Res. Rep. Fac. Eng. Mie Univ., Vol. 18, pp. 73-80 (1993)

Review

A Survery on Probabilistic Relaxation Method in Pattern Recognition

Shinji TSURUOKA, Tetsushi WAKABAYASHI+,
Fumitaka KIMURA+, Yasuji MIYAKE+
(Department of Electrical and Electronic engineering)
(+Department of Information Engineering)

(Received September 16, 1993)

This paper reviews the recent developments in the use of "relaxation method" in pattern recognition. We introduce some representative relaxation procedures, and the applications in image analysis. The application of these methods include histogram modification, edge detection, and Character recognition. These applications are briefly described.

Key Words: relaxation method, iterative method, probabilistic approach, image analysis, character recognition

1. Introduction

Pattern recognition almost always involves discriminating or classifying of the part of the image. When we recognize the character image, we classify the strokes of the input image according to the dictionary in the brain, which is learned by the learning samples since we were born. When we detect edgs in an image, we classify each point as an edge or non-edge point according to the value of some local property (for example, the magnitude of the gradient) evaluated at that point; and similarly when we detect angles on a curve. We classify the edge with reference to some classification decisions at other points. The criterion for accepting a point P as an edge point can depend on whether or not P extends an edge that has already been detected. But, the results of sequential classification will in general depend on the order in which the points are examined, and this is undesirable.

This paper discusses "relaxation method" (iterative approach) to classifying the parts of the input image. The decisions are allowed to depend on the decisions made at the previous iteration, the iterative approach shares an advantage of the sequential approach, but it does not share the disadvantage of order-dependence, since the decision at a given itteration does not depend on any previous decisions at same iteration - i.e. each iteration is performed in "parallel (at the same time)".

In the next chapter, we formulate general models for this probabilistic relaxation method, and in the third chapter we give the three applications of its uses.

2. Princeples for probabilistic relaxation

There are two type "Relaxation methods", which are the "Discrete Relaxation" and "Probabilistic Relaxation". In "Discrete Relaxation", we assume two states, which are possible and impossible, between objects and classes of the each objects. This approach is described in reference (1) in detail and is assured the convergence. "Probabilistic Relaxation" is the generalization of "Discrete Relaxation", the possibility is represented by the probability of the classifying, which is between 0 and 1. In image analysis, the probabilistic approach is used more than the discrete approach, because ambiguousness with low-level output from the image can be hold in the probability. We introduce this approach in the following.

2.1 Concept of probabilistic relaxation (2)

The probabilistic relaxation can use probabilistic classifications rather than making firm classification decisions immediately. Initially, for each point P_i, we estimate the probability p_{ih} that it belongs to each of the possible classes (C_h). We then examine the probabilities at neighboring points P_j, and we increase p_{ih} if supporting evidence is found for it or decrease it if contradictory evidence is found. For example,we increase p_{ih} if there exist high probabilities that the neighbors of P_i have compatible classifications, or decrease it if there exist high probabilities of incompatible classification at P_i's neighbors. This process is done simultaneously for every P_i and every p_{ih}; and it can be iterated as many times as desired. What often happens when this is done is that the probablities tend to become less ambiguous according to the number of the iteration. Thus classification of P_i becomes very easy, if only one probability p_{ih} is high, that is, the point P_i is classified into the class that has maximum probability.

2.2 General models for probabilistic relaxation (2)

This section describes general models for probabilistically classifying a set of objects $A_1,...,A_n$ into a set of classes $C_1,...,C_m$. The objects may be image points, or may be entities such as lines in a character, or regions in an image. We assume that each A_i ($i = 1,2,...n$) has a specified set of A_j ($j = 1,2,...n$) 's as neighbors.

With each object Ai we associate a probability vector $(p_{i1},...,p_{im})$, where p_{ih} ($h = 1,2,...m$; $-1 \leq p_{ih} \leq 1$) is an estimate of the probability that A_i belongs to class C_h. Initially, this estimate is derived by some convientional type of analysis of A_i; for example, we might estimate the probability that the point A_i is an edge as proportional to the magnitude of the gradient of the grey-level image at A_i ($\sum_j |f(Ai) - f(Aj)|$ where f(Ai) is the grey-level at Ai).

We next assume that we are given , for each pair of neighboring objects (Ai, Aj) and each pair of classes (Ch, Ck), a measure of the compatibility between object Ai belonging to class Ch and object Aj belonging to class Ck. For example, let Ai, Aj be vertically adjacent points, and let Ch,Ck be the classes of vertical edges having their dark sides on the left and on the right, respectively. Then Ai ε Ch and Aj ε Ch are highly compatible (and similarly for Ai, Aj ε Ck), but Ai ε Ch and Aj ε Ck (or vice versa) are highly incompatible. We will denote the compatibility of Ai ε Ch with Aj ε Ck by r (Ai, Ch, Aj, Ck) ("Compatibility Coefficients"), and we will assume that $-1 \le r \le 1$, where -1 is complete incompatibility, +1 is complete compatibility, and 0 is "don't care".

If we are given the initial p_{ih}'s and the r's, we can now adjust the p_{ih}'s on the basis of the neighboring p_{jk}'s and their associated r (A_i, C_h,A_j,C_k)'s. The adjudtment process should satisfy the following properties:

(a) If p_{jk} is high and r (A_i, C_h,A_j,C_k) is close to +1, then p_{ih} should be increased.

(b) If p_{jk} is high and r (A_i, C_h,A_j,C_k) is close to -1, then p_{ih} should be decreased.

(c) If p_{jk} is low, or r (A_i, C_h,A_j,C_k) is close to 0, then p_{ih} should not change significantly.

A simple expression that satisfies these properties is the product p_{jk}*r (A_i, C_h,A_j,C_k). Thus we can use these products, for all Aj and C_k, to increment p_{ih} . The contribution from each neighbor A_j (the sum of p_{jk}* r(A_i, C_h,A_j,C_k).about j) is between -1 and +1, since the sum of p_{jk} about j is 1 and $-1 \le$ r (A_i, C_h, A_j, C_k)≤ 1. To insure that p_{ih} remains nonnegative, we can weight the neighbors' contributions using weights w_j that sum to 1, and we can define the net increment to pih by an expression of the form

$$p_{ih}\{1+\sum_j w_j \sum_k p_{jk} r(Ai,Ch,Aj,Ck)\},$$

where the double sum is always between -1 and +1, so that p_{ih} is always multiplied by a nonnegative quantity. When all the p_{ih}'s have been adjusted in this way, we can renormalize by dividing each pih by the sum of p_{ih} about h (=1,...,m), to insure that the pih's sum to 1 for each i.

In summary, let $p^{(t)}_{ih}$ be the value of pih at the i-th iteration, and let the net increment to $p^{(t)}_{ih}$ be

$$q_{ih}^{(t)} = \sum_j w_j \sum_k p_{jk}^{(t)} r(Ai,Ch,Aj,Ck)$$

Then the new value of pih is

$$p_{ih}^{(t+1)} = p_{ih}^{(t)}(1+q_{ih}^{(t)}) \,/\, \sum_{h=1}^{m}(1+q_{ih}^{(t)})$$

A more detailed discussion of probability adjustment rules of this type can be found(1), and a study of the convergence properties of one such rule is given(3). This approarch is sometimes referred to as a "relaxation" process, because of its resemlbance to certain iterative techniques used in numerical analysis. A general discussion of these processes and their applications can be found(4,5).

2.3 Another models for probabilistic relaxation

There are another relaxation models, which are Kirby(6)'s and Peleg(7)'s relaxation. The former is better than the latter in probabilistic theory(8,9). But in practice, Peleg's relaxation is better than the Kirby's relaxation in the classification rates and the convergence on multispectram image(10-13).

(1) Kirby's relaxation(6)

The new value of $p^{(t+1)}_{ih}$ is given by the following form.

$$p_{ih}^{(t+1)} = \frac{p_{ih}^{(t)} \prod_{j} q_{ijh}^{(t)}}{\sum_{h=1}^{m} p_{ih}^{(t)} \prod_{j} q_{ijh}^{(t)}}$$

(2) Peleg's relaxation(7)

The new value of p(t+1)ih is given by the following form.

$$p_{ih}^{(t+1)} = \sum_{j} c_j \frac{p_{ih}^{(t)} q_{ijh}^{(t)}}{\sum_{h=1}^{m} p_{ih}^{(t)} q_{ijh}^{(t)}}$$

where c_j is the positive real coefficient, and the sum of cj about j is 1.

Kirby's relaxation is the products of the neighbor's probability, but Peleg's relaxation is the sum of the neighbor's probability

(3) Suzuki's relaxation(11)

H. Suzuki et al. proposed the modified Peleg's relaxation, which is used the judgement of the convergence. The criterion is given as follows.

[Criterion for convergence]

Let $p^{(t-1)}_{ih}$ be the probability at the (t-1)-th iteration.

$$p_{ih}^{(t-1)} = \max_{H}\{p_{iH}^{(t-1)}\}$$

If the following relation is satisfied at the next iteration,

$$p_{ih}^{(t)} = \max_{H}\{p_{ih'}^{(t)}\} > p_{ih}^{(t-1)}, \text{ and } p_{iH}^{(t)} \le p_{iH}^{(t-1)} \ (h' \neq h)$$

then $p^{(t)}_{ih'}$ is replaced by the following equation,

$$p_{iH}^{(t)} = \begin{cases} 1 & (H = h) \\ 0 & (H \neq h) \end{cases}$$

and this $p^{(t)}_{ih'}$ is not changed after this iteration. Because the evidence that the pixel Ai belongs to Ch is supported by the neighbors.

3. Applications

In this section, we briefly describe the application of iterative probabilistic classification and parameter estimation techniques to a number of image analysis tasks. Further details about each application can be found in the cited references.

3.1 Histogram modification(14)

We first describe a simple iterative peak enhancement scheme, as applied to peak detection on histograms and to adaptive image quantization. Let the objects $A_i,\ldots, A_n$ be the histogram bins, and let us associate with each bin, as a parameter value, the number n_i stored in that bin. (We could, if desired, also associate with each bin the probabilities p and 1 - p that the bin is or is not a peak, where the peak probability is initially proportional to the bin height; but we will dispense with these probabilities in the present discussion.) The bin parameter values n_i can then be adjusted as follows; n_i is incremented if it has a neighbor with a smaller value n_j, and decremented if it has a neighbor with a larger value. For example, we could add to n_i a quantity proportional to n_i - n_j for each neighbor A_j of A_i. It is not hard to see that this process will sharpen peaks and eliminate ramps. If we use only A_j's two immediate neighbors A_{i+1}, any local peak will be sharpened; but if we use a larger set of neighbors, we will enhance only the major peaks. Given a set of peaks in an image's gray level histogram, we can requantize the image by mapping each gray level into the nearest peak. Thus our histogram peak enhancement scheme can be used for adaptive image quantization.

3.2 Edge detection(15)

The edge detection has already been mentioned by various methods. Let $A_1,\ldots, A_n$ be the image points, and let the classes correspond to edges in specific orientations $\theta_1,\ldots, \theta_{m-1}$, plus an additional class for "no edge" -- or for simplicity, we can use only the two classes "edge" and "no edge", together with an orientation parameter θ. Initially we estimate the edge probability p_i at point A_i as proportional to the gradient magnitude at A_i, and we estimate the edge orientation θ_i as perpendicular to the gradient direction at A_i. We can now adjust p_i in accordance with the probabilities p_j and orientations θj of edges at neighboring points A_j that extend or contradict the edge in orientation θ_i at A_i. For example, if θ_{ij} is the direction from A_i to A_j, we can use an increment proportional to $|\cos(\theta_i - \theta_{ij})|$ and to $\cos(\theta_i - \theta_j)$. In defining the orientation of an edge, we should use a convention in which the dark side of the edge is always on the clockwise side, in order to distinguish between edges that have the same direction but opposite "senses"; for such edges we then have $\cos(\theta i - \theta j) = \cos\pi = -1$, so that they decrement one anothers' probabilities. Examples of this approach to edge detection are given(15). In these examples, iterating the reinforcement process suppresses noisy edge responses and enhances edges that form long, smooth boundaries. A specific example of an edge reinforcement process(15) is as follows: We use two classes, "edge" and "no edge". The initial "edge" probability at point P is defined to be

$$p_e^{(0)} \equiv \text{gradientmagnitude at } P \,/\, \max_Q \{\text{gradient magnitude at } Q\}\,;$$

the max is taken over the entire image.

For any point (x,y) and neighboring point (u,v), let α be the edge slope at (x,y), β the edge slope at (u,v), γ the slope of the line joining (x,y) to (u,v), and $D= \max(|x-u|, |y-v|)$ the "chessboard distance " from (x,y) to (u,v). Then we define the compatibilities between pairs of class assignments of the points (x,y) and (u,v) as some rules. It can be verified that these compatibilities have the following properties:

(a) Parallel or perpendicular edges have no effect on one another.

(b) Collinear edges reinforce each other, and anti-collinear edges weaken each other (by an amount that decreases with distance).

(c) Nonedges collinear with edges weaken them; nonedges alongside edges have no effect on them.

(d) Edges alongside nonedges strengthen them; edges collinear with nonedges have no effect on them.

(e) Nonedges reinforece one another.

A more elaborate scheme can be used to classify both edge and interior points of an image simultaneously. Here interior probabilities reinforce neighboring interior probabilities; edge probabilities reinforce edge probabilities that extend them; and edge probabilities reinforce interior probabilities that lie alongside them (and vice versa). A scheme of this type has been used to detect interior and border points of dot clusters in an image(5).

3.3 Character Recognition(16-18)

Shape of input patterns is represented by polygonal approximation. A line segment Ok is a side of the polygon of input pattern, and .φ is the number of line segments of polygon;

Okφ =(Xsk, Ysk, θk, Lk, Xek, Yek)φ

Where (Xsk, Ysk) and (Xek, Yek) are the positions of the start point and the end point for the line segment k; and θk and Lk are the direction and the length of the line segment k.

Each line segment k has two kinds of neighboring segments for start and end points. Each category has only one mask then a dictionary is composed of 3036 masks. Each mask is composed of Φ_M segments. Each feature of mask segments is evaluated by the set of associated segments of learning data as indicated as follow.

Feature (Xsi, Ysi, θi, Li, Xei, Yei) of mask segment Mi is calculated from the average value of input segment k for the learning set of same category. The neighboring segments list of the mask is assigned from the neighboring segments list of the first datum of the learning set. Association function F between mask segment i and input segment k is evaluated on success of follow condition:

$$|\theta i - \theta k| \le c_1 \quad \text{and} \quad |Li - Lk| \le c_2$$

$$\text{and} \quad \sqrt{(x_{si} - x_{sk})^2 + (y_{si} - y_{sk})^2} \le c_3 \quad \text{and} \quad \sqrt{(x_{ei} - x_{ek})^2 + (y_{ei} - y_{ek})^2} \le c_3$$

Similarity p'i(k) of the pair (i,k) is assigned as follows:

$$p'_{i(k)} = \max\{1 - W_1 \times \max(|\theta i - \theta k| - c_{21}, 0) - W_2 \times \max(|Li - Lk| - c_{22}, 0) - W_3 \times \max(\sqrt{dx_{sik}^2 + dy_{sik}^2} - c_{23}, 0) - W_4 \times \max(\sqrt{dx_{eik}^2 + dy_{eik}^2} - c_{23}, 0), 0\}$$

where, $dx_{sik} = x_{si} - x_{sk}$, $dy_{sik} = y_{si} - y_{sk}$, $dx_{eik} = x_{ei} - x_{ek}$, $dy_{eik} = y_{ei} - y_{ek}$

The initial mask side probability $P^{(0)}{}_{i(k)}$ is evaluated as follows:

$$P^{(0)}_{i(k)} = p'_{i(k)} / \sum_{k''} p'_{i(k'')}$$

where, k" 's are all the input segments which are associated with i-th segment in the dictionary.

The relaxation operation is used to dynamically change the probability according to the compatibility between the configurations of mask segments and the input segments. They assume that there are spring connections between the start point (x_{si},y_{si}) of M_i and the end point (x_{ej},y_{ej}) of Mj. Where segment M_j is the neighboring segments of the start point of the segment Mi. Let $r^s{}_{ij}(k,l)$ be the start point tension in spring connection between Mi and Mj. $r^s{}_{ij}(k,l)$ is evaluated as follows:

$$r^s_{ij}(k,l) = \max\{\min(1 - W_5 \times |dx_{sik} - dx_{ejl}| + c_{24}, 1), 0\} + \max\{\min(1 - W_5 \times |dy_{sik} - dy_{ejl}| + c_{24}, 1), 0\}$$

The probability is evaluated as follows:

$$P^{(t+1)}_{i(k)} = \{q_{i(k)} \times P^{(t)}_{i(k)}\} / \{\sum_{k''} q_{i(k'')} \times P^{(t)}_{i(k'')}\}$$

$$q_{i(k)} = \{q^s_{i(k)} + q^e_{i(k)}\} / \{\sum_{j=1}^{n_s} Lj + \sum_{j'=1}^{n_e} Lj'\}$$

$$q^s_{i(k)} = \sum_{j=1}^{n_s} \{Lj \times \max_l (r^s_{ij}(k,l) \times P^{(t)}_{j(l)})\} \qquad q^e_{i(k)} = \sum_{j'=1}^{n_e} \{Lj' \times \max_{l'} (r^e_{ij'}(k,l') \times P^{(t)}_{j'(l')})\}$$

We select the ^k, where Pi(k) is the maximum (P'i(k)) about all k, corresponding to i. The distance Dj between the mask J and the input and the confidence Ri(k) of the mask segmrnt i are defined as follows :

$$R_{i(\tilde{k})} = \frac{\sum_{j=1}^{n_s}\{Lj \times r_{ij}^{s}(\tilde{k},\tilde{l}) \times P'_{j(\tilde{l})}\} + \sum_{j'=1}^{n_e}\{Lj' \times r_{ij'}^{e}(\tilde{k},\tilde{l}') \times P'_{j'(\tilde{l})}\}}{(\sum_{j=1}^{n_s} Lj + \sum_{j'=1}^{n_e} Lj') \times 2} \times P'_{i(\tilde{k})}$$

$$D_j = \frac{\sum_{i=1}^{\phi}\{(1 - R_{i(\tilde{k})}) \times Li\} + \sum_{i'} Li' + \sum_{k'} Lk'}{\sum_{i=1}^{\phi} Li} \times c_6$$

where, Li' is the length of the non-corresponding segments in the mask, Lk' is the length of the non-corresponding segments in the input.

The recognition rate for the handprinted "Kanji" and "Hiragana" character database ETL-9 (3036 categories) is 98.55% for unknown subsets with 10 learning subsets.

4. Conclusion

A probabilistic relaxation approach has used in another applications, for example, noise clearning(19,20), curve detection(21,22), angle detection(23), curve thinning(24), template matching(25), identifing regions(26,1), separation of overrapping patterns(27,28), resroring a polyhedron(29). We cannot describe these application, but they are very interesting. Further details about each application can be found in the cited references.

References

1) A. Rosenfeld, R. Hummel and S. W, Zucker : Scene Labelling by Relaxation Operation, IEEE Trans. Syst., Man, Cybern. SMC-6, pp.420-433 (1976).

2) A. Rosenfeld : Iterative Methods in Image Analysis, Pattern Recognition, 10, pp.181-187 (1978).

3) S. W. Zucker, E. V. Krishnamurthy and R. Haar : Relaxation Processes for Scene Labelling: Convergence, Speed, and Stability, IEEE Trans. Syst., Man, Cybern. SMC-8, pp.41-48 (1978).

4) S. W. Zucker : Relaxation Labelling, Local Ambiguity, and Low-Level Vision in C. H. Chen, ed.: Pattern Recognition and Artificial Intelligence. Academic Press, New York, pp.593-616 (1976).

5) S. W. Zucker : Relaxation Labelling and the Reduction of Local Ambiguities, in Proc. 3rd Int. Jnt Conf. on Pattern Recognition (IEEE Publ. 76CH1140-3C), pp.852-861 (1976).

6) R. L. Kirby : A Product Rule Relaxation Method, Comput. Vision, Graphics & Image Process., 13, pp.158-189 (1980).

7) S. Peleg : A New Probabilistic Relaxation Scheme, IEEE Trans. Pattern Anal. Mach. Intell., PAMI-2, 4, pp. 362-369 (1980).

8) P. M. Haralick : An Interpretation for Probabilistic Relaxation, Comput. Vision, Graphics & Image Process., 22, pp. 388-395 (1983).

9) J.Kittler and J. Foeglein : On compatibility and support functions in probabilistic relaxation, Comput. Vision, Graphics & Image Process., 34, pp. 257-267 (1986).

10) T. Watanabe, H. Suzuki, S. Tanba, R. Yokoyama : On the Validities of Kirby's and Peleg's Relaxation

Procedure [in Japanese], IEICE(D-II), J76-D, 4, pp.904-907 (1993).

11) H. Suzuki, T. Watanabe, S. Tanba, R. Yokoyama : An Accelerating Method of Probabilistic Relaxation Procedure [in Japanese], IEICE(D-II), J75-D, 8, pp.1474-1477 (1992).

12) T. Watanabe, H. Suzuki : An Experimental Evaluation of the Classifiers Using Spatial Context for Multispectral Images [in Japanese], IEICE(D), J70-D,4,pp.750-759 (1987).

13) T. Watanabe, H. Suzuki : A Study on Compound Decision Theory and Adaptive Classification for Multispectral Image Data [in Japanese], IEICE(D), J71-D, 6,pp.1083-1090 (1988).

14) A. Rosenfeld and L. S. Davis : Image Segmentation and Image Models,, Proc. IEEE, 67, pp.764-772 (1979).

15) B. R. Schachter, A. Lev, S.W. Zucker and A. Rosenfeld : An Application of Relaxation Methods to Edge Reinforcement, TEEE Trans. Sys., Man, Cybern, SMC-7, pp. 813-816 (1977).

16) K. Yamamoto : Recognition of Handprinted KANJI Characters by Relaxation Matching [in Japanese], IEICE(D), J65-D, 9, pp.1167-1174 (1982).

17) K. Yamamoto, H. Yamada, T. Saito, I. Sakaga : Recognition of Handprinted Characters in the First Level of JIS Chinese Charcters, 8-th Int. Conf. on Pattern Recgnition, pp.570-572 (1986).

18) K. Toraichi, S. Ishiuchi, T. Horiuchi, K. Yamamoto, H. Yamada : Recognition of Handwritten KANJI and HIARGANA Characters by Relaxation Matching Method Using Straight and Curved Lines Approximation of Boundary Lines [in Japanese], IEICE(D-II), J73-D, 9, pp.1448-1457 (1990).

19) A. Lev, S. W. Zucker and A. Rosenfeld : Iterative Enhancement of Noisy images, IEEE Trans. Syst., Man, Cybern. SMC-7, pp.435-442 (1977).

20) L. S. Davis and A. Rosenfeld : Noise Cleaning by Iterated Local Averaging, TR-520, Computer Science Center, Univ. of Maryland, College Park, April (1977).

21) S. W. Zucker, R. A. Hummel and A. Rosenfeld : An Application of Relaxation Labelling to Line and Curve Enhancement, IEEE Trans. Comput.., C-26, pp.394-403 (1977).

22) G. J. VanderBrug : Experiments in Iterative Enhancement of Linear Features, Computer Graphics Image Processing, 6, pp.25-42 (1977).

23) L. S. Davis and A. Rosenfeld : An Application of Relaxation Labelling to Spring-Loaded Template Matching, Proc. 3rd Int. Jnt. Conf. on Pattern Recognition, pp.591-597 (1976).

24) R. B. Eberlein : An Iterative Gradient Edge Detection Technique, Computer Graphics Image Processing, 5, pp.245-253 (1976).

25) L. S. Davis and A. Rosenfeld : Curve Segmentation by Relaxation, IEEE, Trans. Comput. C-26, pp.1053-1057 (1977).

26) J. M. Tenenbaum and H. G. Barrows : IGS: A Paradigm for Integrating Image Segmentation and Interpretation, Pattern Recognition and Artificial Intelligence (C.H.Chen,ed.), Academic Press, New York, pp.472-507 (1976).

27) K. Sakaue : Application of Iterative Method in Image Processing [in Japanese], J. IPS Japan, 23, 7, pp.641-650 (1982).

28) K. Sakaue, M. Takagi : Separation of Overlapping Bar-Like Patterns by Iterative Method [in Japanese], IEICE(D), J66-D, 5, pp.585-592 (1983).

29) T. Horiuchi, K. Toraichi, K. Yamamoto, H. Yamada, Y.Ohtaki : On a Method of Restoring a Polyhedron from Three Orthographic Views by Relaxation Matching [in Japanese], IEICE(D-II), J76-D, 1,pp.9-19 (1993).