ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 273-278

Regularization and Neural Computation: Application to aerial images analysis

E. SCHAEFFER, P. BOURRET, S. MONTROZIER
ONERA-CERT, 2 Avenue Edouard Belin. 31055 TOULOUSE CEDEX, FRANCE
email : schaeffe@tls-cs.cert.fr, bourret@tls-cs.cert.fr,
montroz@tls-cs.cert.fr,

Abstract. The main feature of this paper is to show that the key point of two different problems tackled by neural approaches - pairing pattern and function approximation - lies in the choice of the regularization term in the function which is minimized by the neural approach. After the description of links between approximation of functions, combinatorial optimization and regularization theory, we describe a neural approach for solving the hard problem of change detection in aerial images.

1. Introduction

A wide part of problems which are solved by neural computation are ill posed problem according to Hadamard definition. A problem is said ill posed if there is not only one solution, and if the solution is not a continuous function of the imput data. This is the case in Combinatorial optimization and in Approximation of functions. A combinatorial optimization consists in minimizing C(X) w.r.t some constraints T(X) where $X=\{x_1,,...x_n\}$, $x_i \in [0,1]$ binary variables. It is well known that these problems frequently have a very large number of solutions but even if they are transformed in a problem with one solution it cannot be said that the binary outputs are a continuous function of the input data. Nevertheless a lot of neural approaches for solving them are based on the transformation of x_i into $y_i \in [0,1]$ and in exhibiting a function F(Y) which can only be minimum if $y_i \in [0,1]$. The minimizing process is based on the computation of partial derivatives. The y_i moves from a random starting value to y_i^* in such a way that Y_i^* is (or is supposed to be) the solution. Usually when the y_i converges towards y_i^* there is a bowl around y_i^* such that when y_i enters this bowl, it can only converges towards y_i^*. When the value of y_i is assigned then x_i is assigned. If we considered the radius of these bowls they are continuous functions of the input data and thus for some transformations of a combinatorial optimization to the minimization of a function F we can considered that we have regularized an ill posed problem.
In Approximation of functions, The approximation problem amounts to find F(X) such as F $\in \mathcal{F}$ a set of functions and which minimizes:

$$E = \sum_{i=1}^{n} (F(x_i) - f(x_i))^2$$

Obviously pattern recognition by the mean of backpropagation based neural network is exactly the same problem by adding to the «error term» E a «sensitivity» term which is supposed to be minimum when the ratio between variation of the outputs and the variation of the inputs is minimum. Some solutions are proposed in [1], [2] which are all based on the minimization of a so called sensitivity criterion. But a more general solution is presented in [3]. The basic idea is to define a functional H(F) with an error term and a constraint operator P such that: $H(F) = E + \lambda \| PF \|$ where P stands for the constraint which must be met by the solution F. The structure of P embodies the a priori knowledge about the solution and then is problem dependant. It is proved that for each P there is a function G such that the solution F* is given by:

$$F^*(X) = \frac{1}{\lambda}\sum_{i}^{N} (X_i - F(X_i))\, G(X_i)$$

Obviously the difficulty is to find the function G related to the operator P.
Regularization theory can be used for solving not only academic problems but also real world problem by introducing well adapted regularization terms in functions which have to be minimized. The aim of this paper is to show that well chosen regularization terms allow to solve hard problems like the analysis of differences between two outlook aerial images of the same area taken at two different times.

2. Presentation of the application

While in satellite images differences analysis the registration is not too difficult since there is only a quite isometric transformation between the two images, the problem is much more difficult for aerial images because the two images can have been taken from two different points, at different levels, and with different angles. In order to achieve our goal, we have split the problem into four sub-problems: The selection of characteristic points in the two images, then the mapping of these characteristic points of one image with characteristic points of the other one, on this basis the design of functions allowing a pixel to pixel mapping between the two images, and last exhibiting the differences (new building for instance). When comparing images we have pursued a natural approach for the first problem, in that we have looked for common points in both images. Since each pair of images usually contains millions of pixels, our approach has been to identify a hundred or so feature points in each image and since our test consists of images of an urban area with a few buildings, we have chosen to select the corners of buildings. These points are not frequently hidden and several methods (corner detector) have already been proposed in the literature. In the part three, we present a new algorithm alowing the matching between two sets of points with a nonuniform distribution in the plane. These correctly matched pairs are the control points and are going to represent our future learning set to modelise the global registration presented in the part four. After a presentation of the results in part five, we gave some tracks for the best use of regularization principles.

3. Mapping of characteristic points

3.1 Labelling by adaptative mapping

The problem for discussion in these part can be stated as follows. We have two sets of points in the plane. The second set is similar to the first set, except that some of the points from the first set are missing and some new points, not in the first set, are present. The positions of the points in the second set are not the same as common points in the first set. There is large local distortions. The problem is to find all common points in the two sets and find the correct match. We are going to show that the mapping between characteristic points can be solved by using the Yuille [4] principles if we take into account the nonuniform distribution of points in the plane. A deformable template is a model described by a set of parameter so that variations in the parameter space correspond to the different requests of the model [5]. Let $\{Q_i\}$, i=1...n and $\{P_j\}$, j=1...m the position of points in the two sets. Matching $\{Q_i\}$ and $\{P_j\}$ can be formulate as the minimization of an energy function:

$$E(\{V_{ij}\},\alpha) = \sum_{i,j} A_{ij} V_{ij} \frac{(Q_i - P_j(\alpha))^n}{\lambda^n} + E_p(\alpha)$$

A_{ij} is a compatibility matrix. $A_{ij}=1$ if the features labelled by i and j are totally compatible and $A_{ij}=0$ if they are totally incompatible. The V_{ij} are binary matching elements, λ is a constant and $E_p(\alpha)$ imposes prior constraints on the model. If we want our energy to preserve distances between a point and its neighbors during the matching process, we can defined:

$$E_P(\alpha) = \sum_{i,j} \psi(j) \qquad \text{where} \qquad \psi(j) = \sum_{k \in neighbour(j)} \frac{\left| d_{kj}^2 - |P_j - P_k|^2 \right|^2}{d_{kj}^2}$$

If we now impose constraints on the possible matches by summing over the configurations of $\{V_{ij}\}$, and calculating the sum over these configurations (we eliminate a non valid part of the space) we obtain the effective energy:

$$E_{eff}(\alpha) = \frac{-1}{\beta} \sum \sum_j \log\left(\left(\sum_i e^{-\beta A_{ij} \frac{(Q_i - P_j(\alpha))^n}{\lambda^n}}\right) + E_p(\alpha)\right)$$

Minimizing $E_{eff}(\alpha)$ with respect to α corresponds to deforming the template in the parameter space to find the best match. With these new energy function, we can do an optimization on α by a gradient descent. After some calculation we obtain the displacement suffered by the points P_j:

$$\Delta P_j = \alpha \sum_i \omega_{ij} (Q - P) + \beta\lambda \sum_j \psi(j) \, (P_j - P_k)$$

where ω_{ij} is the normalized influence of the point Q_i on the point P_j and λ a scale parameter. The values of α and β regulate the relative strengths of the two type of force acting on P_j, respectively the attraction toward the point Q_i and the pull toward its neighbors. The normalized influence ω_{ij} is given by:

$$\omega_{ij} = \frac{\Phi(|Qi - Pj|, \lambda)}{\sum_k \Phi(|Qi - Pj|, \lambda)} \qquad \text{where} \qquad \Phi(d, K) = exp\left(\frac{-d^2}{2\lambda^2}\right)$$

The algorithm uses a fixed number m>n of points and a decreasing sequence of values of λ. For each value of λ, a number of iterations is performed and in each iteration all points moved according to Eq (1). The algorithm terminates when at least one point is close enough to every point Q_j according to a pre-specified tolerance. But displacing the points P_j according to Eq (1) brings about two important problems: The first problem is that for large values of m (and n consequently) this algorithm may become excessively time-consuming. The computation of every single displacement depends on the position of all points Q_i. In every iteration the ω_{ij}'s must be re-computed as a function of the new positions of all points.
The second problem is the initialization of the two parameters α and β to have a correct and fast matching when there is a nonuniform distribution of all the points in the space. At each iteration we have to increase the parameter α which regulate the attraction toward the points of the pattern Q_i and decrease β which regulate the attraction toward the neighbors of P_i. If we describe the P_i's behavior during the iterations, we can see that aggregats of P_js are going

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 273-278

irremediably toward the inertia center of Q_i's aggregats. During the same time, isolated points P_j try to find isolated Q_i. When the rigidity β decrease and α increase, element P_j of the aggregats goes toward the nearest Q_i. The result is usually a bad matching.
Boeres [6] proposes a filtering mechanism that solve the first problem. Their notion of λ-filter is based on the identification for each point j and each value of λ, of a subset $Q_\Upsilon(j,\lambda) \subseteq Q$

such that: $$\sum_{i \in Q_\Upsilon(j,\lambda)} \Phi(|Q_i - P_j|, \lambda) = \Upsilon \sum_{i \in Q} \Phi(|Q_i - P_j|, \lambda)$$

Finding these subset for an arbitrary arrangement of points Q_i on the plane is not easy, so Boeres used the approximation that the points Q_i are infinitely many and scattered throughout the entire plane uniformly with density ρ. This approximation can't be apply successfully in many application where the selection of control points in the reference and sensed images depends on the response of features extracting operators.
We present now a filtering mechanism based on the definition of a specific neighborhoods that accelerate considerably the matching algorithm and can cope with a nonuniform distribution of points in the plane

3.2 The filtering Mechanism

We formalized the problem of mapping two sets of points $\{P_j, i=1..n\}$ and $\{Q_i, j=1...m\}$ into the minimization of:

$$E(\{V_{ij}\}, \{P_j\}) = \sum_{ij} V_{ij}(Q_i - P_j)^2 + \lambda_t F_s(Q_i, P_j)$$

V_{ij} are binary variables with $V_{ij}=1$ if Q_i is associated with P_j, $V_{ij}=0$ otherwise. A point Q_i must be matched with one and only one P_j of the model then $\sum V_{ij} = 1$. $F_s(Q_i,P_j)$ stands for the function representing the difference between the structure of $\{P_i\}$ and that one of $\{V_{ij}Q_i\}$ where $V_{ij}Q_i$ is the Q_i which is mapped with P_j. The problem can be seen like the mapping of P_j to the closest Q_i with respect to the "structure" of the two sets. Thus we have an error term and a regularization term. The term F_s is problem dependent. In our application the constraint is that a set of points close together will only be mapped with points which are also close together. The structure F_s represents this constraint by a term which looks like the error term but which is related to barycenters of sets of points which are in a neighborhood. For defining the neighborhood we have used the Delaunay and Voronoi tessellations which provide a fundamental order for a set of multivariate data. Let now W_k (respectively W'_k in the second image) be these neighborhoods and B_k the barycenter of W_k (respectively B'_k) then:

$$F_s = \sum_{k,l} V_{kl}(B_k - B'_l)^2$$

Matching $\{Q_i\}$ and $\{P_j\}$ can be now formulate as the minimization of the energy function:

$$E(\{V_{kl}\}, \{P_j\}) = \sum_{ij} V_{ij}(Q_i - P_j)^2 + \lambda_t \sum_{k,l} V_{kl}(B_k - B'_l)^2$$

λ_t is a decreasing parameter according to the convergence of B_k toward B'_k in such way that after these convergence all the characteristic points, belonging or not to a neighborhood can be mapped. These correctly matched pairs are the control points and are going to represent our learning set to modelise the global registration

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 273-278

4. Mapping pixels and regularization neural networks

To solve the next problem, let us assume that the positions of k corresponding control points in the image $((x_{iP}, y_{iP}),(x_{iQ}, y_{iQ}), i=1...k)$ are given and their mutual correspondence is established by the previous step.We would like to find two functions f and g in such a way that the image registration represented by:

$$x_{iQ} = f(x_{iP}, y_{iP}) , y_{iQ} = g(x_{iP}, y_{iP}) \text{ with:} i=1...k$$

would be as accurate as possible. Because the measurements are sparse, the problem is ill posed and requires adding a smoothing or regularizing term to obtain a solution in areas away from measured points.We can formulate our problem as the problem of minimizing a functional depending on two piecewise-smooth functions f and g given a sparce set of noisy measurements:

$$H(f,g) = \sum_{i=1}^{k} (x_Q - f(x_P, y_P))^2 + \sum_{i=1}^{k} (x_Q - g(x_P, y_P))^2 + \lambda_m (\|Pf\|^2 + \|Pg\|^2)$$

where $\{(x_{iP}, y_{iP}, x_{iQ}), i=1...k\}$ and $\{(x_{iP}, y_{iP}, x_{iQ}), i=1...k\}$ form the learning sets and λ_m a positive real number. The operator P is a constraint operator usually called stabilizer and $\|.\|$ is a norm on the function space to which Pf (respectively Pg) belongs (usually the L^2 norm).The structure of the operator P embodies the a priori knowledge about the solution, and therefore depends on the nature of the particular problem that as to be solved. P measure the smoothness of f (respectively g). The regularization parameter controls the trade-off between the two terms. The minimization can be realized by a neural network [3] with one layer of hidden units. The first layer of this network consists of input units whose number is equivalent to the number k of examples that we have defined. The second layer is composed of non-linear hidden units fully connected to the first layer. There is one hidden unit for each example. The activation function of the hidden units is the Green's function.

5. Results and application to change detection

We present a pair of images (in fact part of images) of the same area taken at a time interval of several years (Fig 1). The control points are building corners. For simplicity's sakes, we have shown only few iterations of the matching process. Control points of the model goes toward control points of the pattern and we can see a part of their trajectory (Fig.2a). The result of the pixel to pixel mapping is shown by the result of the transformation obtained on the first image. The presented result (Fig 2b) is the first image where each pixel (x_{Pj}, y_{Pj}) has been located at $f(x_{Pj}, y_{Pj})$, $g(x_{Pj}, y_{Pj})$ with the same grey level. The change detection is at this stage of our study very simple: we compute the correlation coefficient of local histograms computed on a window centered on each pixel. The correlation number expresses a similarity between two laws (the histogram for example). When the coefficient C(H1,H2) is null, laws are orthogonal, in others words H1 and H2 are overlapping and then their scalar product is null. Histograms H1 and H2 are then representative of two different textures. On the other hand, when C(H1,H2) is very close to 1, the two histograms overlap and are representative of a same region.

6. Conclusion

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 273-278

We have shown in this paper that the key point of two different problems tackled by neural approaches (function approximation and pairing patterns) is in fact the choice of the regularization term in the function which minimized by the neural approach. We believe that this choice is as important as the choice of the neighbourhood function in any local search based method in combinatorial optimization (it is generally agreed that in this case the neighbourhood choice is more important than the «hill climbing» method. Therefore the issue is now to define which are the properties which make efficient the regularization term. We will try to address this problem in further studies but we have first to solve a remaining open question in our application: How to select significant changes among those which are exhibited. It is clear that the significance of the differences in our application is task specific though man-made changes are more important than changes caused by natural factors such as seasonal changes.

References

[1] Fu L., Chen.T., Sensitivity analysis for input vectors in multilayer feedforward neural network. IEEE Int Conf on N Network, San Francisco, 1993

[2] Dimitropoulos, Bourret, Lek. Use of some sensitivity criteria for choosing networks with good generalization ability. Neural Processing Letters, Vol 2 N°6 pp1-4, 1995.

[3] Girosi F., Jones M., Poggio T. Regularization Theory and Neural Networks A rchitectures. Neural Comp. 7, 219-269, 1995.

[4] Yuille A. Generalized deformable models, statistical physics, and matching problem. Neural Comp. 2, 1-24 (1990).

[5] A. Bouju. Labelling by Adaptative Mapping. Neuro-Nimes93. 291-298.

[6] Boeres M.C, A faster elastic-net algorithm for the traveling salesman problem, IJCNN, Baltimore, pp215-220, 1992.

Fig.1. Original images (512x512)

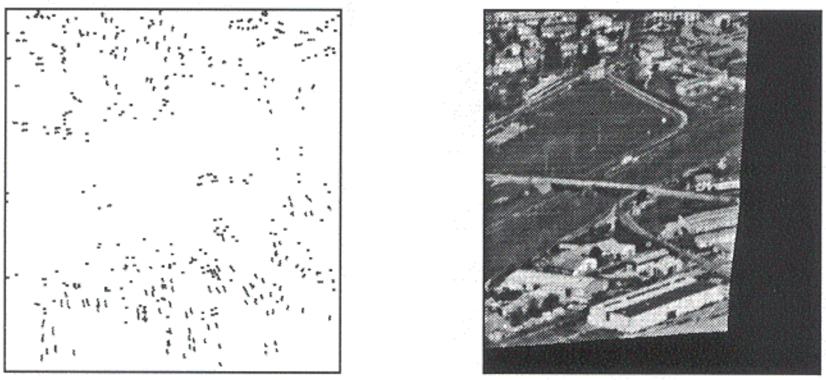

Fig.4. a) Some iterations of the matching process, b) Registration of the first image