COMPUTER VISION, GRAPHICS, AND IMAGE PROCESSING **22**, 388–395 (1983)

NOTE

An Interpretation for Probabilistic Relaxation

ROBERT M. HARALICK

Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

Received October 10, 1981; revised February 24, 1982

Probabilistic relaxation has been the basis of one of the popular cooperative processing mechanisms used in image analysis. It has been a mechanism whose theory has not been well understood. In this paper, some general conditional independence conditions are stated which give probabilistic relaxation the interpretation that each iteration computes the conditional probability of each local label given a new context which is the context of the previous iteration enlarged by one neighborhood width. This interpretation implies that relaxation iterations must only continue until the conditional independence assumptions no longer hold, or until the entire context is taken into account, whichever comes first.

1. INTRODUCTION AND REVIEW

In image analysis there have been numerous papers on the effective use of cooperative processing through the mechanism of probabilistic relaxation. The idea was first introduced by Rosenfeld, *et al* [4]. In cooperative processing, neighboring information positively or negatively reinforces the weights for each local unit of information, depending on the compatibility of the neighboring information with the local information. After each relaxation iteration, the resulting values are more consistent with the prior knowledge of information dependencies and global context. The next paragraph describes the original Rosenfeld *et al.* formulation.

Let $\{1,\dots, I\}$ be the set of units and $N(i)$ be the set of neighbors for unit i. Each unit i has a weight function associated with it which specifies the strength of label x_i for unit i. By $P(x_i, t)$ we denote the strength of label x_i for unit i after iteration t. The initial weight functions $P(x_i, 0)$ are typically determined by some local measurement process on unit i and are normed to be nonnegative and sum to 1.

For any unit j, a neighbor of unit i, we denote by $r_{ij}(x, y)$ the compatibility of unit i having label x and unit j having label y. The compatibility coefficients have magnitude less than 1; $-1 \leq r_{ij}(x, y) \leq 1$ and satisfy $-1 \leq \sum_{j\in N(i)}\sum_y r_{ij}(x, y) \leq 1$. The original relaxation iterations described by Rosenfeld *et al.* [4] have the form

$$P(x_i, t+1) = \frac{P(x_i, t)\left[1 + \sum_{j\in N(i)} \sum_{y_j} r_{ij}(x_i, y_j) P(y_j, t)\right]}{\sum_{z_i} P(z_i, t)\left[1 + \sum_{j\in N(i)} \sum_{y_j} r_{ij}(z_i, y_j) P(y_j, t)\right]}. \tag{1}$$

It is quickly seen that the conditions on the compatibility coefficients, combined with the fact that the weight functions are normed like probabilities, imply that

0734-189X/83 $3.00

$\Sigma_{j \in N(i)} \Sigma_y r_{ij}(x, y) P_j(y, t) < 1$. Hence the resulting weight function will be nonnegative and sum to 1.

Zucker and Mohammed [6] suggested rewriting and modifying the relaxation in such a way that the compatibility coefficients had the meaning of conditional probabilities; they rewrote $r_{ij}(x, y)$ as $P_{ij}(x|y)$ with the interpretation that $P_{ij}(x, y)$ is the conditional probability that unit i takes the label x, given that unit j takes the label y. Of course it is required that $P_{ij}(x|y) \geq 0$ and

$$\sum_x P_{ij}(x|y) = 1.$$

The modified probabilistic relaxation equation takes the form

$$P(x_i, t + 1) = \frac{P(x_i, t) \prod_{j \in N(i)} \sum_{y_j} P_{ij}(x_i|y_j) P(y_j, t)}{\sum_{z_i} P(z_i, t) \prod_{j \in N(i)} \sum_{y_j} P_{ij}(z_i|y_j) P(y_j, t)}. \tag{2}$$

This form is certainly more suggestive of a probability interpretation for $P(x_i, t)$. Whatever the interpretation might be, however, it is not immediately apparent from the relaxation equation.

In either of these forms, the relaxation iteration was considered useful regardless of the number of times it had been previously iterated. The reason is that after each iteration the weights or probabilities were thought to be more consistent with prior expectations of neighboring label dependencies. Thus, the question of whether the iterations converge arose naturally, as did the question about the meaning of the fixed point. Aspects of these questions were answered by Zucker, *et al.* [5], and Haralick *et al.* [1].

Peleg [3] attempted to give a more solid meaning to the question about what the probabilities were in a modified Zucker relaxation. He suggested that the Zucker compatibility coefficients r_{ij} should take the form

$$r_{ij}(x_i, y_j) = \frac{P_{ij}(x_i, y_j)}{P(x_i) P(y_j)}$$

and that the probabilities $P(x_i, t)$ were really just estimates that the unit i took the label x given that the previous estimate for this probability was $P(x_i, t - 1)$. The justification given for this interpretation was based in part on a probability derivation with some conditional independence assumptions followed by some approximations. Section 4 gives a detailed discussion of the problems with the Peleg interpretation. Kirby [2] also gives an analysis of the Peleg relaxation equation; the product rule equation he gives is similar to the one we develop here but the interpretation is different.

In our notation, the relaxation equation which Peleg gives is

$$P(x_i, t+1) = \sum_{j \in N(i)} c_{ij} \frac{P(x_i, t) \sum_{x_j} P(x_j|t) r_{ij}(x_i, x_j)}{\sum_{z_i} P(z_i, t) \sum_{y_j} P(y_j, t) r_{ij}(z_i, y_j)} \tag{3}$$

where the c_{ij} are weights which are nonnegative and which, for each index i, sum on the j index to 1.

In this note, we use a derivation having some similarities to the Peleg derivation and show that, with some appropriate conditional independence assumptions, each succeeding iteration of the relaxation equation produces a conditional probability that a unit takes a label given the context which is one neighborhood width larger than the unit's context at the previous iteration. With this meaning, the question about convergence is irrelevant. The iterations can continue until the entire context is taken into account. Further iterations than this will no longer create interpretations of conditional probability of the label given the entire context. This meaning also explains a behavior sometimes noticed in relaxation experiments: the emerging probabilities sometimes appear to be getting better for the first few iterations, after which they appear to get worse. In these instances, what is happening is that the required conditional independence assumptions are getting further and further away from a good modeling of reality. The error introduced eventually overtakes the benefit produced by the larger context and the probabilities get worse.

Section 2 makes this interpretation of probabilistic relaxation precise and Section 3 states the conditional probability assumptions and shows that these assumptions lead to the interpretation given in Section 2.

2. AN INTERPRETATION OF PROBABILISTIC RELAXATION

In this section we develop an interpretation for the relaxation equation

$$P(q_i, t+1) = \frac{P(q_i, t) \prod_{j \in N(i)} \sum_{q_j} P(q_j, t) r_{ij}(q_i, q_j)}{\sum_{s_i} P(s_i, t) \prod_{j \in N(i)} \sum_{q_j} P(q_j, t) r_{ij}(s_i, q_j)} \tag{4}$$

where

$$r_{ij}(q_i, q_j) = \frac{P(q_i, q_j)}{P(q_i) P(q_j)}.$$

Our interpretation states that $P(q_i, t)$ is the conditional probability that unit i takes label q_i given the tth level context. Furthermore, the context at each iteration grows by an entire neighborhood width surrounding the previous level context.

To make these remarks precise, we will have to make a change in the notation in which the context is explicitly written. Context means the units and their corresponding measurements where the units come from some general neighborhood. Initially a measurement is made of each unit. We denote by d_i the measurement made of unit i. This is its immediate context. The neighborhood context for unit i is the measurement d_i plus all the measurements of units in the neighborhood of unit i. The next larger context for unit i is measurement d_i plus all the measurements of units in the neighborhood of unit i plus all the measurements of units in the neighborhood of the neighbors of unit i. The global context consists of all the units $1,\ldots, I$.

We denote by $Z_i(t)$ the units in the tth level context for unit i and by $N(i)$ the set of neighbors for unit i. The $Z_i(1) = \{i\}$. The units in the successive level contexts can be defined iteratively by $Z_i(t+1) = \{j \mid \text{for some } k \in Z_i(t),\ j \in N(k)\}$.

The purpose of the probabilistic relaxation is to compute, for each unit i and label q_i, the conditional probability $P(q_i|d_1,\ldots, d_I)$, where it is understood that a subscript n on a label or measurement designates that the label or measurement is for unit n. Thus $P(q_2)$ designates a generally different probability value than $P(q_3)$, even if $q_2 = q_3$. A more complete notation would write $P_2(q_2)$ for $P(q_2)$. We use the shorter notation to avoid writing unnecessarily complex expressions.

We will need to write conditional probabilities such as $P(q_i|d_1,\ldots, d_I)$, but only where the condition is on measurements for some arbitrary subset S of units whose names are not explicitly known. We denote this kind of conditional probability by $P(q_i|d_k\colon\ k \in S)$. Thus if $S = \{1, 3, 6, 7\}$, we write $P(q_i|d_k\colon\ k \in S)$ for $P(q_i|d_1, d_3, d_6, d_7)$. Likewise, if $T = \{2, 3, 4\}$ we write $P(q_n\colon\ n \in T|d_k\colon\ k \in S)$ for $P(q_2, q_3, q_4|d_1, d_3, d_6, d_7)$.

In this notation, the relaxation begins with $P(q_i|d_k\colon\ k \in Z_i(1))$ and terminates with the probabilities $P(q_i|d_k\colon\ k \in \{1,\ldots, I\})$. Letting

$$r_{nm}(q_n, q_m) = \frac{P(q_n, q_m)}{P(q_n)P(q_m)}$$

we have the interpretation for relaxation equation (4):

$$P(q_i|d_k\colon k \in Z_i(t+1)) = \frac{P(q_i|d_k\colon k \in Z_i(t)) \prod_{j\in N(i)} \sum_{q_j} P(q_j|d_k\colon k \in Z_j(t)) r_{ij}(q_i, q_j)}{\sum_{t_i} P(t_i|d_k\colon k \in Z_i(t)) \prod_{j\in N(i)} \sum_{q_j} P(q_j|d_k\colon d \in Z_j(t)) r_{ij}(t_i, q_j)}. \tag{5}$$

3. BASIS FOR THE INTERPRETATION

In this section, we state the two conditional probability assumptions which make relaxation equation (5) a valid equation. The assumptions are

$$P(q_i, q_k\colon k \in N(i)) = \frac{\prod_{j\in N(i)} P(q_i, q_j)}{P(q_i)^{|N(i)|-1}} \tag{6}$$

$$P(d_k\colon k \in Z_i(t+1)|q_i, q_k\colon k \in N(i)) = \alpha P(d_k\colon k \in Z_i(t)|q_i) \prod_{j\in N(i)} P(d_k\colon k \in Z_j(t)|q_j). \tag{7}$$

To understand how these generalized conditional independence assumptions may even be realistic, let us consider a one-dimensional example.

Let the units be $\{1, \ldots, I\}$ and the neighborhood of unit i be its immediate predecessor and successor: $N(i) = \{i - 1, i + 1\}$. Then (6) becomes

$$P(q_{i-1}, q_i, q_{i+1}) = \frac{P(q_{i-1}, q_i)P(q_i, q_{i+1})}{P(q_i)}$$
$$= P(q_{i+1}|q_i)P(q_i|q_{i-1})P(q_{i-1}).$$

In this form we immediately recognize (6) as a Markov dependence assumption on neighboring labels. For context level $t + 1 = 2$, eq. (7) becomes

$$P(d_{i-1}, d_i, d_{i+1}|q_{i-1}, q_i, q_{i+1}) = \alpha P(d_i|q_i)P(d_{i-1}|q_{i-1})P(d_{i+1}|q_{i+1}).$$

In virtually all signal and image processing situations this is true with proportionality constant $\alpha = 1$. The assumption simply states that, conditioned on all the units' labels, the measurements made on the units are independent (the no memory channel assumption) and that the measurements depend solely on the true label of the unit being measured (the local measurement process assumption). For context level $t + 1 = 3$, Eq. (7) becomes

$$P(d_{i-2}, d_{i-1}, d_i, d_{i+1}, d_{i+2}|q_{i-1}, q_i, q_{i+1})$$
$$= \alpha P(d_{i-1}, d_i, d_{i+1}|q_i)P(d_{i-2}, d_{i-1}, d_i|q_{i+1})P(d_i, d_{i+1}, d_{i+2}|q_{i+1}).$$

This generalized conditional independence assumption is not typically used and one may begin to think about its validity in practice. It is probably safe to state that for low level contexts ($t + 1 \leq 3$) the assumption is good. For higher level contexts ($t + 1 = 4$), it is probably beginning to be at slight variance with reality. And the larger the context is beyond ($t + 1 = 4$), the less we can expect the assumption to match reality. This implies that the relaxation is good for the first few iterations, after which the error of the assumptions begins to start dominating the result.

To see how (6) and (7) lead to (5), consider the conditional probability that unit i takes label q_i, given its level $t + 1$ context. By definition of conditional probability,

$$P(q_i|d_k\colon k \in Z_i(t+1)) = \frac{P(q_i, d_k\colon k \in Z_i(t+1))}{P(d_k\colon k \in Z_i(t+1))}$$
$$= \frac{\sum_{j \in N(i)} \sum_{q_j} P(q_i, q_k\colon k \in N(i), d_k\colon k \in Z_i(t+1))}{P(d_k\colon k \in Z_i(t+1))}$$
$$= \frac{\sum_{j \in N(i)} \sum_{q_j} P(d_k\colon k \in Z_i(t+1)|q_i, q_k\colon k \in N(i))P(q_i, q_k\colon k \in N(i))}{P(d_k\colon k \in Z_i(t+1))}. \tag{8}$$

Upon using (6) and (7), by substituting into (8) there results

$$P(q_i|d_k: k \in Z_i(t+1)) = \frac{\alpha P(d_k: k \in Z_i(t)|q_i)}{P(d_k: k \in Z_i(t+1))P(q_i)^{|N(i)|-1}} \times \sum_{j\in N(i)} \sum_{q_j} \prod_{n\in N(i)} [P(d_k: k \in Z_n(t)|q_n)P(q_i, q_n)]. \tag{9}$$

Again using the definition of conditional probability, we may rewrite (9) as

$$P(q_i|d_k: k \in Z_i(t+1)) = \left[\frac{\alpha P(d_k: k \in Z_i(t)) \prod_{n\in N(i)} P(d_k: k \in Z_n(t))}{P(d_k: k \in Z_i(t+1))}\right] P(q_i|d_k: k \in Z_i(t)) \times \sum_{j\in N(i)} \sum_{q_j} \prod_{n\in N(i)} \left[P(q_n|d_k: k \in Z_n(t)) \frac{P(q_i, q_n)}{P(q_i)P(q_n)}\right]. \tag{10}$$

The sums of products in (10) can be simplified. The products contain terms, each of which depends simply on n. All other variables involved are constant with respect to sums and product. Hence

$$\sum_{j\in N(i)} \sum_{q_j} \prod_{n\in N(i)} \left[P(q_n|d_k: k \in Z_n(t)) \frac{P(q_i, q_n)}{P(q_i)P(q_n)}\right] = \prod_{n\in N(i)} \sum_{q_n} \left[P(q_n|d_k: k \in Z_n(t) \frac{P(q_i, q_n)}{P(q_i)P(q_n)}\right]. \tag{11}$$

Finally, noticing that

$$\sum_{q_i} P(q_i|d_k: k \in Z_i(t+1)) = 1 \tag{12}$$

we can divide both sides of (10) by the sum in (12). The first term in square brackets on the right-hand side of (10) is a constant with respect to the summation and, therefore, cancels in the division. Thus, upon making the substitution of (11) and the division of (12), there results the relaxation equation

$$P(q_i|d_k: k \in Z_i(t+1)) = \frac{P(q_i|d_k: k \in Z_i(t)) \prod_{j\in N(i)} \sum_{q_j} P(q_j|d_k: k \in Z_j(t))(P(q_i, q_j)/P(q_i)P(q_j))}{\sum_{s_i} P(s_i|d_k: k \in Z_i(t)) \prod_{j\in N(i)} \sum_{q_j} P(q_j|d_k: k \in Z_j(t))(P(s_i, q_j)/P(s_i)P(q_j))}.$$

4. THE PELEG DERIVATION

Peleg [3] gives the following derivation as part of his derivation of relaxation equation (3). He lets P_i^k denote the probability distribution of the labels for unit i at the kth iteration. He gives the interpretation that P_i^k is the probability distribution of the labels for unit i given the $k-1$th iteration distributions for all units influencing unit i. If units 1 to N influence unit i, we can explicitly write this interpretation as

$$P\left(q_i | P_1^k, \ldots, P_N^k\right) = p_i^{k+1}(q_i) \tag{13}$$

and it is with this identification that there is a fundamental problem.

Consider the situation in which only unit j influences unit i. Peleg makes the conditional independence assumption

$$P\left(P_i^k, P_j^k | q_i, q_j\right) = P\left(P_i^k | q_i\right) P\left(P_j^k | q_j\right)$$

for each k, which is a specialization of the same kind of conditional independence assumption as in (7). From this it follows that

$$P\left(q_i | P_i^k, P_j^k\right) = \frac{P\left(q_i | P_i^k\right) \sum_{q_j} P\left(q_j | P_j^k\right) r_{ij}(q_i, q_j)}{\sum_{s_i} P\left(s_i | P_i^k\right) \sum_{t_j} P\left(t_j | P_j^k\right) r_{ij}(s_i, t_j)}.$$

Peleg then generalizes to having more than one unit influence unit i by taking an average of all the pairwise influences. This step does not correspond to any kind of probability assumption and does not result from any model. It is an unjustified heuristic and using it does not allow one to claim that the resulting probability distribution has any kind of meaningful interpretation. Nevertheless, upon using this averaging idea, Peleg writes

$$P\left(q_i | P_1^k, \ldots, P_N^k\right) = \sum_{j=1}^{N} c_j P\left(q_i | P_i^k, P_j^k\right) \tag{14}$$

where we understand that units 1 to N influence unit i and the c_j are nonnegative weights which sum to 1 and which may be different for each unit i since the set of units as well as their influences may be different for each different unit i.

To understand the difficulty in the Peleg interpretation, consider the identification which Peleg must make to have a relaxation equation. Suppose that before the $k+1$th iteration only unit i influences unit i, and at iteration $k+1$, units 1 to N influence unit i. Then, upon making the identifications required to write the equation

$$P^{k+1}(q_i) = \sum_{j=1}^{N} c_j \frac{P^k(q_i) \sum_{g_j} P^k(q_j) r_{ij}(q_i, q_j)}{\sum_{s_i} P^k(s_i) \sum_{t_j} P^k(t_j) r_{ij}(s_i, t_j)} \tag{15}$$

there is the immediate difficulty of

$$P\left(q_i|P_1^k,\ldots,P_N^k\right) = P^{k+1}(q_i) \tag{16}$$

$$P\left(q_i|P_j^k\right) = P^{k+1}(q_i) \tag{17}$$

when he really intended to make the identification

$$P\left(q_i|P_1^k,\ldots,P_N^k\right) = P^{k+1}(q_i) \tag{18}$$

$$P\left(q_i|P_j^k\right) = P^k(q_i). \tag{19}$$

The necessity of the identification in (16) arises because of (14) and (15). The necessity of the identification in (17) arises because of (13). However, the identifications he really wants are (18) and (19). Identification (18) is the same as (15). Identification (19) is needed so that the left-hand side of (15) can be plugged into the right-hand side of (15) for the next iteration. Notice that the difference between (17) and (19) is that (17) has the superscript $k + 1$ and (19) has the superscript k. This difference invalidates the interpretation Peleg gives to (15).

The cause of this difficulty is that the conceptual framework in which Peleg casts the interpretation is incorrect. What the relaxation equation does in each iteration is not to update estimates of probability distributions but to increase the conditioning context of the distributions used to start the relaxation.

5. CONCLUSION

We have shown that with the conditional independence assumptions of (6) and (7), relaxation equation (5) results. This equation states that the probability of a label given the $(t + 1)$-level context for any unit can be computed from the same kind of t-level probabilities of the unit and its neighbors. By iterating (5) until the entire context is taken into account, it becomes possible to compute the probability that a unit has a label given the entire context.

The relaxation iterations can be continued until the entire context has been taken into account or until a context level is reached where the conditional independence assumption (7) no longer holds.

The consequence of this explanation of cooperative processing and relaxation is that we now must begin to determine for each application the precise context level at which the assumption (7) no longer holds, then iterate to this level and stop. Determining the level at which to stop is a statistical question which we hope to answer in a future paper.

REFERENCES

1. R. M. Haralick, J. C. Mohammed, and S. W. Zucker, Compatibilities and the fixed points of arithmetic relaxation processes, *Computer Graphics and Image Processing* **13**, 1980, 242–256.
2. R. L. Kirby, A product rule relaxation method, *Computer Graphics and Image Processing* **13**, 1980, 158–189.
3. S. Peleg, A new probabilistic relaxation scheme, *IEEE Trans. Pattern Anal. Mach. Intelligence* **PAMI-2**, 1980, 362–369.
4. A. Rosenfeld, R. Hummel, and S. W. Zucker, Scene labeling by relaxation operations, *IEEE Trans. Syst. Man, Cybern.*, **SMC-6**, 1976, 420–433.
5. S. W. Zucker, E. V. Krishnamurthy, and R. L. Haar, Relaxation processes for scene labeling: Convergence, speed, and stability, *IEEE Trans. Syst. Man, Cybern.*, **SMC-8**, 1978, 41–48.
6. S. W. Zucker, and J. L. Mohammed, Analysis of Probabilistic Relaxation Labeling Processes, *Conference, Pattern Recognition and Image Processing*, Chicago, Ill. June 1978.