An Energy Function and Continuous Edit Process for Graph Matching

Andrew M. Finch
Richard C. Wilson
Edwin R. Hancock

Department of Computer Science, University of York, York, YO1 5DD, U.K.

The contributions of this article are twofold. First, we develop a new nonquadratic energy function for graph matching. The starting point is a recently reported mixture model that gauges relational consistency using a series of exponential functions of the Hamming distances between graph neighborhoods. We compute the effective neighborhood potentials associated with the mixture model by identifying the single probability function of zero Kullback divergence. This new energy function is simply a weighted sum of graph Hamming distances. The second contribution is to locate matches by graduated assignment. Rather than solving the mean-field saddle-point equations, which are intractable for our nonquadratic energy function, we apply the soft-assign ansatz to the derivatives of our energy function. Here we introduce a novel departure from the standard graduated assignment formulation of graph matching by allowing the connection strengths of the data graph to update themselves. The aim is to provide a means by which the structure of the data graph can be updated so as to rectify structural errors. The method is evaluated experimentally and is shown to outperform its quadratic counterpart.

1 Introduction

Graph matching is a topic of pivotal importance (Mjolsness, Gindi, & Anandan, 1989; Gold & Rangarajan, 1996; Gold, Rangarajan, & Mjolsness, 1996; Sanfeliu & Fu, 1983; Cross & Hancock, 1995, 1997; Wilson & Hancock, 1995; Wilson, Evans, & Hancock, 1995) in pattern interpretation. The problem invariably arises whenever a pattern recognition task is given a relational abstraction. It was the classical work on structural pattern recognition of the 1980s that first identified the main technical difficulties associated with relational graph matching (Sanfeliu & Fu, 1983; Shapiro & Haralick, 1981). Although providing a conceptually attractive structural abstraction, relational graphs are notoriously fragile to noise and segmentation error. For this reason practical graph matching must be realized by inexact means.

Effective inexact graph matching is critically dependent on the availability of two computational ingredients. The first of these is an accurate
means of computing the distance between structurally corrupted relational descriptions (Shapiro & Haralick, 1985; Sanfeliu & Fu, 1983). Here several measures have been investigated, including the idea of defining an edit distance between graphs (Shapiro & Haralick, 1985) and the use of inter-graph entropy (Wong & You, 1985). The second ingredient is a means of controlling structural errors introduced by noise or segmentation error (Shapiro & Haralick, 1981). Classical solutions to this problem have included constraint filtering applied to the association graph (Barrow & Popplestone, 1971), labeling clutter with a null label (Boyer & Kak, 1988), and graph editing (Sanfeliu & Fu, 1983).

The first of these two issues has recently stimulated considerable interest in the connectionist literature (Simic, 1991; Mjolsness et al., 1989; Gold & Rangarajan, 1996; Gold et al., 1996; Suganathan, Teoh, & Mital, 1995). Here the aim has been to deploy continuous optimization methods such as the relatively heuristic graduated assignment (Gold & Rangarajan, 1996; Blake & Zisserman, 1987) or the more principled mean-field theory (Hoffmann & Buhmann, 1997; Yuille & Kosowsky, 1994; Rangarajan, Gold, & Mjolsness, 1996; Yuille, Stolorz, & Utans, 1994; Peterson & Soderberg, 1989) to update a set of assignment variables representing the matching process. Specifically, several authors (Simic, 1991; Mjolsness et al., 1989; Suganathan et al., 1995; Gold & Rangarajan, 1996; Gold et al., 1996) have addressed the issue as to how to capture the relational matching problem using an energy function. However, the basic formulation invariably revolves around an energy function that is quadratic in the assignment variables. Performance has been demonstrated to be enhanced if additional, sometimes nonquadratic, terms are added. Examples include node self-amplification term that encourages binary solutions (Gold & Rangarajan, 1996) and the addition of a logarithmic barrier entropy that convexifies the energy (Yuille & Kosowsky, 1994). Despite this effort expended in designing energy functionals with desirable global features, the second issue, of how to handle structural errors, has received little attention. The favored method is to introduce a null attractor, requiring additional energy terms that must be carefully controlled (Suganathan et al., 1995). A more elegant alternative is provided by the slack variables of Gold and Rangarajan (1996) that endow the null process with a degree of robustness. However, in both cases there is no mechanism for correcting structural errors due to spurious edge or node insertions. A recent addition to the literature that falls closer to the mark is the continuous embedding of Meila and Jordan (1997) that allows triangulated graphs to be moralized so as to produce a tree structure.

In a recent series of papers we have developed a Bayesian framework for relational graph matching (Cross & Hancock, 1995, 1997; Wilson et al., 1995; Wilson & Hancock, 1995, 1997). The novelty resides in the fact that relational consistency is gauged by a probability distribution that uses Hamming distance to measure structural differences between the graphs under match. The probability distribution has a mixture structure that sums a series of
exponential functions of the Hamming distances between neighborhoods of the model and data graphs. From a practical perspective, this new framework has been used to match complex infrared (Cross & Hancock, 1995) and radar imagery (Wilson et al., 1995, 1997). Of particular relevance to the study reported here, it has also been used successfully to overcome structural errors using a graph-edit process (Wilson & Hancock, 1995, 1997) of the sort originally proposed by Sanfeliu and Fu (1983). The optimization of this relational consistency measure and the control of the associated graph edit process have hitherto been confined to the use of discrete update procedures (Hancock & Kittler, 1993; Wilson et al., 1995; Cross & Hancock, 1995, 1997). Examples include discrete relaxation (Sanfeliu & Fu, 1983, Wilson et al., 1995), simulated annealing (Aarts & Korst, 1989; Geman & Geman, 1984; Kirkpatrick, Gelatt, & Vecchi, 1983; Cross & Hancock, 1995), and genetic search (Cross & Hancock, 1997).

Our aim in this article is to consider how the process of updating assignment variables can be realized by continuous means (Gold & Rangarajan, 1996; Peterson & Soderberg, 1989; Yuille, 1994; Suganathan et al., 1995). Specifically we consider how the matching process can be effected using the graduated assignment algorithm of Gold and Rangarajan (1996), which is based on the soft-max idea of Bridle (1990). We begin from the probability distribution for the assignment variables recently developed by Wilson and Hancock (1997) The resulting distribution is defined over a set of discrete assignment labels. The development of a graduation algorithm poses dual demands. The first of these revolves around identifying the energy function that encapsulates the mixture distribution for the discrete matching errors. The second task is to develop continuous update equations that can be used to soften the discrete assignment representation while allowing structural errors in the data graph to be identified.

There are two novel outcomes of our study. First we arrive at a new graph-matching energy. Here we replace the mixture distribution by a single probability function that corresponds to zero Kullback divergence. According to our analysis, the effective matching energy is in fact a weighted sum of Hamming distances between the model graph and the data graph neighborhoods. The quantities of interest in iteratively updating the state of match are the derivatives of the global energy function with respect to the softened assignment variables. In the case of our new matching energy, these derivatives take on a particularly interesting form, which provides an intuitive insight into the minimization of the energy. The second novel contribution is to introduce a new update process which allows the edge set of the data graph to be updated with the aim of controlling structural corruption due to spurious edges or nodes. An experimental evaluation of the technique reveals not only that it is successful in matching noise-corrupted graphs, but that it significantly outperforms the optimization of the standard quadratic energy function.

The optimization process underpinning our matching process is the
graduated assignment algorithm of Gold and Rangarajan (1996) The basic idea is to update assignment variables by exponentiating the derivatives of the energy function. Although this is a relatively heuristic optimization method, it has recently been formally demonstrated to exhibit desirable convergence properties (Rangarajan, Yuille, Gold, & Mjolsness, 1997). It is the nonquadratic nature of our energy function that draws us to graduated assignment. The more rigorous framework that underpins mean-field theory proves to be intractable for our energy function. The reason is that mean-field equations must be located by solving the set of saddle-point equations that minimize the Kullback divergence. This approach is most tractable when the underlying cost function is quadratic in the assignment variables, although Hoffmann and Buhmann (1997) have solved the mean-field equations in a nonquadratic case.

In section 2 we review how relational consistency can be measured by appealing to the concept of a label error process. Section 3 shows how the Bayesian consistency measure can be used to construct a global graph-matching energy. Issues of energy minimization are the subject of section 4. Here we describe how matches are iteratively updated using the soft-assign ansatz. Section 5 details the experimental evaluation of our matching process, focusing on the issue of noise sensitivity by investigating the systematic effects of parameter choice, structural error, and edge density. Finally, section 6 offers some conclusions.

2 Relational Consistency

Our overall goal here is to formulate a nonlinear optimization technique for matching relational graphs. We use the notation $G = (V, E)$ to denote the graphs under match, where $V$ is the set of nodes and $E$ is the set of edges. Our aim in matching is to associate nodes in a graph $G_D = (V_D, E_D)$ representing data to be matched against those in a graph $G_M = (V_M, E_M)$ representing an available relational model. Formally, the matching is represented by a function $f : V_D \to V_M$ from the nodes in the data graph $G_D$ to those in the model graph $G_M$. We capture the structure of the two graphs using a pair of connection matrices. The connection matrix for the data graph consists of the binary array

$$D_{ab} = \begin{cases} 1 & \text{if } (a, b) \in E_D \text{ or } a = b \\ 0 & \text{otherwise}, \end{cases} \quad (2.1)$$

while that for the model graph is

$$M_{\alpha\beta} = \begin{cases} 1 & \text{if } (\alpha, \beta) \in E_M \text{ or } \alpha = \beta \\ 0 & \text{otherwise}. \end{cases} \quad (2.2)$$

Since the current state of match between the two graphs is represented by the function $f : V_D \to V_M$, the statement $f(a) = \alpha$ means that the node $a \in V_D$
Graph Matching

is matched to the node $\alpha \in V_M$. The binary representation of the current state of match is captured by a set of assignment variables that convey the following meaning:

$$s_{\alpha} = \begin{cases} 1 & \text{if } f(a) = \alpha \\ 0 & \text{otherwise.} \end{cases} \quad (2.3)$$

The basic goal of the matching process is to optimize a consistency measure that gauges the structural similarity of the matched data graph and the model graph. In a recent series of papers, we have shown how consistency of match can be modeled using a Bayesian framework (Wilson et al., 1995; Wilson & Hancock, 1995, 1997). The basic idea is to construct a probability distribution that models the effect of memoryless matching errors in generating departures from consistency between the data and model graphs. Suppose that $S_\alpha = \alpha \cup \{\beta | (\alpha, \beta) \in E_M\}$ represents the set of nodes that form the immediate contextual neighborhood of the node $\alpha$ in the model graph. The aim in gauging relational consistency is to measure the structural differences between the set of model-graph neighborhoods and their matched counterparts in the data graph. Further, suppose that $\Gamma_a = f(a) \cup \{f(b) | (a, b) \in E_D\}$ represents the set of matches assigned to the contextual neighborhood of the node $a \in V_D$ of the data graph. Basic to Wilson and Hancock’s modeling of relational consistency is to regard the complete set of model-graph relations as states over which to construct a mixture model for the corrupt matches observed on the data graph. In other words,

$$P(\Gamma_a) = \sum_{\alpha \in V_M} P(\Gamma_a | S_\alpha) P(S_\alpha). \quad (2.4)$$

The component densities appearing in this mixture distribution are modeled under the assumption that discrete matching errors are uniform and memoryless. Accordingly, the confusion probability $P(\Gamma_a | S_\alpha)$ can be factorized over the component matches,

$$P(\Gamma_a | S_\alpha) = \prod_{b \in C_a} P(f(b) | \beta). \quad (2.5)$$

The confusion probabilities are modeled under the assumption that the individual misassignment errors occur with a uniform probability $P_e$. As a result, the single-node match probabilities are distributed as follows:

$$P(f(b) | \beta) = \begin{cases} (1 - P_e) & \text{if } f(b) = \beta \\ P_e & \text{if } f(b) \neq \beta. \end{cases} \quad (2.6)$$

Under this distribution rule, the mixture components acquire the following exponential form:

$$P(\Gamma_a | S_\alpha) = K_a \exp \left[ -\mu H(a, \alpha) \right]. \quad (2.7)$$
The exponential constant is related to the uniform error probability as follows:

$$\mu = \ln \frac{1 - P_e}{P_e}. \quad (2.8)$$

The distribution of misassignment errors is in fact measured by the Hamming distance $H(a, \alpha)$ between the current matching configuration $\Gamma_a$ residing on the data graph and the configuration $S_\alpha$ from the model graph. In terms of our discrete representation of the matching process, the Hamming distance is computed using a series of Kronecker delta functions defined over the data graph node matches and the assignments demanded by the model-graph configurations:

$$H(a, \alpha) = \sum_{b \in C_a} (1 - \delta_{f(b), \beta}). \quad (2.9)$$

According to our binary representation of the matching process, the distance measure is computed using the connectivity matrices and the assignment variables in the following manner:

$$H(a, \alpha) = \sum_{b \in V_D} \sum_{\beta \in V_M} M_{\alpha \beta} D_{ab} (1 - s_{b\beta}). \quad (2.10)$$

With these ingredients and under the assumption of a uniform prior for the model-graph constraints, that is, $P(S_\alpha) = \frac{1}{|V_M|}$, the probability of the matched neighborhood $\Gamma_a$ reduces to

$$P(\Gamma_a) = \frac{K_a}{|V_M|} \sum_{\alpha \in V_M} \exp[-\mu H(a, \alpha)]. \quad (2.11)$$

The probability distribution $P(\Gamma_a)$ may be regarded as providing a natural way of modeling departures from consistency at the neighborhood level. Matching consistency is graded by Hamming distance, and controlled hardening may be induced by reducing the label error probability $P_e$ toward zero.

3 The Effective Potential for Discrete Relaxation

We would like to construct an energy function that can be used to establish continuous update equations for the assignment variables of our matching process. The route to computing the energy function is to identify the probability distribution $q(\Gamma_a)$ that replaces the compound exponential mixture distribution appearing in equation 2.11 with a single exponential function. We locate this distribution by computing the Kullback-Leibler divergence.
Graph Matching

between the components of the mixture distribution, $P(\Gamma_a|S_a)$ and the equivalent distribution function $q(\Gamma_a)$ over the space of model graph label configurations. Although the idea of finding mean-field equations that minimize the divergence between a target probability distribution and a parameterized version is central to some of the classical work on mean-field theory (Saul & Jordan, 1995; Yuille et al., 1994), our approach is somewhat different in philosophy. The classical approach is to find saddle-point equations that minimize the divergence. However, in the case of our compound exponential mixture distribution appearing in equation 2.11 this is not tractable in closed form. For this reason we commence by locating an equivalent energy that results in zero divergence. We then soften the assignment representation to obtain a continuously defined energy function. The assignment variables are updated by applying the soft-assign ansatz to the gradients of the resulting global energy function.

We commence our development by considering the divergence between the mixture components $P(\Gamma_a|S_a)$ of $P(\Gamma_a)$ and the single effective distribution function $q(\Gamma_a)$,

$$J(\Gamma_a) = \sum_{a \in V_M} P(\Gamma_a|S_a) \ln \frac{P(\Gamma_a|S_a)}{q(\Gamma_a)}.$$  (3.1)

When the divergence vanishes, then the distribution $q(\Gamma_a)$ models the mixture distribution $P(\Gamma_a)$ in the maximum entropy sense. This condition is satisfied when

$$\ln q(\Gamma_a) = \sum_{a \in V_M} P(\Gamma_a|S_a) \ln \frac{P(\Gamma_a|S_a)}{P(\Gamma_a)}.$$  (3.2)

Substituting for $P(\Gamma_a|S_a)$ from equation 2.7,

$$\ln q(\Gamma_a) = \ln K_a - \mu \sum_{a \in V_M} H(a, \alpha) \exp[-\mu H(a, \alpha)]$$  (3.3)

The internal energy associated with the matching configuration $\Gamma_a$ is given by

$$U(\Gamma_a) = -\ln q(\Gamma_a).$$  (3.4)

Since the quantity $\ln K_a$ is an additive constant that does not depend on the assignment variables and $\mu$ is a global scaling constant, we omit them from
the definition of energy. We therefore focus on the quantity

\[ U(\Gamma_a) = \frac{\sum_{a \in V_M} H(a, \alpha) \exp[-\mu H(a, \alpha)]}{\sum_{a \in V_M} \exp[-\mu H(a, \alpha)]}. \]  

(3.5)

In other words the neighborhood potentials are simply weighted sums of Hamming distance between the data and model graphs. In fact, the potentials display an interesting barrier property. The potential is concentrated at Hamming distance \( H \approx \frac{1}{\mu} \). Both very large and very small Hamming distances contribute insignificantly to the energy function, that is, \( \lim_{H \to 0} H \exp[-\mu H] = 0 \) and \( \lim_{H \to \infty} H \exp[-\mu H] = 0 \).

Finally, with the neighborhood matching potentials at hand, we construct a global internal matching energy \( \mathcal{E} \) by summing the contributions over the nodes of the data graph, that is,

\[ \mathcal{E} = \sum_{a \in V_D} U(\Gamma_a). \]  

(3.6)

In passing, it is interesting to consider the relationship between our internal energy and the free energy of statistical physics. In statistical physics, it is the free energy that is the focus of attention. There are two physical versions of the free energy. The Helmholtz form is appropriate when there is a fixed number of states. When the volume of the system changes, the Gibbs form is appropriate. For our problem, the number of discrete states is fixed, so it is to the global Helmholtz free energy that we turn:

\[ \mathcal{F} = \sum_{a \in V_D} U(\Gamma_a) + T \sum_{a \in V_D} P(\Gamma_a) \ln P(\Gamma_a). \]  

(3.7)

From this expression it is clear that minimizing the free energy \( \mathcal{F} \) is equivalent only to minimizing the total internal energy \( \mathcal{E} \) when the computational temperature \( T \) is zero. However, at nonzero temperatures, minimizing free energy can also be viewed as minimizing internal energy subject to the maximum entropy constraint, where the computational temperature plays the role of a Lagrange multiplier. Quite informally, since our internal energies have been derived to minimize Kullback divergence, then this additional constraint can be viewed as superfluous. At a more formal level Hoffmann and Buhmann (1997) have invoked a result due to Peirls, which shows that the upper bound on the free energy is equal to the zero-point free energy plus the change in internal energy. Since the zero-point internal energy is zero in our case (recall that \( \lim_{\mu \to \infty} H \exp[-\mu H] = 0 \)), then minimizing internal
energy corresponds to minimizing the upper bound on the free-energy. The same conclusion can be reached by noting that our internal energy corresponds to zero divergence and appealing to the argument of Dayan, Hinton, Neal, and Zemel (1995). Here it is shown that the difference between the free energy and internal energy is proportional to the Kullback divergence between nonequilibrium and equilibrium probability distributions, that is, $P(\Gamma_0)$ and $q(\Gamma_0)$. In section 4, we therefore develop continuous update equations for the assignment variables by applying graduated assignment to the gradients of the total internal energy $\mathcal{E}$.

In the next section we will describe how optima of this essentially discrete global energy function may be located using continuous optimization (Gold & Rangarajan, 1996; Saganathan et al., 1995; Peterson & Soderberg, 1989; Yuille, 1994). However, before we develop our update equations we pause to consider the relationship with the energy function proposed by Gold and Rangarajan (1996). This comparison is quite informal, but under conditions in which $\mu \rightarrow \infty$, that is, $P_\epsilon \rightarrow 0$, then we can identify the confusion probability $P(S_\alpha | \Gamma_0)$ with the assignment variable $s_{\alpha a}$. Substituting for the Hamming distance $H(a, \alpha)$, we find

$$E_{\text{approx}} = \sum_{a \in \mathcal{V}_D} \sum_{\alpha \in \mathcal{V}_M} \sum_{b \in \mathcal{V}_D} \sum_{\beta \in \mathcal{V}_M} D_{ab} M_{\alpha \beta} (1 - s_{\beta | \alpha}) s_{\alpha a}. \tag{3.8}$$

We can rewrite this energy to make the relationship with Gold and Rangarajan’s (1996) energy function more transparent:

$$E_{\text{approx}} = -\sum_{a \in \mathcal{V}_D} \sum_{\alpha \in \mathcal{V}_M} \sum_{b \in \mathcal{V}_D} \sum_{\beta \in \mathcal{V}_M} D_{ab} M_{\alpha \beta} s_{\beta | \alpha} s_{\alpha a} + \sum_{a \in \mathcal{V}_D} \sum_{\alpha \in \mathcal{V}_M} \sum_{b \in \mathcal{V}_D} \sum_{\beta \in \mathcal{V}_M} D_{ab} M_{\alpha \beta} (s_{\alpha a} - 1). \tag{3.9}$$

The first term is the familiar quadratic assignment energy. The second term is also common to Gold and Rangarajan. Moreover, in the case of our approximate energy function, there are no Lagrange multipliers that need to be determined. However, Gold and Rangarajan’s full energy contains two additional terms. The first of these is a node self-amplification term that was introduced as an empirical means for encouraging convergence to a binary solution. The second addition is the logarithmic entropy or barrier term originally introduced in Yuille and Kosowsky (1994). This term convexifies the energy function. There has recently been a formal proof of convergence for both the full soft-assign algorithm and a number of simplified variants. This proof is based on the positivity of the Kullback divergence (Rangarajan et al., 1997). In other words, although we offer no formal proof of binary convergence for our algorithm, this literature points to the fact that algorithms with the same limiting form of the energy function do possess this property (Yuille & Kosowsky, 1994; Rangarajan et al., 1997).
4 Optimising the Global Cost Function

We are now in a position to develop a continuous update algorithm by softening the discrete ingredients of our graph matching potential. The idea is to compute the derivatives of the global internal energy given in equation 3.6 and to effect the softening process using the soft-max idea (Bridle, 1990). Although this update process bears many of the hallmarks of mean-field annealing, it is couched in less rigorous terms. Mean-field equations are invariably derived from a quadratic cost function and are obtained by solving the set of saddle-point equations that minimize the Kullback-Leibler divergence with respect to the assignment variables. However, in the case of our nonquadratic cost function, the closed-form solution of the saddle-point equations is less tractable. For reasons of computational expediency, we therefore adopt the soft-assign idea to constrain the assignment variables to lie within the unit simplex.

In practice we realize the updating of the matched version of the data graph using a two-stage update process. The first update mode involves softening the assignment variables—the $s_{\alpha\beta}$ that indicate the current state of match. The second update process aims to modify the connection strengths associated with the edges in the data graph. Data graph nodes that are unmatchable become disjoint by virtue of having weak connection weights and cease to play any significant role in the update process. It is this second update process that provides the basic mechanism for controlling relational inexactness in our technique. From the implementational standpoint, it obviates the need for an explicit null label and the requisite energy terms in the cost function. In fact, this process can be viewed as an implicit way of performing the explicit graph-edit operations of the type originally suggested in Sanfeliu and Fu (1983) and recently used to match cluttered scenes by Wilson and Hancock (1995). However, we provide no proof of convergence. The dual update steps are heuristically inspired. Suffice it to say that experiments reveal that the algorithm appears to be both numerically stable and capable of controlling a significant population of clutter.

4.1 Updating Assignment Variables. The energy function represented by equations 3.5 and 3.6 is defined over the discrete matching variables $s_{a\alpha}$. The basic idea underpinning this article is to realize a continuous process for updating the assignment variables. The step size is determined by computing the partial derivatives of the global matching energy with respect to the assignment variables. We commence by computing the derivatives of the contributing neighborhood energy,

$$\frac{\partial U(V_a)}{\partial s_{\hat{a}\hat{b}}} = \sum_{a \in V_{\alpha}} \left[ 1 - \mu \left( H(a, \alpha) - U(V_a) \right) \right] \xi_{a\alpha} \frac{\partial H(a, \alpha)}{\partial s_{\hat{a}\hat{b}}}. \quad (4.1)$$
where
\[ \xi_{a\alpha} = \frac{\exp[-\mu H(a, \alpha)]}{\sum_{\alpha' \in V_M} \exp[-\mu H(a, \alpha')]} . \tag{4.2} \]

To develop this result further, we must compute the derivatives of the Hamming distances. From the definition given in equation 2.10, it follows that
\[ \frac{\partial H(a, \alpha)}{\partial s_{b\beta}} = -M_{ab} D_{ab}. \tag{4.3} \]

It is now a straightforward matter to show that the derivative of the global matching energy is equal to
\[ \frac{\partial E}{\partial s_{b\beta}} = -\sum_{a \in V_D} \sum_{\alpha \in V_M} D_{ab} M_{ab} \left[ 1 - \mu \left( H(a, \alpha) - U(\Gamma_a) \right) \right] \xi_{a\alpha}. \tag{4.4} \]

We would like our continuous matching variables to remain constrained to lie within the range [0, 1]. For this reason, rather than using a linear update rule, we exploit the soft-max ansatz (Bridle, 1990):
\[ s_{a\alpha} \leftarrow \frac{\exp \left[ -\frac{1}{T} \frac{\partial E}{\partial s_{a\alpha}} \right]}{\sum_{\alpha' \in V_M} \exp \left[ -\frac{1}{T} \frac{\partial E}{\partial s_{a\alpha'}} \right]} . \tag{4.5} \]

The value of the temperature \( T \) in the update process has been controlled using a slow exponential annealing schedule of the form suggested by Gold and Rangarajan (1996). We initialize the matching algorithm with assignment variables computed using a gaussian error model. Details are application specific and are deferred until we outline our experiments in section 5.

The mathematical structure of the update process is important and deserves further comment. The quantity \( \xi_{a\alpha} \) defined in equation 4.2 naturally plays the role of a matching probability. The first term appearing under the square bracket in equation 4.4 can therefore be thought of as analogous to the optimal update direction for the standard quadratic cost function (Suganathan et al., 1995; Gold & Rangarajan, 1996); we will discuss this relationship in more detail in section 4.3. The second term modifies this principal update direction by taking into account the weighted fluctuations in the Hamming distance about the effective potential or average Hamming distance. If the average fluctuation is zero, there is no net modification to the update direction. When the net fluctuation is nonzero, the direction of update is modified so as to compensate for the movement of the mean value of the effective potential. Moreover, since the fluctuation term is itself proportional to \( \mu \), this has an insignificant effect for \( P_e \approx \frac{1}{2} \) but dominates the update process when \( P_e \to 0 \).
4.2 Updating the Connection Weights. The updating of assignment variables described in the previous subsection can be used to rectify initial matching errors. However, in most realistic graph-matching problems, there are also structural errors due to the presence of contaminating noise or clutter. There are several ways in which clutter can be controlled. For instance, in a recent comparative study of discrete relaxation, we compared the relative efficacy of three alternative strategies for identifying structural errors (Wilson & Hancock, 1996b, 1997). These include the labeling of clutter using null-category (Wilson & Hancock, 1996a; Finch, Wilson, & Hancock, 1997), the explicit removal of clutter using iterative graph-edit operations (Sanfeliu & Fu, 1983), and, finally, the application of a constraint filtering operation to remove clutter nodes from the final match (Barrow & Popplestone, 1971). The main conclusion of this study was that iterative graph-editing outperforms null labeling and constraint filtering as a means of controlling structural corruption.

When a continuous update process is used rather than a discrete one, then the scope for clutter control is more limited. In fact, the process is usually effected by augmenting the model-graph nodes with a null attractor (Wilson & Hancock, 1996a; Finch et al., 1997). This is exactly the strategy adopted by Gold and Rangarajan (1996), who introduce slack variables into their soft-assign algorithm to accommodate unmatchable nodes.

Here, our approach is different in motivation. Based on our experiences in the discrete graph-matching domain, we opt to develop a continuous counterpart of the graph-edit process. The basic idea is to update the elements of the data graph connection matrix, and hence provide a continuous mechanism for controlling clutter and relational inexactness. By weakening connections in the data graph, the matches of extraneous nodes become disjoint.

To develop our continuous graph-edit process, we note that the soft-assign ansatz outlined in the previous section can be applied to update the elements of the connectivity matrix. In order to proceed, we will require the derivative of the Hamming distance with respect to the relevant element of the connection matrix, that is,

$$\frac{\partial H(a, \alpha)}{\partial D_{ab}} = \sum_{\beta \in V_M} M_{\alpha \beta}(1 - s_{\beta \beta}). \quad (4.6)$$

As a result the gradient of the global matching energy with respect to the elements of the data graph connection matrix is equal to

$$\frac{\partial \mathcal{E}}{\partial D_{ab}} = \sum_{a \in V_M} \sum_{\beta \in V_M} M_{\alpha \beta}(1 - s_{\beta \beta}) \left[1 - \mu \left(H(a, \alpha) - U(\Gamma_a)\right)\right] \xi_{aa}. \quad (4.7)$$

The soft-assign update equation for the elements of the data graph
connection matrix is

\[
D_{ab} \leftarrow \frac{\exp\left(-\frac{1}{T} \frac{\partial E}{\partial D_{ab}}\right)}{\sum_{(a',b') \in E_D} \exp\left(-\frac{1}{T} \frac{\partial E}{\partial D_{a'b'}^T}\right)}.
\] (4.8)

From equation 4.8 it is interesting to note that the derivatives of the matching energy are zero when the entire neighborhood of the node \(a\) is consistently matched onto the model graph.

4.3 Quadratic Assignment Problem. Before we proceed to experiment with the new graph-matching process, we briefly review the standard quadratic formulation of the matching problem investigated by Simic (1991), Suganathan et al., (1995), and Gold and Rangarajan (1996). The common feature of these algorithms is to commence from the quadratic cost function,

\[
E_H = -\frac{1}{2} \sum_{a \in V_D} \sum_{\alpha \in V_M} \sum_{b \in V_D} \sum_{\beta \in V_M} D_{ab} M_{\alpha \beta} s_{\alpha a} s_{\beta b}.
\] (4.9)

In this case the derivative of the global cost function is linear in the assignment variables, that is,

\[
\frac{\partial E_H}{\partial s_{\beta b}} = -\frac{1}{2} \sum_{a \in V_D} \sum_{\alpha \in V_M} D_{ab} M_{\alpha \beta} s_{\alpha a}.
\] (4.10)

This step size is equivalent to that appearing in equation 4.4 provided that \(\mu = 0\), that is, \(P_c \rightarrow \frac{1}{2}\). The update is realized by applying the soft-assign ansatz of equation 4.5. When the quantity \(\zeta_{\alpha a}\) defined in equation 4.2 is replaced by the assignment variable \(s_{\alpha a}\) and we take the limit \(\mu = 0\), that is, \(P_c = \frac{1}{2}\), in equation 4.4, then the soft-assign update (Peterson & Soderberg, 1989). In the next section we will provide some experimental comparison with the resulting matching process. However, it is important to stress that the update process adopted here is very simplistic and leaves considerable scope for further refinement. For instance, in Gold and Rangarajan (1996), Sinkhorn matrices have been used to impose a permutation structure on the final solution. We could clearly apply the same apparatus to our matching process. It is also important to stress that null matches are handled differently. Rather than adjusting the data graph connection weights, Gold and Rangarajan control noise by introducing slack variables into the rows and columns of their match matrix. Finally, as we have already pointed out in section 3, the basic quadratic assignment energy is augmented with various additional terms. Examples include barrier potentials and self-amplification terms (Gold & Rangarajan, 1996).
5 Experiments

Our aims in this section are twofold. First, we investigate the sensitivity of our matching process. We focus on three aspects of the matching process: sensitivity to noise and structural error, the effect of edge density, and the best choice of operating parameter $\mu$. The second goal is to offer some comparative evaluation of the matching scheme. We investigate the relative noise sensitivity when we apply the soft-assign algorithm to both the weighted Hamming distance energy and the quadratic assignment energy. Our evaluation of the graph-matching process is based on synthetic data. Specifically, we study randomly connected graphs of various edge densities.

5.1 Randomly Connected Graphs. Our first set of experiments is concerned with matching randomly connected graphs. For a node set of predetermined size, we generate a random set of connections. The number of connections is controlled to give a predefined value of the average edge density per node. The resulting structures are used as the model graphs in our experiments. In order to generate data graphs, we introduce structural errors. These errors are introduced by randomly deleting a predetermined fraction on nodes from the model graphs. If a node is deleted, so are its connecting edges.

We generated initialization errors by randomly misassigning a predetermined fraction of the initial winner-take-all matches between the data and model graphs. These initialization errors are simulated as follows. Associated with each node in the graph is a uniformly distributed random unary attribute; for the node $a \in V_D$, we denote the variable by $x_{\bar{a}}$. This unary attribute is used to compute an initial assignment variable. Uncertainties in the measurements are simulated by randomly perturbing the unary attributes of the data graph with a gaussian error distribution so as to produce a specified fraction of initial matching errors. The initial assignment variables are computed from exponential distributions of the distance between attribute vector pairs using an estimate of the measurement noise-variance $\sigma$, that is,

$$s_{\text{init}} = \frac{\exp\left[-\frac{1}{2\sigma^2}(x_{\bar{a}} - x_{\bar{\alpha}})^2\right]}{\sum_{\alpha' \in V_M} \exp\left[-\frac{1}{2\sigma^2}(x_{\bar{a}} - x_{\bar{\alpha'}})^2\right]}.$$  

(5.1)

Our default algorithm settings are as follows. For the majority of our experiments, the fraction of initial winner-take-all errors is 50%. The graphs used in our study have random connections with an average of six edges per node. The graphs are nonplanar. We operate with the parameter $\mu$ set to 2.0; the motivation for this choice will be provided in section 5.2. The temperature of the soft-assign update is annealed from an initial setting of 10. The schedule for annealing the temperature between the epochs $n$ and
$n + 1$ of the algorithm is of the form $T^{(n+1)} = 0.95T^{(n)}$. This is somewhat slower than that employed by Gold and Rangarajan (1996), who reduce the temperature by a factor of 0.925 with each iteration. There are four updates at each temperature epoch. Typically, we reach a stable winner-take-all matching configuration after 50 iterations. In other words, the final temperature is 0.77.

5.2 Sensitivity Study. In this subsection we study the sensitivity systematics of our matching process. To commence, Figure 1 illustrates the effect of edge density on the matching process. The figure shows the fraction of correctly matched nodes as a function of the fraction of nodes deleted from the model graph. The different curves are for average edge densities of two, four, six, and eight edges per node. In each case, the graphs contain 50 nodes, and the fraction of initialization errors is 50%. Each data point is averaged over 100 random experiments. The main point to be drawn from these curves is that the accuracy of match increases with edge density. When the fraction of deleted nodes is less than 0.3, there is little systematic difference between the matching fractions. It is only when the fraction of structural errors exceeds 0.5 that the effect of edge density becomes marked. This is in accord with Gold and Rangarajan’s (1996) experience with graduated assignment, where it was found that performance improves with increasing edge density.

The curves shown in Figure 1 were obtained under conditions in which the fraction of initialization errors is 50%. Our next sequence of experiments is aimed at illustrating the effects of initialization error on the matching process. Here we have used random graphs in which there are six edges per node. Figure 2 shows the final fraction of correct matches as a function of the fraction of initialization errors. The straight line is the break-even case in which there is no net improvement in the matching process. The curves show the net improvements obtained when the fractions of structural errors are 5%, 10%, and 20%. Clearly, as the fraction of structural errors increases, then so the ability to recover from initialization error degrades. For the highest level of structural corruption, the iterative improvements are marginal. However, there are only a few cases in which the data points dip below the break-even line. In other words, the updating process rarely disrupts the configuration of correctly assigned labels in the initial match.

The parameter $\mu$ controls the balance between the two terms in the step size for assignment variable update. As pointed out in section 4, the first term moves the assignment variables in the direction of maximum neighborhood support. This term can be thought of as ensuring maximum structural congruency between the matched graphs. The second term modifies direction of update to minimize the Hamming distance fluctuations. Figure 3 shows a plot of the fraction of correct matches as a function of $\mu$. Here there are no structural errors. The only source of error is initialization error, which is set at 50%. The plot shows an interesting threshold structure in
Figure 1: Effect of edge density on matching performance: The graph shows the final fraction of correct matches as a function of the fractional structural corruption in the graphs for different edge densities. The solid, dotted, dot-dashed and dashed curves respectively correspond to two, four, six and eight edges per node.

the matching errors. When $\mu < 1$, the matching performance is very poor. When $\mu > 1$, the matching process is essentially insensitive to parameter choice. It is also important to note that the performance falls off rapidly when $\mu = 0$. This corresponds to the case when our matching algorithm implements the standard Potts glass realization of the quadratic assignment problem.

To investigate the role of the control parameter in more detail, we studied the noise sensitivity of the matching algorithm for various choices of $\mu$. The curves in Figure 4 show the noise sensitivity of the matching process when $\mu = 0.5$ and $\mu = 2.0$. These two values span the onset threshold exhibited in Figure 3. Here the noise is structural, and we commence with 50% initialization error. The main point to note from these curves is that the higher value of $\mu$ gives superior performance at low levels of structural corruption. The lower value of $\mu$ gives relatively poor performance at low levels of structural corruption, but performs better when there is significant corruption. The main empirical conclusion to be drawn from these observations is as follows. The Hamming-distance fluctuation term appears to play an important role in controlling the matching process when initializa-
Figure 2: Effect of initialization error. The plot shows the fraction of correct matches in the final match as a function of the fraction of initially correct matches. The different curves are for varying degrees of structural corruption.

5.3 Algorithm Comparison. The main conclusion of the sensitivity study reported in the previous subsection was that the matching process offered optimal performance when the graphs have an average density of six edges per node. Now we offer a comparative evaluation of the matching process using synthetic Delaunay graphs. We selected this structure since it is typical of the neighborhood graphs widely exploited in computer vision problems. Importantly, it also has a density of approximately six edges per node. However, it differs from a randomly connected graph in one critical respect. In a randomly connected graph, unless the edge density is very high, there are few first-order cycles. By contrast, the Delaunay graph is composed entirely of first-order triangular faces. In other words, every node in a Delaunay graph belongs to at least one first-order cycle. Since the graph triangulates the nodes, the average node is shared by six first-order faces.

We commence our construction of the synthetic Delaunay graphs by generating random dot patterns. Each random dot is used to seed a Voronoi cell. The Delaunay triangulation is the region adjacency graph for the Voronoi cells. Structural corruption is simulated by deleting and adding controlled
numbers of random dots. For a series of different corruption levels, we have generated a sample of 100 random graphs, each containing 50 nodes. The random corruption level process is such as to maintain the overall size of the graphs. For each graph, we measure the quality of match by computing the fraction of the surviving nodes for which the assignment variables indicate the correct match.

We have compared the results obtained by applying the soft-assign process to the weighted Hamming distance energy and the standard quadratic assignment energy. Figure 5 shows the final fraction of correct matches for each of the algorithms. The data curves show the correct matching fraction averaged over the graph samples as a function of the corruption fraction. The main conclusions that can be drawn from these plots is that the new matching technique described in this article significantly outperforms its conventional quadratic counterpart (described in section 4.3). The main difference between the two techniques resides in the fact that our new method relies on updating with derivatives of the energy function that are nonlinear in the assignment variables.

Finally, we compare the results obtained using our matching method with those reported by Gold and Rangarajan (1996). First, it is important to note that Gold and Rangarajan’s analysis proceeds only to noise levels of 20%. Several algorithm variants are reported. However, optimal performance is
Figure 4: Effect of varying the parameter $\mu$. The plot shows the effect of the parameter $\mu$ on the noise sensitivity curves. The solid curve is for $\mu = 1.0$ while the dotted curve is for $\mu = 2.0$.

Figure 5: Experimental comparison: softened discrete relaxation (dotted curve); matching using the quadratic cost function (solid curve).
achieved with attributed relational graph matching. This method augments
the quadratic assignment energy with a binary attribute difference term.
When three binary attributes per edge are used, then the performance at the
20% noise level is comparable to that obtained with our algorithm. However,
this version of Gold and Rangarjan’s algorithm contrasts with our own in
that it draws very heavily on attribute structure. After the initialization step,
our algorithm relies purely on graph structure.

6 Conclusions

Our main contribution here has been to develop a new graduation algorithm
for relational graph matching. There are two novel contributions. The first is
the development of a new energy function for graph matching. The second
is to update the connection structure of the data graph with the aim of
controlling structural corruption.

The development of the energy function commences from the Bayesian
consistency measure developed in Wilson et al. (1995). We show how the
discrete assignment representation underpinning the consistency can be
softened. Specifically, we use the apparatus of statistical physics to com-
pute a graph-matching energy. Formally, the energy is a weighted sum of
Hamming distances between the graphs being matched.

The gradient of the energy function has an interesting structure. In par-
ticular, the step size for assignment variable update has a two-component
structure. The first component is in the direction of maximum local gradient.
The second term retains memory of the structure of the energy landscape.
In particular, it modifies the direction of update so as to ensure that local
fluctuations in Hamming distance do not become excessively large.

In the second novel contribution, we address the problem of rectifying
structural errors by developing a graduation step for modifying the con-
nection structure of the data graph. The idea is to update the elements of
the data graph connection matrix so that clutter nodes can become disjoint
if they persist in failing to find a consistent match.

The operating limits of the matching algorithm have been evaluated on
simulation data. Here we show the algorithm to be effective under severe
structural error. The net effect of increasing structural error is to limit the
extent to which the algorithm can recover from initialization errors. Finally,
we have shown how the method relates to the standard quadratic assign-
ment algorithm extensively studied in the connectionist literature (Gold &
Rangarajan, 1996; Simic, 1991; Suganathan et al., 1995). Moreover, an exper-
imental comparison reveals that the method offers superior performance in
terms of noise control.

There are clearly a number of shortcomings in the reported work. In the
first instance, we fall well short of developing rigorous mean-field equations
(Hoffmann & Buhmann, 1997). There is also no formal proof of conver-
gence under the dual update operations. Nonetheless, the experimentation
demonstrates that the conceptual framework offers encouraging potential in the graph-matching domain. Suffice to say that the ideas presented here represent the first steps. We are considering the application of statistical physics to the energy function developed in this article and the conclusions will be reported in due course.

References


Received March 31, 1997; accepted January 6, 1998.