ELSEVIER

Pattern Recognition Letters 23 (2002) 493–499

Pattern Recognition Letters

www.elsevier.com/locate/patrec

A note on core research issues for statistical pattern recognition

Robert P.W. Duin [a,*], Fabio Roli [b], Dick de Ridder [a]

[a] Pattern Recognition Group, Department of Applied Physics, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, Netherlands

[b] Department of Electrical and Electronic Engineering, University of Cagliari, Cagliari, Italy

1. Introduction

Is statistical pattern recognition a collection of topics and techniques borrowed from other fields? Or has it an identity of its own, generating specific research questions, producing its own theoretical framework? This question arises occasionally between participants of the workshops and conferences focusing on the statistically oriented procedures and applications in pattern recognition.

Although many pattern recognition scientists seem pretty sure about the cultural identity of their research field, the authors think that the recent developments of closely related disciplines (e.g., machine learning and neural networks) and the increasing number of research issues that pattern recognition shares with such disciplines makes the answer to the above question more difficult than in past times.

Therefore, this paper aims to stimulate discussion in the pattern recognition community on the structural differences between statistical pattern recognition and closely related disciplines in order to clarify the present cultural identity and the core research issues.

According to computer science encyclopaedia and some famous textbooks, pattern recognition (PR) can be defined as the discipline that studies theories and methods for designing machines that are able to recognise patterns in noisy data (Srihari and Govindaraju, 1993; Duda et al., 2001; Fukunaga, 1990; Devijver and Kittler, 1982). Although this is only one of the possible definitions for PR, it points out well the "engineering" nature of this scientific discipline, because it says that the final goal of PR is the design of "machines". In this engineering perspective on PR, it is quite obvious that PR theory has multi-disciplinary roots, because engineering disciplines aim to bridge the gap between real-world applications and the so-called pure disciplines (mathematics, statistics, physics, etc.). It is worth noting that this multi-disciplinary character of the theoretical foundations of PR was already evident in the seminal book by Watanabe (1985). Watanabe pointed out that pattern recognition can be regarded as a statistical decision-making task, as a structure analysis problem, as an induction process, as a perception task, etc., thereby acknowledging the multiple cultural roots of PR. With regard to the strong relations between PR and artificial intelligence, Tveter recently stated that many artificial intelligence methods can be

* Corresponding author. Tel.: +31-15-278-6143; fax: +31-15-278-6740.

E-mail addresses: duin@ph.tn.tudelft.nl (R.P.W. Duin), roli@diee.unica.it (F. Roli), dick@ph.tn.tudelft.nl (D. de Ridder).

0167-8655/02/$ - see front matter
PII: S0167-8655(01)00181-7

regarded as different ways of doing PR (Tveter, 1998).

At present, the multi-disciplinary roots of pattern recognition theory and the influence of other research communities on the PR field are evident. As examples, we note that:
- Statistical decision and estimation theories commonly used in PR have been almost entirely developed by the statistics community.
- The machine learning community has developed pattern classifiers, such as multi-layer perceptrons and decision trees, that are widely used for pattern recognition now.
- Methods of syntactic and structural pattern recognition exploit theories and concepts originating from the disciplines of formal languages and graph theory.
- Methods for knowledge representation and symbolic processing used in pattern recognition systems are based on theories previously developed in the artificial intelligence field.

In addition, with reference to statistical pattern recognition, one can note that many of the newer and promising theoretical developments have been proposed outside the PR community (e.g., support vector machines, graphical models, bagging and boosting of multiple classifiers).

These observations raise the question towards the core research issue for PR. Is it about applying in practice techniques that are developed elsewhere, or are there problems unique to PR that need to be solved at a theoretical level as well? The main goal of this paper is to stimulate discussion in the PR community on the point of the identity of their research field. In particular, we would like such a discussion to provide some answers to the following question:

> *What are the theoretical research issues that should constitute the "core business" of the PR community because they are fundamental to PR and are not (or only partially) addressed by other communities?*

In the following, we open the discussion from the viewpoint of statistical pattern recognition. As our goal is to stimulate the discussion within the PR community a clear or final answer is out of the scope of this paper. In addition, it should be noted that the discussion from the viewpoints of other main PR approaches (i.e., syntactic, structural, symbolic/knowledge-based, and neural approaches) remains to be addressed by others within the PR community.

2. Statistical pattern recognition as a complex information processing task

Traditionally, the discipline of statistical pattern recognition (SPR) studies how we can generalise from a set of examples of objects ("observations") to some classifier or description. It thereby includes stages like sensing, data preprocessing (e.g., image filtering and segmentation), feature extraction, and classification (Fig. 1). Sometimes the sensors are determined by the application and the initial feature vector representation is defined by an expert. In such simple cases, the SPR process starts with the classification task

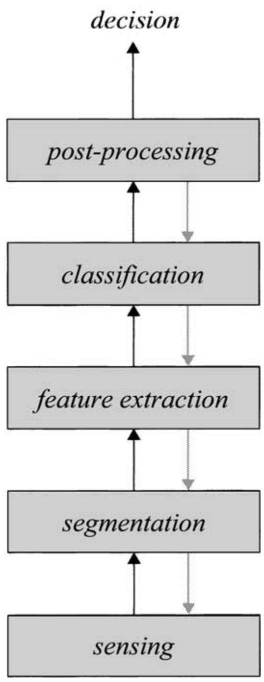

Fig. 1. Traditional description of the pattern recognition system.

and may be considered a direct application of machine learning or statistics methods.

In other words, we are saying that SPR can be considered as a direct application of supervised learning or statistics when the feature-vector-based "modelling" of the pattern recognition task is simple and appropriate, and classification methods provided by such disciplines can be directly applied to solve the problem at hand.

On the other hand, it is well known that the particular difficulties in the design of an SPR system concerns exactly this modelling of the pattern recognition problem and the effective application of available classification tools. Therefore, if we want to study the cultural identity of SPR, all steps between the real world objects and how they are initially represented in a feature space should definitely be taken into account. In the following, we will clarify our opinion about the importance of the modelling stage in the SPR process.

For the sake of our discussion, it should be also noted that real SPR systems often need feedback from higher processing levels to lower ones. As an example, the well-known "chicken-and-egg" problem related to image segmentation and pattern classification can require feedback between these two processing levels. In addition to such feedback among processing levels, the complexity of real-world applications requires that very particular issues be addressed within each level. As examples, we cite the following specific issues that a researcher constructing SPR systems must address:

- design of sensors and sensing modalities tailored to application requirements;
- choice of (invariant) features and feature selection;
- classifier training using small and unbalanced data sets;
- the need for error reject options, taking into account the application requirements;
- exploitation of prior knowledge (e.g., application knowledge and design expertise);
- development of control mechanisms for handling feedback among processing levels;
- systematic criteria and guidelines for choosing the classification tool most appropriate for the application at hand.

It should be quite evident that SPR deals with complex information-processing tasks involving various processing levels, feedback among levels, and specific issues. How to characterize, however, the SPR discipline short and clearly?

It is common among scientists to distinguish scientific disciplines from each other by their topics, and, depending on these topics, by their methods (Poser, 1998). It seems to the authors that the disciplines of artificial intelligence, machine learning, neural networks, and statistics develop theories and methods useful for some of the above processing levels (e.g., for the pattern classification level) and for some of the issues specific to SPR. However, they do not (or only partially) address many issues that are fundamental to SPR, like the way objects may be represented. In addition, due to the different goals of such disciplines, they do not address the engineering questions related to the construction of an SPR system, i.e., the system and application issues. Therefore, one could say that SPR and closely related disciplines share some, but not all, research issues and develop some similar methods.

What, then, is the core issue that SPR research should address? In our opinion, it should try to answer the questions of how to relate and/or to modify theories and methods developed in other disciplines, taking into account the requirements of real world applications.

According to the above-mentioned engineering perspective, SPR discipline should focus on the issues related to the modelling of real PR tasks and on all those issues that limit the effective application of methods developed by other disciplines (e.g., the limits of support vector machines related to multi-class tasks and its possibility to handle rejects). SPR scientists should formulate/model carefully the PR task at hand. Then, they should be able to evaluate the effectiveness or the limits of the tools developed by other disciplines. Adaptations or substantial modifications can be requested for applying effectively the selected tool. According to the engineering paradigm, new tools should be invented when adaptation is clearly not effective. Finally, SPR scientists should devote substantial efforts to the development of methods for the effective use of available PR tools (e.g., methods for effective training of classification algorithms).

This calls upon questions in (at least) three stages of the SPR process: representation, adaptation and generalisation (Fig. 2). As an example, consider the application of a tool like discriminant analysis to some object recognition problem:

- *Representation*: how to represent an object by numbers (e.g., using a CCD camera to create a pixel representation, but perhaps also finding a contour description to represent the shape).
- *Adaptation*: how to change the representation such that it is suitable for the desired tool (e.g., the determination of a small set of features).
- *Generalization*: Training and evaluation (e.g., of a discriminant for solving the recognition problem at hand).

This more general description of the pattern recognition system leaves room for solutions that do not include segmentation (but, for instance, classify an image as a whole) or are not feature based, but use other representations like shapes, dissimilarities, etc. Fig. 2 distinguishes the pattern recognition area clearly from artificial intelligence where the step of *representation* is integrated with *reasoning*.

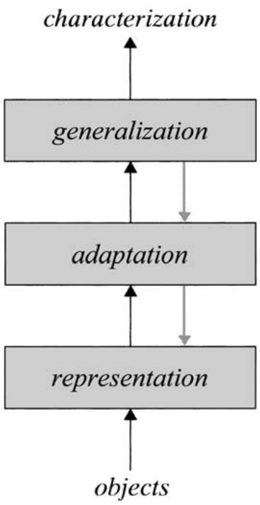

Fig. 2. A more general description of the pattern recognition system.

We like to stress that the primary question, how to represent physical objects, is included in the responsibility of designing a PR system. The PR expert should take care of the physical world and should not start in some "abstract" feature space. As pointed out in the Introduction of the book by Duda et al. (2001), the use of the "abstract" feature space representation allows the development of elegant and domain-independent theories and methods. However, the SPR system designer should carefully avoid the risk of using abstract representations that do not model the characteristics of real PR tasks. A careful trade-off between domain-independence and modelling of physical world should be pursued.

In the following, we briefly discuss these three general problems and the above-mentioned system and application problems. In Section 3, we will list specific issues related to these general problems.

2.1. Representation

In a practical application, we have real world objects. In order to use tools developed within other disciplines, they have to be represented in an appropriate way, e.g., in a feature space. Other representations, however, are possible as well, such as contours or distances to a representation set. Domain knowledge should be used here. For example, the representation should be insensitive to invariants. An early attempt to study other types of representations was made by Gelsema et al. when he introduced correspondence analysis in the field of pattern recognition (Queiros et al., 1983; Gelsema et al., 1982).

The search for an adequate representation requires knowledge of the application as well as of the set of tools one likes to use. Moreover, an optimization of the representation over a set of examples and for a given tool-set may be included in the SPR task. As far as these tools are learning from examples, they belong to the domain of statistical pattern recognition. Some examples of the optimization of a representation are feature selection, feature extraction or prototype selection based on very general criteria like the mutual information or the 1-NN performance. Such a criterion may make use of solution tools to be investigated further in the recognition process. At this stage, however, they are not optimized and even the use of other ones (e.g., a neural network) may be decided later.

2.2. Adaptation

The representation is often not directly suited for the tools that are desired. For tools that have to be trained, the most common mismatch is between the size of the set of examples and the complexity of the representation and/or the procedure. There are various ways to adapt these: simplifying the representation (e.g., feature reduction), simplifying the tool (e.g., less neurons in a neural network), artificially generating new observations, etc. In order to do this properly knowledge on the representation as well as on the tools is necessary. Another example of adaptation is (nonlinear) rescaling of features such that a classification tool can be used that makes certain assumptions on the distribution of the data. In this case this adaptation should really be focused on the solution tools that are finally used. It is the challenge for the PR analyst to become familiar with both worlds, the application area as it is primarily represented and the possible tool-sets. She/he has to select a generalisation tool and adapt the representation to its properties such that the performance is optimized.

2.3. Generalization

Finally, the tool itself has to be applied, e.g., a classifier has to be trained. This has to be followed by an evaluation from which previous steps can be judged, or by which a comparison between different tools can be made. For this, knowledge on the tools as well as on evaluation procedures is necessary. Often, tools may be used which have been developed within other domains, like statistics or machine learning. Classifiers like decision trees and neural networks, and many density estimators used in the Bayes decision rule are (almost) entirely developed outside the PR area. The nearest neighbour rule, on the contrary and many of its derivatives have received much attention within this area, as they can often directly be used on the representation, without much need for adaptation. Research on classifiers is primarily of interest for the PR analyst if she/he needs them to be more robust for his application, or if she/he likes to evaluate them over a class of problems. An emerging and very interesting evaluation issue is the so-called "meta analysis" of classification algorithms that aims to define systematic criteria for choosing the best classifier with respect to the task at hand (Sohn, 1999; Ho, 2000).

2.4. System design

The total system belongs definitely to the area of interest for researchers in the pattern recognition field. This includes primarily procedures to test its performance, also when the test set has to be partially used for the optimization. Further, computational aspects may be important: speed, memory demands, possibilities for user interaction, the integration of sensors and a reduction in their number, etc. Generally speaking, methods for obtaining a satisfactory trade-off between recognition accuracy and required computational resources should be investigated. Another system issue of great interest is the development of effective control mechanisms for handling feedback among processing layers.

2.5. Application issues

Many of the pattern recognition applications are related to 2-dimensional (or even multi-dimensional) signals like images. Due to historical reasons, less studied in the PR field are 1-dimensional problems like speech and time-signal processing. The entire PR area is closely related to sensors and to real world objects. Thereby the study of various sensor types and measurement procedures may be important. Consequently, it has to deal with problems like sensor noise, outliers, missing data and object variability. Moreover, in many applications user interaction is allowed or demanded. It may be important to be able to explain the outcome of the total system in terms of the problem.

The use of statistics is generated by the desire to learn from examples. The traditional demand is that the set of examples should be representative for the objects to be recognised. In many applications this is not really possible. As a consequence, one should correct for the probability, or even for the existence, of objects. Finally, the ever-returning question in statistical pattern recognition is how existing model knowledge (prior knowledge) can be combined with statistical procedures that gain new knowledge. The logical extreme related to this issue is the so-called "analysis by synthesis", that is, the creation of a model of how each pattern is generated.

3. Core research issues for statistical pattern recognition: an open list

This list is meant to fill in the issues discussed in the previous section and give examples of practical problems:

Representation: how can objects be represented (features, point sets, characteristic curves, similarities), how can spatial or temporal information be used, choice of (invariant) features and feature selection, how should missing data be handled (parts of objects, missing features).

Adaptation: feature reduction, artificial data generation (e.g., by noise injection), semi-parametric and adaptive PR tools, practical trade-offs between representation complexity and classifier complexity.

Generalization: classifier training using very small and unbalanced data sets with ambiguous, missing, or wrong training data, need of error reject options taking into account the application requirements, systematic criteria and guidelines for choosing the classification tool most appropriate for the application at hand, combining multiple classifiers and constructing hybrid classification systems.

System design: development of control mechanisms for handling feedback among processing levels, special purpose hardware and software (e.g., special purpose parallel architectures and visual languages), trade-off between recognition accuracy and computational resources, investigating how the PR system scales as a function of the number of features, number of patterns, etc.

Application issues: design of sensors and sensing modalities tailored to application requirements, analysis by synthesis, exploitation of prior knowledge (e.g., application knowledge and design expertise), modelling the pattern generation processes.

4. Outlook

In this paper, we pointed out that statistical pattern recognition involves a number of issues that are not (or only partially) addressed by other research communities. Therefore, we think that the role the SPR community should play is not merely the evaluation of the theories developed by other communities, in the context of real-world applications. Pattern recognition, as an engineering discipline, should mainly study theories and methods that bridge the gap between real-world applications and the pure disciplines. Some specific conferences and workshops in this area are in fact dedicated to this point, e.g., see Gelsema and Veenland (1999).

It is worth noting that Theo Pavlidis, in his King-Sun Fu lecture during ICPR 2000, recommended PR researchers to use the engineering paradigm in their work, because their final goal is the design and construction of PR machines. It is also worth noting that this engineering perspective on PR also has an epistemological explanation, because PR theories are much more aimed at solving real-world problems than explaining laws of physical world (Serra, 2000). This engineering perspective on PR should not worry PR researchers that correctly regard themselves as scientists, as engineering can be considered a science and pure science perspectives on PR are possible as well (Poser, 1998; Duin, 2001).

This brings us to our final observation. Each individual science tries to generalise from observations to theories and laws. This is how our knowledge grows from direct observable facts. Epistemology studies describe how knowledge may grow in general from a philosophical point, independent of a specific scientific field. Is it possible to state that PR develops the set of tools to do this practically, from observation to generalisation, in an automatic way by using computers? As a consequence, the core research business of PR may be related to describing the general principles underlying the engineering tools between object, observation and generalisation. Among the disciplines related with learning, PR is the one with the strongest connection with the physical world. We think thereby that in PR the emphasis of research may shift from pure generalisation issues (how to learn from a given representation) to developing suitable representations and adaptations/modifications to methods developed in areas like statistics and machine learning. New theories and methods should be developed by SPR scientists on the basis of a clear analysis of limits of existing work with respect to requirements of real applications.

Acknowledgements

The authors which to express their gratitude to the anonymous referees for careful reading the manuscript and for their numerous and worthwhile suggestions. We also thank the editors for their stimulating attitude and valuable remarks.

References

Srihari, S.N., Govindaraju, V., 1993. Pattern recognition. In: Encyclopedia of Computer Science. Chapman & Hall, London, pp. 1034–1041.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification. Wiley, New York.

Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition, second ed. Academic Press, New York.

Devijver, P.A., Kittler, J., 1982. Pattern Recognition: A Statistical Approach. Prentice-Hall, Englewood Cliffs, NJ.

Watanabe, S., 1985. Pattern Recognition: Human and Mechanical. Wiley, New York.

Tveter, D.R., 1998. The Pattern Recognition Basis of Artificial Intelligence. IEEE Computer Society Press, Silver Spring, MD.

Sohn, S.Y., 1999. Meta analysis of classification algorithms for pattern recognition. IEEE Trans. PAMI 21 (11), 1137–1144.

Queiros, C.E., Gelsema, E.S., Timmers, T., 1983. Correspondence analysis in the context of pattern recognition. Pattern Recognition Lett. 1 (4), 229–236.

Gelsema, E.S., Queiros, C.E., Timmers, T., 1982. The formalism of correspondence analysis as a means to describe object samples. In: Proc. 6th Internat. Joint Conf. on Pattern Recognition, Munich, October 19–22, pp. 564–568.

Ho, T.K., 2000. Complexity of classification problems and comparative advantages of combined classifiers. In: Kittler, J., Roli, F. (Eds.), Proc. First Internat. Workshop on Multiple Classifier Systems, MCS 2000. Lecture Notes in Computer Science, Vol. 1857. Springer, Berlin, pp. 97–106.

Serra, J., 2000. Is pattern recognition a physical science? In: Proc. ICPR 2000, 15th Internat. Conf. on Pattern Recognition, Barcelona, Spain, September 3–8, pp. 33–40.

Poser, H., 1998. On structural differences between science and engineering. Soc. Philos. Technol. 4 (2). Available from <http://scholar.lib.vt.edu/ejournals/SPT/ v4n2/>.

Duin, R.P.W., 2001. Four scientific approaches to pattern recognition. In: Vossepoel, A.M., Vos, F.M. (Eds.), Fourth Quinquennial Review 1996–2001 Dutch Society for Pattern Recognition and Image Processing, NVPHBV, Delft, pp. 331–337.

Gelsema, E.S., Veenland, J.F., 1999. Pattern Recognition Lett. 20 (11–13) (special issue on Pattern Recognition in Practice VI).