Why Do We Need The Auxiliary Vector Representation for the Metric Pattern Recognition Problem?

by

Lev Goldfarb

TR90-056, September 1990

Faculty of Computer Science
University of New Brunswick
P.O. Box 4400
Fredericton, N.B. E3B 5A3

Phone: (506) 453-4566
Fax: (506) 453-3566

Abstract

"Direct" pattern recognition algorithms in metric spaces, i.e., those that do not use auxiliary vector space, are inferior to the pattern recognition algorithms that rely on the vector approximation of the original, metric, recognition problem. The critical issues associated with the role of the auxiliary vector space representation for the metric pattern recognition problem are considered.

Key Words: Pattern Recognition in Metric Spaces, Efficiency, Auxiliary Vector Representation.

> [Newton] rightly considered that all computations in analysis are more convenient to perform not with the help of repeated differentiations but by computing first members of the [Taylor] series.
>
> V.I. Arnold. Ordinary differential equations (1984)

1. INTRODUCTION

In this short note I would like to address an important question that, on the one hand, was critical in motivating a new analytical framework for pattern recognition proposed in [(1)-(3)], and, on the other hand, has not been fully appreciated by the pattern recognition community. The question is: *Do we need the auxiliary vector representation for the pattern recognition problem in which the patterns are represented in a metric or pseudometric space*? The relevance and even urgency of the question can easily be ascertained by observing the large number of papers dealing with *diverse* pattern recognition problems formulated in various metric or pseudometric spaces. It is enough to point to (a relatively small sample of) such papers and monographs: speech recognition [(4)-(8)], robot navigation [(9)], 2-D and 3-D shape recognition [(10)-(15)], character recognition [(16)-(18)], waveform recognition [(19)-(20)], structural pattern recognition [(21)-(24)]. Indeed, since the above question should have arisen quite often, the very fact that it has not been fully addressed is quite conspicuous. The larger issues related to the dominant role of the metric formulation of the pattern learning problems are considered in a companion paper [(25)] (see also [(26)]).

I propose that the answer to the above question is yes, and I will argue for this answer by presenting what I believe to be decisive reasons why the auxiliary vector representation is necessary. In fact, I have already presented some arguments to that extent in monograph [(2)]. There are reasons to believe, however, that this issue, perhaps one of the most important issues in pattern recognition today, deserves a separate consideration.

By *auxiliary vector representation* I mean a vector representation of patterns that *metrically approximates* the original pattern recognition problem *with any desired accuracy*. The possibility of an *efficient* construction of such vector representations was addressed in [(2)]. (All other vector representation algorithms that have appeared over the last 25 years are based on inadequate heuristics and cannot compete with the latter systematic approach.) Since the metric information is essentially preserved under the vector representation, in considering the above question it is enough to rephrase it as follows: *What are the advantages (for the recognition algorithms) of the vector space representation over the representation in the more abstract , metric or pseudometric , space?* One should note immediately, that this simplified question in no way implies the superiority of the *direct* vector space pattern representation. On the contrary, the sheer number of applications using various non-vector metric spaces for pattern representation testifies to the overriding superiority and flexibility of the metric space model (see also reference 2, Forward and Chapter 1). In fact, if it were otherwise, the original question should not have arisen in the first place. Of course, the class of vector spaces with distances defined on them form a subclass of metric space.

In considering the simplified but equivalent question, we shall address three critical and related to each other issues: advantages associated with 1) the possibility to modify the original metric of the representation space to reflect the discriminatory information, 2) the relative efficiency of the on-line recognition (classification) algorithms, and 3) the relative quality of the corresponding recognition algorithms.

2. ADVANTAGES ASSOCIATED WITH THE POSSIBILITY TO MODIFY THE ORIGINAL METRIC TO REFLECT THE DISCRIMINATORY INFORMATION

To make the basic point of this section absolutely clear, let us consider a simple *metric* two class pattern recognition problem with class distribution sketched in Figure 1.

Insert Fig. 1 about here

Although the figure shows two-dimensional training set, it is useful to keep in mind that in the corresponding metric space we cannot directly observe this configuration, since the only information available in the metric space about the training set is the corresponding symmetric interdistance matrix.

When the above problem is formulated in the vector space, there is a simple transformation of the space that can fundamentally change the quality (reliability) and the efficiency of the resulting problem solution. This transformation may, for example, project the space onto one of the principal axis. The resulting one-dimensional pattern representation has a "simpler" metric structure which leads to a qualitatively better and more efficient recognition algorithm. There are a number of techniques to obtain the desired transformation [(27), (28)], but the main point is that all such techniques depend on the presence of the vector space structure and cannot be transferred directly to the metric space.

To summarize this section, we can say that the ability to *modify the original metric* of the pattern space to reflect more specific, discriminatory, information about the pattern classes, allows one to induce the *new metric* class configuration for which there are both more reliable and more efficient recognition algorithms.

3. ADVANTAGES ASSOCIATED WITH THE RELATIVE EFFICIENCY OF THE ON-LINE RECOGNITION

It appears that in the most successful biological systems most of the critical functions are subordinated to the efficient function of various "on-line recognition mechanisms". And it is not difficult to understand why: the relative speed of these mechanisms to a large extent determines the outcome of "the struggle for survival". To rephrase it in the engineering terms, one can say that in all on-line recognition systems *the efficiency of the on-line (recognition) stage should dominate the design* of the corresponding recognition systems, simply because by definition the on-line is the predominant mode of operation of such systems. Most of the current (and future) pattern recognition systems are (and will be) on-line systems.

The present experience suggests that the computationally dominant type of operation in the on-line metric recognition system is the distance computation. In view of that, to reduce the on-line recognition time, one is often prepared to use in such systems a reasonable number of special purpose hardware devices for computing the corresponding distances. The number of such devices, however, would still be the critical factor.

It remains to mention that the model proposed in [(1)-(3)] allows one to use the auxiliary low-dimensional vector space representation for computing the vector representation of the incoming pattern during on-line recognition stage. This vector representation, *projection*, is constructed as a *linear combination of the basis vectors corresponding to some chosen training patterns, reference patterns.* The projection is computed simply as a product M•b, where M is a square matrix whose columns are the corresponding training coordinate vectors, and b is a vector whose components are directly computed using the distances from the incoming pattern to the chosen training patterns (see, for example, reference 2, section 7.2). Since all the latter distances can be and, if possible at all, should be computed in parallel, we, obviously, obtain the optimal solution to the metric recognition problem.

Without an auxiliary vector space such optimal solution is possible only for a restricted class of recognition problems, when the pattern classes are "well separated", e.g., the diameter of each class is less than the distance to the closest class. The reason for this is simple: in a general case, before the incoming pattern can be classified, a *sequence* of distances from that pattern to some of the training patterns must be computed, and the order in which the distances are to be computed is decided based on the previously computed distances in the sequence. Moreover, the number of such *sequential* distance computations is greater than or equal to the number of the above *parallel* distance computations. This is because the number of the sequential distance computations is greater than or equal to the reduced "global dimension" of the pattern classes, and this dimension cannot be determined without the use of the auxiliary vector representation.

It is also very important to note that with the help of the auxiliary vector space not only the recognition of new patterns but their recollection can be accomplished in the most economical and uniform manner, that can be viewed as a generalization of the Fourier representation: for each new pattern it is enough to store only its "projection coefficients" with respect to the chosen "reference" training patterns. Thus, any new pattern belonging to one of the learned classes can be represented by a *very small* set of its coefficients.

4. ADVANTAGES ASSOCIATED WITH THE RELATIVE QUALITY OF THE RECOGNITION ALGORITHMS

In this section the quality of the recognition algorithms is interpreted to include both the quality, as estimated by the value of the (probalistic) average risk, and the reliability, as estimated by the probability of achieving given average risk (Reference 27, pp. 38-39; see also reference 28, section 1.6).

Thus, the capability to learn is characterized by two concepts:

1) by the quality of the obtained decision rule (by the probability of the wrong answers; the lower this probability, the higher the quality);
2) by the reliability of obtaining the decision rule with the given quality (by the probability of obtaining the given quality; the higher this probability, the higher the reliability of the successful learning).

The problem reduces to the creation of the learning device that from the learning sequence would construct the decision rule whose quality with the given reliability would be not lower than the required [i.e., the required reliability is fixed - L.G.]

So, the designer of the learning devices is faced with two problems:

With which set of the decision rules the learning device is to be supplied;

How to choose the required rule from the set of decision rules.

The main point that I would like to make in this section relates to the absence of any a priori *classes* of global and efficient non-heuristic decision rules (such as linear, quadratic, etc.) in the sufficiently large metric space without the vector space structure. In fact, this is not surprising, since the tremendous growths of all branches of mathematics has been associated with the introduction of "coordinate spaces" by R. Descart and P. Fermat in the 17th century.

The absence of any useful global decision rules is easy to understand if we remember that in a metric space *we do not have direct access to the global information*: the only information directly available to us is the pairwise distances, and *there is no convenient analytical language for specifying non-local decision rules*. Thus, there are no statistically useful *global* decision classes in a metric space, the fact that has the decisive effect on the statistical quality of the recognition algorithms.

5. CONCLUSION

The metric model clearly appears to be the unifying framework for all applications of pattern recognition: speech recognition, shape recognition, character recognition, waveform recognition, machine vision, etc. On the other hand, almost all papers dealing with the metric formulation of various pattern recognition problems do not go beyond the introduction of the metric and the *inefficient* metric nearest neighbor recognition algorithms. This may have contributed to the slow appreciation of the metric model as the superior pattern learning framework.

In this paper I have summarized several important arguments in favour of proceeding to the solution of the metric recognition problems via *the systematic use* of the auxiliary vector space representation proposed in [1]-[3]. The latter analytical model can be viewed as a further elaboration of the leading theme in mathematics -

approximation. Specifically, the distributions in the metric space are efficiently and accurately approximated by the distributions in the low-dimensional vector space.

All the arguments presented in this paper follow from purely analytical considerations. The same considerations should make it quite clear that no heuristic solutions to the metric recognition problem, that are beginning to appear, can compete with the systematic solutions via the auxiliary vector approximation.

Acknowledgement - I would like to thank Bill Knight for a useful discussion.

REFERENCES

1. L. Goldfarb, A unified approach to pattern recognition, *Pattern Recognition* 17, 575-582 (1984).

2. L. Goldfarb, A new approach to pattern recognition, *Progress in Pattern Recognition* 2, L. N. Kanal and A. Rosenfeld, Eds., North-Holland, Amsterdam (1985).

3. L. Goldfarb, Metric data models and associative memories, Proc. IASTED/AFCET *Int. Symp. on Robotics and Artificial Intelligence/Identification and Pattern Recognition*, Vol. 4, pp. 53-73, Universite Paul Sabatier, Toulouse, France (1986).

4. T.K. Vintsiuk, *Analysis, Recognition, and Interpretation of Speech Signals*, Naukova Dumka, Kiev (1987).

5. N.G. Zagoruyko, *Recognition Methods and Their Application*, Sovetskoye Radio, Moscow (1972).

6. H. Sakoe and S. Chiba, Dynamic-programming algorithm optimization for spoken-word recognition, *IEEE Trans.* **ASSP-26**, 43-49 (1978).

7. M.J. Hunt, M. Lenning, and P. Mermelstein, Use of dynamic programming in a syllable-based continuous speech recognition system, *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparisons*, D. Sankoff and J.B. Kruskal, Eds., Addison-Wesley, Reading, Massachusetts (1983).

8. E. Vidal and M.J. Lloret, Fast speaker-independent recognition of isolated words using a metric-space search algorithm (AESA), *Speech Commun.*, No. 7, 417-422 (1988).

9. H. Suzuki and S. Arimoto, Visual control of autonomous mobile robot based on self-organizing model for pattern learning, *Journal of Robotic Systems* 5, 453-470 (1988).

10. T. Pavlidis, Analysis of set patterns, *Pattern Recognition* 1, 165-178 (1968).

11 R.L. Kashyap and B.J. Oomen, A geometric approach to polygonal dissimilarity and shape matching, *IEEE Trans.* **PAMI-4**, 649-654 (1982).

12. S.K. Parui, S.E. Sarma, and D.D. Majumder, How to describe shapes using the shape vector, *Pattern Recognition Letters* 4, 201-204 (1986).

13. P. Cox, H. Maitre, M. Minoux, and C. Ribeiro, Optimal matching of convex polygons, *Pattern Recognition Letters* 9, 327-334 (1989).

REFERENCES con't

14. M. Oshima and Y. Shirai, Object recognition using three-dimensional information, *IEEE Trans.* **PAMI-5**, 353-361 (1983).

15. Y. Sato and I. Honda, Pseudodistance measures for recognition of curved objects, *IEEE Trans.* **PAMI-5**, 362-273 (1983).

16. D.J. Burr, Designing a handwriting reader, *IEEE Trans.* **PAMI-5**, 554-559 (1983).

17. C.C. Tappert, Adaptive on-line handwriting recognition, *Proc. 7th Int. Conf. Pattern Recognition*, 1004-1007 (1984).

18. H. Yamada, Contour DP matching method and its application to handprinted Chinese character recognition, *Proc. 7th Int. Conf. Patterns Recognition*, 389-392 (1984).

19. Y.-C. Cheng and S.-Y. Lu, Waveform correlation by tree matching, *IEEE Trans.* **PAMI-7**, 299-305 (1985).

20. J.J. Liang and V. Clarson, A new approach to classification of brainwaves, *Pattern Recognition* 22, 767-774 (1989).

21 S.-Y. Lu, A tree-to-tree distance and its application to cluster analysis, *IEEE Trans.* **PAMI-1**, 219-224 (1979).

22. K.-S. Fu, A step towards unification of syntactic and statistical pattern recognition, *IEEE Trans.* **PAMI-5**, 200-205 (1983).

23. A. Sanfeliu and K.-S. Fu, A distance measure between attributed relational graphs for pattern recognition, *IEEE Trans.* **SMC-13**, 353-362 (1983).

24. L.G. Shapiro and R.M. Haralick, A metric for comparing relational descriptions, *IEEE Trans.* **PAMI-7**, 90-94 (1985).

25. L. Goldfarb, What is distance and why do we need the metric model for pattern learning?, submitted to *Pattern Recognition.*

26. L. Goldfarb, On the foundations of intelligent processes I: An evolving model for pattern learning, *Pattern Recognition* 23, 595-616 (1990).

27. V.N. Vapnik and Ya. A. Chervonenkis, *Pattern Recognition Theory: Statistical Problems of Learning*, Nauka, Moscow (1974).

28. V. Vapnik, *Estimation of Dependences Based on Empirical Data,* Springer-Verlag, New York (1982).

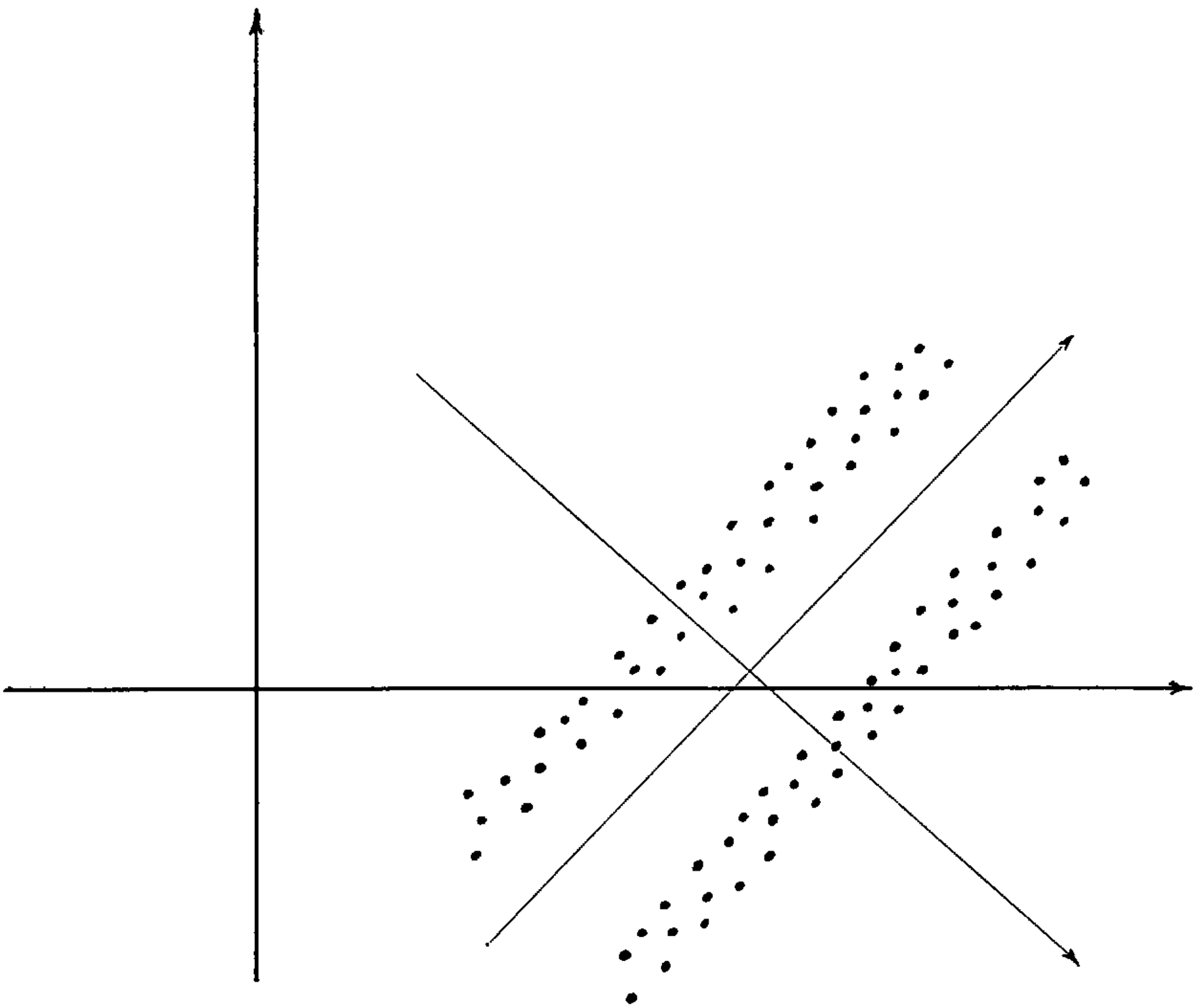

Figure 1. Example of the recognition problem where modification of the metric structure (by means of projection) improves the reliability and efficiency of the transformed problem.