KYBERNETIKA — VOLUME 30 (1994), NUMBER 2, PAGES 159–176

THE CURRENT APPROACHES IN PATTERN RECOGNITION

JIŘÍ KEPKA

The paper presents a brief survey of the current approaches in pattern recognition. In this field the classification processes of patterns are the main subject of interest. The most conventional techniques used to solve pattern recognition problems may be grouped into statistical, structural and hybrid methods of the previous two ones. The artificial intelligence approach to pattern recognition is used for such very complex tasks each solution of which heavily depends on the knowledge of experts. Recently, several attempts have been made to combine pattern recognition system with knowledge-based system so as to build a knowledge-based pattern recognition system of much more sophisticated recognition capability. The links between the structural (syntactic) approach and the artificial intelligence one to pattern recognition are given.

1. INTRODUCTION

Both advantages and drawbacks of the current approaches in pattern recognition and the links between them are described in a very comprehensive and brief form. Their knowledge is necessary for understanding the current trends in pattern recognition field. From this point of view this survey should bring about the benefit to the readers interested in pattern recognition problems. To persons quite new to pattern recognition the complete understanding of the text may be difficult but perhaps the great number of citations in the paper may be very useful for them in their further studies.

There has been a considerable growth of interest in problems of pattern recognition during the past thirty years. Applications of pattern recognition include character recognition, speech recognition and understanding, medical diagnosis, identification of fingerprints and human faces, remote sensing, machine part recognition, etc. Both the development of the new methods for use in design of pattern recognition systems and computer technology advances have made processing of more complex patterns possible. The most conventional techniques so far used to solve pattern recognition problems are based on the one of the following approaches — the statistical, the structural or the hybrid one.

Pattern recognition has been defined by Pavlidis "to involve the identification of the ideal which a given object is made after" [76]. Pattern recognition is mainly related to classification. But in many problems, e. g. identification of fingerprints, recognition of continuous speech, the patterns are complex and/or the number of possible descriptions is very large and therefore, it is usually impossible to regard each description as a separate class. In such cases the information about the pattern structure can not be omitted and/or the requirment of recognition can be satisfied only by founding out a description for each of analyzed patterns rather than by simple task of classification.

One of the necessary conditions for understanding the "signal" which the analyzed pattern is measured on is to obtain its description. Clearly, the description should make the interpretation of the "signal" in terms of the behavior of mechanism, that generate it, possible. In complex tasks the solution heavily depends on the knowledge of experts. That is why many attempts have recently been made to combine pattern recognition system with knowledge-based system. The goal is to build a knowledge based-pattern recognition system of much more sophisticated recognition capability, see e. g. [8, 71, 109, 110].

In the second part of the paper the conventional approaches, namely the statistical, the structural and the hybrid one are briefly reviewed. In the third part the artificial intelligence approach to pattern recognition introduced by Nandhakumar and Aggarwal [71] is discussed and its links to the conventional approaches, especially to the structural (syntactic) one, are shown.

2. CONVENTIONAL APPROACHES

2.1. The Statistical Approach

In this approach the measurements taken from N features are represented in N-dimensional pattern space as one point. Its coordinates characterize the original "signal". The principle of classification is based on the portion of the pattern space into subspaces, each of which corresponds to a particular pattern class. Sometimes, the possibility to reject the classification of some patterns may be also desirable, see [15, 18, 48]. The techniques used to solve such tasks can be subdivided into:

a) parametric classification methods;
b) discriminant functions;
c) clustering analysis;
d) fuzzy set reasoning.

ad a) Each of features is considered to be a random variable. The multivariate probability distribution function is assumed to exist for every pattern class. They are either known a priori or estimated from a training set of patterns.

The best known classifiers are based on Bayes' rule. They are designed with the criterion of minimizing the Bayesian error probability or a cost measure based on it. The input feature vector representing an analyzed pattern is classified into a class C_i, if the likelihood ratio between two pattern classes C_i and C_j is greater than the ratio of the probabilities of occurrences of the pattern classes C_j over C_i. For more details about this group of techniques see e. g. [19, 98, 114].

These techniques come to trouble for the patterns which require a lot of features for recognition. In such cases the feature space is not distributed densely and uniformly, but sparsely and locally. Sometimes the costs of feature measurements and/or their sequential character must be also taken into account. Then it is better to process the pattern features sequentially. The principle of sequential processing is based on a partial decision making in each step. Step by step pattern features are sequentially processed and pattern subcategories are formulated until a subcategory contains only one pattern class. In this manner the often desired trade off between the costs of measurements and the risk of misclassification can be obtained.

Both the selection and the ordering of pattern features has a direct influence on the efficiency of recognition. The Karhunen-Loeve expansion [49] is the well known feature selection technique. It is based on an eigenvector analysis of the sample covariance matrix associated with the input representation vectors. As the result of this analysis the representation vectors are linearly transformed into a new coordinate system in which the coordinate coefficients are mutually uncorrelated. The information from the original representation vectors is concentrated in the first few axes of the new coordinate system.

Probably the best known sequential recognition procedure of parametric classification methods is the Wald's sequential probability ratio test [24]. After each measurements the likelihood ratio is computed and with two parallel boundaries compared. One hypothesis (pattern class) is tested against another. The crossing of each of the boundaries is associated with the acceptance of one of the two hypotheses. The urgency to terminate the Wald's sequential procedure becomes necessary when the cost of taking measurements is found too high or when the process exceeds a certain time limit or when the available measurements are exhausted. Either the truncation of the procedure at a given time [9] or, better, a modified Wald's sequential procedure using time-varying stopping boundaries [13] can be used.

A dynamic programming approach to sequential processing was also tested in the cases where the feature measurements were assumed to be either statistically independent or Markov-dependent, see e. g. [31].

Besides the sequential recognition methods multistage decision making based on various decision trees has been employed in a number of pattern recognition tasks, e. g. [11, 86].

ad b) The classification problem is formulated in terms of discriminant functions. Let X be the feature vector. Then the discriminant function $D_i(X)$ associated with pattern class C_i, $i = 1, ..., K$, is such that if the input pattern represented by the feature vector X is in class C_j, the value of $D_j(X)$ must be the largest one, i. e.

$$D_j(X) > D_i(X), \quad i, j = 1, ..., K, \quad i \neq j.$$

The decision boundary between regions associated with class C_j and class C_i, respectively, is given by the equation

$$D_j(X) - D_i(X) = 0.$$

Many forms can be chosen for $D_j(X)$. The simplest one is the linear discriminant function. If a set of reference vectors, each of which corresponds to a particular pattern class, is given, then a minimum distance classifier may be used. More general situation such that a set of reference vectors, instead of a single one, is given for each pattern class leads to the piecewise-linear discriminant functions. In general, any non-linear discriminant function can be used.

The choice of the form of a discriminant function is influenced by the maximal complexity and misrecognition rate allowed, by the number of training patterns, by the number of features and other a priori knowledge about the given problem. The proper values of the coefficients in discriminant functions are usually not available. They must be learned during the learning process from the training patterns with known classification. For more details see e. g. [114].

ad c) Clustering analysis has become an often used tool for data analysis. It can be thought to be an independent field of pattern recognition approaches. Clustering algorithms may be classified into the hierarchical and the nonhierarchical ones [2].

The k-means method and its variants are the most widely used clustering algorithms from the nonhierarchical ones. For example let us consider k clusters $C_1, \ldots, C_k$ and a criterion

$$J = \sum_{i=1}^{k} \sum_{x_l \in C_i} \|x_l - m_i\|^2,$$

where m_i is the mean value of the i-th cluster. An optimal clustering procedure considers all possible k clusters obtainable from all pattern vectors x_l, $l = 1, \ldots, n$, and evaluates J for each possible combination. In this manner a global minimum of J and the corresponding final grouping can be obtained. This technique is not often used because it is computationally impractical for moderate or large n of pattern vectors. Computationally feasible k-means method seeks the minimum iteratively. To initiate the clustering algorithm the seed points $m_1(1), m_2(1), \ldots, m_k(1)$ have to be chosen. At each step of the algorithm pattern vectors are reclassified and afterwards the mean vectors are updated. This method can be simply and easily implemented, thought, on the other hand, the final grouping computed by it is influenced by the accepted choice of the initial seed points and only a local optimal grouping may be obtained.

The hierarchical clustering algorithms are usually divided into agglomerative and divisive. The agglomerative algorithms operate by step-by-step merging of small clusters into larger ones by the introduction of ultrametric distance criteria. Recently, the efficient agglomerative clustering algorithm using a heap has been proposed in [56]. Its computation time is at most $O(n^2\log(n))$. In contrast to the agglomerative techniques, "the division methods not only describe the situation, but also find a representation and based on that identify the object" [69].

ad d) If there is no a priori knowledge and therefore, the probabilities can not be computed, then the introduction of fuzzy set elements formulated by Zadeh [115] may yield more realistic results. Fuzzy set reasoning creates an alternative to the probabilistic approach given in a), see e. g. [54].

It has to be stressed here that, the statistical approach is also known to have some drawbacks. First, many above mentioned techniques are optimal in classifying the features but there is nothing optimal about their choice. Secondly, quantifying the contribution of a particular feature towards the accuracy of classification is ambiguous. "Many measures have been proposed to evaluate it but there is no universally accepted one. This, compounded with the indeterminacy in the statistical inter-relationship between features, makes features subset specification an exercise in educated guessing" [71]. It should be also noted that, in general, all pattern features can not be represented well in mathematical expressions and the evaluation of the effectiveness of feature ordering largely depends on human subjective judgment. Thirdly, the statistical approach do not utilize structural properties (if any exist) of analyzed patterns and it is not capable to analyze patterns containing the recursion. It does only lead to class descriptions of patterns but it does not provide any descriptions of them.

Nevertheless, the statistical approach is an adequate tool for solutions of many pattern recognition problems. Its effectiveness depends on both the problem domain and the goal to be solved.

2.2. The Structural Approach

The main idea of the structural approach is based on the recursive description of complex patterns in terms of simpler patterns just as sentences are built up by concatenating words, and words are built up by concatenating characters. The application of this approach results in both the classification and the structural description of an analyzed pattern by means of a set of pattern primitives (simplest subpatterns) and their relationships. Of course, the pattern primitives must be much easier to recognize than the patterns themselves.

As the theoretical background formal language theory [35, 40, 85] is usually used and that's why the term "syntactic approach" is also used. Chomsky [14] introduced four types of phrase - structure grammars, namely unrestricted, context-sensitive, context-free and finite-state grammars. They deal only with strings and their rules are successively applied only to a given small part of sentence form derived. In describing patterns using such a string grammar, the only relation between subpatterns and/or primitives is the concatenation. Unfortunately, in many practical problems the pattern description in the form of a simple string of primitives can be cumbersome. The one-dimensional relation has not been very effective in providing efficient structural descriptions of multidimensional patterns such two-dimensional images and three-dimensional scenes. That's the reason why the high-dimensional pattern grammars, e. g. the tree grammars [26, 77, 78], the plex grammars [21], the web grammars [26, 79], etc. are used, although, in principle, any multidimensional grammar can be converted into a one-dimensional one [26].

Classification is performed by syntax analysis which decides whether or not pattern representation is syntactically correct in accordance with the given grammar. In the former case a complete syntactic description - a parsing tree - of the analyzed pattern is produced. Provided the pattern can not be successfully parsed on the basis of any of the grammars describing possible classes of patterns then it is rejected. The parsing tree is constructed either from its root (the starting symbol of the given grammar) towards the bottom (the analyzed description) or from the bottom towards the top. The former technique is called top-down parsing, the latter bottom-up parsing. For the most effective algorithms of syntax analysis see [1, 12, 20, 95]. In most applications a complete syntactic description according to the given pattern class grammar is usually needed. Therefore, only parsing techniques preserving the desired structural description can be used, see [44, 57]. The parsing techniques for tree grammars can be also used to handle the multidimensional concatenation, cf. [27].

In order to obtain a grammar representing the structural information about the patterns under consideration, a grammatical inference machine is required. A grammar inference from a given set of training patterns can be considered to be analogous to the learning processes in the statistical approach. The results obtained so far in this field are not satisfactory enough. The known methods suppose more or less special grammar forms, cf. [4, 26, 30, 42, 53]. It should be also noted that the appropriate trade off between the complexity of the recognizer and the descriptive power of the language must be chosen. Therefore, pattern grammars are primarily determined by the designer or through an interactive procedure in most cases. For the good introduction into the syntactic approach see [23, 26].

The relational graph is an alternative representation of a pattern structure, see [88, 89, 101]. Any useful relation that can be determined from the pattern can be incorporated in it. In contrast to a tree, a graph may contain closed loops and therefore, a richer description can be obtained.[1] The relational model of an unknown pattern obtained by preprocessing and segmentation is matched against a set of prototypes. The interesting approach is to describe the matching procedure of relational structures by means of predicate calculus with the idea to use a logic-programming language, e. g. Prolog [16].

However, serious drawbacks are involved in the pure structural approach. First, although the structural approach is powerful enough to describe the details of structure of an analyzed pattern, it is too weak to noise which may cause structural deformations to it. The result of parsing is: "recognized" or "rejected" with very little tolerance to even small structural changes. Secondly, the structural description obtained is not associated with any usually desirable semantics about recognized subpatterns and their relationships, which may be necessary for a proper interpretation of the analyzed pattern. Thirdly, the pure structural approach is not capable to handle numerical semantic information and to use it for recognition. Also, when a pattern can be generated by more than one pattern grammar, the problem of

[1] "However, the use of tree structures does provide a direct channel for adapting the techniques of formal language theory to the problem of compactly representing and analyzing patterns containing a significant structural content" [26].

ambiguity occurs. Therefore, the pure structural approach is for most practical applications quite unsatisfactory.

2.3. The Hybrid Approach

Both the structural and statistical approaches can yield unsatisfactory results if they are used for solutions of complex pattern recognition problems alone. The former is weak in handling noise patterns and numerical semantic information, the latter is unable to utilize information about patterns structures. That's the reason why the idea of combining both approaches has attracted so much attention.

For the cases, when analyzed patterns descriptions are generated by more than one pattern grammar,[2] the stochastic or fuzzy grammars have been proposed [25, 93, 97]. Each production has a "probability" associated with its application. The analyzed pattern is classified according to the occurrence probabilities of the pattern structure within each pattern class computed during parsing by multiplying the "probabilities" associated with the applied productions. Productions probabilities can be estimated from a given nonstochastic grammar and a sample set which usually consists of a set of distinct sample strings and their associated string probabilities [26].

Another way how to handle noisy and distorted patterns is the use of similarity measures. A similarity measure between two strings is usually interpreted as the weighted Levenshtein distance [59]. Substitution, insertion and deletion error transformations are considered. Each error transformation has a weight associated with its application in dependence which terminal (primitivum) the error transformation is made on. The distance between two strings is defined as the smallest value of the sum of the weights associated with the error transformations required to derive one string from the other one. The techniques used for the calculation of this distance are usually based on the algorithm proposed by Wagner & Fisher [104] and its modifications [5, 22, 34, 55, 67]. The derivation of one string from the other one is often illustrated by an oriented graph (lattice) each path of which corresponds to a sequence of error transformations. In the case of a stochastic model for syntax errors the deformation probability between two strings is defined as the largest value of the product of the deformation probabilities associated with the error transformations, see [62]. Sometimes, other similarity coefficients, e. g. Jaccard or Dice coefficient, may be also used to compute the similarity measure, cf. [81]. The obtained distances between strings can be further processed by an appropriate clustering method. In this manner, a distance between a string and a given language can be introduced, see [63].

If the pattern grammar is given, it can be extended by error productions, each of which corresponds to one type of error transformations. Their weights depend on which terminal symbol (primitivum) the error transformation is made on. In this manner, the deformed structural descriptions can be derived from the nondeformed ones and syntactic distortions caused by noise can be modeled. Then the analyzed pattern descriptions are parsed by error-correcting parsers [33, 58, 66, 94, 97, 102].

[2] Such a situation can often occur as a result of grammatical inference.

The nondeformed description can be obtained from an analyzed (usually deformed) one by eliminating error productions. Likewise, tree error-correcting automata may be used for more complex patterns, cf. [64].

Perhaps the most popular attempt how to overcome the drawbacks of both approaches is based on the introduction of attributed grammars into this field. The introduction of semantics usually reduces the grammar complexity, leads to improvement of both recognition accuracy and capability of recognition of noisy patterns and makes structural guidance of extraction of subpatterns attributes possible. The resulted structural description is associated with the desired semantics, see [17, 28, 45, 47, 50, 74, 75, 80, 90, 99, 103, 106, 113].

In the end of this section an interesting technique proposed by Goldfarb [36] should be mentioned. The two preliminary steps in it are: (1) the choice of an appropriate formal pattern representation, (2) the definition of an appropriate distance function. Taking as an input an interdistance matrix of the training samples, an efficient algorithm constructs a minimal vector representation of the sample. The resulting vector space is no longer a static universal space as in the classical statistical approach, but one determined by the given data. Thus, the space is constructed for a specific problem.

3. THE ARTIFICIAL INTELLIGENCE APPROACH TO PATTERN RECOGNITION

The classification of the approaches to pattern recognition into the conventional approach and the artificial intelligence one appeared probably first in [71]. The artificial intelligence approach was formulated to involve the description of abstract concepts and the recognition of instances of these in "signals". The concept is represented by several (hierarchical) levels of abstraction. At each level of abstraction, knowledge appropriate to that level is used to identify components of the higher level concept. At different levels different sources of knowledge are used to model the mechanism that generate and deform patterns.

The design of an AI-based signal understanding system involves the following major interdependent issues, namely knowledge representation, inference mechanisms, and the control structure.

In [71] the conventional approach is considered to be impoverished for automated extraction of information from "signals" because the approach does not allow the integration of large amounts of diverse information to accurately model the mechanisms that generate and deform the "signal" under study.

The above conclusion seems to be too strict because as it will be shown there are direct links between structural (hybrid) pattern recognition and artificial intelligence. It may be in some cases difficult to decide whether still the structural (hybrid) approach or the artificial intelligence one was used as the disciplines of pattern recognition and artificial intelligence continuously converge.

In the following, several links between structural (syntactic) pattern recognition and artificial intelligence are presented. Some of them already described in [8] are only briefly reviewed at first.

Generally, a syntactic pattern recognition system can be considered a particular instance of a knowledge based system with a formal grammar as knowledge base and a parser as inference engine.

A prototype graph can be interpreted as a simple version of a semantic net, especially when nodes and edges are augmented by attributes. The part-of hierarchy, playing an essential role in semantic networks, corresponds in many applications to a hierarchical scene description, which may also be built up by a parser during syntactic analysis.

An analogy between the different types of control structures in the area of artificial intelligence and structural pattern recognition control procedures can be also observed. Top-down and bottom-up control corresponds to top-down and bottom-up parsing, respectively. A heterarchical control structure can be interpreted as a tree or graph grammar parsers starting at any nonterminal. The blackboard control model corresponds to several grammars which look at a problem using different parsers reporting a series of conclusions to an agenda. Then there is another grammar which selects which parser to apply [8]. Another viewpoint is to regard parsing lists produced by the Earley parsing algorithm [20] as a form of blackboard, where current subparsings are recorded and new subparsings are proposed by predictor operation.

There has been some confusion about how a link between the left hand and right hand side of a grammar production and the promise and conclusion of the corresponding expert's knowledge should be interpreted. A viewpoint that a grammar production $A \rightarrow B$ represent the knowledge $IF \;\; A \;\; THEN \;\; B$ was presented in [7, 8, 29]. But in [109, 110], a context-free grammar production $A \rightarrow B$ has been shown to represent the knowledge $IF \;\; B \;\; THEN \;\; A$, where B is the ordered set of premises and A is the conclusion. Suppose, there are derivations of n sentences, generated by given context-free grammar G

$$S \Rightarrow x_1, \quad S \Rightarrow x_2, \; \ldots, \; S \Rightarrow x_n.$$

It means, if a sentence $x \in L(G)$ appears, the start symbol S appears surely, i.e. if x is true, then S is true. But from the start symbol S a specially designed sentence x is not necessarily formed, since S may derive any sentence of $L(G)$ besides x. Therefore, there exists the only relation IF x THEN S.

Note, that if abstract symbols are regarded as primitives (terminals) and subpatterns (nonterminals), then the theory of formal languages to syntactic pattern recognition can be applied. If the symbols are regarded as phenomena, facts, and conceptions, then the grammar production can be regarded as a form of knowledge representation. Both production systems and formal grammars are of common origin.

The grammar production is more capable of representing knowledge than the production rule because the relation among the symbols on the right side is ordered conjunction while in production system the order of propositions is not important. Therefore, the derivation tree of language generated by a context-free grammar is an ordered AND/OR tree, see also [39]. The relation "OR" exists among the derivation trees of each sentence. The ambiguous derivations of a sentence are also

"OR" related, and the branches of each node are of an ordered "AND" relation. A part of the work reported in [91] is based on this idea. The comparison of this work with the previous one of the same authors reported in [92] also supports the conclusion about the convergence between the disciplines of pattern recognition and artificial intelligence.

There are also analogies between heuristic inference algorithms and the search algorithms in structural (hybrid) pattern recognition. Stochastic, fuzzy grammars and attributed grammars may be considered to be the tools by means of which the experience heuristic information can be represented. Especially, in the attributed grammars there may be many calculations and tests involved in semantic rules to guide the inference process (syntax analysis).

In [109] the knowledge-based pattern recognition system based on the syntactic approach has been proposed. The system consists of two parts. In the first basic part the syntactic (hybrid) pattern recognition system plays the conventional role, in the second inference part it plays the role of knowledge-based system, where its function in storing knowledge and in inference is maintained. For the heuristic search the author presents a depth-first parsing algorithm; see also [111]. Because the strength of the syntactic approach is well founded background from the theory of formal languages, and the syntactic approach itself has the basic characteristics of knowledge-based systems, the syntactic recognition system can sometimes have advantages over some knowledge-based systems. On the other hand, the results of research in the field of knowledge-based systems are also valuable for pattern recognition. For example, the recognition system based on fuzzy set theory and approximate reasoning [68], and methods of combining multiple classifiers based on Dempster-Shafer theory [107] have been recently proposed. Combining syntactic pattern recognition system and knowledge-based system is valuable for both of them [109].

Let's turn back to the formulation of the artificial intelligence approach. In general, on different hierarchical levels of abstraction different sources of knowledge may be used. According to the fact that a grammar production is capable of representing knowledge there is also a link between the artificial intelligence approach and the syntactic (structural) one based on Lindenmayer's systems with interactions and tables (TIL-systems) [41], where on each hierarchical level first a table (source of knowledge) and then the productions (knowledge) from it are chosen. As an more powerful alternative to TIL-systems the hierarchical description systems based on sets of context-dependent substitutions, which may differ on each hierarchical level, were formulated in [82]. As it has been shown in [43] the experience heuristic information can be expressed by means of control sequences of sets of substitutions (tables), stochastic or fuzzy substitutions (productions), or in the best manner, by means of semantic tests and computations associated with each syntactic operation. In the latter case, attributed hierarchical description systems are used to specify the patterns classes under considerations. The patterns are analyzed from the rough level towards the more detailed ones. On each hierarchical level the appropriate set of substitutions associated with corresponding semantic computations (source knowledge) is chosen for next processing of an analyzed pattern according to the results so far obtained. The parallel nature of processing on separate hierarchical levels makes the idea of the introduction of parallel algorithms attractive. For more details see [43, 46, 82].

The results of research in the field of machine learning are also very valuable for pattern recognition. The field of machine learning is very diversified and the same problems are often studied by artificial intelligence researches, psychologists, cognitive scientists and others and treated from many different perspectives. As the result attempts to summarize the principles of learning in a unifying learning theory are still rare, e. g. [52, 70]. Many various algorithms and mechanisms have been found that can be used to assist automatic knowledge acquisition, extraction of relevant knowledge from large knowledge bases, and abstraction of higher level concepts out of data sets. The results reached in subdisciplines of machine learning, namely in learning from samples, learning by analogy, deductive learning, learning by observation, etc. can be often utilize (and can be expected to be utilized) in pattern recognition, namely in the problems of knowledge acquisition, grammatical inference, parameters estimation, abstraction of higher level concepts, inference mechanisms, decision trees, etc., see e. g. [65, 83, 96, 105, 112].

Especially, many machine learning systems have been proposed for constructive decision trees from collection of samples. Such systems can be used besides pattern recognition also in taxonomy, decision table programming, and switching theory. Four essential elements in a decision tree algorithm are 1) a set of features; 2) a feature selection criterion; 3) a stop-splitting rule and 4) a class assigning rule. Decision trees as hierarchical classifiers are especially preferred in the cases when the pattern classes are multimodel in nature [116].

Although treated as an independent discipline of artificial intelligence neural nets should be also mentioned in the frame of machine learning. Neural nets are often able to provide accurate and reliable classification, see e. g. [3, 6, 10, 32, 37, 38, 51, 60, 61, 72, 73, 84, 87, 108]. On the other hand, it should be also noted that they do not give insight into the inherent logic of the results and, as the result, users can have substantial problems with reasonable interpretation of them. Of crucial importance to the successful use of artificial neural networks for pattern classification problems is how the appropriate network size can be automatically determined [112].

The techniques used in the decision-theoretic approach usually involve large amounts of numerical computations. As the result the corresponding algorithms should be computationally inexpensive. The techniques used in the artificial intelligence approach involve intensive numerical computations only in the lowest levels of analysis - in the transformation of the digitized "signal" into a string of (hypothesized) symbols. The artificial intelligence based "signal" understanding scheme involves mainly the making of hypotheses and inferences [71]. Syntactic pattern recognition can be considered a link between these approaches. The conventional approach and the artificial intelligence one to pattern recognition are usually used at different levels of AI-pattern recognition systems. The coupling between the analytically formulated low-level "signal" processing and higher-level inference is the current trend. It has been proposed that the most effective control strategy (to solve difficult pattern recognition problems) would be a combination of top-down and bottom-up processing. The objective of such a control structure in pattern analysis is to develop a smart analysis method which can escape a huge amount of computation at the lowest levels. Feature extraction should be done only at the necessary local parts of the original "signal" with the expected feature extraction programs. But if many hierarchical levels for pattern analysis are used, the control has to go up and down between the levels. The wide use of such systems in practice especially depends on the results of research obtained in the field of machine learning and knowledge-based systems.

4. SUMMARY

The paper presents an overview of the current approaches in pattern recognition. The conventional approaches used so far to solve pattern recognition problems are usually divided into the decision-theoretic (statistical), structural (syntactic) and hybrid approaches. The artificial intelligence approach to pattern recognition is used in such very complex tasks each solution of which heavily depends on the knowledge of experts. With the help of some appropriate artificial intelligence techniques this approach is capable to integrate large amounts of diverse information and to model the mechanisms that generate and deform the signals under study.

The decision-theoretic approach is for solution of many practical problems satisfactory enough. On the other hand, it does not utilize structural properties of analyzed patterns and it does only lead to class descriptions of patterns but it does not provide any descriptions of them. Also, quantifying the contribution of a particular feature towards the accuracy of classification is ambiguous.

The structural approach is capable to describe the details of structure of an analyzed pattern but it is too week to noise which may cause the structural changes to it. It can not handle numerical semantic information and it can not use it for recognition.

The hybrid approach includes various techniques, e. g. stochastic and fuzzy grammars, various similarity measures, error transformations, error-correcting parsers. The attributed grammars are probably the most effective tool of this approach.

The artificial intelligence approach was formulated to involve the description of abstract concepts and the recognition of instances of these in "signals". The concept is represented by several hierarchical levels of abstraction. At each level of abstraction, knowledge appropriate to that level is used to identify components of the higher level concept.

The links between the structural (hybrid) approach and the artificial intelligence one are reviewed in the paper. It may be in some cases difficult to decide whether still the structural approach or the artificial intelligence one was used as the disciplines of pattern recognition and artificial intelligence continuously converge. As a nice example the knowledge-based pattern recognition system based on syntactic approach proposed in [110] is noticed. Because the strength of syntactic (structural) approach is a well founded background from the theory of formal languages, and the syntactic approach itself has the basic characteristics of knowledge-based system, the syntactic recognition system can sometimes have advantages over some

knowledge-based system. On the other hand, the results of research in the field of knowledge-based systems may be also valuable for pattern recognition. The same conclusion holds in the case of machine learning.

From the recently found new links between the artificial intelligence approach and the structural (syntactic) approach the attributed hierarchical description systems [43] should be mentioned.

The conventional approach and the artificial intelligence one to pattern recognition are usually used at different levels of pattern recognition systems. The coupling between the analytically formulated low-level "signal" processing and higher-level inference is the current trend. In practice, constructions of the systems based on a combination of top-down and bottom-up processing especially depend on the progress obtained in the field of machine learning and of knowledge-based systems.

(Received December 8, 1992.)

REFERENCES

[1] A. Aho and J. Ullman: The Theory of Parsing, Translation and Compiling, Volume I. Prentice Hall, Englewood Cliffs, N. J. 1972.

[2] M. R. Anderberg: Cluster Analysis for Applications. Academic Press, New York 1973.

[3] E. Barnard, R. A. Cole, M. P. Vea and F. A. Alleva: Pitch detection with a neural-net classifier. IEEE Trans. Signal Process. *39* (1991), 2, 298–307.

[4] A. W. Biermann and A. J. Feldman: A survey of results in grammatical inference. In: Frontiers of Pattern Recognition, Academic Press, New York – London 1972.

[5] R. Božinovič and S. N. Srihari: A string correction algorithm for cursive script recognition. IEEE Trans. Pattern Recognition Mach. Intellig. *4* (1982), 12, 655–663.

[6] D. E. Brown, V. Corruble and C. L. Pittard: A comparison of decision tree classifiers with backpropagation neural networks for multimodal classification problems. Pattern Recognition *26* (1993), 6, 953–961.

[7] H. Bunke: Hybrid methods in pattern recognition. In: Pattern Recognition Theory and Applications, Springer–Verlag, Berlin 1987.

[8] H. Bunke and A. Sanfeliu: "Introduction" to the special issue: advances in syntactic pattern recognition. Pattern Recognition *19* (1986), 4, 249–254.

[9] J. J. Bussgang and M. B. Marcus: Truncated sequential hypothesis tests. IEEE Trans. Inform. Theory *3* (1967), 3, 512–516.

[10] G. A. Carpenter and S. Grossberg: Pattern Recognition by Self-Organizing Neural Networks. MIT Press, Cambridge, Mass. 1991.

[11] R. L. Chang and T. Pavlidis: Fuzzy decision tree algorithms. IEEE Trans. Systems Man Cybernetics 7 (1977), 1, 28–35.

[12] Z. T. Chiang and K. S. Fu: Parallel parsing algorithms and VLSI implementation for syntactic pattern recognition. IEEE Trans. Pattern Recognition Mach. Intellig. *6* (1984), 3, 302–314.

[13] Y. T. Chien and K. S. Fu: A modified sequential recognition machine using time-varying stopping boundaries. IEEE Trans. Inform. Theory *12* (1966), 2, 206–214.

[14] N. Chomsky: Three models for the description of languages. IEEE Trans. Inform. Theory *2* (1956), 113–124.

[15] C. K. Chow: On optimum recognition error and reject trade off. IEEE Trans. Inform. Theory *16* (1970), 1, 41–46.

[16] W. F. Clocksin and C. S. Mellish: Programming in Prolog. Springer–Verlag, Heidelberg 1984.

[17] P. Deransart, M. Jourdan and B. Lorho: Attribute Grammars. (Lecture Notes in Computer Science.) Springer–Verlag, Berlin – New York – London 1988.
[18] B. Dubuisson and M. Masson: A statistical decision rule with incomplete knowledge about classes. Pattern Recognition *26* (1993), 1, 155–165.
[19] R. O. Duda and P. E. Hart: Pattern Classification and Scene Analysis. Wiley, New York 1973.
[20] J. Earley: An efficient context-free parsing algorithm. Comm. ACM *13* (1970), 2, 94–102.
[21] J. Feder: Plex-languages. Inform. Sci. *3* (1971), 225–241.
[22] N. V. Findler and J. Leeuwen: A family of similarity measures between two strings. IEEE Trans. Pattern Recognition Mach. Intellig. *1* (1979), 1, 116–119.
[23] K. S. Fu: Syntactic models in pattern recognition and applications. In: Pattern Recognition in Practice, Proc. Internat. Workshop, North-Holland, Amsterdam 1980.
[24] K. S. Fu: Sequential Methods in Pattern Recognition and Machine Learning. Academic Press, New York 1968.
[25] K. S. Fu: On syntactic pattern recognition and stochastic languages. In: Frontiers of Pattern Recognition, Academic Press, New York – London 1972.
[26] K. S. Fu: Syntactic Pattern Recognition and Applications. Prentice Hall, New Jersey 1982.
[27] K. S. Fu: Tree languages and syntactic pattern recognition. In: Pattern Recognition and Artificial Intelligence. Academic Press, New York 1976.
[28] K. S. Fu: A step towards unification of syntactic and statistical pattern recognition. IEEE Trans. Pattern Recognition Mach. Intellig. *5* (1983), 2, 200–205.
[29] K. S. Fu: Syntactic Methods in Pattern Recognition. Academic Press, New York 1974.
[30] K. S. Fu and T. L. Booth: Grammatical inference: introduction and survey. IEEE Trans. Pattern Recognition Mach. Intellig. *5* (1975), 1, 95–111 (part I), 4, 409–423 (part II).
[31] K. S. Fu and Y. T. Chien and G. P. Cardillo: A dynamic programming approach to sequential pattern recognition. IEEE Trans. Electronic Computers *13* (1967), 12, 484–492.
[32] M. Fukumi, S. Omatu, F. Takeda and T. Kosaka: Rotation-invariant neural pattern recognition system with application to coin recognition. IEEE Trans. Neural Networks *3* (1992), 2, 272–279.
[33] L. W. Fung and K. S. Fu: Stochastic syntactic encoding for pattern classification. IEEE Trans. Comput. *24* (1975), 6, 662–669.
[34] Z. Galil and K. Park: An improved algorithm for approximate string matching. SIAM J. Comput. *19* (1990), 6, 989–999.
[35] A. V. Gladkij: Formalnyje gramatiki i jazyky (Formal Grammars and Languages). Nauka, Moscow 1973.
[36] L. Goldfarb: A unified approach to pattern recognition. Pattern Recognition *17* (1984), 5, 575–582.
[37] H. Guo and S. B. Gelfand: Classification trees with neural network feature extraction. IEEE Trans. Neural Networks *3* (1992), 6, 923–933.
[38] L. Gupta, J. Wang, A. Charles and P. Kisatsky: Prototype selection rules for neural network training. Pattern Recognition *25* (1992), 1401–1408.
[39] P. A. N. Hall: Equivalence between AND/OR graphs and context-free grammars. Comm. ACM *16* (1973), 7, 444–445.
[40] M. A. Harrison: Introduction to Formal Language Theory. Addison–Wesley, London – Amsterdam – Sydney 1978.
[41] G. T. Herman and G. Rozenberg: Development Systems and Languages. North-Holland–Elsevier, Amsterdam – New York 1975.

[42] T. Grönfers and M. Juhola: Experiments and comparison of inference methods of regular grammars. IEEE Trans. Systems Man Cybernet. *22* (1992), 4, 821–830.
[43] J. Kepka: Hierarchical Approach to the Analysis of Experimental Curves (in Czech). Institute of Information Theory and Automation Research Report, Prague 1991.
[44] J. Kepka: The employment of Prolog for syntax analysis in syntactic pattern recognition applications. Kybernetika *28* (1992), 1, 62–68.
[45] J. Kepka: Several comments on pattern recognition system based on the use of attributed grammars. Kybernetika *28* (1992), 1, 69–76.
[46] J. Kepka: The structural description of self-embedding nature of waveform peaks. Kybernetika *30* (1994), 1, 63–76.
[47] J. Kepka: An example of waveform analysis based on attributed hierarchical descriptions. In: Proc. Czech Pattern Recognition Workshop'93, Temešvár u Písku, November 4–6, Czechoslovak Pattern Recognition Society, 1993, pp. 144–148.
[48] J. Kepka and J. Psutka: A voice controlled system in a noisy environment. In: Proc. Czech Pattern Recognition Workshop'93, Temešvár u Písku, November 4–6, Czechoslovak Pattern Recognition Society, 1993, pp. 90–95.
[49] J. Kittler and P. C. Young: A new approach to feature selection based on the Karhunen–Loeve expansion. Pattern Recognition *5* (1973), 4, 335–352.
[50] A. Klaiber and M. Gokhale: Parallel evaluation of attribute grammars. IEEE Trans. Pattern Recognition Mach. Intellig. *3* (1992), 2, 206–220.
[51] S. Knerr, L. Personnaz and G. Dreyfus: Handwritten digit recognition by neural networks with a single-layer training. IEEE Trans. Neural Networks *3* (1992), 6, 962–968.
[52] Y. Kondratoff: Introduction to Machine Learning. Pitman, London 1988.
[53] M. Kudo and M. Shimbo: Efficient regular grammatical inference techniques by the use of partial similarities and their logical relationships. Pattern Recognition *21* (1988), 4, 401–409.
[54] A. Kumar: A real time system for pattern recognition of human sleep stages by fuzzy system analysis. Pattern Recognition *9* (1977), 1, 43–46.
[55] S. Kuo and G. R. Cross: A two step string-matching procedure. Pattern Recognition *24* (1991), 7, 711–716.
[56] T. Kurita: An efficient agglomerative clustering algorithm using a heap. Pattern Recognition *24* (1991), 3, 205–209.
[57] R. Kurki-Suonio: On top to bottom recognition and left recursion. Comm. ACM *9* (1966), 7, 527–528.
[58] H. C. Lee and K. S. Fu: A stochastic syntax analysis procedure and its applications to pattern classification. IEEE Trans. Comp. *21* (1972), 7, 660–667.
[59] V. I. Levenshtein: Binary codes capable of correcting deletions, insertions and reversals. Soviet Phys. Dokl. *10* (1966), 11, 707–710.
[60] R. Linggard, D. J. Myers and C. Nightingale: Neural Networks for Vision, Speech and Natural Language. Chapmen & Hall, London 1992.
[61] D. Lowe and A. R. Web: Optimized feature extraction and the Bayes decision in feed-forward classifier networks. IEEE Trans. Pattern Recognition Mach. Intellig. *13* (1991), 4, 335–364.
[62] S. Lu and K. S. Fu: Stochastic error correcting syntax analysis for recognition of noisy patterns. IEEE Trans. Comput. *26* (1977), 12, 1268–1276.
[63] S. Lu and K. S. Fu: A sentence to sentence clustering procedure for pattern analysis. IEEE Trans. Systems Man Cybernet. *8* (1978), 5, 381–389.
[64] S. Lu and K. S. Fu: Error correcting tree automata for syntactic pattern recognition. IEEE Trans. Comput. *27* (1978), 12, 1043–1052.
[65] G. Lugosi: Learning with an unreliable teacher. Pattern Recognition *25* (1992), 1, 79–87.

[66] G. Lyon: Syntax-directed least-errors analysis for context free languages: a practical approach. Comm. ACM *17* (1974), 1, 3–14.

[67] M. Maes: Polygonal shape recognition using string-matching techniques. Pattern Recognition *24* (1991), 5, 433–440.

[68] D. P. Mandal and C. A. Murthy and K. S. Pal: Formulation of a multivalued recognition system. IEEE Trans. Systems Man Cybernet. *22* (1992), 4, 607–620.

[69] J. Mantas: Methodologies in pattern recognition and image analysis – a brief survey. Pattern Recognition *20* (1987), 1, 1–6.

[70] R. S. Michalski: Toward a unified theory of learning: an outline of basic ideas. First World Conference on the Fundamentals of Artificial Intelligence, Paris, July 1–5, 1991.

[71] N. Nandhakumar and J. K. Aggarwal: The artificial intelligence approach to pattern recognition – a perspective and an overview. Pattern Recognition *18* (1985), 6, 383–389.

[72] S. C. Newton and J. S. Pemmaraju and S. Mitra: Adaptive fuzzy leader clustering of complex data sets in pattern recognition. IEEE Trans. Neural Networks *3* (1992), 5, 794–800.

[73] S. K. Pal and S. Mitra: Multilayer perceptron, fuzzy sets, and classification. IEEE Trans. Neural Networks *3* (1992), 5, 683–697.

[74] G. K. Papakonstantinov: An interpreter of attributed grammars and its application to waveform analysis. IEEE Trans. Soft. Engineering *7* (1981), 3, 279–283.

[75] G. Papakonstantinov, E. Skordalakis and F. Gritzali: An attribute grammar for QRS detection. Pattern Recognition *19* (1986), 4, 297–303.

[76] T. Pavlidis: Structural Pattern Recognition. Springer–Verlag, New York 1977.

[77] T. Pavlidis: Linear and context-free graph grammars. J. Assoc. Comput. Mach. *19* (1972), 1, 11–22.

[78] T. Pavlidis: Representation of figures by labeled graphs. Pattern Recognition *4* (1972), 1, 5–16.

[79] J. L. Pfaltz and A. Rosenfeld: Web Grammars. Proc. 1st Internat. Joint Conf. Artif. Intell., Washington DC, 1969.

[80] E. Pietka: Feature extraction in computerized approach to the ECG analysis. Pattern Recognition *24* (1991), 2, 139–146.

[81] J. Pik: Structural analysis of experimental curves in numerical taxonomy. Proc. 8th ICPR, Paris 1986.

[82] J. Pik: A hierarchical pattern description in the syntactic approach to pattern recognition. Pattern Recognition *14* (1981), 2, 191–196.

[83] C. E. Priebe and D. J. Marchette: Adaptive mixtures: recursive nonparametric pattern recognition. Pattern Recognition *24* (1991), 12, 1197–1209.

[84] S. Sakaue, T. Hohda, H. Yamamoto, S. Maruno and Y. Shimeki: Reduction of required precision bits for backpropagation applied to pattern recognition. IEEE Trans. Neural Networks *4* (1993), 2, 270–275.

[85] A. Salomaa: Formal Languages. Academic Press, New York – London 1973.

[86] J. Schuermann and W. Doster: A decision-theoretic approach to hierarchical classifier design. Pattern Recognition *17* (1984), 3, 359–369.

[87] I. K. Sethi and A. K. Jains: Artificial Neural Networks and Statistical Pattern Recognition: Old and New Connections. North Holland, Amsterdam 1991.

[88] A. C. Shaw: Picture graphs, grammars, and parsing. In: Frontiers of Pattern Recognition, Academic Press, New York – London 1972.

[89] A. C. Shaw: Parsing of graph-representable pictures. Comm. ACM *17* (1970), 3, 453–481.

[90] E. Skordalakis: Syntactic ECG processing. A review. Pattern Recognition *19* (1986), 4, 305–313.

[91] G. Stockman and L. N. Kanal: Problem-reduction representation for the linguistic analysis of waveforms. IEEE Trans. Pattern Recognition Mach. Intellig. *5* (1983), 3, 287–298.

[92] G. Stockman and L. N. Kanal and M. C. Kyle: Structural pattern recognition of carotid pulse waves using a general waveform parsing system. Comm. ACM *19* (1976), 12, 688–695.

[93] P. H. Swain and K. S. Fu: Stochastic programmed grammars for syntactic pattern recognition. Pattern Recognition *4* (1972), 1, 83–100.

[94] E. Takanaka and K. S. Fu: Error correcting parsers for formal languages. IEEE Trans. Comp. *27* (1978), 7, 605–616.

[95] E. Tanaka and K. Ezure: Direct parsing. Pattern Recognition *19* (1986), 4, 315–323.

[96] G. D. Tecuci: Automating knowledge acquisition as extending, updating, and improving a knowledge base. IEEE Trans. Systems Man Cybernet. *22* (1992), 6, 1444–1460.

[97] M. G. Thomason and R. C. Gonzales: Syntactic recognition of imperfectly specified patterns. IEEE Trans. Comput. *24* (1975), 1, 93–95.

[98] J. T. Tou and R. C. Gonzales: Pattern Recognition Principles. Addison–Wesley, Reading 1974.

[99] P. Trahanias and E. Skordalakis: Syntactic pattern recognition of the ECG. IEEE Trans. Pattern Recognition Mach. Intellig. *12* (1990), 7, 648–657.

[100] W. H. Tsai and K. S. Fu: A pattern deformational model and Bayes error correcting recognition system. IEEE Trans. Systems Man Cybernet. *9* (1979), 745–756.

[101] W. H. Tsai and K. S. Fu: Error correcting isomorphisms of attributed relational graphs for pattern analysis. IEEE Trans. Systems Man Cybernet. *9* (1979), 757–768.

[102] W. H. Tsai and K. S. Fu: A pattern deformational model and Bayes error correcting recognition system. IEEE Trans. Systems Man Cybernet. *9* (1979), 745–756.

[103] W. H. Tsai and K. S. Fu: Attributed grammar – a tool for combining syntactic and statistical approaches to pattern recognition. IEEE Trans. Systems Man Cybernet. *10* (1980), 12, 873–885.

[104] R. A. Wagner and M. J. Fisher: The string to string correction problem. J. Assoc. Comput. Mach. *21* (1974), 1, 168–173.

[105] L. Wang and J. M. Mendel: Generating fuzzy rules by learning from examples. IEEE Trans. Systems Man Cybernet. *22* (1992), 6, 1414–1427.

[106] Q. Xiao and H. Raafat: Fingerprint image postprocessing: a combined statistical and structural approach. Pattern Recognition *24* (1991), 10, 985–992.

[107] L. Xu, A. Krzyžak and Ch.Y. Suen: Methods of combining multiple classifiers and their applications to handwriting recognition. IEEE Trans. Systems Man Cybernet. *22* (1992), 3, 418–435.

[108] H. Yan: Prototype optimization for nearest neighbor classifiers using a two-layer perceptron. Pattern Recognition *26* (1993), 2, 317–324.

[109] G. Z. Yang: A syntactic approach for building knowledge-based pattern recognition system. Proc. 9th ICPR, Rome 1988, pp. 1236–1238.

[110] G. Z. Yang: On the knowledge-based pattern recognition using syntactic approach. Pattern Recognition *24* (1991), 1, 185–193.

[111] G. Z. Yang: The search algorithms stimulated by premise set in the syntactic knowledge system. Pattern Recognition *26* (1993), 1, 17–22.

[112] D. Y. Yeung: Constructive neural networks as estimators of Bayesian discriminant functions. Pattern Recognition *26* (1993), 1, 189–204.

[113] K. C. You and K. S. Fu: A syntactic approach to shape recognition using attributed grammars. IEEE Trans. Systems Man Cybernet. *9* (1979), 6, 334–345.

[114] T. Y. Young and T. W. Calvert: Classification, Estimation and Pattern Recognition. Elsevier, New York 1974.

[115] L. A. Zadeh: Fuzzy sets. Inform. and Control *8* (1965), 3, 338–353.
[116] X. J. Zhou and T. S. Dillon: A statistical-heuristic feature selection criterion for decision tree induction. IEEE Trans. Systems Man Cybernet. *13* (1991), 8, 834–841.

Ing. Jiří Kepka, CSc., Západočeská univerzita (The University of West Bohemia), Univerzitní 22, 306 14 Plzeň. Czech Republic.