Correspondence

Another Move Toward the Minimum Consistent Subset: A Tabu Search Approach to the Condensed Nearest Neighbor Rule

Vicente Cerverón and Francesc J. Ferri

***Abstract*—This paper presents a new approach to the selection of prototypes for the nearest neighbor rule which aims at obtaining an optimal or close-to-optimal solution. The problem is stated as a constrained optimization problem using the concept of consistency. In this context, the proposed method uses tabu search in the space of all possible subsets. Comparative experiments have been carried out using both synthetic and real data in which the algorithm has demonstrated its superiority over alternative approaches. The results obtained suggest that the tabu search condensing algorithm offers a very good tradeoff between computational burden and the optimality of the prototypes selected.**

***Index Terms*—Multiple-prototype classifiers, nearest neighbors, prototype selection, tabu search.**

I. Introduction

Among other nonparametric approaches, distance-based classification rules are especially appealing both to researchers and practitioners because of their interesting properties for performance and implementation issues. In particular, the (k-)nearest neighbor (NN) rules frequently appear in the specialized literature either by themselves, or as a common reference to compare the performance of other approaches. This is due not only to their very good asymptotic behavior—even for the plain NN (1-NN) rule—but also for the convenient trade-off between ease of implementation and performance in practical situations.

The only requirements of these rules are 1) a representative set of labeled samples or prototypes and 2) a procedure to identify the k closest prototypes to an unknown sample. With these two requirements, the most voted class among these neighbors is assigned to the sample. Nevertheless, these requirements may give rise to some drawbacks. Even though it is usually easy to obtain a convenient and representative set of prototypes, sometimes these sets may contain erroneous or noisy prototypes that may produce a decrease in performance. Furthermore, *representative* often implies large, and consequently, the search for neighbors becomes a computationally expensive task.

Many approaches have been suggested to overcome these drawbacks of the NN rules. Among these approaches, prototype selection (PS) techniques aim at obtaining a convenient set of prototypes from an initially given set in such a way that the plain 1-NN rule using these selected (and usually reduced) set of prototypes gives classification results which are good enough. To date, different approaches to PS have been proposed and have been named differently by different authors. A distinction is usually made [1] between *editing* and *condensing* techniques. The main goal of editing techniques is to improve the performance of the resulting classifier by discarding outliers and cleansing the overlap among classes. The main goal of condensing techniques is to reduce the number of prototypes. There are also approaches that share in some way the aims of editing and condensing [2], [3].

In the framework of condensing, it is convenient to consider methods in which the resulting set is constructed or adapted from the initial one. This is referred to as prototype *replacement* [4], [5] instead of prototype *selection* to emphasize the fact that resulting prototypes do not necessarily coincide with any prototype in the initial set. Prototypes obtained using both approaches are then referred to as R-prototypes and S-prototypes, respectively.

Regardless of the way prototypes are obtained, the criterion for guiding the reduction of prototypes (trade-off between size reduction and performance) must be decided. Most of the approaches proposed share the concept of *consistency* [6]. A resulting set of prototypes is said to be consistent with an initial set if it can classify all initial prototypes using the 1-NN rule with no errors. If PS is considered as a training process which uses the initial set of prototypes, consistency can be related to a training result (a set) which produces a zero *resubstitution* error rate.

This work is focused on developing a new method for obtaining consistent sets of S-prototypes. To do this, we face a hard combinatorial problem for which a lot of heuristic and approximate algorithmic solutions have already been proposed. This paper is organized as follows. In Section II, several existing approaches to PS are included and their main properties are put forward. In Section III, the proposed condensing method is introduced. The comparative experimentation carried out using the proposed method is shown in Section IV along with the corresponding discussion. Section V includes some final remarks.

II. Consistency-Based Prototype Selection

A. Condensing Algorithm

Prototype selection techniques were first introduced by Hart [6] who presented the *condensing* algorithm to obtain *reduced* and *consistent* sets of prototypes to be used with the 1-NN rule without significantly degrading its performance. In short, this plain condensing algorithm proceeds by repeatedly selecting prototypes whenever they cannot be correctly classified by the currently selected set. The whole process loops until there are no changes in a complete pass through the initially given set. This straightforward algorithm usually converges in three or four iterations leading to a reduced set which clearly improves the efficiency of the resulting classifier. The resulting condensed NN rule can be seen as an approximation to the plain 1-NN rule (using all the prototypes) in terms of classification performance. The final size and composition of the final condensed set may strongly depend on the order in which the initial set is processed.

Hart left some open questions about this algorithm and, in particular, about the accuracy of the resulting classifier. There is no theoretical evidence about how the consistency of the condensed set relates to the generalization abilities of the corresponding classifier. In fact, even if an optimally or *pure* condensed set (in the sense of preserving even the *same* classification boundaries) is obtained, it may have an arbitrarily poor performance when applied to independent samples (exactly as with the 1-NN rule). It is worth mentioning that this may happen regardless of the representativity and size of the initial set of prototypes if overlapping among classes is large. This is the reason why some authors assume that condensing should always be applied to previously edited or overlapping-free sets [1].

Manuscript received May 10, 1999; revised February 3, 2001. This work was supported in part by Spanish Projects TIC98677-C02-02, 1FD97-279, GV98-14-124, and TIC2000-1703-C03-03. This paper was recommended by Associate Editor B. J. Oommen.

The authors are with the Department Informàtica, Universitat de València, Burjassot, Spain (e-mail: cerveron@uv.es; ferri@uv.es).

Publisher Item Identifier S 1083-4419(01)04853-1.

B. Heuristic Improvements to Plain Condensing

Once consistency has been adopted as the criterion to guarantee the final performance, most proposed approaches aim at obtaining a consistent subset which is as reduced as possible. Regardless of the quality that consistency implies, the obtainment of the *minimal consistent subset* is a challenging combinatorial problem in itself [7], [8].

One initial idea that has been exploited by different approaches (either in an explicit or implicit way) is the fact that Hart's condensing retains less prototypes if the ones near the classification boundaries are taken into account first. Among other approaches, the MNV-based condensing [9] uses the so-called mutual neighborhood value (MNV) to pre-order the prototypes according to their closeness to the classification boundaries. Basically, the MNV of a prototype, p, is computed as follows. Let q be its nearest prototype from a different class—also referred to as *nearest unlike neighbor* (NUN) [8]—and then compute the k-neighborhood of q for a value of k which is large enough. If p is the ith neighbor of q, then the MNV of p is $1 + i$. By its nature, the MNV condensing algorithm leads to a good reduced set in the sense that it preserves the original classification boundaries by retaining only prototypes which are very close to them. Unfortunately, both the final size and the boundary quality are far from optimal and may arbitrarily vary depending on the particular set and the degree of overlapping.

The approach introduced by Dasarathy [8] uses the concept of NUN and a particular measure of prototype representativity based upon the number of samples each prototype is able to classify correctly. At each iteration, every prototype casts a vote for any prototype of the same class which is closer than its NUN. The more votes a prototype receives, the higher its representativity. Prototypes are then selected according to this ranking until consistency is achieved. The algorithm starts with the whole set and iterates while the selected subset size decreases. In this way, it generally obtains high quality subsets with lower cardinality than the previous approaches regardless of the initial ordering of prototypes.

C. Randomized Optimization Applied to the PS Problem

Although the algorithm of Dasarathy was named minimal consistent set (MCS), no proof of this was given. In fact, experimentation has shown that this is not the case [5]. As many other combinatorial problems, the MCS problem would require (implicit or explicit) exhaustive search to obtain optimal solutions in the general case. This has driven some researchers to envisage the problem of PS as a combinatorial optimization problem and to use general techniques that are known to perform well in similar situations. In particular, random search (RS) and genetic algorithms (GAs) have already been proposed to solve this problem [5]. As the problem can be seen as a multiobjective optimization problem (minimizing the subset size and minimizing the number of classification errors), a composite fitness function can be adopted to carry out both minimizations. The proposed fitness function for the problem seen as a maximization problem is [5]

$$f(s) = \text{acc}(S) - \alpha|S| \tag{1}$$

where

S	corresponding subset of prototypes;
$\text{acc}(S)$	accuracy or ratio of correct classified prototypes in the whole set;
α	positive weighting factor.

The corresponding search algorithm can be forced to obtain subsets of a given size T by using the alternative fitness function

$$f(s) = \text{acc}(S) - \alpha(|S| - T)^2. \tag{2}$$

The results reported suggest that GAs compete well with other heuristics. Even the RS approach, which was implemented in [5] using the GA with the genetic operators disabled, outperformed other previously proposed algorithms.

A random restarting Hart approach has also been recently proposed [10]. Hart's algorithm is conveniently (and efficiently) used in this approach to constrain (only to consistent subsets) a kind of random search using the same equations (1) and (2).

III. Alternative Approach to Condensing Based on Tabu Search

Tabu search (TS) [11] can be regarded as a master strategy that guides and modifies other heuristics to produce solutions beyond those that are normally generated in a quest for local optimality. The emphasis on guidance distinguishes TS from other metaheuristics based on a variety of concepts such as GAs and simulated annealing. All these techniques perform a kind of "intelligent search" over the solution space which attempts to avoid local minima on the way to a globally optimal solution.

In the particular case of TS, this search is done in a very efficient and straightforward way. TS has already been successfully applied to other hard combinatorial problems in pattern recognition [12]. Also, a first attempt to apply TS to PS was included in a previous work [10].

Tabu search can be viewed as a way of traveling across a particular solution space and visiting different solutions following a certain neighborhood definition while discouraging some particular ways of exploring this neighborhood. Neighbor solutions are identified with the concept of move which refers to a slight modification of a given solution to obtain another one. Some moves (or, in fact, move attributes) are declared tabu (or undesirable) during a given number of iterations (tabu tenure). TS evaluates all possible moves from a given solution and proceeds to the best one. Tabu moves cannot be considered unless they satisfy an *aspiration criterion* which usually consists of improving the best solution at that point (*improved-best* aspiration criterion). The use of appropriate neighborhoods leads to a convenient way of searching the space while the use of meaningful tabus prevents the algorithm from being trapped in local minima. A generic description of TS is as follows:

```
Generic Tabu Search
Input: a solution space, an objective
  function, a set of possible moves, a
  tabu tenure, an aspiration criterion, a
  termination criterion, and an initial-
  ization procedure.
Output: a (close to) optimal solution.
Method: obtain an initial solution and
  repeat the following steps until the
  termination criterion is satisfied.
1) Evaluate all neighbor solutions.
2) Select the best neighbor solution
  without considering tabu ones unless
  they satisfy the aspiration criterion.
3) Declare tabu the attributes of the move
  that have led to the selected solution
  for a specified number of iterations
  (tabu tenure).
4) Update the best overall solution if the
  selected one is better.
```

In the proposed particularization of TS, all possible prototype subsets constitute the solution space. Possible moves from a particular

subset consist of adding or deleting each one of the n initial prototypes. The attribute used for declaring tabu moves is the prototype which is added or deleted. The improved-best aspiration criterion is used. The objective functions considered are the same as in the GA approach [(1) and (2)].

Additional parameter values and options common to other TS implementations have to be set. In particular, the way of initializing the search may strongly affect the final performance of the algorithm. Two different initializations have been considered in this work: *condensed* and *constructive*. In the first one, Hart's algorithm is applied once to obtain a first subset. In the second one, the initial consistent subset is obtained by applying TS with sample deletion disabled starting from a randomly picked prototype from each class.

In contrast to other PS approaches, the proposed TS algorithm (as RS and GA) offers the possibility of obtaining good subsets of prototypes which are not necessarily consistent. By storing the best slightly inconsistent sets during the search process, the algorithm can identify good subsets of prototypes adding flexibility to this PS approach. Also, the search for the MCS can be substituted by the search for the best subset of a given size [by using (2)] which allows the algorithm to obtain good prototype sets with a certain level of inconsistency.

In our implementation, ties are randomly broken, which in practice implies that our algorithm is a randomized approach regardless of the initialization used (as RS and GA). Also, as TS always performs *moves* from one solution to a neighbor one, subset evaluation is implemented in a more efficient way by incrementally maintaining some information (the NN of each sample).

IV. Experimental Results

A number of experiments were conducted to assess the abilities of the proposed condensing algorithm. A large number of different settings for the different parameters were tried and the following one was selected: $\alpha = 0.002$ in the objective functions, a tabu tenure equal to 10% of the initial set size, and 200 iterations without improvement as a stopping criterion (which, in our experiments, roughly corresponds to 500 iterations). Small variations on these parameters led to approximately the same results. From our results, the condensed initialization led to faster convergence while the constructive one took more time but managed to improve the solution slightly further. In other words, constructive initialization resulted in a more expensive search but with more chances of arriving at a better solution. The results shown in this section correspond to TS using constructive initialization unless otherwise specified.

Other algorithms were also considered for comparison purposes. In particular, Hart and MNV condensings, MCS algorithm, and restarting Hart [10], along with GA and RS approaches [5]. Special care was taken to allow roughly the same time spent to find a solution in the algorithms in which this was possible (GA, RS, and restarting Hart). As a general result, the TS always outperformed any other approach in all our experiments.

A. Synthetic Database

To illustrate the behavior of the algorithm introduced in comparison with the classical ones, some experiments using the synthetic problem proposed by Hart [6] were carried out. Random sets of 482 integer-valued two-dimensional (2-D) vectors were generated (with uniform probability) and then labeled according to the two disjoint support sets that are shown in Fig. 1 as shadowed areas (class 1) and unshadowed areas (class 2). A total of ten different random subsets were considered.

The Hart's condensing, its MNV extension, and the MCS algorithm were considered for this first experiment along with the random restarting Hart procedure and TS condensing with constructive initialization. One of the 482-prototype sets considered along with the corresponding classification boundary induced by the 1-NN rule is shown in Fig. 1(a). The boundaries resulting after applying the five algorithms considered are shown in Figs.1(b)–(f). It can be observed that the subset in Fig. 1(b) is the only one that clearly contains redundant prototypes (in the sense of their contribution to the actual 1-NN boundary). On the other hand, it is worth noting that the MNV-based condensing really selects prototypes which are (very) close to the (ideal) boundaries, and more importantly, this fact does not necessarily imply a better result in terms of the final size or the decision boundaries obtained.

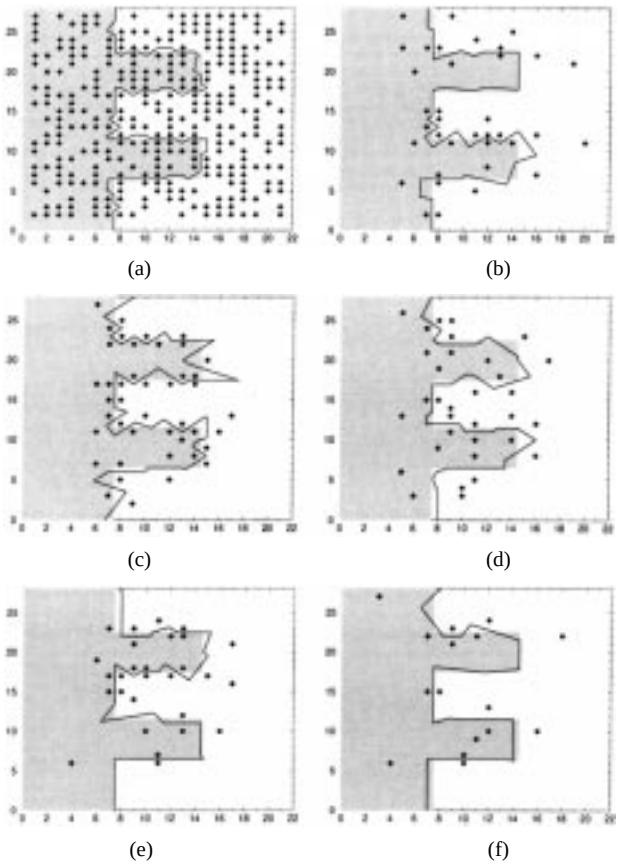


Fig. 1. Decision boundaries induced by different subsets of prototypes. (a) Original set. (b) Hart's condensing. (c) MNV-based condensing. (d) Dasarathy's MCS. (e) restarting Hart. (f) tabu search condensing.

The last three algorithms led to subsets exhibiting roughly the same properties (relatively smooth boundaries, with prototypes in arbitrary locations with regard to the decision boundaries). The key difference was the final size of the subsets. Both restarting Hart and TS condensing algorithms obtained fewer prototypes than the MCS algorithm. In the case of TS condensing, the number of prototypes selected, 16, was much smaller than any other result and, moreover, the corresponding 1-NN boundary was very close to the optimal result.

A slight modification of the previous synthetic experiment was also considered. Sets of 500 *real* valued vectors were randomly drawn using the same procedure as in the integer-valued case. In this way, a number of ties are avoided and the corresponding underlying classification problem, although (essentially) disjoint [6], looks more realistic. The corresponding average sizes that were obtained when applying the algorithms over ten different random sets for each version of the problem is shown in Fig. 2 along with the corresponding standard deviations and the minimal size.

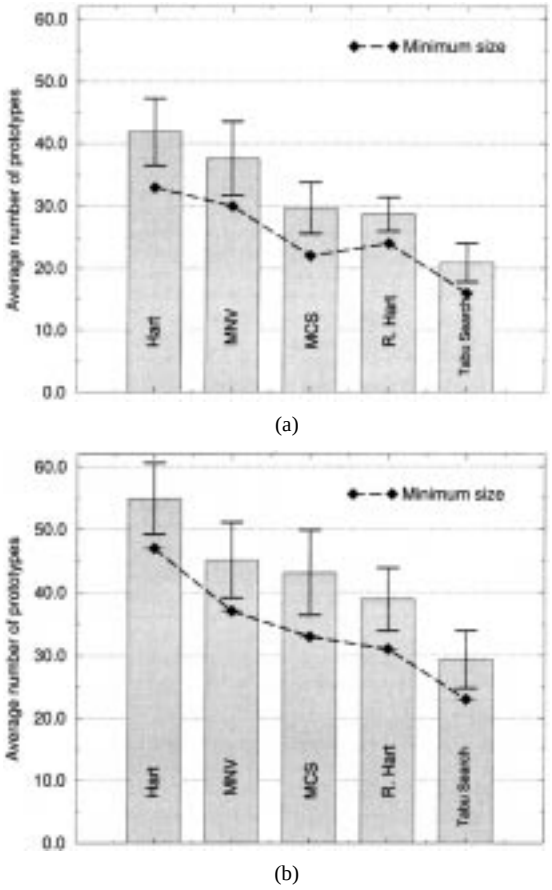


Fig. 2. Minimum and average number of prototypes with standard deviations obtained by different algorithms on (a) integer-valued and (b) real-valued versions of Hart's problem.

The differences among different approaches were reduced in the real-valued version of this experiment. This is due to the fact that as prototypes appear at arbitrary locations there is less chance of *optimal* solutions among the original random samples. Nevertheless, the proposed method was still better by a significant difference.

B. Iris Database

The second experiment dealt with real (instead of synthetic) data. One of the goals was to compare the above approaches to genetic algorithms for prototype selection. As particular implementations of GA may differ and their fine tuning may be done differently, we decided not to implement the GA approach but rather to use the recently published results using the Iris database [5].

The Iris problem consists of three classes of 50 four-dimensional (4-D) vectors each, corresponding to three subspecies of iris flowers [13]. This problem has often been used for benchmarking and, in particular, it has been considered for prototype selection algorithms [5], [8], .

To obtain more reliable results given the random nature of the algorithms proposed, 30 different runs using different random seeds were carried out for random restarting Hart and TS using both condensed and constructive initializations. The (deterministic) algorithms MCS and MNV-ordered condensing were also considered in this experiment. The minimum and average sizes obtained are shown in Fig. 3.

As in the previous experiment, it can be seen that there was a substantial difference in the results between the proposed algorithm (and the GA approach) and the other algorithms considered. Besides its randomness, the relatively small variance obtained (this parameter for the GA

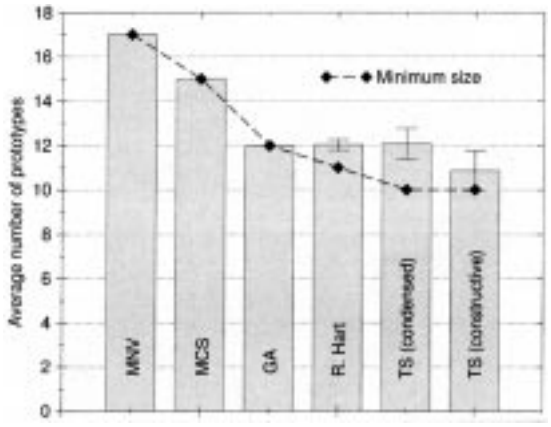


Fig. 3. Minimum and average number of prototypes with standard deviations obtained by different algorithms on the Iris database.

TABLE I
RESULTS OBTAINED WITH THE IMPLEMENTED ALGORITHMS ON THE IRIS DATABASE WITH THE TABLE SHOWING THE NUMBER OF TIMES A CARDINALITY WAS OBTAINED IN 30 INDEPENDENT RUNS ALONG WITH THE CORRESPONDING AVERAGE AND STANDARD DEVIATION

	10	**11**	**12**	**13**	**>13**	Avg (stdev)
R. Hart	0	3	23	4	0	12.03 (0.24)
TS (condensed)	1	3	19	7	0	12.07 (0.69)
TS (constructive)	12	12	4	2	0	10.87 (0.90)

TABLE II
TWO TEN-PROTOTYPE CONSISTENT SUBSETS CORRESPONDING TO THE FOUR-DIMENSIONAL IRIS DATABASE OBTAINED WITH TABU SEARCH WITH CONDENSED AND CONSTRUCTIVE INITIALIZATIONS, RESPECTIVELY

class 1	class 2	class 3
4.6 3.2 1.4 0.2	6.7 3.1 4.4 1.4	6.0 2.2 5.0 1.5
	6.3 2.5 4.9 1.5	6.3 2.7 4.9 1.8
	6.0 2.9 4.5 1.5	6.3 2.8 5.1 1.5
	6.0 2.7 5.1 1.6	6.1 2.6 5.6 1.4
		5.8 2.7 5.1 1.9

class 1	class 2	class 3
5.0 3.3 1.4 0.2	6.3 2.5 4.9 1.5	6.3 2.9 5.6 1.8
	6.0 2.9 4.5 1.5	6.0 2.2 5.0 1.5
	6.0 2.7 5.1 1.6	6.3 2.7 4.9 1.8
	6.7 3.1 4.7 1.5	6.3 2.8 5.1 1.5
		5.8 2.7 5.1 1.9

was not available) suggests that the algorithms robustly achieve close to optimal results. The TS approach was the only one that managed to obtain ten-prototype consistent subsets. In particular, when using the constructive initialization, the chances of obtaining a ten-prototype subset were about 40%. Table I summarizes the results obtained classified by the cardinality of the subsets.

From the 12 subsets of cardinality ten obtained by TS with constructive initialization, only eight were different. The ten-prototype subset obtained with the condensed initialization was also different from the other 12 subsets. Table II shows this subset along with the most frequently obtained subset using constructive initialization.

A similar set of runs was carried out using (2) and storing all good solutions to a certain level of inconsistency. Fig. 4 graphically summarizes the best results (subset sizes) obtained with all the algorithms considered for different levels of inconsistency (number of classification errors in the original set). It is worth mentioning that the results obtained with TS consistently improved the ones previously published in the literature using the Iris database. With two errors, the GA, RS, and TS approaches managed to obtain the same result, but with 1 or 0

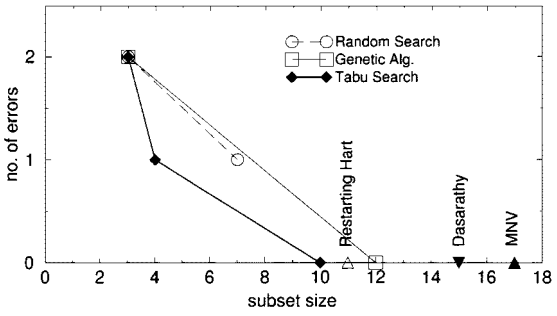
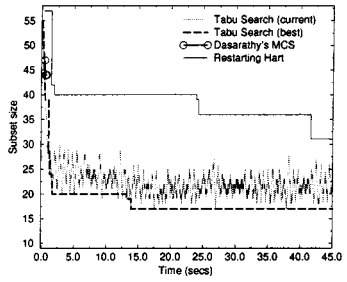
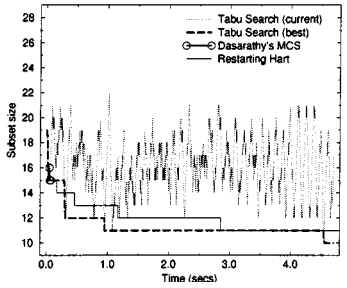


Fig. 4. Subset sizes and number of classification errors in the original prototype set using the different condensing methods considered in the Iris problem.



(a)



(b)

Fig. 5. Subset size versus time spent (in a Pentium II-based workstation at 300 MHz for (a) real-valued version of Hart's problem and (b) Iris database.

(consistent) errors, the results obtained with TS are, to our knowledge, the best to date.

The computation needed by TS and, especially by the restarting Hart procedure, is greater than for the other algorithms implemented. This is the price for obtaining the best solutions. If obtaining a very close to optimal solution is not important, both proposed algorithms can be speeded up by early stopping. The way in which these algorithms improve the current solution is shown in Fig. 5 for a particular run on data from both experiments. It can be seen that both algorithms improved the result of the MCS algorithm (the best from the ones implemented) in the first iterations using about the same time for the Iris problem and twice the time for Hart's problem. From that point on, the algorithms needed relatively much more time to improve the current solution. In the case of TS, it was observed that a very good solution was always obtained in the first iterations. This suggests that by using a conservative termination criterion, this algorithm could compete both in performance and efficiency against most alternative approaches including the GA. Fig. 6 shows the average times necessary for arriving to different cardinalities using the Iris database. The figure shows that restarting Hart and TS with condensed initialization arrived at the same solution as the MCS algorithm in similar time and kept improving the solution faster than TS with constructive initialization. Nevertheless, the constructive initialization is able to obtain better solutions much earlier and much more likely (see also Table I).

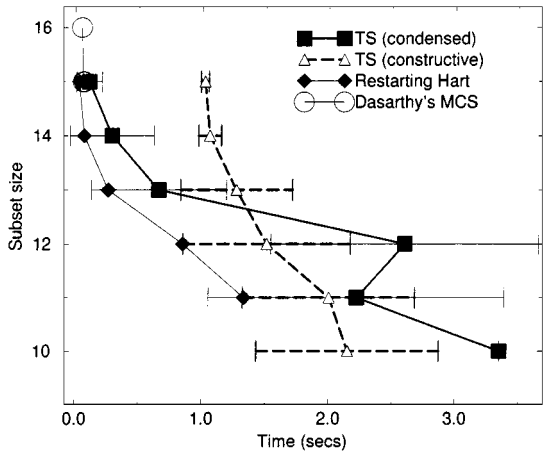


Fig. 6. Average cpu times to arrive at different subset sizes for the algorithms proposed on the Iris problem.

It is possible to approximately compare our algorithms (and TS in particular) to the GA approach using the Iris database in terms of computational burden. From the results reported [5], between 5000 and 10 000 subset evaluations are needed to obtain a 12-prototype set with the GA. Our TS needed about 70 000 subset evaluations to arrive at the (until now) optimal ten-prototype set (and about 50 000 to obtain 12 prototypes). Taking into account that TS almost always evaluates subsets of very low cardinality while the GA needs to start with sets of about 80% the total size [5], the computation needed by both methods for these problems must be similar.

V. Concluding Remarks and Further Work

An attempt to apply TS to a hard combinatorial optimization problem such as the MCS has been presented. According to the comparative experiments carried out, well-adapted metaheuristics can lead to a good trade-off for solving this problem. TS seems to be a very appealing alternative because of its flexibility and its efficiency in the search for global optima.

We believe that TS as implemented here, adapts to this problem in such a way that it can obtain a very good solution in reasonable time. At this point, provided there is enough time, there is a great probability that the search for the (global) minimum consistent subset will be successful.

In this paper, a straightforward implementation of TS has been used. Many specific modifications and improvements have been left for future work. As an immediate challenge, an exhaustive empirical comparison between TS and GAs (simulated annealing could possibly be considered as well) should be done using a set of databases which covers a range of practical situations (overlapping levels, dimensionalities, initial sizes, etc.).

References

[1] P. A. Devijver and J. Kittler, *Pattern Recognition. A Statistical Approach.* Englewood Cliffs, NJ: Prentice-Hall, 1982.

[2] T. Kohonen, *Self-Organization and Associative Memory*, 2nd ed. New York: Springer-Verlag, 1988.

[3] U. Lipowezky, "Selection of the optimal prototype subset for 1-NN classification," *Pattern Recognit. Lett.*, vol. 19, no. 10, pp. 907–918, 1998.

[4] J. C. Bezdek *et al.*, "Multiple prototype classifier design," *IEEE Trans. Syst., Man, Cybern. C*, vol. 28, pp. 67–79, Jan. 1998.

[5] L. I. Kuncheva and J. C. Bezdek, "Nearest prototype classification: Clustering, genetic algorithms, or random search," *IEEE Trans. Syst., Man, Cybern. C*, vol. 2, pp. 160–164, Jan. 1998.

[6] P. E. Hart, "The condensed nearest neighbor rule," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 515–516, May 1968.
[7] G. W. Gates, "The reduced nearest neighbor rule," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 431–433, May 1972.
[8] B. V. Dasarathy, "Minimal consistent set (MCS) identification for optimal nearest neighbor decision systems design," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, pp. 511–517, Mar. 1994.
[9] K. Chidananda-Gowda and G. Krishna, "The condensed nearest neighbor rule using the concept of mutual nearest neighborhood," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 488–490, July 1979.
[10] V. Cerverón and A. Fuertes, "Parallel random search and tabu search for the minimal consistent subset selection problem," *Lecture Notes Comp. Sci.*, vol. 1518, 1999.
[11] F. Glover and M. Laguna, *Tabu Search*. New York: Kluwer, 1997.
[12] K. S. Al-sultan, "A tabu search approach to the clustering problem," *Pattern Recognit.*, vol. 28, no. 9, pp. 1443–1451, 1995.
[13] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.*, vol. 7, pp. 178–188, 1936.

An Elastic Contour Matching Model for Tropical Cyclone Pattern Recognition

Raymond S. T. Lee and James N. K. Liu

***Abstract*—In this paper, an elastic graph dynamic link model (EGDLM) based on elastic contour matching is proposed to automate the Dvorak technique for tropical cyclone (TC) pattern interpretation from satellite images. This method integrates traditional dynamic link architecture (DLA) for neural dynamics and the active contour model (ACM) for contour extraction of TC patterns. Using satellite pictures provided by National Oceanic and Atmospheric Administration (NOAA), 120 tropical cyclone cases that appeared in the period from 1990 to 1998 were extracted for the study. An overall correct rate for TC classification was found to be above 95%. For hurricanes with distinct "eye" formation, the model reported a deviation within 3 km from the "actual eye" location, which was obtained from the aircraft measurement of minimum surface pressure by reconnaissance.**

Compared with the classical DLA model, the proposed model has simplified the feature representation, the network initialization, and the training process. This leads to a tremendous improvement of recognition performance by more than 1000 times.

***Index Terms*—Active contour model (ACM), elastic graph dynamic link model, elastic graph matching, satellite images, tropical cyclone pattern recognition.**

I. Introduction

In the past half century, satellite technologies have been extensively applied in various fields, ranging from the military operations to the search and discovery of natural resources. One of the most important applications of satellite interpretation technology is the identification of tropical cyclones (TC)—including hurricanes and typhoons—which with their remarkable spiral shape and central eye, are the most critical meteorological phenomenon to affect our daily lives. Extensive research has been conducted to estimate the movement and intensity of tropical cyclones from satellite images. One of the most widely accepted techniques is the Dvorak technique [5], [6], which assigns a wind intensity value (called the TC number) based on the size, shape, and vorticity of the dense cloud shield adjacent to the center of the storm.

Owing to the high variation of cloud patterns and lack of efficient scene analysis techniques for the isolation and extraction of cloud systems from satellite pictures, the TC pattern matching jobs in Dvorak analysis are so far all done by subjective human justification. There is no successful alternative technique to support pattern recognition automatically in Dvorak analysis [20], let alone with the automatic identification for the position of the "eye" in hurricanes and typhoons [19].

In this paper, an elastic graph dynamic link model (EGDLM) is used to provide an automated pattern matching solution in Dvorak analysis. Based on the extension of dynamic link architecture (DLA) as a neural framework and its integration with the active contour model (ACM) [3], [9], [19] for the contour extraction of TC patterns, the sophisticated pattern-matching problem is simplified into an elastic graph matching problem of TC contour patterns.

Section II provides a brief discussion of the Dvorak technique for tropical cyclone identification. Section III provides an overview of DLA, the preliminary study on TC recognition and the major limitations. Section IV gives an overview of the framework of the EGDLM. In the implementation process, 120 cases of tropical cyclones during the period from 1990 to 1998 are chosen for system testing. Various test plans are done for system verification; these will be presented in Section V. The paper will conclude with a brief discussion on the overall performance in the final section.

II. TC Identification Using Dvorak Technique

During recent decades, the most important technique for the identification and classification of TC from satellite pictures is the Dvorak technique [5], [6]. Based on Dvorak's theory, each tropical cyclone goes through a life cycle that may be classified into one of several types by its appearance in visible images. Fig. 1 shows the "templates" used in the Dvorak technique.

In addition to classifying the storm, Dvorak's technique can be used to determine TC strength from the satellite images using "T-numbers" (T1–T8) as reference. By comparison with aircraft-observed wind intensity, the Dvorak technique has a rms. error of approximately 6 ms-1 in tropical cyclone intensity.

In 1984, Dvorak [7] introduced a variant of the above technique, called the enhanced infrared EIR technique, which uses specially enhanced infrared images instead of visible ones. This, of course, enables wind intensity to be estimated at night.

Nowadays, the Dvorak technique is still the worldwide-agreed standard for the determination of TC intensity. However, due to the high variation of TC patterns, the visible and enhanced infrared Dvorak techniques are subjective, requiring professional training to be done effectively for good wind estimates.

III. Dynamic Link Architecture—An Overview

A. Introduction

The main idea of DLA was first proposed by von der Malsburg in 1981 as a neuroscience model, namely the "Correlation Theory of the Brain Function" [16], which had been consolidated into a complete neural network framework, namely DLA, in the later years [1], [17].

In short, the DLA model for pattern recognition can be interpreted as the process of elastic graph matching between the memory patterns

Manuscript received January 24, 1999; revised January 9, 2001. This work was supported in part by RGC Grant G-T142 and iJADE Project 4-61-09-Z042 both from the Hong Kong Polytechnic University. This paper was recommended by Associate Editor B. J. Oommen.

The authors are with the Department of Computing, Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: csstlee@comp.polyu.edu; csnkliu@comp.polyu.edu.hk).

Publisher Item Identifier S 1083-4419(01)04859-2.