Conditional Random Fields for Shape Recognition in Images

A project report submitted

By

Akash Chand (axc173730), Deeksha Lakshmeesh Mestha(dxm172630),

Dheeptha Badrinarayanan(dxb171030)

For the course

CS 6347.001- Statistical Methods in AI and Machine Learning

Under the supervision of

Dr. Nicholas Ruozzi

At

The University of Texas at Dallas

**Abstract**

CRFs have been emerging as a powerful tool in the domain of computer vision owing to their modularity in designing complex relationships. This project deals with solving a high-level computer vision task, namely, shape recognition. We present an approach that utilizes a discriminative, undirected graphical model called Conditional Random Field. The probability of a class label for each object is computed given the assignments to the local features. The parameters of the model are computed by employing maximum likelihood estimate and the most likely class label under the model is identified as the shape. The major advantage of the approach lies in the fact that CRFs assume conditional independence of input features given the class labels.

# 1 Introduction

Humans identify a multitude of shapes even when the shapes in the images vary in size and skewed at some angle. This task still has a lot of hurdles to overcome when it comes to computer vision systems. Many approaches have been attempted in the past to tackle this problem. This shape detection project aims to correctly detect the shapes present in a given image dataset by using machine learning approach. We have used Conditional Random Field model for the above-discussed task.

The most widely used approach for shape/object detection is generative approach. This approach involves modelling of dependencies between the observed data. But there is a limitation here. To make the model computationally tractable we may need to assume the independence of the observed data. To overcome this limitation, we are using a discriminative model, Conditional Random Fields(CRFs). CRFs also involve flexibility in terms of representation as it allows to incorporate feature-vector representation.

# 2 Dataset

The dataset considered for this project can be obtained from:

<http://www.iro.umontreal.ca/~lisa/twiki/bin/view.cgi/Public/BabyAIShapesDatasets>

The above-mentioned address lets you download python files through which an image dataset of 5000 images can be generated. The python file which when executed generates an .amat file which contains images of squares, circles, and triangles with random colors, size, position, and orientation.

The .amat file is an ASCII format and is organized as follows:

* The first line is the number of examples and the number of values per line (1031).
* On each subsequent line, the first 1024 values represent the gray tone of each pixel (or the color, depending on how you interpret it) as a floating-point value between 0 and 1, inclusively (32 lines of 32 pixels). The next 7 values are:
* The shape: 0=square, 1=circle, and 2=triangle
* The color of the shape: this is actually an integer between 0 and 7. It must be divided by 7 to get the corresponding gray tone.
* The x coordinate of the centroid of the shape, between 0 (leftmost) and 256 (rightmost).
* The y coordinate of the centroid of the shape, between 0 (top) and 256 (bottom).
* The rotation angle of the shape, between 0 (no rotation) and 256 (full circle). This can probably not be learnt reliably because there the reference point is ambiguous (for instance, there is currently no way to know relatively to which side the triangle was rotated).
* The size of the shape, between 0 (a point) and 256 (the whole area). There is a lower bound and an upper bound.
* The elongation of the shape, between 0 (at least twice as wide as tall) and 256 (at least twice as tall as wide).

Among these details we are using the pixels as features and shape values as the class labels.

# 3 The Model

The main objective of the model is to learn a set of shape labels from the given images. The possible shape labels are {square, circle, triangle}. The images are of size 32x32. We divide each image into 16 blocks of size 8x8 and only the blocks containing the border shape are used to model the CRF. This is done since the borders of each shape is significantly different, for example, circle has a curved border while the border of a square is effectively a straight line.

Each pixel, Xpi forms a local clique and we compute potentials over each clique. Xs also forms a clique with every Xpi. After applying appropriate pre-processing, the random variables for pixels, Xpi take the values of 0 or 1 according to whether or not it forms a part of the shape. The random variable for the class label, Xs takes the values of 0, 1, or 2 depending on the shape it signifies.

Xpi = {0,1}

Xs= {0,1,2}

The potential functions are represented by feature vectors and ), where

and =

= 0, else

) = 1, if =

= 0, else.

Given these definitions for image pixels Xpi and shape labels Xs, the conditional probabilistic model is given by

The potentials are the exponential functions of weighted feature functions, where s are the parameters that have to be learnt.

The final CRF model turns out to be a simple tree structure as shown in figure 1 below:

…..

…

Fig 1. CRF model for shape recognition

## 3.1 Parameter estimation

We perform maximum likelihood estimation to obtain the parameters. Given an unseen test image, we use the parameter values ’ learnt from the training samples to find the most likely shape label using .

The log-likelihood function L( is given by,

L(

We use L( as the objective function in training the parameters. The parameters arg max L( are computed using the training samples by adopting gradient ascent method. The derivatives are as follows:

=

=

It follows that these derivatives yield a local optimum value for that parameters, which can be used to obtain the most likely shape label. Inference sub-routine is trivial since the resulting factor graph of the CRF model is a simple tree structure. Inference is performed by directly plugging in the optimal parameter values in the conditional probability equation.

# 4 Experiments

We ran our experiments on a dataset of 5000 images. The experiment involved training a three-class model to distinguish between the three different shapes (square, circle and triangle). The parameters were learned using Maximum Likelihood Estimation. For the sake of our experiments, we split the dataset into separate training and test data. 80% of the samples were utilized as training instances and the remaining 20% was utilized for testing. The training samples have equal distribution of all the three class labels so that there exists no bias towards the most frequent labels.

Our dataset has 32x32 size images. To improve the performance, we have divided the images into 16 8x8 size blocks. Among these 16 blocks we have considered only the blocks that have the shape borders.

## 4.1 Results

Figure 2 shows a sample image from our dataset. We ran Naïve-Bayesian classifier on our dataset to obtain a baseline accuracy for this task of shape recognition. Naïve-Bayes yielded an accuracy of 33%.

The CRF model performed slightly better than naïve-Bayesian classifier because of better modelling of dependencies between class labels and pixels. It produced an accuracy of 38%. It gives an accuracy of 91% when predicting square shapes, while the other two shapes have a much lower accuracy. This might probably be due to the features favoring the learning of squares. The difference between the CRF model and the naïve-Bayesian classifier is that our model learns local features since the potentials span over local pixel features.

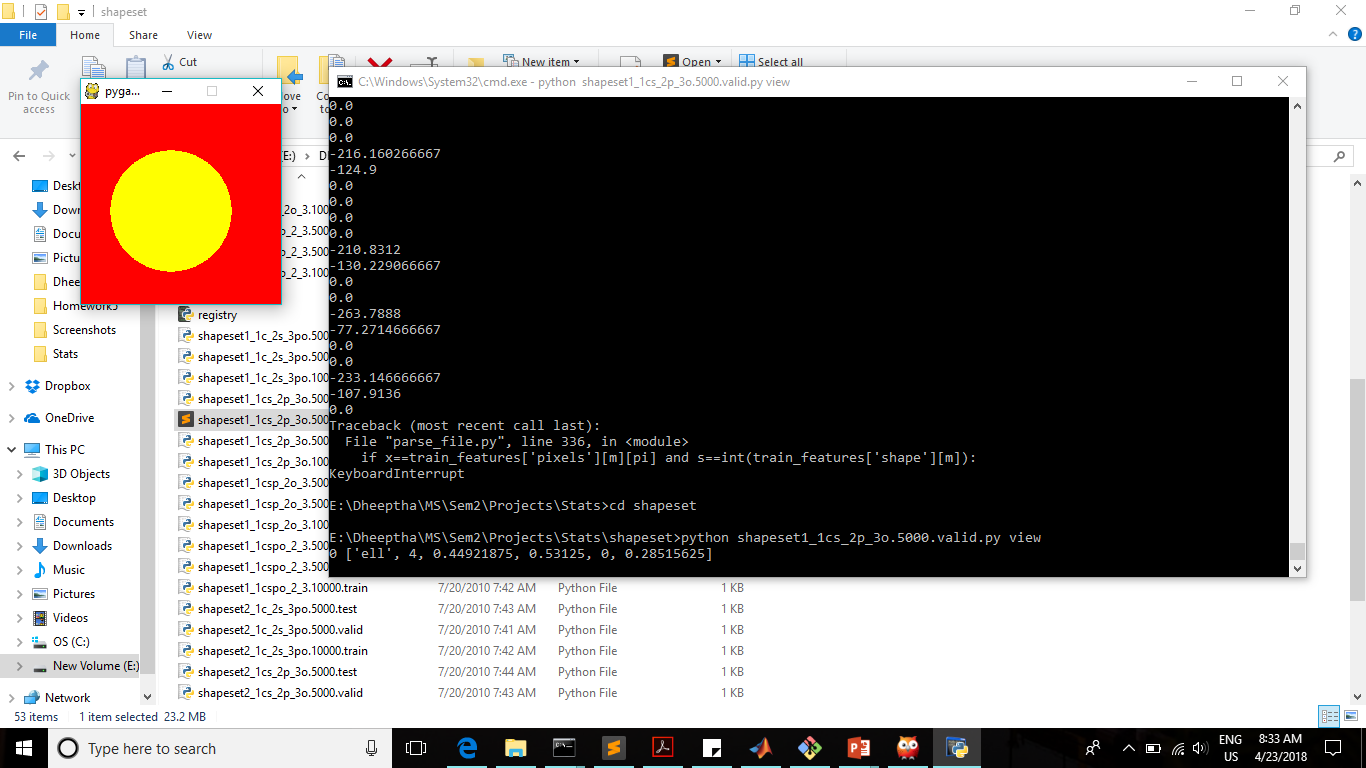


Fig 2. Sample image from the dataset

# 5 Conclusion and Future Work

This project implements a conditional random field to perform the computer vision task of shape recognition. CRFs allows us to model arbitrary feature functions to train the discriminative classifier. The parameters have been computed in the maximum likelihood estimate framework. The most likely class label is identified as the correct shape. To improve the performance of our model, we plan to improve our feature functions as the major shortcoming of our model is that its performance depends on feature selection. So, as future work, we would like to integrate a better feature detection algorithm in our model.

## References

[1] Quattoni, A., Collins, M. and Darrell, T. (2018). Conditional Random Fields for Object Recognition.

[2] S. Kumar and M. Hebert. Discriminative random fields: A framework for contextual interaction

in classification. In *IEEE Int Conference on Computer Vision*,volume 2, pages 1150-1157, June

2003.

[3] Schnitzpan, P. (2010). Conditional Random Fields for Detection of Visual Object Classes A dissertation. Technische Universitat Darmstadt.