Parallel, Finite-Convergence Learning Algorithms for Relaxation Labeling Processes

Xinhua Zhuang

Dept. of Computer Engineering & Computer Science
Univ. of Missouri, Columbia, MO 65211

Yunxin Zhao

Beckman Institute and ECE Department
Univ. of Illinois, Urbana, IL 61801

Abstract

This paper is theoretical. We present sufficient and "almost" necessary conditions for learning compatibility coefficients in relaxation labeling whose satisfaction will guarantee each desired sample labeling to become consistent and each ambiguous or erroneous input sample labeling to be attracted to the corresponding desired sample labeling. The derived learning conditions are parallel and local information based. In fact, they are organized as linear inequalities in unit wise and thus the perceptron like algorithms can be used to solve them efficiently with finite convergence.

1 Introduction

Relaxation labeling processes represent a class of mechanisms originally developed to resolve ambiguities in vision systems, and to correct errors arising in low level image processing. Since its launch [1], the relaxation labeling has found far broader applications and many algorithms have been proposed. Conceptually, those algorithms are parallel offering an appealing qualitative match with the distributed appearance of neural machinery in the early human vision system.

It has been discovered over the years that relaxation labeling provides a philosophically different robust method for *Computer Vision* by its cooperative use of context. As is well known, conventional robust statistical techniques assume sample independence.

Recently, a theoretical connection between relaxation labeling and associative memory models has been explored [2] that reveals relaxation labeling models also represent a biological model for memory, more general and powerful than Hopfield associative memory [3] and the like. By analogy to a healthy human memory, both recalling and learning processes in relaxation labeling models must be made of finite convergence. This justifies the simplex-like updating or recalling rule [4], which caused skepticism when proposed because of its finite convergence. This further motivates development of learning algorithms of finite convergence for relaxation labeling processes.

This work was supported by NSF Grant IRI-9112649 and IRI-9502074, and by NASA Grant NAG 5-2579 and NAGW-4698.

Over decades, a number of studies have been conducted in an attempt to understand compatibilities and to develop methods to derive them [1, 5–13]. Some already addressed the difficult issue of learning compatibility coefficients in terms of given sample labelings [10, 13]. Nevertheless, the proposed learning algorithms were not local information based, nor of finite convergence. In the paper, we develop new sample learning algorithms which are parallel, use only local information, and offer a finite convergence. In brief, they learn the condition which not only ensures consistency of each sample labeling, but also enforces an ambiguous or erroneous labeling being recalled or attracted to a desired sample labeling. Using only local information for learning provides further important evidence for the relaxation labeling processes to have biological plausibility.

The paper is theoretical and organized as follows. In Section 2, we briefly review the *RHZ* relaxation labeling formulation [1, 14] and the simplex-like labeling updating rule [4]. In Section 3, we slightly modify the simplex-like labeling updating rule from simultaneous updatings to sequential updatings, and develop the *Attraction Theorem* which establishes the conditions under which the ambiguity reduction or error correction on an input labeling can be accomplished in a single recalling cycle. In the final section, we present the *Sample Learning Algorithms* which are parallel and local information based providing the demanded finite convergence.

2 *RHZ* Relaxation Labeling Formulation

A consistent labeling problem has many units, each of which has an unknown label. There are n units, $U_1, \ldots, U_n$, and m labels, $L_1, \ldots, L_m$. Each U_i will be assigned a labeling distribution $p(i) = (p_1(i), \ldots, p_m(i))^T$ which is restricted to lie on a (m-1)-D simplex K_{m-1} in R_m, i.e., $p_j(i) \geq 0$ and $\sum_{j=1}^{m} p_j(i) = 1$. As is seen, the labeling distribution $p(i)$ actually ranks likelihood of unit U_i possessing different labels.

There are consistency constraints between label assignments. Let a real number $r(i,j;h,k)$ $(i \neq h)$ called by a compatibility coefficient represent how label L_k at unit U_h influences label L_j at unit U_i. If unit U_i having label L_j is highly compatible with unit U_h having label L_k, then the coefficient $r(i,j;h,k)$ should be large and positive. If the constraints are such that unit U_h having label L_k means that unit U_i having label L_j is highly unlikely, then $r(i,j;h,k)$ should be small. Thus, for fixed $i, h, i \neq h$, the $m \times m$ matrix $R_{ih} = (r(i,j;h,k))$ collectively represents the overall compatibility of unit U_i with unit U_h.

Denote the n labeling distributions as a whole by a labeling assignment matrix, i.e., an $m \times n$ matrix $p = [p(1), \ldots, p(n)]$. For brevity, we simply call p a labeling. Then, the compatibility of unit U_i having label L_j with the rest n-1 labeling distributions $p(h), h \neq i$, is measured by

$$q_j(i|p) = \sum_{h \neq i} \sum_{k=1}^{m} r(i,j;h,k) p_k(h). \qquad (1)$$

Let $q(i|p) = (q_1(i|p), \ldots, q_m(i|p))^T$. Then, the m-D compatibility vector $q(i|p)$ is linear and homogeneous in $R_{ih}, h \neq i$ and ranks the compatibilities of U_i having different labels with the rest n-1 labeling distributions $p(h), h \neq i$. The above formulation follows in spirit the Rosenfeld, Hummel, and Zucker(RHZ) relaxation labeling model [1].

In a fundamental work by Hummel and Zucker [14], a labeling distribution $p(i)$ is defined as consistent if it is most compatible with the rest n-1 labeling distributions $p(h), h \neq i$, i.e.,

$$p(i) = \arg max_{v(i) \in K_{m-1}} (q(i|p), v(i)). \qquad (2)$$

In other words, a consistent labeling distribution $p(i)$ attains the maximum correlation with the compatibility vector $q(i|p)$. A labeling p is defined as consistent if each individual labeling distribution $p(i)$ is consistent.

Let $M(i|p)$ represent a set of labeling distributions at unit U_i each of which is most compatible with the rest n-1 labeling distributions $p(h), h \neq i$, i.e.,

$$\begin{aligned} M(i|p) &= \{v(i) : (q(i|p), v(i)) \\ &= \max_{v'(i) \in K_{m-1}} (q(i|p), v'(i))\}. \end{aligned} \qquad (3)$$

Then, we showed the maximum labeling distribution set $M(i|p)$ is a linear convex set in K_{m-1} and can be calculated by $M(i|p) = \{v(i) : (q(i|p), v(i)) = max_j\ q_j(i|p)\}$. Using the maximum labeling distribution set, the consistency condition for $p(i)$ can be described as $p(i) \in M(i|p)$. The definition of consistency by Hummel and Zucker [14] is somehow different from the original Rosenfeld et al.'s [1]. We characterized that p is consistent in Hummel and Zucker's sense if, and only if, $\forall p_j(i) \neq 0$ $q_j(i|p) = max_k\ q_k(i|p)$ and that p is consistent in Rosenfeld et al.'s if, and only if, $\forall p_j(i) \neq 0$ the corresponding $q_j(i|p)$'s are equal. Based upon the two characterizations, it was shown that a consistent labeling in Hummel and Zucker's sense is a consistent labeling in Rosenfeld et al.'s. But the reverse may often not be true [4]. For instance, every unambiguous labeling is consistent in Rosenfeld et al.'s sense by the second characterization, a truth that can be directly verified by noticing an unambiguous labeling is a fixed point of Rosenfeld et al.'s labeling updating rule. But unambiguity does not warrant consistency in Hummel and Zucker's. Another interesting phenomenon with Rosenfeld et al.'s relaxation labeling is anytime when a labeling distribution component $p_j(i)$ becomes silent, i.e., $p_j(i) = 0$, it will never speak out again, a fact that can be easily checked by using Rosenfeld et al.'s updating rule. This indicates that a true desired labeling will not be perfectly recovered from an ambiguous or erroneous labeling through Rosenfeld et al.'s relaxation labeling if the latter has one labeling distribution component which is improperly set to silence. It appears that Hummel and Zucker's relaxation labeling represents significant progress, even though the new one admits no algebraic updating rules as neat as Rosenfeld et al.'s.

A labeling distribution $p(i)$ is defined as strictly consistent if $p(i)$ is a unique point in K_{m-1} which is most compatible with $q(i|p)$. A strictly consistent labeling distribution must be a vertex of K_{m-1} and hence unambiguous [14]. But the reverse may not be true. Similarly, labeling p is defined as strictly consistent if each individual labeling distribution $p(i)$ is strictly consistent.

Let p^s be the current labeling which has not been consistent yet. Then the relaxation labeling processes shall choose the next labeling p^{s+1} to enhance labeling consistency. Let $M(i|p^s)$ represent the maximum labeling distribution set decided by labeling p^s for unit U_i. We argued [4] that the following dynamic system improves the labeling consistency:

$$\forall i, p^{s+1}(i) = \arg min_{v(i)\in M(i|p^s)} \parallel v(i) - p^s(i) \parallel . \quad (4)$$

As is seen, $p^{s+1}(i)$ represents an orthogonal projection of $p^s(i)$ onto $M(i|p^s)$ which is uniquely determined. We proved a consistent labeling is a fixed point of the dynamic system and vice versa, and showed that a convergent labeling updating sequence $\{p^s\}$ provides a consistent labeling at the limit and a strictly consistent labeling is an attractor. Finally, we substantiated the dynamic system by the simplex-like labeling updating rule which is parallel and of finite convergence.

For the reader's convenience, we describe the simplex-like labeling updating rule as follows.

The Simplex-Like Labeling Updating Rule:

Since $M(i|p)$ is a linear convex set, it can be easily constructed through its vertices which are each a standard unit basis vector in R_m. Let $I(i|p)$ denote the set of indices of aforementioned basis vectors. Then

$$I(i|p) = \{k : \; q_k(i|p) = \max_j q_j(i|p)\}. \quad (5)$$

If p represents an arbitrary input labeling of the dynamic system (4) and p' the corresponding output labeling, then p' can be calculated as follows:

$$p'_j(i) = \begin{cases} p_j(i) + \sum_{k\notin I(i|p)} p_k(i)/\#I(i|p), & \text{if} j \in I(i|p) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

for $j = 1, \cdots, m$ and $i = 1, \cdots, n$.

3 Quantitative Attraction Theorem

The Modified Simplex-Like Labeling Updating Rule:

In the modified updating rule, an input labeling p experiences a series of updatings during a so-called recalling cycle to produce an output labeling p', i.e.,

$$p = p^{(0)} \rightarrow p^{(1)} \rightarrow \cdots \rightarrow p^{(n)} = p'. \quad (7)$$

As is seen, the sequential updatings are used in (7) to replace the simultaneous updatings in (4).

The updating from $p^{(k-1)}$ to $p^{(k)}$ can be formally described as follows.

$$\begin{cases} p^{(k)}(i) = p^{(k-1)}(i), 1 \le i < k, \\ p^{(k)}(k) = \arg min_{v(k)\in M(k|p^{(k-1)})} \|v(k) - p(k)\|, \\ p^{(k)}(i) = p(i), \qquad k < i \le n. \end{cases} \quad (8)$$

The formula (6) can be used to substantiate the operation argmin in (8). Since the output labeling p' is defined by $p^{(n)}$, it is apparent that $p'(k) = p^{(k)}(k)$, $k = 1, 2, ..., n$.

The updating rule by (7)-(8) defines a modified dynamic system. It can be easily shown that a consistent labeling is a fixed point of this modified dynamic system, i.e., $p = p'$, and vice versa. Moreover, the condition for an input labeling p to be attracted to a desired labeling $\tilde{p}$ in a single recalling cycle is identified as $p' = \tilde{p}$ and can be derived in detail as

$$p^{(k)}(k) = \tilde{p}(k), \; k = 1, 2, ..., n \quad (9)$$

due to $p'(k) = p^{(k)}(k)$, $k = 1, 2, ..., n$. This attraction condition is significant in that it clarifies the condition which is needed in learning compatibility coefficients.

Let $p_c^{(k)}$ denote a labeling which is a genetic "crossover" of the input labeling p by the desired labeling $\tilde{p}$, i.e.,

$$\begin{cases} p_c^{(k)}(i) = \tilde{p}(i), & 1 \le i \le k \\ p_c^{(k)}(i) = p(i), & k < i \le n \end{cases} \quad (10)$$

where $0 \le k \le n$. As is easily verified, $p_c^{(0)} = p$ and $p_c^{(n)} = \tilde{p}$. In terms of (10) and those $n+1$ crossovers, the attraction condition (9) can be transformed into the following equivalent condition:

$$p^{(k)}(k) = p_c^{(k)}(k), \quad k = 1, 2, ..., n. \quad (11)$$

Theorem (*Attraction*) The input labeling p will be attracted to the desired labeling $\tilde{p}$ in one recalling cycle if the following n maximum conditions hold

$$\begin{cases} (q(k|p_c^{(k-1)}), \tilde{p}(k)) \ge q_j(k|p_c^{(k-1)}), \\ j = 1, 2, ..., n, \; if \; \tilde{p}(k) = p(k) \\ \tilde{p}(k) = \text{ some } e_j \text{ and} \\ \quad q_j(k|p_c^{(k-1)}) > q_{j'}(k|p_c^{(k-1)}), \\ \forall \; j' \ne j, \; if \; \tilde{p}(k) \ne p(k), \\ k = 1, 2, ..., n. \end{cases} \quad (12)$$

Remark: It can be easily verified that those maximum conditions are also necessary to warrant attraction.

4 Sample Learning Algorithm

Assume a set of observation sequences O^ν, $\nu = 1, 2, ...V$, of the problem we intend to solve are available, where $O^\nu = (o_1^\nu, ..., o_n^\nu)$ and o_k^ν represents an observation at unit U_k. Moreover, assume for each observation sequence O^ν there is a valid unambiguous labeling $\tilde{p}^\nu$ being correctly identified prior to training. Since the relaxation labeling processes are developed to dissect ambiguities and rectify errors caused by some well-defined procedures used in preprocessing, it is thus important to include those flawed labelings during training as well. Suppose for each observation sequence O^ν there is a so-produced ambiguous or erroneous labeling p^ν. Our task is to learn the compatibility coefficients so that each valid sample labeling $\tilde{p}^\nu$ will be consistent and each erroneous sample labeling p^ν will be driven to the corresponding true sample labeling $\tilde{p}^\nu$ in one recalling cycle by the modified dynamic system.

Let $\tilde{p}^\nu(i)$, which is assumed unambiguous, be represented by a unit basis vector $e_{i(\nu)}$. Then the consistency for $\tilde{p}^\nu$ and the attraction for $(p^\nu, \tilde{p}^\nu)$ require that at each unit U_i there hold

$$q_{i(\nu)}(i|\tilde{p}^\nu) \geq q_j(i|\tilde{p}^\nu), \quad \forall j \neq i(\nu) \tag{13}$$

and

$$\begin{cases} q_{i(\nu)}(i|p_c^{\nu(i-1)}) \geq q_j(i|p_c^{\nu(i-1)}), \\ \qquad \forall j \neq i(\nu), \text{ if } \tilde{p}^\nu(i) = p^\nu(i), \\ q_{i(\nu)}(i|p_c^{\nu(i-1)}) > q_j(i|p_c^{\nu(i-1)}), \\ \qquad \forall j \neq i(\nu), \text{ if } \tilde{p}^\nu(i) \neq p^\nu(i). \end{cases} \tag{14}$$

Theorem (*Sample Learning*) At each unit U_i, if the compatibility matrices $R_{ih}, h \neq i$, can be learned so that the so-called quantitative attraction condition is satisfied, i.e.,

$$\begin{cases} q_{i(\nu)}(i|\tilde{p}^\nu) > q_j(i|\tilde{p}^\nu), \\ q_{i(\nu)}(i|p_c^{\nu(i-1)}) > q_j(i|p_c^{\nu(i-1)}), \\ \nu = 1, 2, \cdots, V \text{ and } j \neq i(\nu), \end{cases} \tag{15}$$

then those unambiguous sample labelings $\tilde{p}^\nu$ will each be consistent, and each ambiguous or erroneous labeling p^ν will be attracted to the corresponding true sample labeling $\tilde{p}^\nu$.

Remark: The theorem presents sufficient and "almost" necessary conditions of learning compatibility coefficients in relaxation labeling for multiple sample labelings' consistency/attraction and lets the whole learning processes be decomposed into n independent, parallel learning processes, one at a unit. Independence of learning at U_i from learning at U_h, $h \neq i$, indicates that they are local-information-based. There are a total of $2V \times (m-1)$ linear inequalities involved in (15). The general solution of (15) forms a cone of high dimension which can be obtained by using the *TZW* algorithm [15]. A specific solution can be obtained by using Rosenblatt's perception rule or Ho-Kashyap algorithm [16]. All three algorithms are of finite convergence. Existence of a cone for the general solution explains relative-immunity of relaxation labeling operations to a large variation in compatibility coefficients and solves the mystery why some heuristic methods provide workable compatibility coefficients in certain applications.

References

[1] A. Rosenfeld, R. A. Hummel, and S. W. Zucker, "Scene labeling by relaxation operations," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, no. 6, pp. 420–433, 1976.

[2] X. Zhuang and Y. Zhao, "A unification of Relaxation Labeling and Associative Memory," *Proceedings of 1996 IEEE Int. Conf. on Acoust., Speech, & Signal Processing* , vol. 6, pp. 3406–3409.

[3] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities." *Proc. Nat. Acad. Sci. USA*, vol. 79, pp. 2554–2558, 1982.

[4] X. Zhuang, R. M. Haralick, and H. Joo, "A simplex-like algorithm for the relaxation labeling process," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 12, pp. 1316–1321, 1989.

[5] S. Peleg and A. Rosenfeld, "Determining compatibility coefficients for curve enhancement relaxation processes," *IEEE Trans. Syst., Man. Cybern.*, vol. SMC-8, no. 7, pp. 548–555, 1978.

[6] H. Yamamoto, "A method of deriving compatibility coefficients for relaxation operators," *Comput. Graph. Image Processing*, vol. 10, pp. 256–271, 1979.

[7] S. Peleg, "A new probabilistic relaxation scheme," *IEEE Trans. Patten Anal. Machine Intell.*, vol. PAMI-2, no. 4, pp. 362–369, 1980.

[8] L. S. Davis and A. Rosenfeld, "Cooperating processed for low-level vision: A survey," *Artificial Intell.*, vol. 17, pp. 245–263, 1981.

[9] J. Kittler, "Compatibility and support functions in probabilistic relaxation," in *Proc. 8th Int. Conf. Pattern Recognit.*, Paris, France, 1986. pp. 186–189.

[10] R. A. Hummel, "A design method for relaxation labeling applications," in *Proc. Nat. Conf. Artificial Intell.(AAAI-83)*, Washington, DC, 1983, pp. 168–171.

[11] S. W. Zucker, R. A. Hummel, and A. Rosenfeld, "An application of relaxation labeling to line and curve enhancement," *IEEE Trans. Comput.*, vol. C-26, no. 4, pp. 394–403, 1977.

[12] L. S. Davis and A. Rosenfeld, "Curve segmentation by relaxation labeling," *IEEE Trans. Comput.*, vol. C-26, no. 10, pp. 1053–1057, 1977.

[13] M. Pelillo and M. Refice,"Learning Compatibility Coefficients for Relaxation Labeling Processes,"*IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 9, pp. 933–945, 1994.

[14] R. A. Hummel and S. W. Zucker, "On the foundations of relaxation labeling processes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, no. 3, pp. 267–287, 1983.

[15] Y. Takada, X. Zhuang, and H. J. Wakita, "A Geometric Algorithm Finding Set of Linear Decision Boundaries," *IEEE Trans. Signal Processing*, vol. 42, no. 7, 1994.

[16] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*, Reading, MA: Addison-Wesley, 1981.