P. Perner,H. Perner, and B. Müller, Texture Classification based on Random Sets and its Application to Hep-2 Cells, In: R. Kasturi, D. Laurendeau, and C. Suen (Eds.), ICPR 2002, IEEE Computer Society 2002, vol. II, p. 406-411

Texture Classification based on the Boolean Model and its Application to Hep-2 Cells

Petra Perner and Horst Perner
Institute of Computer Vision and Applied Computer Sciences, August-Bebel-Str. 16-20, 04275 Leipzig
e-mail: ibaiperner@aol.com http://www.ibai-research.de

Abstract *We investigated the Boolean model for the classification of textures. We were interested in three issues: 1. What are the best features for classification? and 2. How does the number of Boolean models which were created from the original image influence the accuracy of the classifier?, and 3. Is decision tree induction the right method for classification?*

1. Introduction

The Boolean model and its application to the characterization of materials, biological objects and geological structure has been described by Serra [1], Stoyan et. al [2], and Lang et. al [3]. It has been studied for texture classification by Garcia et. al [4] under the assumption that the primary grains are lines of random length and orientation. There it has been examined based on the Brodatz texture and by using a similarity-based classifier.

We are working on a real-world application which is the classification of HEp-2 cells. This kind of cells are used in medicine for the identification of antinuclear autoantibodies (ANA). Human experts describe the characteristics of these cells by symbolical texture features. We apply the Boolean model to this problem and assume that the primary grains are regions of random size and shape. We use decision tree induction in order to learn the relevant classification knowledge and the structure of the classifier. We did an advanced study on the behavior of the Boolean model for texture classification.

In Section 2 we are describing the methods and the problems we wanted to study. The material used for our study is described in Section 3. In Section 4 we describe our experiments. Results are given in Section 5. Our observations of the behavior of the model are described in Section 6. Finally, we give conclusions in Section 7.

2. Method

Let us first consider the basic properties of a Boolean model[2] and how it can be used for texture classification.

The simplest possible example of a Boolean model can be described as follows: Suppose points are scattered in the plane according to a stationary Poisson process of intensity λ. On each of these points a disc of a fixed radius r is placed. The union of all these discs is an example of a Boolean model. The points of the Poisson process are called the germs of the model while the discs are the primary grains. The Boolean model Ξ is constructed by using the germs x_n and the primary grains Ξ_n

$$\Xi = \bigcup_{n=1}^{\infty}(\Xi_n + x_n) = (\Xi_1 + x_1) \cup (\Xi_2 + x_2) \cup ...$$

Typical possibilities for the primary grains include: Discs of random radius, segments of random length and orientation, and random finite clusters of points. The primary grains are characterized by their distribution M. As they are random sets, this distribution is a probability measure on K, where K are all nonempty compact random sets of IR^d; it is the mark distribution of the marked point process $\{[x_n; \Xi_n]\}$.

When the primary grains are convex, then statistical averages of various numerical measures of convex sets are important, such as e.g. mean area and mean perimeter for planar sets. In the case of non-convex grains several of the quantities still have a meaning.

When λ is small relative to the size of the grains, then primary grains will not often overlap and so Ξ will consist in the main of separated particles. With increasing λ the number of overlappings increases. This might give a good descriptor for irregular patterns observed in nature.

So the random set in question can be relatively well described by means of the intensity λ of the germ process and the various mean measures of the primary grains.

In a natural generalization of the Boolean model the Poisson point process of germs is replaced by a general point process. This leads to the class of germ-grain models. Random models that are essentially Boolean models, but use general point processes to provide the germs, are called germ-grain models.

Now, let us consider the cells to a be a random closed set Ξ_v with $x(u,w,z)$ where (u,w) are the coordinates and z

P. Perner,H. Perner, and B. Müller, Texture Classification based on Random Sets and its Application to Hep-2 Cells, In: R. Kasturi, D. Laurendeau, and C. Suen (Eds.), ICPR 2002, IEEE Computer Society 2002, vol. II, p. 406-411

is the gray level. The random closed set Ξ_v under study is intersected by two parallel planes separated by a distance t . The slice that is to be examined is the set

$$T^t = \{x(u,w,z): 0 \le z \le t\}$$

lying between these two planes. The experimental data is the image Ξ^t of the projection of $\Xi_v \cap T^t$ onto the *(u,w)*-plane. The projected thick section Ξ^t inherits the properties of stationarity and isotropy from Ξ_v. In stereology [3] this thick section is used in order to estimate the volumetric characteristics of an object from slices. In this work the projected thick section is used for the classification of the texture images. Since we do not know whether one such thick section is enough to classify different textures, we construct several thick sections from one image. Therefore, the gray level is partitioned into several gray level intervals *n*. Each pixel of the original image, whose gray level corresponds to the considered gray level interval, is mapped onto the slice image *i* and labeled by "1". As a result we obtain *n* binary images from the original image. Each of them is a Boolean model.

There are several open questions for the description of the classification knowledge: 1. How many of such slice images are necessary to characterize the different textures?, 2. What are the right features for the description of the primary grains and how should their distribution be described?, and 3. Is decision tree induction the right method for classification?

3. Materials and Image Processing

For our experiment we used a data base of 600 HEp-2 cell images with 6 classes (see table 1 for the 6 different cell patterns) taken under real clinical conditions.

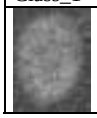
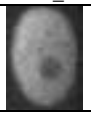
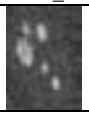
Class_1	Class_2	Class_3	Class_4	Class_5	Class_6
					

Table 1. Image of Cells from 6 different Classes

The color image was transformed into a gray level image without loss of information. Histogram equalization was done to eliminate the influence of the different staining. Automatic thresholding was performed by the algorithm of Otsu [5]. The algorithm can localize the cells with their cytoplasmatic structure very well, but not the nuclear envelope itself. We then applied morphological filters like dilation and erosion to the image in order to get a binary mask for cutting out the cells from the image. Overlapping cells were not considered for further analysis. They were eliminated based on a simple heuristic. Each object having an area bigger than twice the mean area was removed from the image. For each cell in the image the area A_{cell} and the features described in the next section were calculated. Note, the image *f(x,y)* considered for further calculation contains now only one cell. This image was now separated into *n* binary images according to the gray level. The gray level $g=\{0,...,255\}$ was partitioned into *n* intervals by dividing the maximal gray level number by *n* (*256/n*). We used for $n=\{2,4,8,12,16\}$.

The remaining objects in each slice image were regions of different size and shape which were characterized by proper features.

4. Features

From each slice image *i* the following features were calculated: The first one is the marking probability of the slice image *i* which is the number of pixels in the slice image labeled by "1" divided by the area of the cell. When all pixels in an image are labeled by "1", the marking probability is one. It is zero when no pixel in a slice image is labeled by "1". From the objects in the slice image *i* the area and the perimeter are calculated. According to the model, not a single feature of each object is taken for classification, but the mean and the standard deviation of each feature is calculated over all the objects in the image *i*. Since we do not know the real distribution of the object features, we introduced the curtosis, the skewness, and the variation coefficient in order to characterize the real distribution of the area and the perimeter in more detail. We also calculate the irregularity factor see Section 2.

5. Experiment

For decision tree induction we used our data mining tool Decision Master [6]. The decision tree was constructed based on the entropy criteria for attribute selection and reduced-error pruning [7]. The learnt classifier was evaluated by cross validation.

With the increase of the number of slices the number of features will also increase. This may lead to a deterioration of the performance of the decision tree [8]. Therefore, we performed in another experiment feature subset selection before the decision tree induction experiment. Feature subset selection was done by hierarchical clustering with single linkage and using Euclidean distance. The resulting dendrogram showed groups of similar features. The dendrogram was then decomposed into groups of similar features by a cut-off value of 10%, 5%, and 2.5% similarity. From the remaining groups one feature for each group was selected and stored into the feature subset. Based on this feature set a new data base was created. This data base was then used for the decision tree induction experiment. The best results for the different feature subsets are reported in Section 6.

Altogether, we carried out the following experiments:

1. Determination of the number of slices which give the best classification accuracy based on the feature set containing marking probability, mean and standard deviation for area and perimeter,
2. Feature subset selection before decision tree induction for the data sets of experiment 1,
3. Extension of the feature set by curtosis, skewness, variation coefficient, and irregularity factor, create the data sets, and carry out the decision tree induction experiment,
4. Feature subset selection before decision tree induction for the data sets of experiment 4, and
5. Learning of a decision tree from data of one slice and determination of the classification accuracy.

6. Results

6.1. Experiment 1

Figure 1 shows the result of the experiment. The best result is obtained for n=12 (ε=13.02%) for the number of gray level intervals.

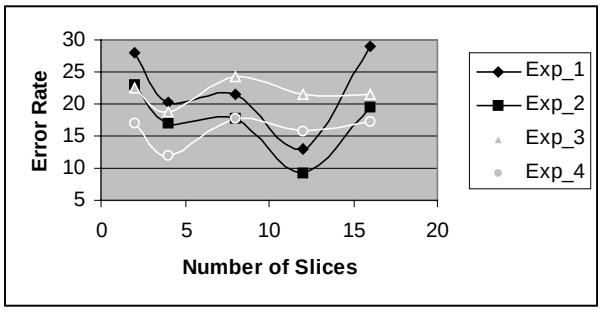


Figure 1. Error Rate versus Number of Slices

6.2. Experiment 2

An improvement in the error rate of 4 % (ε=9.3%) could be reached in case of n=12 and of 10 % in case of n=16 (see Fig. 1) when using feature subset selection before decision tree learning. However, we still get the best result for n=12, although the improvement in accuracy is much higher for n=16. The observations confirm that feature subset selection is a necessary step before decision tree induction.

6.3. Experiment 3

The first feature set contains the marking probability, the mean and the standard deviation of the area and the perimeter. Now we introduce more features for the description of the distribution of the area and the perimeter. It is interesting to note that the best result were obtained for n=4 (ε=18.2%) and not for n=12 (ε=21.4%) (see Fig. 1). However, the result is not better than the one in experiment_1 for n=12.

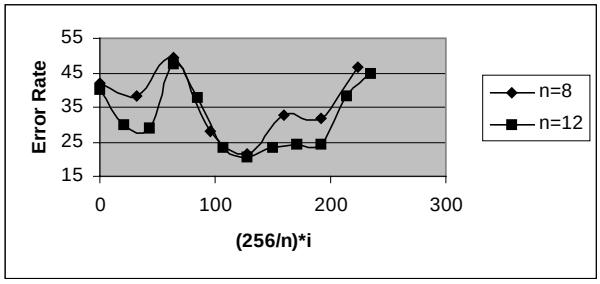
6.4. Experiment 4

Feature subset selection improves the accuracy of the classifier (n=4, ε=12.1%) but the results are not better than those obtained in Experiment 2 (see Fig. 1).

6.5. Experiment 5

The observation in experiment 1 that only features from the middle slices are taken for classification inspired the last experiment. Data sets for each slice were constructed containing all features. In the case of 12 slices we have now 12 different data sets. From each data set a decision tree is learnt and the error rate is calculated. Figure 2 shows the error rate for each slice *i* versus the midpoint of the gray level interval for n=8 and n=12. It is interesting to note that in the best case (i=6, t=127,99) the error rate (20.51 %) is only 7% higher than the error rate obtained when using all slices.

Figure 2. Error Rate versus the gray level of the midpoint of the interval



7. Discussion

Experiment_1 shows that a high number of gray level intervals is necessary to get sufficient accuracy. However, the conclusion that an increasing number of gray level intervals *n* leads to a decreasing error rate is not valid. In case of *n=16* the objects in each slice tend to be only single points and as those they have no difference in the area and the perimeter.

An increasing *n* results into an increasing number of features. The proportion between the number of features and the number of data samples will get worse and worse. The chosen method is therefore only applicable when a large enough number of data samples is available. Otherwise feature subset selection is necessary before the decision tree induction experiment. Experiment_2 shows that a reasonable improvement can be reached by feature subset selection concerning the accuracy of the classifier.

The results also show that the way in which the gray level has been partitioned into several intervals highly influences the error rate (*n=2,4,8, 16* and another *n* is *12* which is not a multiple of 2^i). The error rate shown in Fig. 1 is not a monotone decreasing function. The function has peaks and valleys which leads to the conclusion that there are bad and good partitions. The right partition has to be determined experimentally.

Whereas for a small number of *n* the primary grains in the slice images are big regions, sometimes with holes in it or with another region inside the hole, in the case of a large *n* the regions tend to be single points and compact regions. The model assumption that the primary grains are convex fits better in the latter case. For the first case we need to describe the object gestalt and their relation to each other (overlapping, touching, etc.) in more detail. The computational burden will increase for this case. However, if we are able to describe the objects in each slice in more detail, then it must be possible to use only a few slices for texture classification. Experiment_3 seems to confirm this assumption, since a good accuracy can be reached in this experiment already from the data from 4 slices and by using features that describe the distribution of the area and the perimeter in more detail.

Experiment_5 shows that sufficient accuracy can be reached by using only data from one single slice. Therefore, we conclude that it must be possible to classify the textures based only on one projected thick section. Histogram equalization (see Sect. 3) must play a role in this problem. It ensures a uniform distribution of the gray levels and they are spread over the whole range of gray levels. Whereas at the two ends of the gray level scale only a few pixels are remaining in a slice image, we obtain in the thick sections taken from the middle of the gray level range objects of different size and distribution.

A visual examination of the different slice images has shown that the objects in different slices have different spatial relation among each other. It might be interesting to describe this characteristic in more detail. A good descriptor might be derived from the set covariance function. This will be our next investigation.

As long as we are only describing the regions in each slice by area and perimeter, the computational burden is reasonable. When we have to calculate more detailed features for the gestalt of a region, the complexity of the problem will dramatically increase and it is doubtful if such an approach is applicable under real-time conditions. On the other hand the Boolean model is very flexible and it is very easy to include data from new classes into the data base and derive a classifier with sufficient accuracy.

8. Conclusion

The Boolean model (also known as the Poisson grain model) is an important and relatively simple example of a random set. It is flexible and can be used to describe a particular random set and it approximates more complicated random sets. As such it can be used for texture characterization as well as for texture classification. We have studied the Boolean model based on our problem of HEp-2 cell classification. However, the results might also be valid for other applications. The accuracy of the model is a function of the number of slices. The right number of slices must be determined experimentally. A clue for an automatic determination of the right partition can be that the objects in the slices must be more or less convex objects when describing them by mean and standard deviation of the area and the perimeter.

The large number of features which will be created may deteriorate the performance of the classifier. Therefore, feature subset selection should be used before the construction of the classifier.

Decision tree induction can be used as classification method for this kind of texture description.

According to the theory and partially confirmed by the experiment it must be possible to classify the texture from only one projected thick section. The open question is under which condition this is possible.

Acknowledgment

The research presented in this paper has been founded by the German Ministry of Economy under the grant LERNBILDZELL for Biotechnology applications.

References

[1] Serra, J. The Boolean model and random sets, Computer Vision and Image Processing, 12, 1980

[2] Stoyan D., Kendall W.S., Mecke J., Stochastic Geometry and Its Applications, Akademie-Verlag Berlin 1987

[3] Lang C., Ohser J., and Hilfer R., On the analysis of spatial binary images, Journal of Microscopy, vol. 203, Pt 3, Sept. 2001, pp.303-313.

[4] Garcia P., Petrou M., and Kamata S., The Use of Boolean Model for Texture Analysis, Computer Vision and Image Understanding, vol. 74(3)1999, pp. 227-235

[5] Otsu, N., A threshold selection method from gray-level histograms, IEEE Trans., vol. SMC-9, Jan. 1979, p. 38-52.

[6] Data Mining Tool *Decision Master*, http://www.ibai-solutions.de

[7] Perner P., Data Mining on Multimedia Data, Book Springer Verlag 2001 to appear

[8] P. Perner, Improving the Accuracy of Decision Tree Induction by Feature Pre-Selection, Applied Artificial Intelligence, vol. 15, No. 8, p. 747-760.