TR-562
MCS-76-23763

August 1977

RELAXATION PROCESSES FOR SCENE LABELING

Robert A. Hummel*
Azriel Rosenfeld
Computer Science Center
University of Maryland
College Park, MD 20742

ABSTRACT

A general treatment of "relaxation" processes for probabilistic classification of scene parts is presented. A geometrical framework for studying such processes is defined, in terms of finding paths in a space. General conditions that should be satisfied by path-finding procedures are formulated, and a variety of linear and nonlinear procedures that satisfy these conditions are described.

The support of the National Science Foundation under Grant MCS-76-23763 is gratefully acknowledged, as is the help of Mrs. Shelly Rowe in preparing this paper.

*Present address: Dept. of Mathematics, Univ. of Minnesota, Minneapolis, MN 55455

This paper presents a general treatment of iterative ("relaxation") processes for updating probabilities. A geometrical framework for the updating problem is defined; general conditions that should be satisfied by updating rules are formulated; and the rules treated in [1] are discussed from this standpoint.

1. Introduction

Let $A_1,\ldots,A_n$ be objects in a scene, each of which belongs to one of the classes $C_1,\ldots,C_m$. Suppose that we are given an initial set of probabilistic guesses as to the class assignments of the objects -- in other words, for all $i=1,\ldots,n$ and $\ell=1,\ldots,m$, we are given a number $p_i(\ell)$ which can be thought of as an estimate of the probability that object A_i is in class C_ℓ. We want to iteratively update these estimates until hopefully each A_i is nearly certain to belong to a single class C_ℓ.

In [1] an iterative procedure for updating the probability estimates was defined as follows: At the kth iteration, for every A_i and every C_ℓ we compute

$$q_i^{(k)}(\ell) = \sum_j \sum_{\ell'} r_{ij}(\ell,\ell')p_j^{(k)}(\ell') \tag{1}$$

Here the coefficients $r_{ij}(\ell,\ell')$ are fixed parameters representing the compatibilities between given pairs of class assignments for given pairs of objects. We then set

$$p_i^{(k+1)}(\ell) = \frac{p_i^{(k)}(\ell)(1+q_i^{(k)}(\ell))}{\sum_\ell p_i^{(k)}(\ell)(1+q_i^{(k)}(\ell))} \tag{2}$$

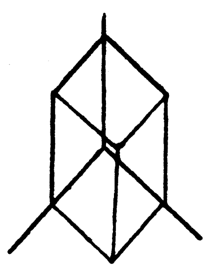
This updating rule has been used in a number of successful and partly successful experiments, particularly in connection with low-level scene features such as edges and curves [2]. Modifications to this rule have also been used in some cases.

2. Geometrical interpretation

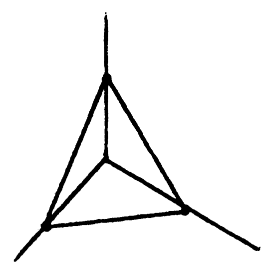
2.1 The assignment space

Let $v_{i\ell}$ be a variable representing the assignment of object A_i to class C_ℓ. The state vector for A_i is defined as $\bar{v}_i = (v_{i1},\ldots,v_{im})$, and the total state vector is the concatenation of these vectors: $\bar{v} = (\bar{v}_1,\ldots,\bar{v}_n)$. The space of all possible values for $\bar{v}$ will be called assignment space.

Various sets of values for the $v_{i\ell}$'s can be used. For example, if each $v_{i\ell}$ can assume only the value 0 or 1, we have the discrete model treated in [1]. Our interest here, however, is in models where the $v_{i\ell}$'s can assume continuous ranges of values -- specifically, values in the interval [0,1], representing our degree of belief in the assertion "A_i is in class C_ℓ". In the fuzzy model of [1], the class memberships are not mutually exclusive, so that any combination of $v_{i\ell}$'s in [0,1] is allowable; thus the assignment space is the unit hypercube $[0,1]^{mn}$. In the probabilistic model, the classes are exclusive, so that each state vector $\bar{v}_i$ must satisfy $\sum_{\ell=1}^{m} v_{i\ell}=1$ (as well as $0\le v_{i\ell}\le 1$ for each ℓ). Here the assignment space consists of n copies of (m-1)-dimensional probability space, which is a surface in m-dimensional space formed by joining the basis vectors (0,...,1,...,0). This is illustrated in Figure 1 for m=3.



Fuzzy assignment
space (unit cube)



Probabilistic assignment
space (triangular surface)

Figure 1. Fuzzy and probabilistic assignment spaces for m=3.

2.2 The tangent space

The goal of relaxation processes is to define and traverse a path through assignment space from the initial state vector toward a less ambiguous vector. It is therefore important to consider the possible directions that such a path can take at each point while remaining in assignment space. In differential geometry, the set of all such directions from a given point is called the tangent space at that point [3].

In the fuzzy case, at any interior point of the unit hypercube, paths in every direction are possible, so that the tangent space at all interior points is just mn-dimensional Euclidean space. In the probabilistic case, paths must remain on the probability surface; thus for any point p in the interior of that surface, an nm-dimensional vector q is in the tangent space at p iff. $\sum_{\ell} q_{i\ell}=0$, $1\leq i\leq n$. In either case, for points on the boundary of the assignment space (hypercube or probability surface), directions pointing out of the space are prohibited; only those pointing back into the space, or along the boundary, are allowed.

In the object assignment problem, any scheme which collects evidence for modifying the $p_i(\ell)$'s defines a direction (of change) for any given state vector $\bar{p}$. However, this direction is not necessarily a tangent vector, since it may point out of the space. We shall now consider how to transform an evidence vector indicating a desired direction of change at a point into a tangent vector at that point. This transformation, which will be referred to as the updating rule, can then be used to determine paths leading to improved assignments, by starting at the given point and moving from point to point along tangent vectors.

2.3 Updating rules

Let $\bar{p}$ be a current assignment vector, and let $\bar{q}$ be an evidence vector that defines a desired direction of updating. We cannot simply update $\bar{p}$ by adding $\bar{q}$ to it, since the vector $\bar{q}$ may not be a tangent vector. In general, we must use $\bar{q}$ to determine an updated assignment vector $\bar{p}'$ for which the following conditions are satisfied:

a) The components of $\bar{p}'-\bar{p}$ and $\bar{q}$ should have the same relative sizes. Ideally, if $q_{i\ell}$ is positive (negative), $p_{i\ell}$ should increase (decrease); or, if this is impossible because $\bar{p}'-\bar{p}$ would then not be a tangent vector, at least the components of $\bar{p}$ should increase or decrease according to the relative sizes of the components of $\bar{q}$. Usually, we would want the components of $\bar{p}'-\bar{p}$ to be smaller in magnitude than those of $\bar{q}$; by making these components small, we can help to insure that $\bar{p}'$ remains inside the assignment space, and to reduce errors.

b) If $\bar{q}=0$, $\bar{p}$ should not change. If $\bar{q}\neq 0$, $\bar{p}$ should change unless $\bar{q}$ is perpendicular to all the tangent vectors at $\bar{p}$ (or unless $\bar{p}$ lies in the boundary of the assignment space, and $\bar{q}$ points away from the space and is perpendicular to all tangent vectors at $\bar{p}$ that lie in the boundary face).

We also require that $\bar{p}'$ be in the assignment space and lie close to $\bar{p}$. Given $\bar{q}$, there are many different ways to define

$\bar{p}'-\bar{p}$ so that these conditions are satisfied. Some examples of such updating rules are given in the Appendix.

In the assignment spaces considered here, the surface is "flat", so that $\bar{p}'-\bar{p}$ is a tangent vector at $\bar{p}$ providing $\bar{p}'$ is in the assignment space. In general, $\bar{p}'-\bar{p}$ is approximately a tangent vector if all its components are small.

By (a-b), $\bar{p}'-\bar{p}$ is approximately the projection of $\bar{q}$ on the tangent space at $\bar{p}$. Thus for a given $\bar{q}$, updating rules that satisfy (a-b) will all tend to behave similarly; the exact choice of a rule may be relatively unimportant.

3. Iterative probability estimation

3.1 Conditional assignment probabilities

Up to now we have not discussed the goal of the updating process; we were concerned with how an "evidence vector" $\bar{q}$ could be used to define an updated vector $\bar{p}'$ lying in the assignment space, but we did not consider in what sense $\bar{p}'$ might be a better assignment vector than $\bar{p}$ was. In this section we present one approach to defining a goal, in the probabilistic case.

Suppose that the given scene is a member of an ensemble of scenes, each containing n objects $A_1,\ldots,A_n$ that can belong to the m classes $C_1,\ldots,C_m$. Let $E_{i\ell}$ be the event that object A_i is in class C_ℓ. We can define the probability of $E_{i\ell}$, either a priori or conditioned on some other event(s). For example, let $\bar{M}_i$ be a vector of measurements made on A_i, $1 \le i \le n$; then we can consider the conditional probabilities $P(E_{i\ell}|\bar{M}_i)$, $1 \le \ell \le m$. The initial probability estimates referred to in Section 1 would ordinarily be obtained in this way. These estimates should be relatively easy to make, since they are made for each A_i independently of the others, based on $\bar{M}_i$ only.

Alternatively, we can consider the more informed set of conditional probabilities $P(E_{i\ell}|\bar{M})$, where $\bar{M} = (\bar{M}_1,\ldots,\bar{M}_n)$; here the probability of each $E_{i\ell}$ is conditioned on the full set of measurements made on all the A's, rather than just on the measurements $\bar{M}_i$ made on A_i. Ordinarily, these probabilities are hard to compute, since it is difficult to describe the dependency of $E_{i\ell}$ on the entire collection of measurements.

Let $\bar{p}^o$ be the assignment vector whose (i,ℓ) component is $P(E_{i\ell}|\bar{M}_i)$, and let $\bar{p}$ be the vector whose (i,ℓ) component is $P(E_{i\ell}|\bar{M})$. We shall assume that $\bar{p}^o$ is known (or has been estimated), and that we want to estimate $\bar{p}$. In the next subsection we discuss the problem of estimating $\bar{p}$, given $\bar{p}^o$, in terms of the geometrical formalism introduced in Section 2.

3.2 Linear approximation

Since we do not know how to compute $\bar{p}$ from the set of measurements, we use the components (or a subset of the components) of the initial assignment $\bar{p}^o$ to estimate $p_{i\ell}$ for each (i,ℓ). Since we do not know the exact functional relationship which finds $\bar{p}$ given $\bar{p}^o$, we may use a linear approximation for $p_{i\ell}-p^o_{i\ell}$ in terms of the initial components of $\bar{p}^o$, given by

$$\bar{\sigma}^o_{i\ell} \equiv [\bar{p}-\bar{p}^o]_{i\ell} \doteq \sum_{j,\ell'} s_{ij}(\ell,\ell')p^o_{j\ell'} \ .$$

Since this is only an approximation, it is not in general a tangent vector, and we must use an updating rule to change $\bar{p}^o$ in the $\bar{\sigma}^o$ direction, as discussed in Section 2.3(a). Furthermore, we change $\bar{p}^o$ by only a small increment, since $\bar{\sigma}^o$ points toward $\bar{p}$, but $\bar{p}^o + \bar{\sigma}^o$ may be very far from $\bar{p}$. This yields a new vector $\bar{p}^1$, and we can now repeat the process, adding a small increment in the direction $\bar{\sigma}^1$ where

$\bar{\sigma}^1_{i\ell} = \sum_{j,\ell'} s_{ij}(\ell,\ell')p^1_{j\ell'}$. In other words, we define an

iterative process for which

$$[\bar{p}^{k+1} - \bar{p}^k]_{i\ell} \doteq \alpha \sum_{j,\ell'} s_{ij}(\ell,\ell')p^k_{j\ell'} \ ,$$

where the scalar α determines the size of the increment, and the approximation is determined by the updating rule which transforms the summation quantity into a tangent vector. This updating process is essentially the same as the linear process considered in Section IV of [1]. For a suitable choice of α, a few iterations of this process should ordinarily yield an assignment close, or at least closer, to the desired $\bar{p}$.

Note that $\vec{\sigma}^k$ is recomputed at each iteration using the same linear approximation, based on the assignment components of that iteration. Each iteration estimates the vector $\bar{p}$ independently of the previous assignments, as though the measurements had led to an initial assignment equal to the current iteration. In this way, the linear approximation must estimate the difference between $\bar{p}$ and the current assignment $\bar{p}^k$ for each iteration. If the classification of objects conditioned by all the measurements gives a $\bar{p}$ which assigns a single class to each object with near certainty, then the estimate of $\bar{p}$ should not change greatly as the iterates move successively closer to $\bar{p}$, since $\bar{p}$ is in a "corner" of the assignment space. Under these conditions, the process should converge to $\bar{p}$.

The coefficients $s_{ij}(\ell,\ell')$ in the linear approximation describe how $E_{i\ell}$ depends on the evidence for $E_{j\ell'}$ that is currently available. This probability in turn reflects the initial measurements made at the object A_j. In other words, $s_{ij}(\ell,\ell')$ is related to the correlation between $E_{i\ell}$ and $E_{j\ell'}$; this is the basis for calling the s_{ij}'s "correlation coefficients" in [1].

If one knew the exact dependence of $\bar{p}$ on the initial assignment $\bar{p}^o$, then the $s_{ij}(\ell,\ell')$ could be computed from the best linear approximation of this assignment-valued function. On the other hand, since these coefficients involve only pairs of (object, class) pairs, they should be relatively easy to estimate, as compared with specifying the dependence of $E_{i\ell}$ on all the (object, class) assignments at once.

3.3 Updating in log probability space

For some purposes it may be advantageous to use an assignment space based on logarithms of probabilities, rather than on the probabilities themselves. The ordinary probability assignment space is bounded by faces at which at least one of the probability components is zero. If a vector is near one of these faces, and we update it, we must be careful to make the component of displacement toward the face very small, to insure that we remain within the space. Thus the dynamic range of the updating rule near the boundaries of the assignment space is severely limited. In log probability space, on the other hand, zero maps into $-\infty$, so that the faces of the space recede to infinity, and updating becomes a less critical process.

If we define the log of a probability vector componentwise, the updating process in log probability space has the form $\log \bar{p}' - \log \bar{p}$. We can define a linear approximation to the log updating vector $\bar{\tau}^{o}$ in the same manner as in Section 3.2:

$$\bar{\tau}^{o}_{i\ell} \equiv [\log \bar{p} - \log \bar{p}^{o}]_{i\ell} \doteq \beta \sum_{j,\ell'} t_{ij}(\ell,\ell') p^{o}_{j\ell'}$$

where the scalar β defines the size of the increment in the $\bar{\tau}^{o}$ direction. By the remarks in the preceding paragraph, this approximation should give a more uniform estimation of $\bar{p}$ for the various $\bar{p}^{(k)}$'s. The updating scheme resulting from this linear approximation rule is equivalent to

$$p_{i\ell} = p^{o}_{ij}\, e^{\beta \Sigma t_{ij}(\ell,\ell') p^{o}_{j\ell'}}$$

$$\doteq p^{o}_{ij}\Big(1+\sum_{j,\ell'} t_{ij}(\ell,\ell') p^{o}_{j\ell'}\Big)^{\beta}$$

where we have further approximated the exponential by the first two terms of its expansion. For $\beta=1$, this is just the nonlinear updating rule defined in [1], provided that we normalize it (by dividing by the sum, over ℓ, of the right hand sides to insure that $\sum_{\ell} p_{i\ell}=1$). For $\beta>1$ we obtain the "accelerated" nonlinear updating rules discussed in [4].

3.4 Higher-order approximations

Nonlinear approximations to $\bar{p}-\bar{p}^o$ or to $\log \bar{p}-\log \bar{p}^o$ can also be used to define updating rules. For example, using a Taylor expansion gives, to a first approximation,

$$[\log \bar{p}-\log \bar{p}^o]_{i\ell} \doteq \sum_{\ell_1,\ldots,\ell_n=1}^{m} c_{i\ell,\ell_1,\ldots,\ell_n} p_{1\ell_1} \cdots p_{n\ell_n}$$

In this expression, the c's express the dependencies between $E_{i\ell}$ and the joint occurrences of the sets of events $E_{1\ell_1},\ldots,E_{n\ell_n}$; this is more general than using the dependencies between $E_{i\ell}$ and the events $E_{j\ell'}$, taken one at a time, as we did in Section 3.2.

The disadvantage of this method of approximating $\bar{p}-\bar{p}^o$ is that it will be much harder to define the dependency of each $E_{i\ell}$ on every set of events $E_{1\ell_1},\ldots,E_{n\ell_n}$ than it was to define its dependency on single events. One way to reduce this problem is to limit the permitted combinations of events; but even if we restrict ourselves to combinations k at a time, we still need to determine $O(m^k)$ dependencies, while in Section 3.2 we needed to know only $O(m^2)$ of them.

Another possibility is to search for a locally supporting configuration of p's corresponding to a large positive coefficient $c_{i\ell,\ell_1,\ldots,\ell_n}$. The search can be guided by large assignment values $p_{j\ell'}$, as well as by models of compatible configurations. In the absence of a supportive configuration, it can be assumed that either the evidence for updating is ambiguous, or that the evidence indicates that $p_{i\ell}$ should be decreased. The latter case occurs when the local configuration supports some other class at object A_i, and will be automatically incorporated into the updating scheme by the normalization of the probabilities. This represents another way in which search techniques may be used in conjunction with iterative probability updating processes; compare [5, 6].

4. Concluding remarks

In this paper we have introduced a convenient geometrical representation for problems involving class assignment vectors -- discrete, fuzzy, or probabilistic. We have formulated conditions that must be satisfied by any process that updates assignment vectors, and have shown how a variety of updating rules can be formulated that satisfy these conditions.

For probabilistic assignments, we have introduced the notion of transforming locally-conditioned into globally-conditioned probabilities, and shown how this can be regarded as the goal of a probability updating process. Finally, we have defined a set of approximate updating processes (linear, log-linear, and polynomial), and discussed their interpretation and implementability.

The concepts introduced in this paper provide a convenient framework for discussing assignment updating processes, and have led to the formulation of several interesting updating rules. However, many basic issues have been left unsettled, notably just how the initial probability estimates, and the coefficients in the approximate updating rules, are to be determined in practice. Nevertheless, the framework provided in this paper should facilitate the further study of assignment updating processes in a variety of problem domains.

Appendix: Examples of updating rules

The following are some examples of updating rules that satisfy the conditions discussed in the Section 2.3. They serve to illustrate the variety of possibilities; it would be of interest to test them in actual applications.

In the fuzzy case, one simple possibility is to set $\bar{p}'-\bar{p} = h\bar{q}/|\bar{q}|$ (where h is small), where $\bar{p}'$ is truncated if it lies outside the unit cube (i.e., any component <0 or >1 is set to 0 or 1). Alternatively, we can let $\bar{p}'$ be defined componentwise using any function $g(p_{i\ell}, q_{i\ell})$ such that $g(p,q)$ is monotonic in q and $g(p,0) = p$, while $g(p,q) \to 1$ as $q \to \infty$, and $\to 0$ as $q \to -\infty$; there are many smooth functions having these properties. The min-max fuzzy updating rule given in [1] provides still another possibility, since it is guaranteed to yield values in the unit cube.

Probabilistic updating rules must satisfy stronger restrictions, $\bar{p}'$ must lie on the probability surface. The rule used in [1] (see (1-2) in Section 1) satisfies our conditions, provided that the $q_{i\ell}$ are greater than -1, to assure that the $p'_{i\ell}$ are nonnegative. More generally, one might use the rule

$$p'_{i\ell} = \frac{p_{i\ell}e^{q_{i\ell}}}{\sum_{\ell} p_{i\ell}e^{q_{i\ell}}}$$

in which nonnegativity is guaranteed. For another viewpoint about such exponential rules see Section 3.3. Another possibility is to find the tangent vector $\bar{v}$ at $\bar{p}$ that is nearest to $\bar{q}$; this can be done as follows:

i) If $p_{i\ell} = 0$ and $q_{i\ell} < 0$, set $v_{i\ell} = 0$

ii) Let k_i be the number of ℓ's for which (i) holds for each i; let $c_i = \sum_{\ell} q_{i\ell}/(n-k_i)$

iii) For each ℓ such that (i) does not hold, set $v_{i\ell} = q_{i\ell} - c_i$

We can then update $\bar{p}$ by moving in direction $\bar{v}$, but adjusting the size of the step taken in each direction to assure that the resulting $\bar{p}'$ remains on the probability surface.

References

1. Rosenfeld, A., R. A. Hummel, S. W. Zucker, Scene labeling by relaxation operations, IEEE Trans. on Systems, Man, and Cybernetics, SMC-6, 1976, pp. 420-433.

2. Zucker, S. W., R. A. Hummel, A. Rosenfeld, An application of relaxation labeling to line and curve enhancement, IEEE Trans. on Computers C-26, 1977, pp. 394-403.

3. O'Neill, B., Elementary Differential Geometry, Academic Press, 1966.

4. Zucker, S. W., E. Krishnamurthy, R. Haar, Relaxation processes for scene labeling: convergence, speed, and stability. University of Maryland Computer Science Center TR-477, 1976.

5. Tenenbaum, J. M. and H. G. Barrow, Experiments in interpretation-guided segmentation, Artificial Intelligence 8, 1977, p. 241-274.

6. Tenenbaum, J. M. and H. G. Barrow, MSYS: A system for reasoning about scenes, Stanford Research Institute project 1187, Annual Report to the Office of Naval Research, 1975.