FUZZY TECHNIQUES IN PATTERN RECOGNITION

Abraham Kandel

FUZZY TECHNIQUES IN PATTERN RECOGNITION

ABRAHAM KANDEL

The Florida State University, Tallahassee

1807 1982

175 YEARS OF PUBLISHING

A Wiley-Interscience Publication

JOHN WILEY & SONS

New York Chichester Brisbane Toronto Singapore

Library of Congress Cataloging Publication Data:

Kandel, Abraham.
Fuzzy techniques in pattern recognition.

"A Wiley-Interscience publication."
Bibliography: p.
Includes index.
1. Pattern perception. 2. Fuzzy sets. I. Title.

Q327.K3 1982 001.53'4 82-8359
ISBN 0-471-09136-7 AACR2

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Preface

Fuzzy techniques represent the most recent aspect of the newly emerging field of pattern recognition, in itself an important component of the relatively new discipline of computer applications to science and engineering. My intent in writing this book is to cover in descriptive terms a good part of the problems in pattern recognition by use of fuzzy techniques. I am not concerned with the substantial body of research in fuzzy set theory and its applications, but restrict the discussion to those subjects having implications to the fundamental nature of the recognition process. During the past decade, there has been a considerable growth of interest in the theory and applications of fuzzy sets as well as in the field of pattern recognition. Most books published so far in the recognition area deal with either the decision-theoretic approach or the syntactic approach. This book treats the design of pattern recognition systems by use of the fuzzy approach. It may be used either (1) in a graduate-level course devoted specifically to fuzzy techniques in pattern recognition, a course planned with the idea that, after covering the basic material of this text, the class can profitably turn to the journal literature on the subject, or (2) as a reference research monograph on fuzzy set theory and its applications in the design of pattern recognition systems. The presentation is kept concise, and the comprehensive bibliography is designed to facilitate the use of the book both as a research compendium and as a graduate-level textbook.

The text is organized in five chapters. Chapter 1 presents a discussion of pattern recognition characterization and cluster analysis. The classical approach to pattern recognition is introduced and several pattern analysis techniques are briefly reviewed. In Chapter 2 the algebra of inexactness and the nature of fuzziness are treated. Fuzzy sets, different kinds of uncertainty, the theory of possibility, and the concept of fuzzy expectation are covered, as well as operations on fuzzy relations and estimated similarities. Various recognition techniques for pattern description that include syntactic and semantic methods, decision trees, fuzzy partitions, and fuzzy ISODATA are presented in Chapter 3. Chapter 4 is a treatment of inexact hierarchical clustering via the use of fuzzy relations, inexact matrices, similarity relations, and convex decompositions. Several important classes of imprecise operations are defined and their properties investigated. Chapter 5 begins with some specific applications of the previous concepts. These applications include phonetic and phonemic learning of speech, speech identification, and fuzzy filters. The final sections of the chapter treat the topics of clustering via unimodal fuzzy sets, fuzzy covariance matrix, and cluster validity. An updated bibliography listing more than 3000 publications in the field of fuzzy sets and its applications is included at the end of the book.

Some of the material contained in this book has been used in courses at The Florida State University. I am indebted to my students and peers for their help and encouragement throughout the development of this book, especially I want to express my lasting gratitude to Douglas H. Schlak for assistance in putting together the extensive bibliography, and to Pamela L. Flowers and Lawrence O. Hall for proofreading an early version of the manuscript. Portions of the material in this text are derived from many early investigations listed in the bibliography.

I would like to acknowledge my indebtedness to all those researchers and the many scientists who have contributed to the fast growing field of research known as "fuzzy set theory and its applications." It is my pleasure to particularly acknowledge the constant encouragement and consistent support of Lotfi A. Zadeh, King-Sun Fu, Madan M. Gupta, Ramesh Jain, Arnold Kaufmann, Ebrahim H. Mamdani, Ronald R. Yager, Hans J. Zimmermann, and William J. Byatt. I owe a debt of gratitude to all of them.

ABRAHAM KANDEL

Tallahassee, Florida
June 1982

Contents

CHAPTER ONE

Pattern Characterization

1.1 CLUSTER ANALYSIS THEORY

It is frequently stated that the process of recognition and classification is one of the most fundamental of human activities. As a matter of fact, one of the most primitive and common activities of animals (human beings included) consists of sorting like items into groups. These groups are described by patterns and what we perform is the act of recognition of certain patterns and then classification of them into groups. The word "pattern" follows the root of the word "patron" and reflects the concept of an ideal model of a set of objects or structures. This perfect pattern brings to mind the Platonic philosophy of perfect structures; the structures of the "real world" are considered to be imperfect replicas of the ideal. This deep question of the imperfection of "real world" patterns and their classification is the subject of this work.

Without pretending to offer a general theory of pattern recognition and classification, we try to suggest a new point of view about the subject of clustering, a subject representing a mental activity in which we formulate, select, modify, and adjust our frames of reference so that we can relate to a certain structure. This is a psychophysiological process that involves a relationship between a recognizer (person) and a physical stimulus. In psychology pattern recognition is defined as the process by which "external signals arriving at the sense organs are converted into meaningful perceptual experiences" (Lindsay and Norman, 1972).[†] Following our intuitive search for computerized recognition and classification, we obviously have to quantize the structural representation of patterns. For example, when we say that

[†] References in this chapter are found at the end of the chapter.

$Q(s)$ holds, the predicate Q plays the role of a quantizer on an item s. If s is a picture, for example, representing raw information, the process of recognition and classification becomes more complex (Pavlidis, 1977) and the quantization Q might involve the transformation of the scene to a numerical array, texture identification via scene segmentation and analysis, and quantitative description of shapes and objects.

The Greek word $\beta o \tau \rho \upsilon s$ means a cluster of grapes. In his excellent paper, Good (1977) proposed again the adoption of this origin of the English prefix botryo, and used the term "botryology," meaning the theory of clusters. To quote Good:

> It seems to me that the subject of clustering is now wide enough and respectable enough to deserve a name like those of other disciplines, and the existence of such a name enables one to form adjectives and so on. For example, one can use expressions such as "a botryological analysis" or "a well-known botryologist said so and so." There is another word that serves much the same purpose, namely "taxonomy," but this usually refers to biological applications whereas "botryology" is intended to refer to the entire field, provided that mathematical methods are used.

Adopting this concept, this work can be labeled as "fuzzy botryology."

1.2 NEED FOR CLUSTERING METHODS

Classification is the process of assigning an item or an observation to its proper place; the problem of cluster analysis is frequently stated as one of finding the "natural groups." In a more concrete sense, the objective is to sort a data set into categories such that the degree of "natural association" is high among members of the same category and low between members of different categories. The essence of cluster analysis might be viewed as assigning appropriate meaning to the terms "natural groups" and "natural association," where "natural" usually refers to homogeneous and "well-separated" structures.

We have used the word "theory" with relation to the above stated problem. The word "theory" has two rather different general connotations. One has the character of speculation, an hypothesis or guess; this kind of theory is the opposite of "practice" and sometimes is criticized as being "idle." The other kind of "theory" is the scientific one, an explanation (as opposed to a catalog of facts) from general principles. In mathematics there is also a technical meaning: a body of definitions, axioms, and theorems that are a systematic presentation of a subject.

As a field that encompasses many diverse techniques for discovering structure within complex bodies of data, cluster analysis occupies an interesting position with regard to theory and practice.

On the one hand, the field has practical problems, of importance not only to itself, but of economic and social interest as well. Applications of pattern recognition and classification exist in many important areas such as computer-assisted medical diagnosis and treatment, multiphasic screening and analysis, image processing and scene analysis, weather prediction, process control, neurobiological signal processing, analysis of aerial photography, earth-resource analysis, sonar detection and classification, speech and fingerprint recognition, and many, many more. So, as in a physical or social science, phenomena exist in cluster analysis that require understanding and may even be controlled through that understanding.

At the same time, the fundamental entities of the field, digital computers, are themselves entirely the product of human design, and in a sense represent an idealization of the way certain applied mathematicians performed calculations. In this aspect, the field is like a combination of engineering, mathematics, and computer science in that the subject of study is under explicit control.

The last decade has witnessed considerable interest and rapid advances in both the development and research of automatic pattern recognizers and classifiers, but many questions are still open.

One responsibility of theory in cluster analysis is to explain practical matters, and another to meet conventional mathematical criteria as systematic abstraction. These two goals are not incompatible, but neither are they close together.

Exact problems can be fully stated by means of mathematics, for example, in the form of equations. Algorithms for the solution of exact problems need only be translated into a computer language and the deed is done; the mathematician has acquired a strong and untiring helper.

The situation is different with respect to the problems that we have conditionally called *inexact*, problems for which the mathematical methods have not been defined, so that there is very little to translate into machine language. The inexact tasks include a wide variety of possibilities. Human beings handle them more or less satisfactorily, but they do it subconsciously, so to speak, not knowing how they really solve them. Cluster analysis is one of these inexact problems.

The variety of applications listed above may cause us to question whether a unified approach to pattern recognition and classification is at all possible. I believe that there is a general approach to the problem but the techniques under this general approach still vary, based on the judgment as well as the imagination of the scientist or the engineer.

In earlier studies, such as the optical character recognition device that reads the code characters on bank checks, one advantage of the technique has been the simplicity of the algorithm. In the era of networking and parallel systems, we do not measure the success of an algorithm by its simplicity. Rather we will encourage a complex scheme, as long as the implementation of the algorithm is on a parallel system with k processors and with a reasonable computational complexity (linear or low order polynomial function of n, the data size). The complexity of computations is especially important, as in most computer-implemented algorithms that are applied to complex and large data sets.

Even though little or nothing about the category structure can be stated in advance, the analyst usually is informed sufficiently about the problem that he or she can distinguish between "good" and "bad" category structures when confronted with them. Hence, Anderberg (1973) asks the question: Why not enumerate all the possibilities and simply choose the most appealing?

The number of ways of sorting n observations into m groups is a Stirling number of the second kind:†

$$S_n^{(m)} = \frac{1}{m!} \sum_{j=0}^{m} (-1)^{m-j} \binom{m}{j} j^n$$

The problem is compounded by the fact that the number of groups is usually unknown, so that the number of possibilities is a sum of Stirling numbers. In the case of 25 observations it is

$$\sum_{k=1}^{25} S_{25}^{(k)} > 4 \times 10^{18}.$$

Hence it is generally the intent of cluster analysis techniques to emulate human efficiency and find an acceptable solution while considering only a small number of the alternatives.

1.3 PATTERN ANALYSIS

The common problem in data analysis is the lack of homogeneity among attributes, which leads to the inexact meaning of the word "cluster." Most cluster analysis methods require a measure of similarity to be defined for every pairwise combination of the entities to be clustered. When clustering

† For even the relatively tiny problem of sorting 25 observations into 5 categories, the number of possibilities is the astounding quantity $S_{25}^{(5)} = 2{,}436{,}684{,}974{,}110{,}751$.

data units, the proximity of individuals is usually expressed as a distance. The clustering of variables generally involves a correlation or other such measure of association, many of which are discussed in the literature. Some have operational interpretations while others are rather difficult to describe. The measures interact with cluster analysis criteria so that some measures give identical results with one criterion and distinctly different results with another. The combined choices of variables, transformations, and similarity measures give operational meaning to the term "natural association." The choices are often made separately and should be reviewed for their composite effect to make sure the result is satisfactory.

The term "cluster" is often left undefined and taken as a primitive notion in much the same manner as "point" is treated in geometry. Such treatment is fine for theoretical discussions, but when it comes to finding clusters in real data, the term must bear a definite meaning. The choice of a clustering criterion is tantamount to defining a cluster. It may not be possible to say just what a cluster is in abstract terms, but it can always be defined constructively through statement of the criterion and an implementing algorithm.

Many criteria for clustering have been proposed and used. In some problems there is a natural choice, while in others almost any criterion might have status as a candidate. It should not be necessary to choose only a single criterion because clustering the data set several times with different criteria is a good way to reveal various facets of the structure. On the other hand, the expense of using cluster analysis prohibits trying out everything that is available.

In spite of the lack of a complete theory of recognition and classification, extensive study of these problems has led to some excellent treatments of the subject in the nonfuzzy way. Many of the books and papers in the brief reference list at the end of this chapter provide a systematically treated overview of classical (nonfuzzy) pattern recognition and classification. The purpose of this work is to give a systematic discussion of some important principles of pattern recognition and classification that involve fuzzy set theory.

The pattern space is essentially the domain that is defined by the data observed by a sensory device. A way of characterizing an object q is to assign to it the values of a finite set of parameters considered relevant to the object. A column vector $\mathbf{x}$ in the pattern space will have scalar elements

$$\mathbf{x} = (x_1, \ldots, x_n)^t \mid t = \text{transpose},$$

where each x_i, $1 \leq i \leq n$, represents the particular value associated with the ith dimension, or the value associated with feature i of object q. Namely, object q is represented by pattern $\mathbf{x}$ where

$$\mathbf{x} = f(q) = (f_1(q),\ldots,f_n(q)) = (x_1,\ldots,x_n)'$$

and f_i, $1 \leq i \leq n$, is the measurement procedure associated with feature i. $f_i(q)$ is obviously the feature value. The first problem deals with feature extraction; this is the selection of a small set of measurement procedures $\{f_i\}_{i=1}^n$ and/or a set of primitives into which q is decomposed in order to formulate the mathematical description x. The other problem is that of cluster analysis (also referred to as classification theory categorization), which is concerned with the problem of partitioning a given set of entities into homogeneous and well-separated subsets.

The concepts of homogeneity and of separation can be made precise when a precise measure of dissimilarity between the entities is given. In his extensive "Review of Classification," Cormak (1971) notes that, "A classification, as usually understood, allocates entities to initially undefined classes so that individuals are in some sense close to one another." In *Cluster Analysis*, Duran and Odell (1974) state that the determination of clusters should be such that, "Those individuals which are assigned to the same cluster are *similar*, yet individuals from different clusters are different (*not similar*)."

Cluster analysis is generally divided into nondeterministic pattern classification and deterministic pattern classification. The first approach usually fails when it is impossible to represent the highly joint and scattered data set by some known distribution functions. The deterministic approach cannot be implemented in a satisfactory manner whenever the data set consists of highly joint and scattered patterns. Both approaches tend to separate the concept of discrimination from the notion of clustering.

Discrimination techniques usually begin with either some data base separated into *a priori* known categories or *a priori* conceptual distinctions, and proceed to develop algorithms by which to separate incoming data into those *a priori* categories, whereas clustering algorithms use *a priori* selection of a *measure of similarity* in order to find an inherent structure in the data. The term "homogeneous" is applied in the sense that all elements in some category are similar (according to some predefined measure of similarity) to each other and dissimilar to elements that belong to other categories.

It is very interesting to visualize the different aims and goals that various users of recognition and classification techniques have used. Once it is realized that these goals are so different, it is easier to understand why we have such a variety of classification techniques and also to avoid somewhat fruitless arguments as to which techniques are the *most* useful in some global sense. The goals might vary from data exploration to data reduction, model fitting, hypothesis testing, data investigation to find useful hypotheses or a true typology, prediction based on categories of data, and many more. It must be clear to the users of clustering algorithms how and why they should consider quite consciously the interaction between technique and discipline context. The techniques, based on certain mathematical models, influence the selection and preprocessing of the data. The explicit definition of what the users mean intuitively by similarity between elements becomes embodied in their algorithms, which in essence lead to an interpretation of the structure of the data. In short, implicit and explicit judgments regarding data variables, scaling, selective criteria, and selection of elements will affect the outcome of a clustering method.

It should be clear that *explicitness is subjective*; just because Dr. Smith chose one method of classification does not mean that he has found Truth. He exemplified his reasoning why he has grouped elements into a given category, and therein lies the value of explicitness, but metaphysical questions regarding the *nature of reality* cannot be settled by making *a priori* methodological choices before analyzing a given data set (Anderberg, 1973). The problem of selecting an "appropriate" method for a particular data set will remain problematic even when results are obtained through clustering analysis, and it is our feeling that they will not be resolved through methodological decisions made out of context.

It is necessary for the users of classification techniques to regard several important questions that should provide them with a guide for developing a clustering method:

1 What are the goals of the classification?

2 How are my goals affected by my particular research problems and specific paradigms?

3 What are the basic assumptions as to the format and underlying structure of the data set?

4 Is there available a training sample or a relatively undifferentiated set of elements whose underlying structure has to be explored by classification on the basis of some similarity relation between the elements?

5 Feature vector analysis:

- **(a)** How can I best determine the elements' variables?
- **(b)** How shall I scale and count the variables?
- **(c)** Can individual variables or elements be combined into a single measure of similarity?

6 Can classes of classification techniques be identified? What are the monetary advantages of each class in connection with the resources of the user and the specific answers to questions 1–5.

7 How can the results be evaluated?

The resolution of these questions will enable the users to understand the relations between their classification problem and the relevant technique of clustering.

1.4 SIMILARITY MEASURES AND CLUSTERING

Let us now discuss the general ideas behind the question: How shall we devise some manner of grouping?

Figure 1.1, with its emphasis on the imprecision of the environment, shows a general route from the data collection to its processing. A more detailed diagram is presented in Figure 1.2.

The scientist selects an inexact environment and a set of tools to measure the data set. This data set is then preprocessed, normalized, and inverted into an associative set of descriptive numbers. Then the data is clustered or discriminated and modifications are fed back to improve data collection by selecting slightly different environments, improving instruments, or modifying the preprocessing by removing variations irrelevant to the scientist's clustering interest.

Throughout the clustering process, it is apparent that similarity measures can be constructed not only between pairs of elements, but also from an individual object to an entire category. Three primary forms of category description can be discussed:

1 Extracting descriptive states (average, etc.) by weighting the element in several ways. The most common one is the probability of an element belonging to that category.
2 Identification numbers of each element in a given category (used largely in single-linkage and graph-theoretic methods).
3 Functional forms (normal distribution with a standardized mean and covariance matrix).

The main problem in all these examples is boundary specification. Boundaries between categories can be specified implicitly by utilizing the

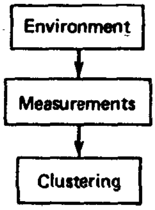

Figure 1.1 Environment and data processing.

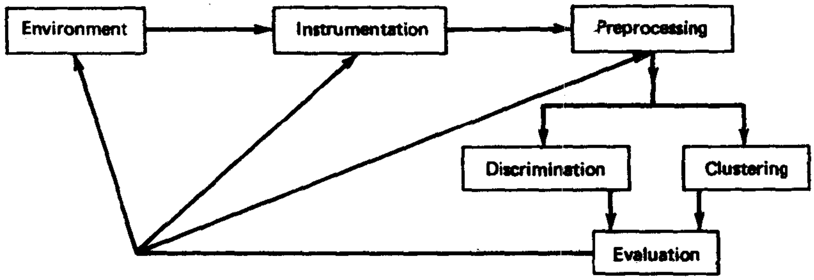

Figure 1.2 Context of clustering.

concept of "category position," or explicitly by parameterizing the boundaries themselves. Once we have resolved this problem, we can select the procedure for seeking optional partitions. However, these should be implemented with regard to the type of boundaries around the classification centers. These could be either soft boundaries (Figure 1.3*a*) or hard boundaries (Figure 1.3*b*).

If a description has hard boundaries, something either is or is not similar to a given concept. In soft boundaries we see a piece of information as varying in the degree of relevance to various concepts that are already internalized.

Even though a complete guiding philosophy of cluster analysis utilization is not available, some individual principles can be stated following Anderberg (1973):

1 Any given data set may admit many different but meaningful classifications. Each classification may pertain to a different aspect of the data. It is unnecessarily narrowing to seek a single "right" classification. Several different clustering procedures will be needed to discover multiple classifications. New insight and understanding might result from alternative classifications suggested through a cluster analysis and totally unexpected aspects of structure might be revealed in the process.

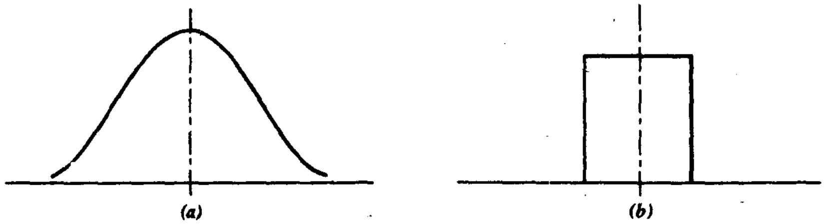

Figure 1.3 Classification boundaries. (*a*) Soft boundaries and (*b*) hard boundaries.

2 Cluster analysis is a device for suggesting hypotheses. The classification of data units or variables obtained from a cluster analysis procedure has no inherent validity. The analyst should not feel any pressure to embrace a particular classification, nor should he or she feel bound to the details of a classification that seems generally interesting. The worth of a particular classification and its underlying explanatory structure is to be justified by its consistency with known facts and without regard to the manner of its original generation.

3 A set of clusters is not itself a final result but only a possible outline. It follows then that there is little justification for using excessively detailed and expensive algorithms when results of the same general character can be achieved with low cost and intuitive procedures.

4 Cluster analysis methods involve a mixture of imposing a structure on the data and revealing that structure that actually exists in the data. The notion of finding "natural groups" tends to imply that the algorithm should passively conform like a wet tee shirt. Unfortunately, practical procedures involve fixed sequences of operations, which systematically ignore some aspects of structure while intensively dwelling on others. Such properties are often quite inadvertent and therefore are discovered only by observing their effects on data. To a considerable extent, a set of clusters reflects the degree to which the data set conforms to the structural forms embedded in the clustering algorithm.

5 In view of the preceding four points, it should come as no surprise that the results of a cluster analysis method rarely suggest a satisfactory structure for the total set of data. More commonly, one or more interesting clusters lead to inferences about part of the data. Along this line, certain clusters may be so natural and self-evident (once discovered) as to constitute "features" of the data likely to be revealed by almost any method. Rather than continue to recover the obvious, it is economical and relatively riskless to remove such features from the data set as they are found and concentrate attention on the more confused residue. Of course, adding or deleting variables alters the measurement space, and consequently such features may lose their distinctive character or alter their composition in response to such a change.

6 In the quest for clusters, two possibilities are often overlooked.

- (a) The data may contain no clusters. When clustering variables, nearly complete independence or orthogonality would lead to such a result. When clustering data units, an absence of discriminating variables and a more or less uniform distribution of points in the measurement space would lead to a distinct lack of cohesion.