Methodologies of Pattern Recognition

Methodologies of Pattern Recognition

Edited by

Satosi Watanabe

University of Hawaii
Honolulu, Hawaii

Academic Press • New York • London • 1969

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
Berkeley Square House, London W.1

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 69-13489
AMS 1968 Subject Classifications 6230, 6241, 68XX, 9310, 9401

PRINTED IN THE UNITED STATES OF AMERICA

CONTRIBUTORS

M. A. Aiserman (U. S. S. R.)

E. Andreewsky (France)

G. H. Ball (United States)

J. P. Benzecri (France)

D. B. Brick (United States)

P. W. Cooper (United States)

T. M. Cover (United States)

M. J. B. Duff (England)

T. G. Evans (United States)

K. S. Fu (United States)

E. E. Gose (United States)

A. Grasselli (Italy)

J. K. Hawkins (United States)

J. H. Holland (United States)

B. Julesz (United States)

L. Kanal (United States)

T. Kasvand (Canada)

V. A. Kovalevsky (U.S.S.R.)

P. F. Lambert (United States)

H. M. Lipp (Germany)

D. M. MacKay (England)

J. H. Munson (United States)

S. Noguchi (Japan)

T. Sakai (Japan)

S. D. Shapiro (United States)

J. T. Tou (United States)

I. Vajda (Czechoslovakia)

S. Watanabe (United States)

E. Wong (United States)

PREFACE

To the layman's ear, the term pattern recognition sounds like a very narrow esoteric field of electronic computer applications. But, actually, it is a vast and explicit endeavor at mechanization of the most fundamental human function of perception and concept formation. It represents, indeed, one of the most ambitious scientific ventures of this century, requiring collaboration of electronic engineers, physicists, physiologists, psychologists, logicians, mathematicians, and philosophers.

In recent years, pattern recognition has definitely emerged from the status of a programming stunt to become a respectable branch of scientific art. This implies that from now on the progress in the field will no longer be as dazzling as in the past, but that it has to become more solid and systematic. We can no longer congratulate ourselves because "it works," but we should ask why it works and explore the limits beyond which it will not work. Today, therefore, is a good time for researchers in this field to sit back awhile to evaluate the significance of their past achievements and to dream, prospect, and plan their future paths. This will require above all a soul-searching methodological examination of the entire enterprise of pattern recognition. Such was the motivation behind compilation of the volume you are holding in your hand at this moment.

I am very happy to have been instrumental in making this unique collection of papers on pattern recognition available to those who are interested in this fascinating new field of science. The articles in this volume were written by 29 outstanding authorities of 10 different national origins presently active in the field, who have critically reviewed their own methods and projected their present activities into the future. Each author was requested to place emphasis on the "philosophy" of his approach rather than on mathematical derivations and experimental data, and to provide the reader with a self-contained survey of lasting value.

These papers were presented, or were intended to be presented, at the International Conference on Methodologies of Pattern Recognition which was held on January 24-26, 1968 at the University of Hawaii, Honolulu, under the co-sponsorship of the University of Hawaii and the Air Force Office of Scientific Research with program participation by the Systems Science and Cybernetics Group of I.E.E.E. From the inception of the idea of the conference to the final redaction of these proceedings, Mrs. Rowena Swanson of the A.F.O.S.R. has tirelessly helped this project with wise advice and "motherly" care. Dr. Hans Oestreicher, representing I.E.E.E., participated actively in preparing the pro-

gram. Dr. Norman Abramson and Col. Harold P. Brown of the University of Hawaii were always at hand, ready to help us in organizing the conference. The Organizing Committee of the Conference consisted of the four names mentioned above, myself (who served as the chairman), and Mr. Timothy Ewald. If there was a single person who contributed most to the successful completion of this project, it was the secretary of the Committee, Miss Janet H. Yamamoto. She not only typed up everything from the original letters of invitation to all the articles which are photocopied and reproduced in this volume, but also actively participated in planning, organizing, and carrying out the project. The expert collaboration by the University of Hawaii Conference Center considerably alleviated the burden of our work in organizing the Conference.

I appreciate particulary the friendly act of Dr. M. Aiserman and Dr. V.A. Kovalevsky of U.S.S.R. who kindly contributed their papers to this book although the insufficiency of time between the invitation and the conference did not permit them to be personally present at the meeting. Heavy daily schedules prevented Dr. C. Masson (France), Dr. G. Sebestyen (United States), and Dr. L. Zadeh (United States) from preparing their written manuscripts. This is so much more regretted as their presentations aroused considerable interest and discussion at the Conference. I may also add that I deeply regret the absence of many other valuable papers which could have been contributed by equally outstanding scholars had the scale of the Conference only been large enough to allow us to invite them all. To everybody who in one way or another has helped us in organizing the conference and preparing its proceedings, I repeat

Mahalo and Aloha

Satosi Watanabe

Honolulu, Hawaii
April 1969

CONTENTS

CONTENTS

REMARKS ON TWO PROBLEMS CONNECTED WITH PATTERN RECOGNITION

M.A. Aiserman

INSTITUTE OF AUTOMATICS AND TELEMECHANICS
ACADEMY OF SCIENCES OF THE U.S.S.R.
MOSCOW, U.S.S.R.

To my regret, I was not able to take advantage of Professor Watanabe's kind invitation to participate in the "round table" discussion on questions of pattern recognition organized by him at the University of Hawaii. So it is with the greatest satisfaction that I take advantage of his kind proposal that I send a written copy of my presentation so that it can be published with the rest of the papers presented at this discussion. I decided to use this occasion to answer the following two questions, which in recent years have been asked of me by audiences at universities in the U.S.A. (in 1964), Canada (1967), Italy (1967) and other countries.

First Question: How do you evaluate the two different approaches which are developing in the area of pattern recognition—the geometrical approach, associated with the automatic construction of a partitioning surface, and the structural approach, associated with the automatic construction of a pattern description?

Second Question: What is the connection between the Robbins-Monro procedures of the method of stochastic approximation and the method of potential functions, developed collectively by you?

Since these questions probably interest not only the small circles of engineers and mathematicians who happened to be present at my lectures, I will try to answer them briefly.

I. Geometric and Structural Approaches to Problems of Machine Instruction in Pattern Recognition

The geometric approach to problems of pattern recognition in Western countries developed gradually along with an understanding of the principle of operation of Rosenblatt's perceptron. In the U.S.S.R. this approach was investigated and reported by E.M. Braverman, resulting in a great number of scientific publications. The essence of this approach can be described thus: We consider a space X of objects lying within a partition, such that every object is a point (vector) of this space; and the collection of objects is related

to some pattern-region in it. The partition of patterns is interpreted as the construction of a characteristic function assigned on X, that takes on numerical values of different signs on points of X belonging to different patterns. The problem lies in the construction of a characteristic function, when its signs are known only at a comparatively small number of points randomly chosen from the regions subjected to the partition.

Such an interpretation of the pattern recognition problem allowed us to relate it to a distinct extrapolation problem. The usual methods of extrapolation function theory turned out to be of little use in its solution mainly because of the high dimensionality of the space X and because of the random method used for the selection of points, for which information about the extrapolating function is known. However, a strict statement of the problem allowed us to introduce implicit bounds, with which it became possible to not only propose extrapolating algorithms, but also to prove their convergence to the extrapolating characteristic function; and sometimes to even estimate the speed of convergence.

Such a geometric approach was successfully used in a number of countries to solve important applied problems. However, it appeared weak in solving some quite elementary examples. So, for example, V.M. Glushkov studied the extremely difficult uses of the geometrical approach in cases where objects of one class are different collections of horizontal straight lines, and objects of another class are different collections of vertical straight lines. Another example of this type is the separation of the handwritten Latin letters "capital O" and "capital Q", or the Russian letters Ш and Щ.

The reason for these difficulties is understandable. In the first case (horizontal and vertical lines), the boundaries of the region with which the space X must be divided are extremely complex and devious, and therefore the form of the characteristic function is highly complex (for example, in its series representation by any "usual" system of functions, a great many harmonics must be present). In the second case (the letters O and Q or Ш and Щ) it is difficult to divide the regions due to another reason: they are too closely adjoined to one another and can only be well partitioned on subspaces of small dimensionality. Nevertheless, man easily discerns the differences between these forms; moreover, many animals (for example, rats) can be easily taught to perceive this difference, for example, by means of evoking their various reflexes.

Dissatisfaction with this situation led to the appearance of another important approach – the structural approach. In the West it was successfully investigated by R. Narasimhan, M. Eden and R. Kirsch and many others. In the U.S.S.R. this approach is associated with the name of M.M. Bongard.

The basic ideas of the structural approach are these: initially, one considers a given set of simpler fragments of objects of the partitioned collections (for example, for representation, the simpler details are intersections, curves, angles, line ends, etc.). This initial set is used as letters of the language; and the grammar of the language is given (the rules of construction of letters give words, rules of words give sentences, and so on). The problem then lies in the composition of a description for previously shown objects of the given space in this language. During the observation of a new object

the program checks the conformity of its construction with its description. It is easy to see that in the above examples (horizontal and vertical straight lines and letters O and Q) the application of the structural approach does not cause any difficulties.

Actually, in the first example, everything is straightforward, since horizontal and vertical lines can be simply the first letters of a language. The second example, is not much more complicated; thus, to describe the pattern of letters O and Q, it is sufficient to have such simple concepts as the oval, "intersection", "left" and "right", "up" and "down".

As far as I know, the structural approach has not yet been used in practice to solve any complex problems.* But, as the International Symposium at Sukhanovo (USSR), showed, the number of its adherents is growing rapidly.

In my opinion, any of these methods is of itself of little use in solving complex problems and the future will demand their combination.

Having understood this idea, it is not difficult to construct any number of examples that are not realizable in practice by the geometrical method and almost trivial for the structural method and vice versa. However, such examples are of no interest to us. We will try to give a structural description of a topographical or meteorological map, geological section or such a pattern as a "man's portrait" – this is enough to show how extremely difficult it is to describe an intricate pattern as in these cases. This intricacy is usually associated neither with the presence of an excessively large number of initial letters in the language alphabet nor with the excessive number of words that can be formed. The intricacy of the description arises on higher "grammatical levels" and is closely associated with the initial complexity of the pattern.

In order to understand the nature of this difficulty we will now proceed to consider the space Y of descriptions constructed in the chosen language. Every description of the pattern is a point (vector) of this space. Thus the whole process of composing the description is only a process of constructing a transformation which carries a region of space X into a point of space Y. It is clear that a transformation of this type can be simple in particular cases, but in general it is complex and cumbersome.

Naturally the thought arises – is it necessary to transform a region into a point. Would it not be better to consider an "intermediate space" Z, for example, the space of phrases or other fragments of the language not complete enough to fully describe it, and to choose this space so that in the considered problem, the regions of X that are difficult to divide would transform themselves into well-partitioned (e.g. distantly spaced from one another) regions of Z. Assignment of traits and partitioning, not in the initial space X, but in the space of traits is an example of this type. By such an approach the use of structural linguistical ideas does not lead to absurdity, and the language emerges as a means of transforming space or forming a space adequate for the problem. In this case both approaches, the geometrical and the structural do not compete, but are used to their full extent, with a common purpose and means to reach their goals.

* Perhaps the symposium at Honolulu will show the opposite. The participants should then forgive and correct me.

In the structural approach there is yet one more difficulty. To overcome it the geometrical method can be used. The question lies in forming the initial language alphabet, that is, those initial concepts to which grammatical rules apply from the first, for the construction of words and sentences. In all the programs known to me, this alphabet is introduced "from without" by the programmer. Now there are two alternatives. First the programmer can every time "creatively" find a specific alphabet, adapted to a given particular problem, and introduce it into the program. Thus, for example, he can notice that four or five concepts are sufficient (oval, stroke, right-left, up-down), to give a description of the pattern of the letters O and Q. But if the person deals with the situation thus, the problem for the machine vanishes. A person with such success could introduce into the program the completed description of the pattern also. As the second alternative, some "standard" alphabet could be proposed, sufficient for a wide class of problems (e.g. for all line drawings). But programs using such an alphabet, as is known, excessively overload the machine memory, complicate all steps in the construction of the description and always leave a feeling of uncertainty: will symbols for the understood description of the pattern be included in the alphabet for every new, arbitrary, complex problem? The issue is as follows: the program itself in every problem, considering the underlying distribution of the objects, should build an alphabet, convenient for their description.

In our laboratory a program of this type was realized.* The idea of this program consists of looking over the representation to assign "information fragments" (this concept can be formalized) and then using the geometrical approach to automatically divide these fragments into several groups according to their similarity. Typical patterns of each group of fragments thus assigned are taken as letters of the alphabet for describing the representation.

Thus the controversy between geometric and structural approaches for problems of pattern recognition seems to me historically inevitable, but temporary. There are problems to which the geometric approach is ideally suited. Also there are some well known problems which, though solvable by the geometric method, are more easily solvable by the structural approach. But any difficult problems require a combination of these approaches, and methods are gradually crystallizing to combine them; the structural approach is the means of construction of a convenient space; the geometric is the partitioning in it.

II. The Method of Potential Functions and the Robbins-Monro Procedure

We will now consider one part of the question associated with the geometric method in pattern recognition.

The question which is often asked of me is usually stated as follows: what is there in common between the method of potential functions and

* See article by I.B. Muchnik: "Algorithms for forming local traits for visual patterns" in "Automatics and Remote Control", No. 10, 1966.

their application to the same problems treated by the Robbins-Monro procedure of the method of stochastic approximation? Is there a difference between them, and of what does it consist?

To a great extent this question is terminological: it contains two indistinct concepts (the "method of potential functions" and the "method of stochastic approximation") and the answer to this depends only on how these concepts are defined.

Attempting to clarify what my questioners in fact had in mind when talking about the different methods, I was almost always convinced that they understood, strictly speaking, not a comparison of methods, but a comparison of procedures: the recursive procedures of the method of potential functions and the Robbins-Monro procedures.

In connection with this, we will note, first of all, that the idea of "method" is a significantly wider concept than that of "utilized procedure", so that a method, besides containing a procedure, contains facts and considerations, allowing one in every concrete problem to first choose a procedure in a reasonable fashion (such as, parameters entering into it; the form of the functions contained in it; its sequence, etc.) and then to establish that it converges to the solution of the considered problem. Usually, these initial and final steps of the application of the method are considerably shorter but more complex than the corresponding steps of an actual realization of the procedure.

The term "method of potential functions" was introduced through me, and it is easier for me to start with it. We call any recursive procedure of solving approximation problems a method of potential functions when it satisfies the following two conditions:

1) It can be represented in the form:

$$f^{n+1}(x) = q^n f^n(x) + r^n K(x, x^{n+1}) . \qquad (i)$$

or, if it is possible to assume that $f^n(x)$ can be represented in a series $f^n(x) = \sum_{i=1}^{\infty} c_i^n \phi_i(x)$ by a certain system of functions $\phi(x)$, then the procedure takes the form:

$$\tilde{c}_i^{n+1} = q^n \tilde{c}_i^n + r^n \psi_i(x^{n+1}) \qquad i = 1, 2, \ldots \qquad (ii)$$

In formulas *(i)* and *(ii)* we designate: x as a finite dimensional or infinite dimensional vector; x^{n+1} as the vector "indexed" on the (n+1)-st step of the procedure; $f^n(x)$ as the n-th degree estimate of the approximating function $f^*(x)$; q^n and r^n as numerical series, which differ with the choice of the method of potential function; and the potential function $K(x, x^{n+1}) = \sum_{i=1}^{\infty} \lambda_i^2 \phi_i(x) \phi_i(x^{n+1})$ and designate $\tilde{c}_i = \frac{c_i}{\lambda_i}$; $\psi_i(x) = \lambda_i \phi_i(x)$.

2) This procedure with any choice of $f^0(x)$ (or correspondingly c_i^0, $i = 1, 2, \ldots$) with $n \to \infty$ in this or another sense converges to the approximating function $f^*(x)$.

In the method of potential functions, the very establishment of the procedure *(i)* or *(ii)* in itself already suggests that in every concrete problem preliminary work is done for the selection of a potential function $K(x, x^{n+1})$ or of a system $\phi_i(x)$, without yet considering a choice of a series q^n and r^n. For this reason the method includes not only exact facts but many other intuitive considerations which as a group are an important component part of the method.† Another component part of the method is the group of concepts and theorems serving to establish the convergence of the procedure.

We will now consider the method of stochastic approximation. This method is used in mathematical statistics to solve the so-called regression equations.

We consider N functions $\Phi_i(c,x)$, $i = 1, \ldots, N$, where x is a random quantity and $c = \{c_1, \ldots, c_N\}$ is some vector. The system of regression equations is

$$M_x \{\Phi_i(c,x)\} = 0; \quad i = 1, \ldots, N \tag{1}$$

in which the unknowns are the components c_i of vector c, and $M_x\{\cdot\}$ is the mathematical expectation.

It is assumed that the probability distribution of the random variable x is unknown beforehand, and consequently the left hand sides of equation (1) evidently cannot be calculated. Let us suppose, however, that at consecutive moments of time $(1, 2, \ldots, n)$ there appear points $x^1, x^2, \ldots, x^n$ with probability distribution P; then with any c the quantities $\Phi_i(x^n,c)$ can be calculated. For this case Robbins and Monro proposed the following recurrent procedure.*

$$c_i^{n+1} = c_i + \gamma_n \Phi_i(c^n, x^{n+1}) \quad i = 1, \ldots, N \tag{2}$$

---

† See for example, M. A. Aiserman, E. M. Braverman, and L. I. Rozonoer, "On the Selection of Potential Functions in Symmetrical Spaces". "Automatics and Remote Control", No. 10, 1967.

* In the work of Robbins and Monro, procedures (2) and (3) were proposed for the scalar case; where c is a scalar. In the case where c is an n-dimensional vector it was extended by Bloom. I think that my audience under the term "Robbins-Monro procedures" understood procedures (2) and (3) in the case where c is a vector. Recently several works have appeared in which the terms "Robbins-Monro procedures" are understood in a broader sense. Undoubtedly my audience did not have in mind this widened terminology, simply because the work containing it appeared after the basic results of the method of potential functions had been published.

where γ_n is a series of non-negative numbers satisfying the condition

$$\sum_1^\infty \gamma_n = \infty \; ; \quad \sum_1^\infty \gamma_n^2 < \infty \tag{3}$$

It was shown that within certain limits on the form of function Φ_i, the procedures (2) and (3) converge in probability to the root of the system of equations (1). The method of stochastic approximation, aside from procedure (2), includes also theorems on convergence which define the bounds on the choice of functions Φ_i.

If the Robbins-Monro procedure (2) is compared with the general procedure *(i)* and *(ii)* of the method of potential functions, then these procedures turn out to be essentially different, even if only because expression *(i)* and *(ii)* admit, generally speaking, any dependence q^n and r^n on the number n, which evidently enters into consideration from the time that (2) and (3) depended on n and (in this special case) only on the "convergence factor" γ_n. Besides this, the procedures of the method of potential functions can be applied both in cases where vector c is finite dimensional and where vector c is infinite dimensional (in this case the procedure *(i)* is used). The Robbins-Monro procedures (2) and (3) are introduced for finite vectors c.* Therefore, only one special form of procedures *(i)* and *(ii)*, and not the general procedures, can be compared to the Robbins-Monro procedure.**

$$c_i^{n+1} = c_i + \gamma_n [r \left(\sum_{i=1}^N c_i \phi_i(x^{n+1}), x^{n+1} \right) + \xi^{n+1}] \; \phi_i^{n+1}(x) \tag{iii}$$

where c is a finite dimensional vector and ξ^n is a disturbance with zero mathematical expectation.

Procedure *(iii)* can be considered*** as a Robbins-Monro procedure (2), adapted to the solution of the following system of regression equations:

$$M_{x,\xi} \{ [r \sum_i^N c_i \phi_i(x), x) + \xi] \; \phi_i(x) \} = 0 \tag{4}$$

If we consider as a given condition that the mathematical expectation of disturbance ξ is zero, then for every fixed x, the system (4) assumes the form:

---

* In recent works (see the note on page 11) the term "Robbins-Monro procedure" was used in the case where c was an infinite vector. It is not immediately clear, however, how one can realize in this case procedures (2) and (3) in practice, without restricting oneself to the specific case of procedures (2) and (3) which lead to procedure *(i)* of the method of potential functions (see further).

** Many algorithms, considered in our publications had the form *(iii)*, but it is understood, that this is only a special case of the general procedure *(i)*.

*** If we do not consider the here non-essential circumstances that in *(iii)* it is not assumed (in contrast to (2)), that γ_n satisfies condition (3).

$$M_x \{ r(\sum_i^N c_i\phi_i(x), x) \phi_i(x) \} = 0 \qquad (5)$$

We will now consider that the sought after parameters c_i enter in the regression equation (4) and (5) in quite a special form. In the same way, procedure *(iii)* of the method of potential functions defines a specific subclass of regression equations (4) and (5) and a corresponding specific subclass of Robbins-Monro procedures. Such procedures have a series of special properties, because of which they can be studied in more detail than Robbins-Monro procedures of the general form. We will next consider some of these properties.

1. The possibility of machine realization of the procedure: Procedure *(iii)* is obtained from the assumption that the building functions $f^n(x)$ can be represented by a series of the system $\phi(x)$. It is understood, that an inverse transformation from the correspondence *(iii)* to procedure *(i)* of the special form is always possible. Indeed, this circumstance allows one to realize the procedure *(iii)* with the help of calculating machines, and in those cases where vector c is infinite dimensional--that is, where functions $f^n(x)$ can be represented only by an infinite series. It is not immediately clear how for an infinite dimensional case, the general Robbins-Monro procedure can be brought to form *(i)* and thus realized in practice.

2. Existence of an extremizing functional: For procedures of the form *(iii)* a functional can be written extremizing this procedure. On the assumption that vector c^n is finite dimensional, procedure *(iii)* can be represented in the form of a stochastic gradient procedure and the regression equation (5) assumes the form:

$$M_x \{ \frac{dG(c,x)}{dc_i} \} = \frac{dJ(c)}{dc_i} = 0 \quad i = 1, \ldots, N \qquad (6)$$

where the function G and the functional J are defined by the formulas:

$$G(c,x) = Q[f(x), x] = \int_0^{f(x)} r[u,x]\,du \qquad (7)$$

$$J(c) = M_x\{G(c,x)\} \qquad (8)$$

It is understood that in the general case the regression equation (1) does not have a gradient form (6). The existence of an extremizing functional for procedure *(iii)* allows one to give a specific meaning to the extrapolating problem that is being solved*. For such a particular form of the Robbins-

* The first to consider such a possibility was Ya.Z. Tzypkin. He found an extremizing functional for the series of procedures of the method of potential functions of the form *(iii)* establishing by this that these procedures can be interpreted as Robbins-Monro