Algebraic Methods in Recognition and Classification Problems

Y. I. Zhuravlev

Computer Center of the USSR Academy of Sciences, Moscow, USSR

The problem of recognition has been the focus of interest for experts in the field of applied mathematics and, later, computer science for quite some time. It is relevant to mention the pioneering studies of R. Fisher in the 1920s which resulted in the creation of discriminant analysis, the setting up by A.N. Kolmogorov and A.Y. Khinchin in the early 1940s of a problem of separating the mixture of two distributions, the theory of statistical solutions, and numerous other papers published over 1950-60 which dealt with the search for, and application of, algorithms providing a classification of an object as belonging to a given class or the separation of a certain set of objects into several non-intersecting classes.

In the mid-1970s, the role of recognition as a distinct scientific direction changed somewhat, since its development made it possible to create an adequate mathematical theory of recognition.

One of the prerequisites of this possibility was the formulation and processing of several models of recognition algorithms - families of algorithms, to solving classification problems. The following models were widely accepted by that time.

1. Models based on the use of the principle of separation (R-models). These differ basically in setting of the class of surfaces, among which the surface (or a set of surfaces), separating elements of different classes are being determined in the most accurate way.

2. Statistical models. This type of models of recognition algorithms, based on mathematical statistics, is applied in cases where the probability characteristics of classes, e.g. the respective distribution functions, are known or can be determined easily.

3. Models developed on the basis of the so-called method of potential functions (P-models). This involves the concept of potential taken from physics. Potential is defined for any point of a space and it depends only on the location of the potential source. The potential function, which is positive everywhere and monotonously decreases with distance, is taken in this model as a membership function for the object of study.

4. Models of calculation of estimations (polls) (G-models). These are derived from the principle of partial precedence. One has to analyze the "proximity" of the descriptive sections of the object of studies with those of previously classified objects. The closeness detected is considered to be the partial precedent and is evaluated according to a specific rule (through the numerical estimate). A criterion of proximity estimates is used to elaborate the general estimate for the object to be recognized, which is considered to be the value of the object's membership function with respect to a given class.

5. Models which are based on statement counting and on the apparatus of logic algebra (L-models) in particular. In these models, classes and features of objects are considered as logical variables, whereas the description of classes in terms of feature language is given as Boolean relationships.

It would be natural to assume that the central problem of the theory of pattern recognition is the development of efficient computational means to attribute the formalized descriptions of recognized objects to respective classes. Such a classification (recognition) is based on the derivation of a certain aggregated estimation of an object derived from its description. It is also natural to assume that recognition problems are discrete analogues of the tasks of the search for optimum solutions. In this case one can distinguish a wide class of problems for which one should establish whether the studied situations (objects and phenomena which are extremely complicated and, in a sense, "complex") possess a fixed finite array of properties enabling one to attribute them to a specific class using a certain information which is often highly inhomogeneous. These are problems of recognition and classification. On the other hand, one can use the same type of information concerning the finite set of similar processes to predict to what area of the finite number of areas these processes evolve over a certain period of time. This is a problem of forecasting.

The advantages, achievements and perspectives in recognition are initially perceived from this point of view. However, there is another side to the matter, which is that the development of recognition provides a good model for refining mathematical theory to process and transform information, the continuous process in which essentially heuristic methods enjoy rigorous substantiation and are being applied within the framework of quite formalized regular procedures. It is interesting to note that recognition itself is now a sufficiently developed variant of such a theory since it enables one to solve the major problem: to synthesize and choose algorithmic means capable of obtaining useful information from the type of initial data described above.

It is well known that one should resort to setting up a recognition problem in cases where it is difficult to build up formal theories and apply classical mathematical methods. In such a situation one of the following cases is confronted:

Received December, 1990.

Pattern Recognition and Image Analysis, Vol. 1, No. 1, 1991.

a) the formalization level corresponding to the object field and/or the accessible information cannot provide the basis of synthesis of a mathematical model compatible with classical mathematical or mathematical-physical standards which would allow study of the phenomenon using classical analytical and numerical techniques;

b) the mathematical model can be developed in principle, but its synthesis or study involves expenditures (the collection of necessary information, computing resources, time) which are either disproportionate to the value of the result obtained by the solution in question, or lie beyond the capacity of modern technology, or make a solution of the initial problem senseless.

Thus, the "duality" of recognition was already manifested in the fact that the solution of such a problem necessitated the invention of a large number of non-correct (heuristic) algorithms. Over a fairly long period, most applications of recognition theory dealt with poorly formalized sciences - medicine, geology, chemistry, etc. Even now it is not easy to develop a formalized theory and apply standard mathematical methods in the above scientific areas. At best, one can succeed only in producing a mathematical formulation using certain intuitive principles, and then applying the obtained "empirical formalisms" to the solution of specific problems. The result of this was that at the first stage of development of the theory and practice of recognition, a great many different methods and algorithms applicable to practical tasks were originated which had not been seriously substantiated. While examining a problem or class of problems using so-called "true looking" reasoning, a loose but sensibly consistent solution technique, and the algorithm based on it, was suggested. Substantiation was then achieved immediately in experiments involving specific problems. The algorithms complying with such experimental verification, i.e., resulting in successive solutions to specific practical problems, were then widely applied despite the missing mathematical substantiation.

It became evident that the development of each heuristic algorithm of such a kind can be considered as an experiment allowing us to consider the set of these experiments and their results as a novel set of mathematical objects, i.e. to study a set of non-correct procedures aimed at a solution of poorly formalized problems, using rigorous mathematical methods. Therefore, the second stage of development of recognition theory was characterized, on the one hand, by attempts to put forward and solve the problem of the choice of the best (in a certain sense) algorithm applicable to any specific situation. On the other hand, there were also attempts to switch from the description of specific incorrect algorithms to the description of the principle of their formation, i.e., by attempts to obtain a unified description for sets of procedures which are heuristic but, nevertheless, successfully applicable to the real problems. This set is specified by the assignment of variables, objects, functions, parameters and an exact definition of the fields of the variation. Fixing these variables, objects, functions and parameters enables one to distinguish a specific algorithm from a corresponding set (model). Initially, the class of algorithms to compute estimations was represented as a model; later the descriptions of other models appeared.

The need to synthesize models of recognition algorithms was dictated, at first, by a need to fix the class of algorithms in a certain way while choosing the optimal, or at least acceptable, procedure to solve a specific problem. Attempts to develop such models in turn stimulated interest in the proper "mathematical" features of recognition algorithms and, in particular, in the problem of their rigorous substantiation. It was found that description of the class of recognition algorithms is a problem compatible with the development of the classical definition of the algorithm. Consequently, the necessary *condition for developing recognition theory lies* in classical algorithmic studies of the notion of "recognition algorithm".

Analyzing an array of non-correct recognition algorithms, (as the amount of the latter increases), enables one to distinguish and describe not only specific algorithms, but also the principles of their creation. These principles, already acting on subsets of algorithms and initially formulated in poorly formalized shape, are capable of being realized in the future as precise mathematical descriptions. At this development stage, the heuristic character should be attributed to the principle choice, whereas algorithms originated on the basis of a correspondent principle may be derived in a standard manner. In this very sense, the formalization of various principles to develop recognition algorithms results in the appearance of recognition algorithms.

Transition to the model of recognition algorithms did not result in development of a universal model. Nor did it lead to formalization of a process of choosing a specific model to solve a specific recognition problem. But the formulation of models enabled one to put forward and solve within the framework of specific models, the problem of choosing an algorithm which is extremal with respect to the classification or forecast quality functional. The derivation of such optimal algorithms usually results in the investigation, implementation and development of computational schemes for non-standard extremal problems.

Indeed, the parameterization of several algorithms (models) of recognition and capability to determine parametric values using the information concerning the class of models enables one to choose the correct algorithms for several subclasses of problems. In most practical cases, however, this subclass is quite narrow since, in the opposite case of synthesis of the models of recognition algorithms, and in the description of classes and the choice of features of recognition objects, one would have had to use a large amount of a priori information, the latter being accessible only if one possesses a sufficiently precise model of objects and phenomena studied. Besides, the development of an optimal algorithm in a multiparameter model leads to the solution of difficult extremum problems (frequently NP-complete ones). As a rule, one fails to specify the global extremum which significantly decreases the recognition quality, impeding realization of the model's total potential. Sometimes one faces a situation where the use of models with small parameters, allowing determination of the global extremum, is more effective than implementation of local-extremum algorithms of multiparameter models. However, one can hardly expect that the algorithm of optimum model will remain the same when applied to objects not participating in the learning procedure.

In the second stage, the substantiation involves one of the following techniques:

1) Experimental approach - the possibility of obtaining a "Solution" to a problem under consideration (one which would be acceptable from the user's point of view) using an appropriate recognition algorithm, is seen as a substantiation of its applicability in solving that problem;

2) By means of a solution to an optimization problem and the use of a recognition algorithm which is optimal within the framework of the chosen model; substantiation involves application of the best possible recognition algorithm in the model used;

3) Substantiation may be carried out in the same manner as in technique (2), but it is proved additionally that for the satisfaction of several "natural" hypotheses (conditions) valid for the class of problems under examination, the algorithms optimal in the model used do indeed guarantee a high accuracy of recognition, i.e., both the choice of algorithm and the choice of the model are substantiated.

The next stage of development of the recognition area was related to an overall study of the structure of a set of non-correct algorithms. It became evident that improvement of a model often does not result in adequate refining of results and there is a natural complexity limit for any model. Thus, the idea was put forward to choose algorithms from among existing families and to apply the respective operation to algorithms (correction operations) resulting in building up an optimal algorithm from an initial ones.

An early version of this concept was the so-called corrector with respect to results based on the creation of the solution to a recognition problem involving results of processing of initial information by separate algorithms. It was shown that, generally speaking, there are no "good" simple operations resulting in necessary correction, only in the case where the answers "yes", "no", "no idea" are considered allowable. The fact is that the domain of initial information and the set of possible answers is defined by the consistent setting of a problem. The former consists, therefore, of elements with sufficiently complex organization (usually vectors with high dimensionality), whereas the latter is highly void ($\{0, 1\}$).

As a possible solution of this situation, we proposed a technique to define the recognition algorithm containing all possible types of algorithms and the so-called algebraic approach to recognition and classification problems, ensuring efficient study and adequate description of classes of recognition algorithms. This approach involves updating the initial heuristic families of algorithms through algebraic operations and building up a family which ensures derivation of correct algorithms resulting in solution of the class of problems being studied.

The algebraic approach is based on the idea of inductive generation of mathematical objects through a generalized inductive definition. The basic algorithms and recognition models are introduced, followed by the introduction of operations on them which enable one to generate consecutively new algorithms and models. The conditions influencing a given family of algorithms are basic with respect to the operations introduced and are being elucidated, as are the properties a model ought to contain in order to possess an algorithm capable of classifying correctly all objects of arbitrary finite samples. The methods of derivation of such algorithms are being formed. The main idea of this approach is that the family of such algorithms is considered as an algebra the operations of which enable one, using as a basis the family of algorithms, to build an expansion of this family containing the correct algorithms which correctly classify the finite sample in all classes.

The algebraic approach makes use of peculiarities of the structure inherent in any recognition procedure. It also involves an introduction of the so-called estimation space, which is intermediate with respect to initial descriptions and allowable answers. The recognition algorithm in this case is considered as a superposition of two operators. The first is a recognition operator and it forms elements called estimates as answers, whereas the second (decision rule) defines these final answers using the estimates obtained. Therefore, instead of considering "inconvenient" spaces of initial descriptions and allowable answers, one can introduce a correction in space of estimates (which is usually a set of real numbers).

The notion of completeness relating separate problems with models of algorithms is important in an algebraic approach; completeness of a certain problem with respect to a model implies that one can build up an algorithm resulting in the correct answer for an arbitrary array of a priori classifications for the object considered within the model's framework. Completeness of a certain problem concerning a model has as its direct consequence the existence of an algorithm within the model which guarantees absolute accuracy on learning material. It is important that in most cases the derivation of the extremal algorithm is a problem which can be solved easily using standard mathematical methods.

Several studies dealing with the examination and substantiation of developed methods have been carried out within the framework of the algebraic approach. It was found that the problem of the boundary of the set of correcting operations, the crossing of which in the expansion process does not result in any real effect, is related to the choice of an allowable way of processing information by applied algorithms. Formalization and subsequent study of a consistent idea concerning an permissible method of processing information by recognition algorithms led to several final estimates for the models of algorithms and sets of correcting operations. Thus, it became possible to derive a universal upper degree boundary for sets of operations of polynomial type; the lower complexity boundaries for the models of recognition operators were established of estimation calculation. The same boundaries were determined for models of recognition operators based on the separation principle.

It was shown that families of algorithms encountered while applying the algebraic approach are of restricted capacity, providing a correctness of applicability of such families in cases where several fairly general hypotheses of statistical nature are fulfilled. The extremum algorithms formed within the framework of an algebraic approach were shown to have a non-zero stability radius. This implies that in the presence of a small (in a certain sense) variation of initial information the classification generated by extremum algorithm is retained, i.e., when the fairly general suggestions concerning compactness almost everywhere are fulfilled, the convergence of classifications generated by extremum algorithms to a true classification takes place. Studies were also carried out to investigate the possibility of the simplest representation of extremum algorithms.

Simultaneously with the process of the transfer from separate algorithms to models, another branch of investigations (dealing with the use of algebraic techniques to expand the types of initial information suitable for recognition problems) enjoyed a successful development. It is relevant to mention here U. Grenander's pattern theory, as well as the descriptive theory of image analysis developed within the framework of an algebraic approach constituting a basis of the new research direction in recognition.

Summing up, we should like to point out that the methodology of recognition is used in computer science in two separate ways:

- for the straightforward solution to problems of recognition in a classical sense;

- as a tool for the direct study of poorly defined problems.

In the latter case, this methodology is implemented roughly as follows. Let there be, for example, certain data obtained during a physical or simulated experiment. These data characterize the object of studies or phenomenon under investigation in a certain highly restricted sense. One should try to combine all experimental information in order to establish what regular features are contained in the material collected. This can be done by putting forward a certain simple hypothesis formulated in mathematical terms. At the next stage one should try to "explain" the material collected using the above model. Consecutive use of several heuristics (realizations of the hypothesis) may be helpful in guessing the true model. In the opposite case, one should make a choice within the framework of a model originated by heuristics and then search for the optimum (adequate) heuristic principle (model). In cases where there is no adequate principle, or where such a principle cannot be implemented, one has to formulate an array of principles which guarantee the isolation of a certain "federal" principle. It is this upper level which corresponds to the possibilities and implementation of an algebraic approach.

The mathematical description of the algebraic approach will be given in full detail in articles to be published in this journal later this year.

YURI I. ZHURAVLEV, Professor, Doctor of Science (Physics and Mathematics), and Corresponding-Member of the USSR Academy of Sciences, was born in Voronezh, USSR, in 1935 and graduated with a BSc in Computer Science from Moscow State University (Department of Mechanics and Mathematics) in 1957. He undertook postgraduate studies at Moscow State University and submitted his Ph. D. thesis in 1959.

His postgraduate study involved an examination of the normal forms of logic functions. Between 1959 and 1969 he was a head of department then deputy-director of the Institute of Mathematics at the Siberian Division of the USSR Academy of Sciences (Novosibirsk). From 1962 to 1969 he was a senior scientist, then professor at Novosibirsk State University. In 1965 he submitted his doctoral thesis "Local algorithms of calculation of information". Since 1969, he has been head of laboratory, head of department and deputy-director of the Computation Center of the USSR Academy of Sciences in Moscow and professor at the Moscow Institute of Physics and Engineering.

Professor Zhuravlev's main areas of interests concern mathematical logics, algebra, discrete optimization, theory of algorithms, and pattern recognition as well as application of mathematical methods and computation facilities to solve applied problems of processing and evaluation of information in automation of scientific studies, industry, medicine, biology and sociology.

He has published more than 150 papers on informatics and applied mathematics. He is chief editor of the annual journal *Recognition. Classification. Forecasting. Mathematical Methods and Their Application* (Nauka, Moscow). He is also a member of the editorial boards of numerous international and Soviet journals on computer science and applied mathematics.

Professor Zhuravlev is a laureate of the USSR State Prize, which was awarded to him for a comprehensive study in the field of discrete mathematics (1966). He was also awarded a Prize of the USSR Council of Ministers for the solution of applied problems using methods of pattern recognition (1986).