting a subset that performs to satisfaction the given classification task.

We present details of the proposed learning paradigm, and discuss issues such as object representations, selection of features, termination criteria, and reprocessing effort. Finally we present some initial ideas for testing the paradigm and our conclusions.

Integrated Feature Extraction and Classifier Induction

The proposed learning paradigm integrates feature extraction, object representation and classifier induction, exploiting their close interrelationship in an attempt to improve classification performance; see Figure 2. The underlying principle is to iteratively extract additional/alternate features until the desired classification goal is attained. Note the feedback from the **Learning Element** to the **Feature Extractor** via the **FEP/param Adaptor**.

The input to the system consists of image/label pairs. The **Feature Extractor** applies the initial feature extraction procedures(FEP) with selected parameter values, giving a feature vector (of one or more tuples). This forms the input to the **Learning Element**. During successive iterations, in addition to the new features extracted, previously extracted but not discarded features, are provided to the **Learning Element**.

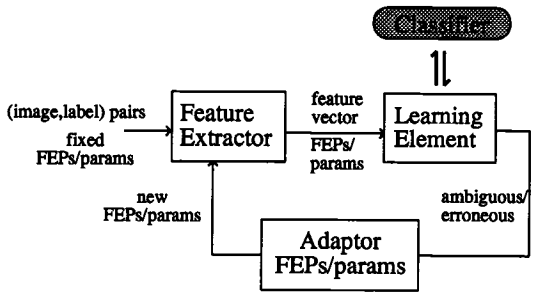

Figure 2: Integrated Feature Extraction and Classifier Induction

The **Learning Element** can be any type of classifier, including an artificial neural network, decision tree, or an instance based classifier. Its goal is to induce a classifier using the object representations provided by the feature extractor. Training terminates when the classification goal is met. For instance, classification goals such as minimizing misclassification cost, or maximizing classification accuracy may be used. If the classification goal is not met, the **Learning Element** declares a failure and passes control to the **FEP/param Adaptor** along with information about which instances and/or classes of objects were poorly classified, and the feature extraction procedures and parameters used.

The goal of the **FEP/param Adaptor** is to suggest alternate IFEPs to achieve the classification goal. It does so based on the information provided by the **Learning Element**, either recommending alternate parameters for the current FEP or a new FEP (with parameters). In the absence of domain knowledge, a gradient approach may be used. Gradient approaches vary slightly one or more of the independent variables in directions which cause the value of the dependent variable to move in the desired direction. The gradient method may be generalized to include features. For instance, consider a goal of maximizing classification accuracy. If a given feature provides improved performance, a related feature would be a candidate for further exploration. Related could be in terms of the physical quantity being measured or the method of manipulation. One or both could be varied slightly. Consider a two dimensional color image. Let us assume that the intensity in a particular spectral band has significantly contributed to classification success. A gradient heuristic is to explore the spectral content of adjoining bands. This corresponds to measuring a related physical quantity. Alternately, texture information could be extracted from the image focusing on the promising band. This would correspond to varying the manipulation technique. On the other hand, if there is no perceived improvement in meeting the classification goal, either an unrelated or random feature could be explored, which might dislodge the system from a possible local minima.

Yet another approach would be to use Case Based Reasoning techniques to aid in the search for appropriate features (IFEPs). As always, if domain knowledge is available, it is used in recommending alternate features, i.e., IFEPs.

Discussion

Object Representations

In the context of adaptive feature extraction systems, the features extracted are a function of the IFEPs used. As a result, three different approaches to object representation are possible: a canonical representation, i.e., a representation that encompasses all object signatures in a comprehensive manner, uniform feature vectors (features and their number not prefixed, but uniform), and finally variable feature vectors (features and their number variable).

To construct canonical representations domain knowledge is necessary. It involves combining object signatures obtained using different IFEPS. Such a representation would be more comprehensive, with the advantage of transparency to interpretation and prediction tasks. Such a representation would lend itself to conventional matching approaches.

A uniform feature vector representation would have the advantage of ease of matching without significant domain knowledge requirements. On the other hand it is computationally inefficient because it requires that the union of all the features, deemed relevant during training, be computed and stored for each instance. This is regardless of the pertinence of the features to individual object classes. For efficiency and conciseness, a variable feature vector representation would be more attractive. Object matching using conventional distance metrics though becomes infeasible. Tentatively, we are exploring a distance metric that takes into consideration only shared features. When used in the recognition phase such a metric would direct further feature extraction, in order to disambiguate possible classification alternatives.

The choice of object representation is also a function of the type of classifier used. For instance, neural net classifiers require fixed length feature vectors, especially during training. Variable length feature vectors can be employed by both decision tree and instance based classifiers. During the object recognition phase, in the context of decision tree and instance based classifiers, feature extraction would be guided by the tests at the nodes and the distance metric respectively. For recognition tasks that are complicated by object occlusion and unpredictable activity, canonical models are more appropriate. Our intuition is that the choice of classifier and object representation may be application dependent.

Features

The emphasis of the learning paradigm is to identify a satisficing set of features with respect to the recognition task and classification goal. Two key questions are how large should the feature set be, and what criteria should be used to include or discard a feature:

Feature set size: Larger feature sets carry with them the cost of increased computational requirements. This can be significant when neural net classifiers are used. When a decision tree classifier is used irrelevant features tend to carry lower weight or do not get tested at all [Brodley93]. Large feature sets also require larger training sets, which may be an issue if training data is difficult to obtain.

Discarding a feature: Features that are irrelevant can be identified by considering the improvement in classification accuracy as a result of including them consecutively. Consider a feature set [f] and the inclusion of a new feature f'. If the new feature set yields little or no improvement, it implies that either f' is subsumed by one or more features in [f] or that f' is irrelevant to the recognition task. Yet another possibility is to discard one or more "unused" features from a set of features by examining the induced classifier. This translates to identifying features that either carry low weights in the case of artificial neural network classifiers. With respect to decision tree classifiers, these could be features tested for only a small proportion of training instances, or carrying a low weight in a test.

Termination Criteria

The quality of a classifier is defined with respect to the classification goal. Several criteria are possible, such as reducing misclassification cost, maximizing classification accuracy, minimizing classification cost with respect to a fixed minimum classification accuracy and combinations of these. When no domain specific termination criterion is given, training may be terminated when it becomes apparent that a local optima has been reached, or a certain number of features have been tested or computational limits such as time or space exceeded.

Experiments

Within this paradigm, we are first looking at simplified problems and domains, before addressing the problem in its full generality. As a first step, we are considering the problem of pixel classification in a remote sensing domain. Features can be as simple as the intensity of a spectral channel or more complex, such as a spatial property governed by mask parameters (window size, measure etc). Although any classification technique can be used within this framework, we are using a Linear Machine Decision Tree (LMDT) [Brodley93]. We use a the thermal training rule which allows for multiple object classes. We shall use a gradient descent approach in our search of the feature space.

We are testing the paradigm on the Wien landsat data [Wein]. The Wien data consists of 512×512 sized images in 7 spectral channels. Also provided is a 512×512 ground truth table. Each pixel belongs to

one of four classes: water, agriculture, built-up-land or forest. Our goal is to automatically suggest and test features to give improved performance on identifying pixel class.

Conclusions

A novel aspect of our approach is that it recognizes the close interrelationship between feature extraction and object representation. By learning the classifier and the feature extraction procedure together, reliable and efficient classifiers can be induced, as opposed to systems that are limited to the use of fixed feature extraction procedures that might not necessarily be good.

Acknowledgements

We thank Carla Brodley for providing us with the LMDT code and for her patience. We also thank Axel Pinz for providing us the Landsat data of Vienna.

References

[Brodley93] C. Brodley, P. Utgoff, "Multivriate Decision Trees", *Machine Learning*, to appear.

[Bhanu90] B. Bhanu, S. Lee, J. Ming, "Self-Optimising Control System for Adaptive Image Segmentation", *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA., Sept. 1990. pp. 583-595.

[Draper93] B. Draper, "Learning Object Recognition Strategies", PhD Thesis, Dept. of Computer Science, University of Massachusetts, CMPSCI TR93-50, 1993.

[Kohl87] C. Kohl, A. Hanson, E. Riseman, "A Goal-Directed Intermediate Level Executive for Image Interpretation", *Proc. 1987 of the Joint International Conference on Artificial Intelligence.*

[Lesser93] V. Lesser, H. Nawab, I. Gallastegi, F. Klassner, "IPUS: An Architecture for Integrated Signal Processing and Signal Interpretation in Complex Environments", *AAAI-93*, Washington, USA, 1993.

[Ming90] J. Ming, B. Bhanu, "A Multistrategy Learning Approach for Target Model Recognition, Acquisition, and Refinement", *Proc. of the DARPA Image Understanding Workshop*, Pittsburgh, PA., Sept. 1990. pp. 742-756.

[Wein] H. Bischof, W. Schneider, and A.J. Pinz, "Multispectral Classification of Landsat-Images Using Neural Networks", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 30, No. 3, May 1992.