14

FUZZY SETS AND FUZZY LOGIC

Chapter Objectives

- Explain the concept of fuzzy sets with formal interpretation in continuous and discrete domains.
- Analyze the characteristics of fuzzy sets and fuzzy-set operations.
- Describe the extension principle as a basic mechanism for fuzzy inferences.
- Discuss the importance of linguistic imprecision and computing with them in decision-making processes.
- Construct methods for multifactorial evaluation and extraction of a fuzzy rule-based model from large, numeric data sets.
- Understand why fuzzy computing and fuzzy systems are an important part of data-mining technology.

In the previous chapters, a number of different methodologies for the analysis of large data sets have been discussed. Most of the approaches presented, however, assume that the data are precise, that is, they assume that we deal with exact measurements for further analysis. Historically, as reflected in classical mathematics, we commonly seek

Data Mining: Concepts, Models, Methods, and Algorithms, Second Edition. Mehmed Kantardzic.

a precise and crisp description of things or events. This precision is accomplished by expressing phenomena in numeric or categorical values. But in most, if not all, real-world scenarios, we will never have totally precise values. There is always going to be a degree of uncertainty. However, classical mathematics can encounter substantial difficulties because of this fuzziness. In many real-world situations, we may say that fuzziness is reality, whereas crispness or precision is simplification and idealization. The polarity between fuzziness and precision is quite a striking contradiction in the development of modern information-processing systems. One effective means of resolving the contradiction is the fuzzy-set theory, a bridge between high precision and the high complexity of fuzziness.

14.1 FUZZY SETS

Fuzzy concepts derive from fuzzy phenomena that commonly occur in the real world. For example, rain is a common natural phenomenon that is difficult to describe precisely since it can rain with varying intensity, anywhere from a light shower to a torrential downpour. Since the word rain does not adequately or precisely describe the wide variations in the amount and intensity of any rain event, "rain" is considered a fuzzy phenomenon.

Often, the concepts formed in the human brain for perceiving, recognizing, and categorizing natural phenomena are also fuzzy. The boundaries of these concepts are vague. Therefore, the judging and reasoning that emerges from them are also fuzzy. For instance, "rain" might be classified as "light rain," "moderate rain," and "heavy rain" in order to describe the degree of raining. Unfortunately, it is difficult to say when the rain is light, moderate, or heavy, because the boundaries are undefined. The concepts of "light," "moderate," and "heavy" are prime examples of fuzzy concepts themselves. To explain the principles of fuzzy sets, we will start with the basics in classical-set theory.

The notion of a set occurs frequently as we tend to organize, summarize, and generalize knowledge about objects. We can even speculate that the fundamental nature of any human being is to organize, arrange, and systematically classify information about the diversity of any environment. The encapsulation of objects into a collection whose members all share some general features naturally implies the notion of a set. Sets are used often and almost unconsciously; we talk about a set of even numbers, positive temperatures, personal computers, fruits, and the like. For example, a classical set A of real numbers greater than 6 is a set with a crisp boundary, and it can be expressed as

$$A = \{x \mid x > 6\}$$

where there is a clear, unambiguous boundary 6 such that if x is greater than this number, then x belongs to the set A; otherwise, x does not belong to the set. Although classical sets have suitable applications and have proven to be an important tool for mathematics and computer science, they do not reflect the nature of human concepts and thoughts, which tend to be abstract and imprecise. As an illustration, mathematically we can express a set of tall persons as a collection of persons whose height is more than 6 ft; this is the set denoted by the previous equation, if we let A = "tall person" and x = "height." Yet, this is an unnatural and inadequate way of representing our usual concept of "tall person." The dichotomous nature of the classical set would classify a person 6.001 ft tall as a tall person, but not a person 5.999 ft tall. This distinction is intuitively unreasonable. The flaw comes from the sharp transition between inclusions and exclusions in a set.

In contrast to a classical set, a fuzzy set, as the name implies, is a set without a crisp boundary, that is, the transition from "belongs to a set" to "does not belong to a set" is gradual, and this smooth transition is characterized by membership functions (MFs) that give sets flexibility in modeling commonly used linguistic expressions such as "the water is hot" or "the temperature is high." Let us introduce some basic definitions and their formalizations concerning fuzzy sets.

Let X be a space of objects and x be a generic element of X. A classical set A, $A \subseteq X$, is defined as a collection of elements or objects $x \in X$ such that each x can either belong or not belong to set A. By defining a *characteristic function* for each element x in X, we can represent a classical set A by a set of ordered pairs (x, 0) or (x, 1), which indicates $x \notin A$ or $x \in A$, respectively.

Unlike the aforementioned conventional set, a fuzzy set expresses the degree to which an element belongs to a set. The characteristic function of a fuzzy set is allowed to have values between 0 and 1, which denotes the degree of membership of an element in a given set. If X is a collection of objects denoted generically by x, then a fuzzy set A in X is defined as a set of ordered pairs:

$$A = \{(x, \mu_A[x]) \mid x \in X\}$$

where $\mu_A(x)$ is called the membership function (MF) for the fuzzy set A. The MF maps each element of X to a membership grade (or membership value) between 0 and 1.

Obviously, the definition of a fuzzy set is a simple extension of the definition of a classical set in which the characteristic function is permitted to have any value between 0 and 1. If the value of the MF $\mu_A(x)$ is restricted to either 0 or 1, then A is reduced to a classic set and $\mu_A(x)$ is the characteristic function of A. For clarity, we shall also refer to classical sets as ordinary sets, crisp sets, non-fuzzy sets, or, simply, sets.

Usually X is referred to as the universe of discourse, or, simply, the universe, and it may consist of discrete (ordered or non-ordered) objects or continuous space. This can be clarified by the following examples. Let X = {San Francisco, Boston, Los Angeles} be the set of cities one may choose to live in. The fuzzy set C = "desirable city to live in" may be described as follows:

$$C = \{(\text{San Francisco}, 0.9), (\text{Boston}, 0.8), (\text{Los Angeles}, 0.6)\}.$$

The universe of discourse X is discrete and it contains non-ordered objects: three big cities in the United States. As one can see, the membership grades listed above are quite subjective; anyone can come up with three different but legitimate values to reflect his or her preference.

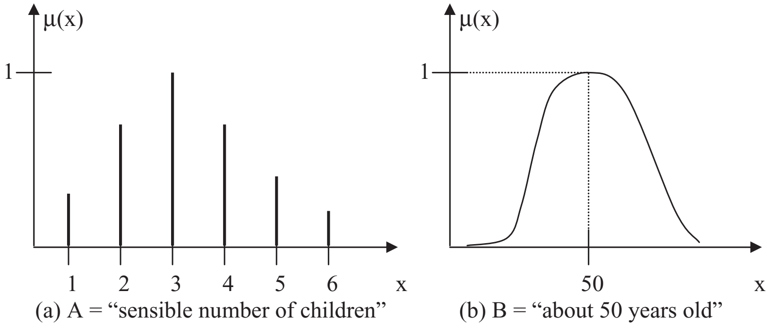

Figure 14.1. Discrete and continuous representation of membership functions for given fuzzy sets. (a) A = "sensible number of children"; (b) B = about 50 years old.

In the next example, let X = {0, 1, 2, 3, 4, 5, 6} be a set of the number of children a family may choose to have. The fuzzy set A = "sensible number of children in a family" may be described as follows:

$$A = \{(0, 0.1), (1, 0.3), (2, 0.7), (3, 1), (4, 0.7), (5, 0.3), (6, 0.1)\}$$

Or, in the notation that we will use through this chapter,

$$A = 0.1/0 + 0.3/1 + 0.7/2 + 1.0/3 + 0.7/4 + 0.3/5 + 0.1/6$$

Here we have a discrete-order universe X; the MF for the fuzzy set A is shown in Figure 14.1a. Again, the membership grades of this fuzzy set are obviously subjective measures.

Finally, let $X = R^+$ be the set of possible ages for human beings. Then the fuzzy set B = "about 50 years old" may be expressed as

$$B = \{(x, \mu_B[x]) \mid x \in X\}$$

where

$$\mu_B(x) = 1/(1 + ((x - 50)/10)^4)$$

This is illustrated in Figure 14.1b.

As mentioned earlier, a fuzzy set is completely characterized by its MF. Since many fuzzy sets in use have a universe of discourse X consisting of the real line R, it would be impractical to list all the pairs defining an MF. A more convenient and concise way to define an MF is to express it as a mathematical formula. Several classes of parametrized MFs are introduced, and in real-world applications of fuzzy sets the shape of MFs is usually restricted to a certain class of functions that can be specified with only

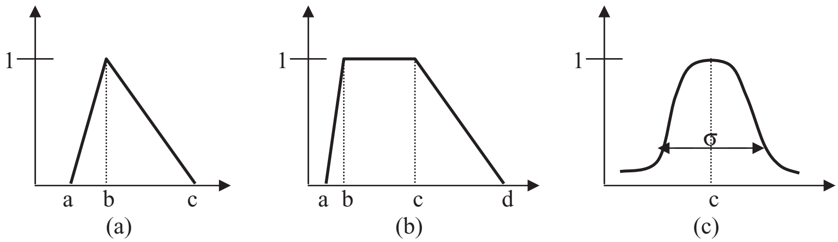

Figure 14.2. Most commonly used shapes for membership functions. (a) Triangular; (b) trapezoidal; (c) Gaussian.

few parameters. The most well known are triangular, trapezoidal, and Gaussian; Figure 14.2 shows these commonly used shapes for MFs.

A triangular MF is specified by three parameters {a, b, c} as follows:

$$\mu(x) = \text{triangle }(x, a, b, c) = \begin{cases} 0 & \text{for} \quad x \le a \\ (x-a)/(b-a) & \text{for} \quad a \le x \le b \\ (c-x)/(c-b) & \text{for} \quad b \le x \le c \\ 0 & \text{for} \quad c \le x \end{cases}$$

The parameters {a, b, c}, with a < b < c, determine the x coordinates of the three corners of the underlying triangular MF.

A trapezoidal MF is specified by four parameters {a, b, c, d} as follows:

$$\mu(x) = \text{trapezoid }(x, a, b, c, d) = \begin{cases} 0 & \text{for} \quad x \le a \\ (x-a)/(b-a) & \text{for} \quad a \le x \le b \\ 1 & \text{for} \quad b \le x \le c \\ (d-x)/(d-c) & \text{for} \quad c \le x \le d \\ 0 & \text{for} \quad d \le x \end{cases}$$

The parameters {a, b, c, d}, with a < b ≤ c < d, determine the x coordinates of the four corners of the underlying trapezoidal MF. A triangular MF can be seen as a special case of the trapezoidal form where b = c.

Finally, a Gaussian MF is specified by two parameters {c, σ}:

$$\mu(x) = \text{gaussian }(x, c, \sigma) = e^{-1/2\,((x-c)/\sigma)2}$$

A Gaussian MF is determined completely by c and σ; c represents the membership-function center, and σ determines the membership-function width. Figure 14.3 illustrates the three classes of parametrized MFs.

From the preceding examples, it is obvious that the construction of a fuzzy set depends on two things: the identification of a suitable universe of discourse and the

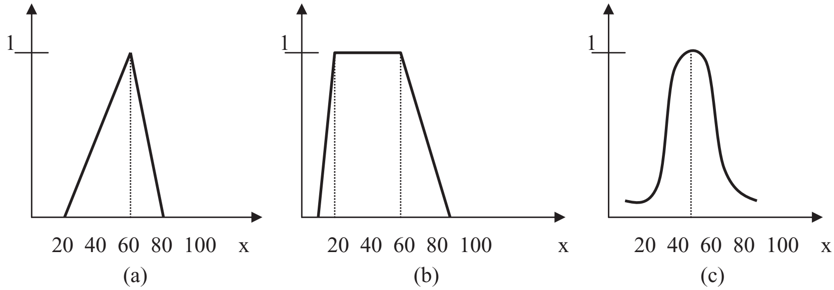

Figure 14.3. Examples of parametrized membership functions. (a) Triangle(x, 20, 60, 80); (b) trapezoid(x, 10, 20, 60, 90); (c) Gaussian(x, 50, 20).

specification of an appropriate MF. The specification of an MF is subjective, which means that the MFs for the same concept (e.g., "sensible number of children in a family") when specified by different persons may vary considerably. This subjectivity comes from individual differences in perceiving or expressing abstract concepts and has little to do with randomness. Therefore, the *subjectivity* and *nonrandomness* of fuzzy sets is the primary difference between the study of fuzzy sets and the probability theory, which deals with the objective treatment of random phenomena.

There are several parameters and characteristics of MF that are used very often in some fuzzy-set operations and fuzzy-set inference systems. We will define only some of them that are, in our opinion, the most important:

1. *Support*. The *support* of a fuzzy set A is the set of all points x in the universe of discourse X such that $\mu_A(x) > 0$:

$$\textit{Support}\ (A) = \{x\mu_A(x) > 0\}$$

2. *Core*. The *core* of a fuzzy set A is the set of all points x in X such that $\mu_A(x) = 1$:

$$\textit{Core}\ (A) = \{x \mid A(x) = 1\}$$

3. *Normalization*. A fuzzy set A is *normal* if its core is nonempty. In other words, we can always find a point $x \in X$ such that $\mu_A(x) = 1$.
4. *Cardinality*. Given a fuzzy set A in a finite universe X, its cardinality, denoted by Card(A), is defined as

$$\textit{Card}(A) = \Sigma\mu_A(x), \text{ where } x \in X$$

Often, *Card*(X) is referred to as the scalar cardinality or the count of A. For example, the fuzzy set A = 0.1/1 + 0.3/2 + 0.6/3 + 1.0/4 + 0.4/5 in universe X = {1, 2, 3, 4, 5, 6} has a cardinality *Card*(A) = 2.4.
5. *α-cut*. The *α-cut* or *α-level set* of a fuzzy set A is a crisp set defined by

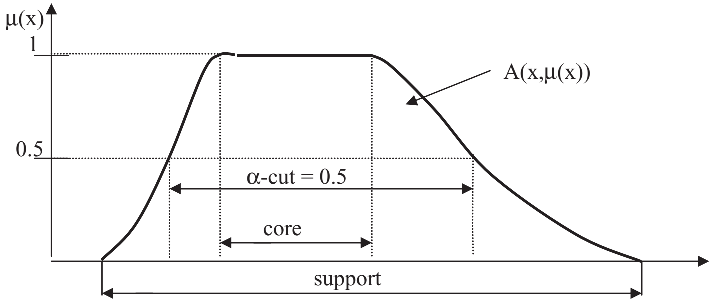

Figure 14.4. Core, support, and α-cut for fuzzy set A.

$$A_\alpha = \{x \mid \mu_A(x) \geq \alpha\}$$

6. *Fuzzy number*. Fuzzy numbers are a special type of fuzzy sets restricting the possible types of MFs:
   (a) The MF must be normalized (i.e., the core is nonempty) and singular. This results in precisely one point, which lies inside the core, modeling the typical value of the fuzzy number. This point is called the modal value.
   (b) The MF has to monotonically increase the left of the core and monotonically decrease on the right. This ensures that only one peak and, therefore, only one typical value exists. The spread of the support (i.e., the nonzero area of the fuzzy set) describes the degree of imprecision expressed by the fuzzy number.

A graphical illustration of some of these basic concepts is given in Figure 14.4.

14.2 FUZZY-SET OPERATIONS

Union, intersection, and complement are the most basic operations in classic sets. Corresponding to the ordinary set operations, fuzzy sets too have operations, which were initially defined by Zadeh, the founder of the fuzzy-set theory.

The *union* of two fuzzy sets A and B is a fuzzy set C, written as $C = A \cup B$ or $C = A$ *OR* B, whose MF $\mu_C(x)$ is related to those of A and B by

$$\mu_C(x) = \max(\mu_A[x], \mu_B[x]) = \mu_A(x) \vee \mu_B(x), \quad \forall x \in X$$

As pointed out by Zadeh, a more intuitive but equivalent definition of the union of two fuzzy sets A and B is the "smallest" fuzzy set containing both A and B. Alternatively, if D is any fuzzy set that contains both A and B, then it also contains $A \cup B$.

The *intersection* of fuzzy sets can be defined analogously. The intersection of two fuzzy sets A and B is a fuzzy set C, written as $C = A \cap B$ or $C = A$ *AND* B, whose MF is related to those of A and B by

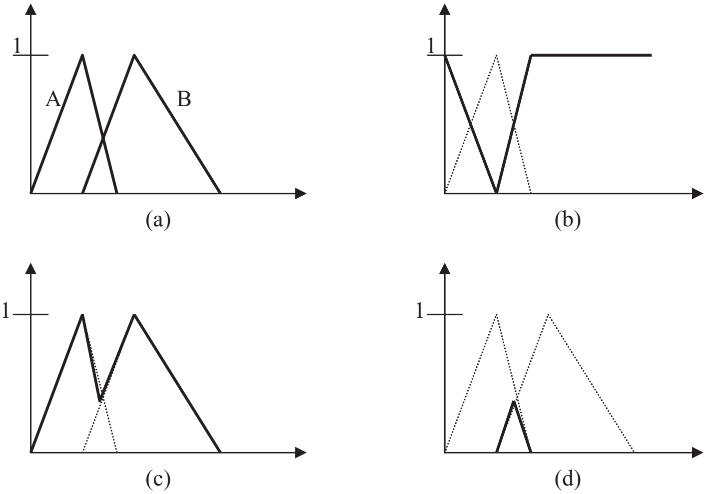

Figure 14.5. Basic operations on fuzzy sets. (a) Fuzzy sets A and B; (b) C = A′; (c) C = A ∪ B; (d) C = A ∩ B.

$$\mu_C(x) = \min(\mu_A[x], \mu_B[x]) = \mu_A(x) \wedge \mu_B(x), \quad \forall x \in X$$

As in the case of the union of sets, it is obvious that the intersection of A and B is the "largest" fuzzy set that is contained in both A and B. This reduces to the ordinary-intersection operation if both A and B are non-fuzzy.

The *complement* of a fuzzy set A, denoted by A′, is defined by the MF as

$$\mu_{A'}(x) = 1 - \mu_A(x), \quad \forall x \in X$$

Figure 14.5 demonstrates these three basic operations: Figure 14.5a illustrates two fuzzy sets A and B; Figure 14.5b is the complement of A; Figure 14.5c is the union of A and B; and Figure 14.5d is the intersection of A and B.

Let A and B be fuzzy sets in X and Y domains, respectively. The Cartesian product of A and B, denoted by A × B, is a fuzzy set in the product space X × Y with an MF

$$\mu_{A\times B}(x, y) = \min(\mu_A[x], \mu_B[y]) = \mu_A(x) \wedge \mu_B(y), \quad \forall x \in X \text{ and } \forall y \in Y$$

Numeric computations based on these simple fuzzy operations are illustrated through one simple example with a discrete universe of discourse S. Let S = {1, 2, 3, 4, 5} and assume that fuzzy sets A and B are given by:

$$A = 0/1 + 0.5/2 + 0.8/3 + 1.0/4 + 0.2/5$$

$$B = 0.9/1 + 0.4/2 + 0.3/3 + 0.1/4 + 0/5$$

Then,

$$A \cup B = 0.9/1 + 0.5/2 + 0.8/3 + 1.0/4 + 0.2/5$$

$$A \cap B = 0/1 + 0.4/2 + 0.3/3 + 0.1/4 + 0/5$$

$$A^C = 1/1 + 0.5/2 + 0.2/3 + 0/4 + 0.8/5$$

and the Cartesian product of fuzzy sets A and B is

$$\begin{aligned} A \times B = {} & 0/(1,1) + 0/(1,2) + 0/(1,3) + 0/(1,4) + 0/(1,5) \\ & + 0.5/(2,1) + 0.4/(2,2) + 0.3/(2,3) + 0.1/(2,4) + 0/(2,5) \\ & + 0.8/(3,1) + 0.4/(3,2) + 0.3/(3,3) + 0.1/(3,4) + 0/(3,5) \\ & + 0.9/(4,1) + 0.4/(4,2) + 0.3/(4,3) + 0.1/(4,4) + 0/(4,5) \\ & + 0.2/(5,1) + 0.2/(5,2) + 0.2/(5,3) + 0.1/(5,4) + 0/(5,5) \end{aligned}$$

Fuzzy sets, as defined by MF, can be compared in different ways. Although the primary intention of comparing is to express the extent to which two fuzzy numbers match, it is almost impossible to come up with a single method. Instead, we can enumerate several classes of methods available today for satisfying this objective. One class, distance measures, considers a distance function between MFs of fuzzy sets A and B and treats it as an indicator of their closeness. Comparing fuzzy sets via distance measures does not place the matching procedure in the set-theory perspective. In general, the distance between A and B, defined in the same universe of discourse X, where $X \in R$, can be defined using the Minkowski distance:

$$D(A, B) = \{\Sigma |A(x) - B(x)|^p\}^{1/p},\ x \in X$$

where $p \geq 1$. Several specific cases are typically encountered in applications:

1. Hamming distance for $p = 1$,
2. Euclidean distance for $p = 2$, and
3. Tchebyshev distance for $p = \infty$.

For example, the distance between given fuzzy sets A and B, based on Euclidean measure, is

$$D(A, B) = \sqrt{(0-0.9)^2 + (0.5-0.4)^2 + (0.8-0.3)^2 + (1-0.1)^2 + (0.2-0)^2} = 1.39$$

For continuous universes of discourse, summation is replaced by integration. The more similar the two fuzzy sets, the lower the distance function between them. Sometimes, it is more convenient to normalize the distance function and denote it $d_n(A,B)$, and use this version to express similarity as a straight complement, $1 - d_n(A, B)$.

The other approach to comparing fuzzy sets is the use of possibility and necessity measures. The possibility measure of fuzzy set A with respect to fuzzy set B, denoted by *Pos*(A, B), is defined as

$$Pos(\mathrm{A}, \mathrm{B}) = \max[\min(\mathrm{A}(x), \mathrm{B}(x))],\ x \in \mathrm{X}$$

The necessity measure of A with respect to B, *Nec*(A, B) is defined as

$$Nec(\mathrm{A}, \mathrm{B}) = \min[\max(\mathrm{A}(x), 1 - \mathrm{B}(x))],\ x \in \mathrm{X}$$

For the given fuzzy sets A and B, these alternative measures for fuzzy-set comparison are

$$Pos(\mathrm{A}, \mathrm{B}) = \max[\min\{(0, 0.5, 0.8, 1.0, 0.2), (0.9, 0.4, 0.3, 0.1, 0)\}] = \max[0, 0.4, 0.3, 0.1, 0] = 0.4$$

$$Nec(\mathrm{A}, \mathrm{B}) = \min[\max\{(0, 0.5, 0.8, 1.0, 0.2), (0.1, 0.6, 0.7, 0.9, 1.0)\}] = \min[0.1, 0.6, 0.8, 1.0, 1.0] = 0.1$$

An interesting interpretation arises from these measures. The possibility measure quantifies the extent to which A and B overlap. By virtue of the definition introduced, the measure is symmetric. On the other hand, the necessity measure describes the degree to which B is included in A. As seen from the definition, the measure is asymmetrical. A visualization of these two measures is given in Figure 14.6.

A number of simple yet useful operations may be performed on fuzzy sets. These are one-argument mappings, because they apply to a single MF.

1. *Normalization*: This operation converts a subnormal, nonempty fuzzy set into a normalized version by dividing the original MF by the height of A

$$NormA(x) = \{(x, \mu_A[x]/\mathrm{hgt}[x] = \mu_A[x]/\max \mu_A[x]), \text{ where } x \in \mathrm{X}\}$$

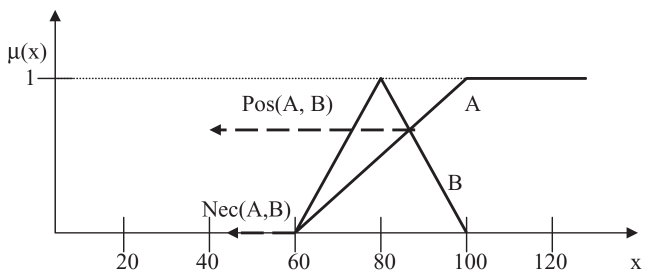

Figure 14.6. Comparison of fuzzy sets representing linguistic terms A = high speed and B = speed around 80 km/h.

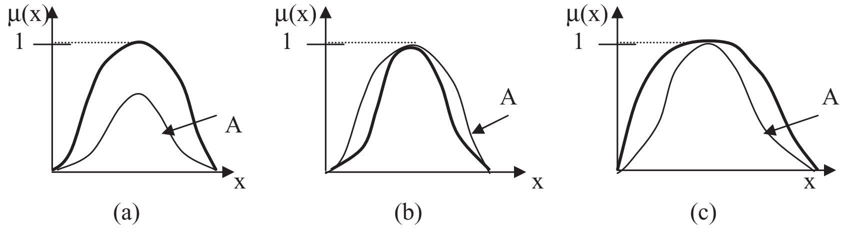

Figure 14.7. Simple unary fuzzy operations. (a) Normalization; (b) concentration; (c) dilation.

2. *Concentration*: When fuzzy sets are concentrated, their MFs take on relatively smaller values. That is, the MF becomes more concentrated around points with higher membership grades as, for instance, being raised to power two:

$$ConA(x) = \{(x, \mu_A^2[x]), \text{ where } x \in X\}$$

3. *Dilation*: Dilation has the opposite effect from concentration and is produced by modifying the MF through exponential transformation, where the exponent is less than 1

$$DilA(x) = \{(x, \mu_A^{1/2}[x]) \text{ where } x \in X\}$$

The basic effects of the previous three operations are illustrated in Figure 14.7.

In practice, when the universe of discourse X is a continuous space (the real axis R or its subset), we usually partition X into several fuzzy sets whose MFs cover X in a more-or-less uniform manner. These fuzzy sets, which usually carry names that conform to adjectives appearing in our daily linguistic usage, such as "large," "medium," or "small," are called *linguistic values* or linguistic labels. Thus, the universe of discourse X is often called the linguistic variable. Let us give some simple examples.

Suppose that X = "age." Then we can define fuzzy sets "young," "middle aged," and "old" that are characterized by MFs $\mu_{young}(x)$, $\mu_{middleaged}(x)$, and $\mu_{old}(x)$, respectively. Just as a variable can assume various values, a linguistic variable "age" can assume different linguistic values, such as "young," "middle aged," and "old" in this case. If "age" assumes the value of "young," then we have the expression "age is young," and so also for the other values. Typical MFs for these linguistic values are displayed in Figure 14.8, where the universe of discourse X is totally covered by the MFs and their smooth and gradual transition from one to another. Unary fuzzy operations, concentration and dilation, may be interpreted as linguistic modifiers "very" and "more or less," respectively.

A linguistic variable is characterized by a quintuple (x, T(x), X, G, M) in which x is the name of the variable; T(x) is the term set of x—the set of its linguistic values; X is the universe of discourse; G is a syntactic rule that generates the terms in T(x); and M is a semantic rule that associates with each linguistic value A its meaning M(A),

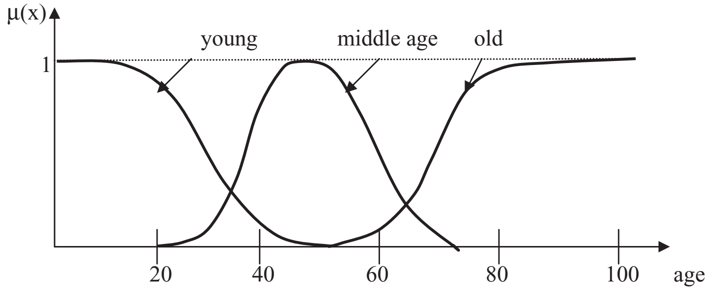

Figure 14.8. Typical membership functions for linguistic values "young," "middle aged," and "old."

where M(A) denotes an MF for a fuzzy set in X. For example, if age is interpreted as a linguistic variable, then the term set T(age) could be

$$\begin{aligned} T(\text{age}) = \{&\text{very young, young, not very young, not young, \ldots, middle aged, not} \\ &\text{middle aged, not old, more-or-less old, old, very old}\} \end{aligned}$$

where each term in T(age) is characterized by a fuzzy set of a universe of discourse X = [0, 100]. The syntactic rule refers to the way the linguistic values in the term set T(age) are generated, and the semantic rule defines the MF of each linguistic value of the term set T(age), such as the linguistic values in Figure 14.8.

14.3 EXTENSION PRINCIPLE AND FUZZY RELATIONS

As in the set theory, we can define several generic relations between two fuzzy sets, such as *equality* and *inclusion*. We say that two fuzzy sets, A and B, defined in the same space X are equal if and only if (iff) their MFs are identical:

$$A = B \quad \text{iff} \quad \mu_A(x) = \mu_B(x), \forall x \in X$$

Analogously, we shall define the notion of *containment*, which plays a central role in both ordinary and fuzzy sets. This definition of containment is, of course, a natural extension of the case for ordinary sets. Fuzzy set A is *contained* in fuzzy set B (or, equivalently, A is a subset of B) if and only if $\mu_A(x) \le \mu_B(x)$ for all x. In symbols,

$$A \subseteq B \Leftrightarrow \mu_A(x) \le \mu_B(x), \forall x \in X$$

Figure 14.9 illustrates the concept of $A \subseteq B$.

When the fuzzy sets A and B are defined in a finite universe X, and the requirement that for each x in X, $\mu_A(x) \le \mu_B(x)$ is relaxed, we may define the degree of subsethood DS as

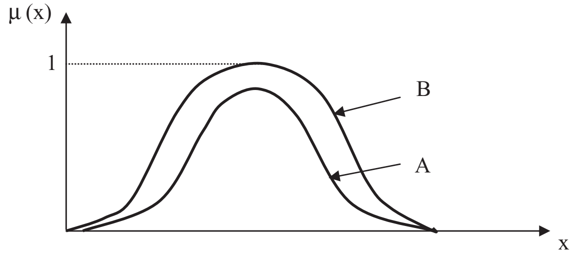

Figure 14.9. The concept of $A \subseteq B$ where A and B are fuzzy sets.

$$DS(A, B) = (1/Card(A))\{Card(A) - \Sigma \max[0, A(x) - B(x)]\}, x \in X$$

DS(A,B) provides a normalized measure of the degree to which the inequality $\mu_A(x) \le \mu_B(x)$ is violated.

Now we have enough background to explain one of the most important concepts in formalization of a fuzzy-reasoning process. The *extension principle* is a basic transformation of the fuzzy-set theory that provides a general procedure for extending the crisp domains of mathematical expressions to fuzzy domains. This procedure generalizes a common point-to-point mapping of a function f between fuzzy sets. The *extension principle* plays a fundamental role in translating set-based concepts into their fuzzy counterparts. Essentially, the extension principle is used to transform fuzzy sets via functions. Let X and Y be two sets, and F is a mapping from X to Y:

$$F: X \to Y$$

Let A be a fuzzy set in X. The extension principle states that the image of A under this mapping is a fuzzy set B = f(A) in Y such that for each $y \in Y$:

$$\mu_B(y) = \max \mu_A(x), \text{subject to } x \in X \text{ and } y = f(x).$$

The basic idea is illustrated in Figure 14.10. The extension principle easily generalizes to functions of many variables as follows. Let X_i, i = 1, . . . , n, and Y be universes of discourse, and $X = X1 \times X2 \times \ldots \times Xn$ constitute the Cartesian product of the X_is. Consider fuzzy sets A_i in X_i, i = 1, . . . , n and a mapping y = f(x), where the input is an n-dimensional vector $x = (x_1, x_2, \ldots, x_n)$ and $x \in X$. Fuzzy sets A1, A2, . . . , An are then transformed via f, producing the fuzzy set B = f(A1, A2, . . . , An) in Y such that for each $y \in Y$:

$$\mu_B(y) = \max_x\{\min(\mu_{A1}[x_1], \mu_{A2}[x_2], \ldots, \mu_{An}[x_n])\}$$

subject to $x \in X$ and y = f(x). Actually, in the expression above, the min operator is just a choice within a family of operators called triangular norms.

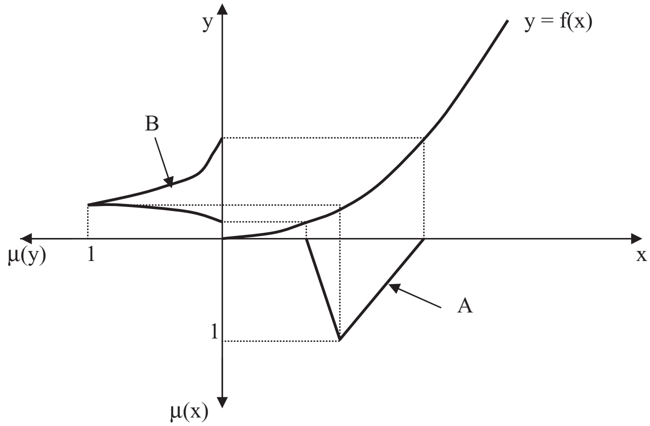

Figure 14.10. The idea of the extension principle.

More specifically, suppose that f is a function from X to Y where X and Y are discrete universes of discourse, and A is a fuzzy set on X defined as

$$A = \mu_A(x_1)/x_1 + \mu_A(x_2)/x_2 + \mu_A(x_3)/x_3 + \ldots + \mu_A(x_n)/x_n$$

then the extension principle states that the image of fuzzy set A under the mapping f can be expressed as a fuzzy set B:

$$B = f(A) = \mu_A(x_1)/y_1 + \mu_A(x_2)/y_2 + \mu_A(x_3)/y_3 + \ldots + \mu_A(x_n)/y_n$$

where $y_i = f(x_i)$, i = 1, . . . , n. In other words, the fuzzy set B can be defined through the mapped values x_i using the function f.

Let us analyze the extension principle using one example. Suppose that X = {1, 2, 3, 4} and Y = {1, 2, 3, 4, 5, 6} are two universes of discourse, and the function for transformation is y = x + 2. For a given fuzzy set A = 0.1/1 + 0.2/2 + 0.7/3 + 1.0/4 in X, it is necessary to find a corresponding fuzzy set B(y) in Y using the extension principle through function B = f(A). In this case, the process of computation is straightforward and a final, transformed fuzzy set is B = 0.1/3 + 0.2/4 + 0.7/5 + 1.0/6.

Another problem will show that the computational process is not always a one-step process. Suppose that A is given as

$$A = 0.1/-2 + 0.4/-1 + 0.8/0 + 0.9/1 + 0.3/2$$

and the function f is

$$f(x) = x^2 - 3$$

Upon applying the extension principle, we have

$$\begin{aligned}B &= 0.1/1 + 0.4/-2 + 0.8/-3 + 0.9/-2 + 0.3/1 = \\ &= 0.8/-3 + (0.4 \vee 0.9)/-2 + (0.1 \vee 0.3)/1 = \\ &= 0.8/-3 + 0.9/-2 + 0.3/1\end{aligned}$$

where $\vee$ represents the max function. For a fuzzy set with a continuous universe of discourse X, an analogous procedure applies.

Besides being useful in the application of the extension principle, some of the unary and binary fuzzy relations are also very important in a fuzzy-reasoning process. Binary fuzzy relations are fuzzy sets in X × Y that map each element in X × Y to a membership grade between 0 and 1. Let X and Y be two universes of discourse. Then

$$R = \{([x, y], \mu_R[x, y]) \mid (x, y) \in X \times Y\}$$

is a binary fuzzy relation in X × Y. Note that $\mu_R(x, y)$ is in fact a two-dimensional (2-D) MF. For example, let $X = Y = R^+$ (the positive real axis); the fuzzy relation is given as R = "y is much greater than x." The MF of the fuzzy relation can be subjectively defined as

$$\mu_R(x, y) = \begin{cases}(y - x)/(x + y + 2), & \text{if } y > x \\ 0 & \text{if } y \le x\end{cases}$$

If X and Y are a finite set of discrete values such as X {3, 4, 5} and Y = {3, 4, 5, 6, 7}, then it is convenient to express the fuzzy relation R as a relation matrix:

$$R = \begin{bmatrix}0 & 0.111 & 0.200 & 0.273 & 0.333 \\ 0 & 0 & 0.091 & 0.167 & 0.231 \\ 0 & 0 & 0 & 0.077 & 0.143\end{bmatrix}$$

where the element at row i and column j is equal to the membership grade between the i*th* element of X and the j*th* element of Y.

Common examples of binary fuzzy relations are as follows:

1. x is close to y (x and y are numbers).
2. x depends on y (x and y are categorical data).
3. x and y look alike.
4. If x is large, then y is small.

Fuzzy relations in different product spaces can be combined through a composition operation. Different composition operations have been suggested for fuzzy relations; the best known is the max–min composition proposed by Zadeh. Let R_1 and R_2 be two fuzzy relations defined on X × Y and Y × Z, respectively. The max–min composition of R_1 and R_2 is a fuzzy set defined by

$$R_1 \circ R_2 = \{[(x, z), \max_y \min(\mu_{R1}(x, y), \mu_{R2}(y, z))] \mid x \in X, y \in Y, z \in Z\}$$

or equivalently,

$$R_1 \circ R_2 = \vee_y[(\mu_{R1}(x, y) \wedge \mu_{R2}(y, z))]$$

with the understanding that $\vee$ and $\wedge$ represent max and min, respectively.

When R1 and R2 are expressed as relation matrices, the calculation of $R_1 \circ R_2$ is similar to the matrix-multiplication process, except that $\times$ and $+$ operations are replaced by $\vee$ and $\wedge$, respectively.

The following example demonstrates how to apply the max–min composition on two relations and how to interpret the resulting fuzzy relation $R_1 \circ R_2$. Let R_1 = "x is relevant to y" and R_2 = "y is relevant to z" be two fuzzy relations defined on $X \times Y$ and $Y \times Z$, where $X = \{1, 2, 3\}$, $Y = \{\alpha, \beta, \gamma, \delta\}$, and $Z = \{a, b\}$. Assume that R_1 and R_2 can be expressed as the following relation matrices of μ values:

$$R_1 = \begin{bmatrix} 0.1 & 0.3 & 0.5 & 0.7 \\ 0.4 & 0.2 & 0.8 & 0.9 \\ 0.6 & 0.8 & 0.3 & 0.2 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.3 \\ 0.5 & 0.6 \\ 0.7 & 0.2 \end{bmatrix}$$

Fuzzy relation $R_1 \circ R_2$ can be interpreted as a derived relation "x is relevant to z" based on relations R_1 and R_2. We will make a detailed max–min composition only for one element in a resulting fuzzy relation: (x, z) = (2, a).

$$\begin{aligned} \mu_{R1 \circ R2}(2, a) &= \max(0.4 \wedge 0.9, 0.2 \wedge 0.2, 0.8 \wedge 0.5, 0.9 \wedge 0.7) \\ &= \max(0.4, 0.2, 0.5, 0.7) \\ &= 0.7 \end{aligned}$$

Analogously, we can compute the other elements, and the final fuzzy matrix $R_1 \circ R_2$ will be

$$R_1 \circ R_2 = \begin{bmatrix} 0.7 & 0.5 \\ 0.7 & 0.6 \\ 0.6 & 0.3 \end{bmatrix}$$

14.4 FUZZY LOGIC AND FUZZY INFERENCE SYSTEMS

Fuzzy logic enables us to handle uncertainty in a very intuitive and natural manner. In addition to making it possible to formalize imprecise data, it also enables us to do arithmetic and Boolean operations using fuzzy sets. Finally, it describes the inference systems based on fuzzy rules. Fuzzy rules and fuzzy-reasoning processes, which are the most important modeling tools based on the fuzzy-set theory, are the backbone of any fuzzy inference system. Typically, a fuzzy rule has the general format of a conditional proposition. A *fuzzy If-then rule*, also known as *fuzzy implication,* assumes the form

$$\text{If x is A, then y is B}$$

where A and B are linguistic values defined by fuzzy sets on the universes of discourse X and Y, respectively. Often, "x is A" is called the antecedent or premise, while "y is B" is called the consequence or conclusion. Examples of fuzzy if-then rules are widespread in our daily linguistic expressions, such as the following:

1. If pressure is high, then volume is small.
2. If the road is slippery, then driving is dangerous.
3. If a tomato is red, then it is ripe.
4. If the speed is high, then apply the brake a little.

Before we can employ fuzzy if-then rules to model and analyze a fuzzy reasoning-process, we have to formalize the meaning of the expression "if x is A then y is B," sometimes abbreviated in a formal presentation as $A \rightarrow B$. In essence, the expression describes a relation between two variables x and y; this suggests that a fuzzy if-then rule be defined as a binary fuzzy relation R on the product space $X \times Y$. R can be viewed as a fuzzy set with a 2-D MF:

$$\mu_R(x, y) = f(\mu_A[x], \mu_B[y])$$

If we interpret $A \rightarrow B$ as A *entails* B, still it can be formalized in several different ways. One formula that could be applied based on a standard logical interpretation, is

$$R = A \rightarrow B = A' \cup B.$$

Note that this is only one of several possible interpretations for fuzzy implication. The accepted meaning of $A \rightarrow B$ represents the basis for an explanation of the fuzzy-reasoning process using if-then fuzzy rules.

Fuzzy reasoning, also known as approximate reasoning, is an inference procedure that derives its conclusions from a set of fuzzy rules and known facts (they also can be fuzzy sets). The basic rule of inference in a traditional two-valued logic is *modus ponens*, according to which we can infer the truth of a proposition B from the truth of A and the implication $A \rightarrow B$. However, in much of human reasoning, *modus ponens* is employed in an approximate manner. For example, if we have the rule "if the tomato is red, then it is ripe" and we know that "the tomato is more or less red," then we may infer that "the tomato is more or less ripe." This type of approximate reasoning can be formalized as

Fact: x is A′
Rule: If x is A then y is B
Conclusion: y is B′

where A′ is close to A and B′ is close to B. When A, A′, B, and B′ are fuzzy sets of an approximate universe, the foregoing inference procedure is called approximate reasoning or fuzzy reasoning; it is also called *generalized modus ponens*, since it has *modus ponens* as a special case.

Using the composition rule of inference, we can formulate the inference procedure of fuzzy reasoning. Let A, A′, and B be fuzzy sets on X, X, and Y domains, respectively. Assume that the fuzzy implication A → B is expressed as a fuzzy relation R on X × Y. Then the fuzzy set B′ induced by A′ and A → B is defined by

$$\begin{aligned}\mu_{B'}(y) &= \max_x \min[\mu_{A'}(x), \mu_R(x, y)] \\ &= \vee_x[\mu_{A'}(x) \wedge \mu_R(x, y)]\end{aligned}$$

Some typical characteristics of the fuzzy-reasoning process and some conclusions useful for this type of reasoning are

1. $\forall A, \forall A' \rightarrow B' \supseteq B$ (\or $\mu_{B'}(y) \geq \mu_B(y)$)
2. If $A' \subseteq A$ (or $\mu_A(x) \geq \mu_{A'}(x)$) $\rightarrow B' = B$

Let us analyze the computational steps of a fuzzy-reasoning process for one simple example. Given the fact A′ = “x is above average height” and the fuzzy rule “if x is high, then his/her weight is also high,” we can formalize this as a fuzzy implication A → B. We can use a discrete representation of the initially given fuzzy sets A, A′, and B (based on subjective heuristics):

A′:	x	μ(x)	A:	x	μ(x)	B:	y	μ(y)
	5’6”	0.3		5’6”	0		120	0
	5’9”	1.0		5’9”	0.2		150	0.2
	6’	0.4		6’	0.8		180	0.5
	6’3”	0		6’3”	1.0		210	1.0

$\mu_R(x, y)$ can be computed in several different ways, such as

$$\mu_R(x, y) = \begin{cases} 1 & \text{for } \mu_A(x) \leq \mu_B(y) \\ \mu_B(y) & \text{otherwise} \end{cases}$$

or as the Lukasiewicz norm:

$$\mu_R(x, y) = \{1 \wedge (1 - \mu_A(x) + \mu_B(y)\}$$

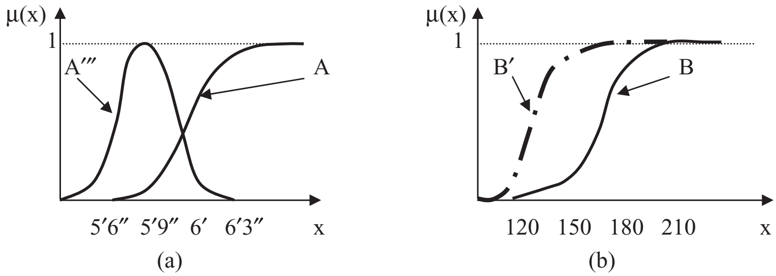

Figure 14.11. Comparison of approximate reasoning result B′ with initially given fuzzy sets A′, A, and B and the fuzzy rule A → B. (a) Fuzzy sets A and A′; (b) fuzzy sets B and B′ (conclusion).

Both definitions lead to a very different interpretation of fuzzy implication. Applying the first relation for $\mu_R(x, y)$ on the numeric representation for our sets A and B, the 2-D MF will be

$$\mu_R(x, y) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0.2 & 0.5 & 1 \\ 0 & 0.2 & 0.5 & 1 \end{bmatrix}$$

Now, using the basic relation for inference procedure, we obtain

$$\mu_{B'}(y) = \max_x \min[\mu_{A'}(x), \mu_R(x, y)]$$

The resulting fuzzy set B′ can be represented in the form of a table:

B′:	y	μ(y)
	120	0.3
	150	1.0
	180	1.0
	210	1.0

or interpreted approximately in linguistic terms: "x's weight is more-or-less high." A graphical comparison of MFs for fuzzy sets A, A′, B, and B′ is given in Figure 14.11.

To use fuzzy sets in approximate reasoning (a set of linguistic values with numeric representations of MFs), the main tasks for the designer of a system are

1. represent any fuzzy datum, given as a linguistic value, in terms of the codebook A;
2. use these coded values for different communication and processing steps; and
3. at the end of approximate reasoning, transform the computed results back into its original (linguistic) format using the same codebook A.

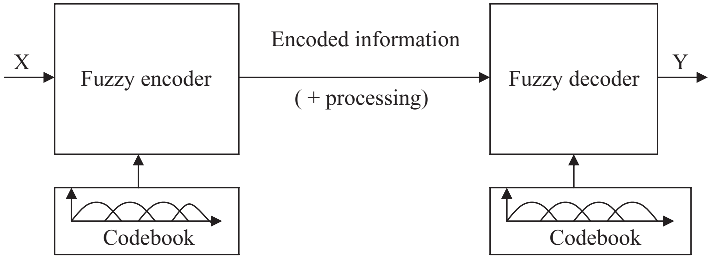

Figure 14.12. Fuzzy-communication channel with fuzzy encoding and decoding.

These three fundamental tasks are commonly referred to as encoding, transmission and processing, and decoding (the terms have been borrowed from communication theory). The encoding activities occur at the transmitter while the decoding take place at the receiver. Figure 14.12 illustrates encoding and decoding with the use of the codebook A. The channel functions as follows. Any input information, whatever its nature, is encoded (represented) in terms of the elements of the codebook. In this internal format, encoded information is sent with or without processing across the channel. Using the same codebook, the output message is decoded at the receiver.

Fuzzy-set literature has traditionally used the terms fuzzification and defuzzification to denote encoding and decoding, respectively. These are, unfortunately, quite misleading and meaningless terms because they mask the very nature of the processing that takes place in fuzzy reasoning. They neither address any design criteria nor introduce any measures aimed at characterizing the quality of encoding and decoding information completed by the fuzzy channel.

The next two sections are examples of the application of fuzzy logic and fuzzy reasoning to decision-making processes, where the available data sets are ambiguous. These applications include multifactorial evaluation and extraction of fuzzy rules-based models from large numeric data sets.

14.5 MULTIFACTORIAL EVALUATION

Multifactorial evaluation is a good example of the application of the fuzzy-set theory to decision-making processes. Its purpose is to provide a synthetic evaluation of an object relative to an objective in a fuzzy-decision environment that has many factors. Let $U = \{u_1, u_2, \ldots, u_n\}$ be a set of objects for evaluation, let $F = \{f_1, f_2, \ldots, f_m\}$ be the set of basic factors in the evaluation process, and let $E = \{e_1, e_2, \ldots, e_p\}$ be a set of descriptive grades or qualitative classes used in the evaluation. For every object $u \in U$, there is a single-factor evaluation matrix R(u) with dimensions $m \times p$, which is usually the result of a survey. This matrix may be interpreted and used as a 2-D MF for fuzzy relation $F \times E$.

With the preceding three elements, F, E, and R, the evaluation result D(u) for a given object $u \in U$ can be derived using the basic fuzzy-processing procedure: the

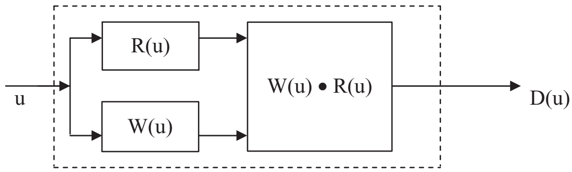

Figure 14.13. Multifactorial-evaluation model.

product of fuzzy relations through max–min composition. This has been shown in Figure 14.13. An additional input to the process is the weight vector W(u) for evaluation factors, which can be viewed as a fuzzy set for a given input u. A detailed explanation of the computational steps in the multifactorial-evaluation process will be given through two examples.

14.5.1 A Cloth-Selection Problem

Assume that the basic factors of interest in the selection of cloth consist of f_1 = style, f_2 = quality, and f_3 = price, that is, F = {f_1, f_2, f_3}. The verbal grades used for the selection are e_1 = best, e_2 = good, e_3 = fair, and e_4 = poor, that is, E = {e_1, e_2, e_3, e_4}. For a particular piece of cloth u, the single-factor evaluation may be carried out by professionals or customers by a survey. For example, if the survey results of the "style" factor f_1 are 60% for the best, 20% for the good, 10% for the fair, and 10% for the poor, then the single-factor evaluation vector $R_1(u)$ is

$$R_1(u) = \{0.6, 0.2, 0.1, 0.1\}$$

Similarly, we can obtain the following single-factor evaluation vectors for f_2 and f_3:

$$R_2(u) = \{0.1, 0.5, 0.3, 0.1\}$$

$$R_3(u) = \{0.1, 0.3, 0.4, 0.2\}$$

Based on single-factor evaluations, we can build the following evaluation matrix:

$$R(u) = \begin{bmatrix} R_1(u) \\ R_2(u) \\ R_3(u) \end{bmatrix} = \begin{bmatrix} 0.6 & 0.2 & 0.1 & 0.1 \\ 0.1 & 0.5 & 0.3 & 0.1 \\ 0.1 & 0.3 & 0.4 & 0.2 \end{bmatrix}$$

If a customer's weight vector with respect to the three factors is

$$W(u) = \{0.4, 0.4, 0.2\}$$

then it is possible to apply the multifactorial-evaluation model to compute the evaluation for a piece of cloth u. "Multiplication" of matrices W(u) and R(u) is based on the max–min composition of fuzzy relations, where the resulting evaluation is in the form of a fuzzy set $D(u) = [d_1, d_2, d_3, d_4]$:

$$D(u) = W(u) \cdot R(u) = [0.4 \quad 0.4 \quad 0.2] = \begin{bmatrix} 0.6 & 0.2 & 0.1 & 0.1 \\ 0.1 & 0.5 & 0.3 & 0.1 \\ 0.1 & 0.3 & 0.4 & 0.2 \end{bmatrix}$$
$$= [0.4 \quad 0.4 \quad 0.3 \quad 0.2]$$

where, for example, d_1 is calculated through the following steps:

$$\begin{aligned} d_1 &= (w_1 \wedge r_{11}) \vee (w_2 \wedge r_{21}) \vee (w_3 \wedge r_{31}) \\ &= (0.4 \wedge 0.6) \vee (0.4 \wedge 0.1) \vee (0.2 \wedge 0.1) \\ &= 0.4 \vee 0.1 \vee 0.1 \\ &= 0.4 \end{aligned}$$

The values for d_2, d_3, and d_4 are found similarly, where $\wedge$ and $\vee$ represent the operations min and max, respectively. Because the largest components of D(u) are $d_1 = 0.4$ and $d_2 = 0.4$ at the same time, the analyzed piece of cloth receives a rating somewhere between "best" and "good."

14.5.2 A Problem of Evaluating Teaching

Assume that the basic factors that influence students' evaluation of teaching are f_1 = clarity and understandability, f_2 = proficiency in teaching, f_3 = liveliness and stimulation, and f_4 = writing neatness or clarity, that is, $F = \{f_1, f_2, f_3, f_4\}$. Let $E = \{e_1, e_2, e_3, e_4\}$ = {excellent, very good, good, poor} be the verbal grade set. We evaluate a teacher u. By selecting an appropriate group of students and faculty, we can have them respond with their ratings on each factor and then obtain the single-factor evaluation. As in the previous example, we can combine the single-factor evaluation into an evaluation matrix. Suppose that the final matrix R(u) is

$$R(u) = \begin{bmatrix} 0.7 & 0.2 & 0.1 & 0.0 \\ 0.6 & 0.3 & 0.1 & 0.0 \\ 0.2 & 0.6 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.6 & 0.2 \end{bmatrix}$$

For a specific weight vector $W(u) = \{0.2, 0.3, 0.4, 0.1\}$, describing the importance of the teaching-evaluation factor f_i and using the multifactorial-evaluation model, it is easy to find

$$D(u) = W(u) \cdot R(u) = [0.2 \quad 0.3 \quad 0.4 \quad 0.1] \cdot \begin{bmatrix} 0.7 & 0.2 & 0.1 & 0.0 \\ 0.6 & 0.3 & 0.1 & 0.0 \\ 0.2 & 0.6 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.6 & 0.2 \end{bmatrix}$$
$$= [0.2 \quad 0.4 \quad 0.1 \quad 0.1]$$

Analyzing the evaluation results D(u), because $d_2 = 0.4$ is a maximum, we may conclude that teacher u should be rated as "very good."

14.6 EXTRACTING FUZZY MODELS FROM DATA

In the context of different data-mining analyses, it is of great interest to see how fuzzy models can automatically be derived from a data set. Besides prediction, classification, and all other data-mining tasks, understandability is of prime concern, because the resulting fuzzy model should offer an insight into the underlying system. To achieve this goal, different approaches exist. Let us explain a common technique that constructs grid-based rule sets using a global granulation of the input and output spaces.

Grid-based rule sets model each input variable usually through a small set of linguistic values. The resulting rule base uses all or a subset of all possible combinations of these linguistic values for each variable resulting in a global granulation of the feature space into rectangular regions. Figure 14.14 illustrates this approach in two dimensions: with three linguistic values (low, medium, high) for the first dimension x_1 and two linguistic values for the second dimension x_2 (young, old).

Extracting grid-based fuzzy models from data is straightforward when the input granulation is fixed, that is, the antecedents of all rules are predefined. Then, only a matching consequent for each rule needs to be found. This approach, with fixed grids, is usually called the *Mamdani model*. After predefinition of the granulation of all input variables and also the output variable, one sweeps through the entire data set and determines the closest example to the geometrical center of each rule, assigning the closest fuzzy value output to the corresponding rule. Using graphical interpretation in a 2-D space, the global steps of the procedure are illustrated through an example in

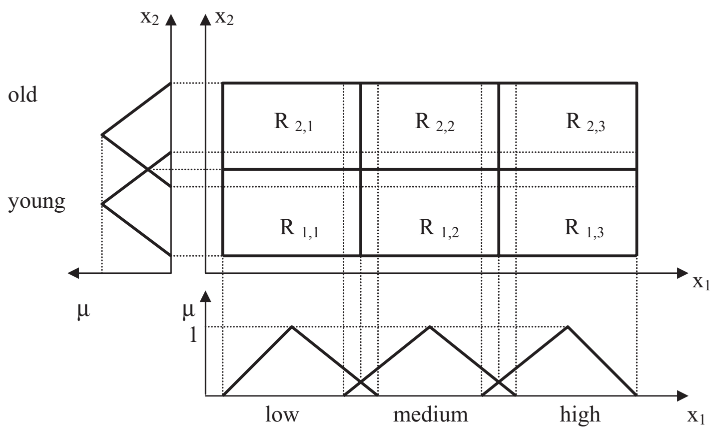

Figure 14.14. A global granulation for a two-dimensional space using three membership functions for x_1 and two for x_2.

which only one input x and one output dimension y exist. The formal analytical specification, even with more than one input/output dimension, is very easy to establish.

1. *Granulate the Input and Output Space.* Divide each variable x_i into n_i equidistant, triangular, MFs. In our example, both input x and output y are granulated using the same four linguistic values: low, below average, above average, and high. A representation of the input–output granulated space is given in Figure 14.15.
2. *Analyze the Entire Data Set in the Granulated Space.* First, enter a data set in the granulated space and then find the points that lie closest to the centers of the granulated regions. Mark these points and the centers of the region. In our example, after entering all discrete data, the selected center points (closest to the data) are additionally marked with x, as in Figure 14.16.

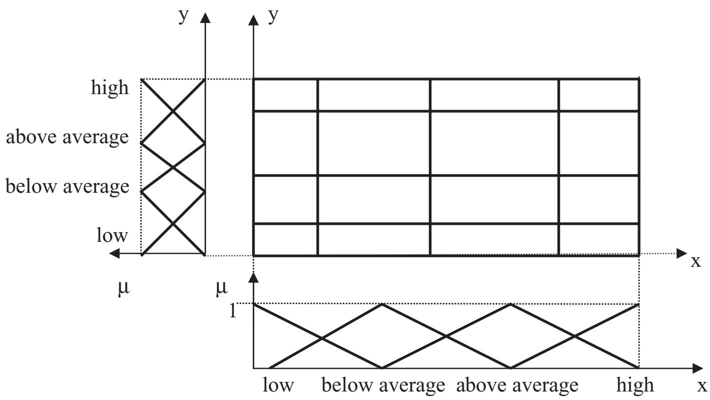

Figure 14.15. Granulation of a two-dimensional I/O space.

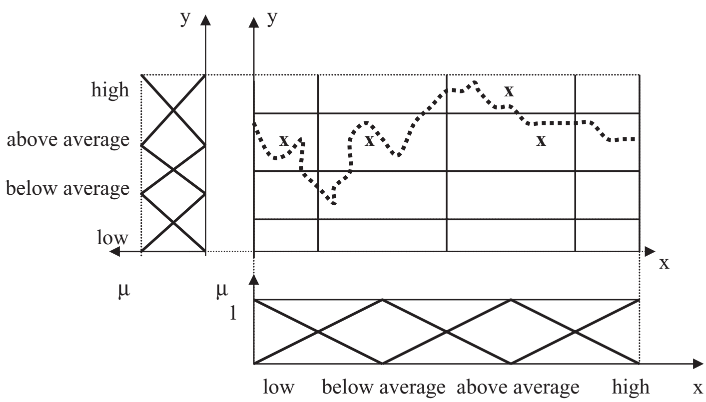

Figure 14.16. Selection of characteristic points in a granulated space.

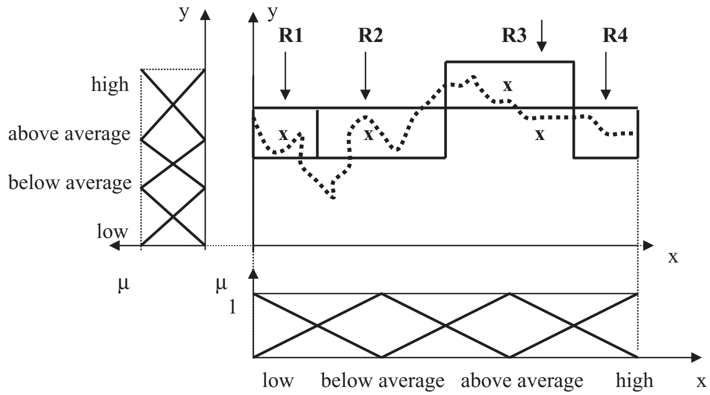

Figure 14.17. Graphical representation of generated fuzzy rules and the resulting crisp approximation.

3. *Generate Fuzzy Rules from Given Data.* Data representative directly selects the regions in a granulated space. These regions may be described with the corresponding fuzzy rules. In our example, four regions are selected, one for each fuzzy input linguistic value, and they are represented in Figure 14.17 with a corresponding crisp approximation (a thick line through the middle of the regions). These regions are the graphical representation of fuzzy rules. The same rules may be expressed linguistically as a set of IF-THEN constructions:

R_1: IF x is *small*, THEN y is *above average*.
R_2: IF x is *below average*, THEN y is *above average*.
R_3: IF x is *above average*, THEN y is *high*.
R_4: *IF* x *is high,* THEN y *is above* average.

Note how the generated model misses the extremes that lie far from the existing rule centers. This behavior occurs because only one pattern per rule is used to determine the outcome of this rule. Even a combined approach would very much depend on the predefined granulation. If the function to be modeled has a high variance inside one rule, the resulting fuzzy rule model will fail to model this behavior.

For practical applications it is obvious, however, that using such a predefined, fixed grid results in a fuzzy model that will either not fit the underlying functions very well or consist of a large number of rules because of small granulation. Therefore, new approaches have been introduced that automatically determine the granulations of both input and output variables based on a given data set. We will explain the basic steps for one of these algorithms using the same data set from the previous example and the graphical representation of applied procedures.

1. Initially, only one MF is used to model each of the input variables as well as the output variable, resulting in one large rule covering the entire feature space.

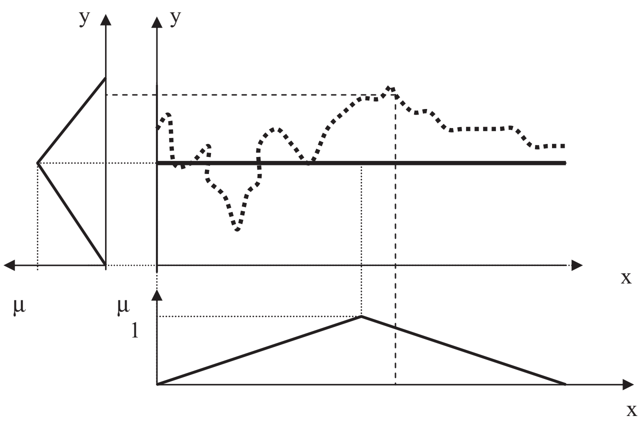

Figure 14.18. The first step in automatically determining fuzzy granulation.

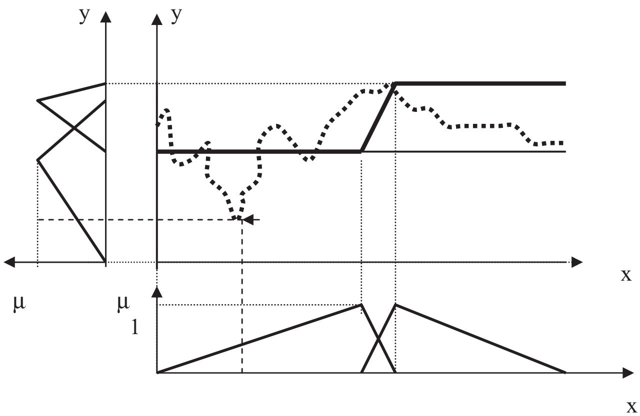

Figure 14.19. The second step (first iteration) in automatically determining granulation.

Subsequently, new MFs are introduced at points of maximum error (the maximum distance between data points and the obtained crisp approximation). Figure 14.18 illustrates this first step in which the crisp approximation is represented with a thick line and the selected point of maximal error with a triangle.

2. For the selected point of maximum error, new triangular fuzzy values for both input and output variables are introduced. Processes of granulation, determining fuzzy rules in the form of space regions, and crisp approximation are repeated for a space, with additional input and output fuzzy values for the second step—that means two fuzzy values for both input and output variables. The final results of the second step, for our example, are presented in Figure 14.19.

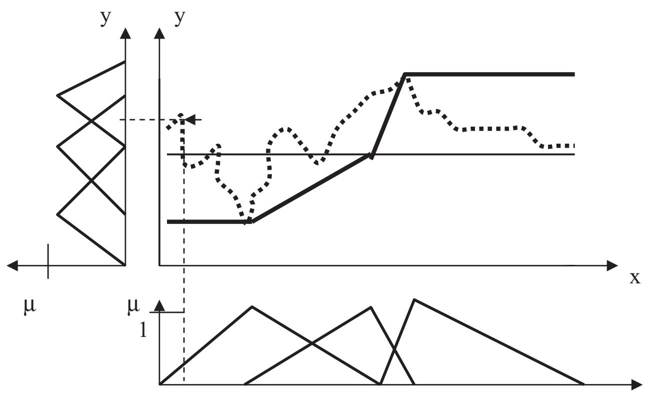

Figure 14.20. The second step (second iteration) in automatically determining granulation.

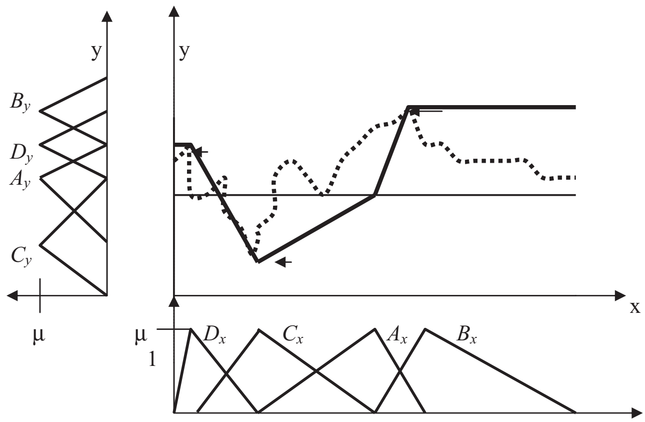

Figure 14.21. The second step (third iteration) in automatically determining granulation.

3. Step 2 is repeated until a maximum number of divisions (fuzzy values) is reached, or the approximation error remains below a certain threshold value. Figures 14.20 and 14.21 demonstrate two additional iterations of the algorithm for a data set. Here granulation was stopped after a maximum of four MFs was generated for each variable. Obviously this algorithm is able to model extremes much better than the previous one with a fixed granulation. At the same time, it has a strong tendency to favor extremes and to concentrate on outliers. The final set of fuzzy rules, using dynamically created fuzzy values A_x to D_x and A_y to D_y for input and output variables, is

R_1: IF x is A_x, THEN y is A_y.
R_2: IF x is B_x, THEN y is B_y.
R_3: IF x is C_x, THEN y is C_y.
R_4: IF x is D_x, THEN y is D_y.

14.7 DATA MINING AND FUZZY SETS

There is a growing indisputable role of fuzzy set technology in the realm of data mining. In a data mining process, discovered models, learned concepts, or patterns of interest are often vague and have non-sharp boundaries. Unfortunately, the representation of graduality is often foiled in data-mining applications, especially in connection with the learning of predictive models. For example, the fact that neural networks are often used as data-mining methods, although their learning result (weight matrices of numbers) is hardly interpretable, shows that in contrast to the standard definition the goal of understandable models is often neglected. In fact, one should recognize that graduality is not only advantageous for expressing concepts and patterns, but also for modeling the qualifying properties and relations. Of course, correctness, completeness, and efficiency are important in data-mining models, but in order to manage systems that are more and more complex, there is a constantly growing demand to keep the solutions conceptually simple and understandable. Modern technologies are accepted more readily, if the methods applied and models derived are easy to understand, and the results can be checked against human intuition.

The complexity of the learning task, obviously, leads to a problem: When learning from information, one must choose between mostly quantitative methods that achieve good performances, and qualitative models that explain to a user what is going on in the complex system. Fuzzy-set theory has the potential to produce models that are more comprehensible, less complex, and more robust. Fuzzy information granulation appears to be appropriate approach for trading off accuracy against complexity and understandability of data-mining models. Also, fuzzy-set theory in conjunction with possibility theory, can contribute considerably to the modeling and processing of various forms of uncertain and incomplete information available in large real-world systems.

The tools and technologies that have been developed in fuzzy-set theory have the potential to support all of the steps that comprise a process of knowledge discovery. Fuzzy methods appear to be particularly useful for data pre- and postprocessing phases of a data-mining process. In particular, it has already been used in the data-selection phase, for example, for modeling vague data in terms of fuzzy sets, to "condense" several crisp observations into a single fuzzy one, or to create fuzzy summaries of the data.

Standard methods of data mining can be extended to include fuzzy-set representation in a rather generic way. Achieving focus is important in data mining because there are too many attributes and values to be considered and can result in combinatorial explosion. Most unsupervised data-mining approaches try to achieve focus by recognizing the most interesting structures and their features even if there is still some level of ambiguity. For example, in standard clustering, each sample is assigned to one cluster in a unique way. Consequently, the individual clusters are separated by sharp boundaries. In practice, such boundaries are often not very natural or even counterintuitive. Rather, the boundary of single clusters and the transition between different clusters are usually "smooth" rather than abrupt. This is the main motivation underlying fuzzy extensions to clustering algorithms. In fuzzy clustering an object may belong to different clusters at the same time, at least to some extent, and the degree to which it belongs to a particular cluster is expressed in terms of a membership degree.

The most frequent application of fuzzy set theory in data mining is related to the adaptation of rule-based predictive models. This is hardly surprising, since rule-based models have always been a cornerstone of fuzzy systems and a central aspect of research in the field. Set of fuzzy rules can represent both classification and regression models. Instead of dividing quantitative attributes into fixed intervals, they employ linguistic terms to represent the revealed regularities. Therefore, no user-supplied thresholds are required, and quantitative values can be directly inferred from the rules. The linguistic representation leads to the discovery of natural and more understandable rules.

Decision tree induction includes well-known algorithms such as ID3, C4.5, C5.0, and Classification and Regression Trees (CART). Fuzzy variants of decision-tree induction have been developed for quite a while and seem to remain a topic of interest even today. In fact, these approaches provide a typical example for the "fuzzification" of standard predictive methods. In the case of decision trees, it is primarily the "crisp" thresholds used for defining splitting attributes, such as size > 181 at inner nodes. Such thresholds lead to hard decision boundaries in the input space, which means that a slight variation of an attribute (e.g., size = 182 instead of size = 181) can entail a completely different classification of a sample. Usually, a decision in favor of one particular class label has to be made, even if the sample under consideration seems to have partial membership in several classes simultaneously. Moreover, the learning process becomes unstable in the sense that a slight variation of the training samples can change the induced decision tree drastically. In order to make the decision boundaries "soft," an obvious idea is to apply fuzzy predicates at the nodes of a decision tree, for example, size = *LARGE*, where *LARGE* is a fuzzy set. In that case the sample is not assigned to exactly one successor node in a unique way, but perhaps to several successors with a certain degree. Also, for fuzzy classification solutions the consequent of single rules is usually a class assignment represented with a singleton fuzzy set. Evaluating a rule-based model thus becomes trivial and simply amounts to "maximum matching," that is, searching the maximally supporting rule for each class.

A particularly important trend in the field of fuzzy systems are *hybrid methods* that combine fuzzy-set theory with other methodologies such as neural networks. In the *neuro-fuzzy* methods the main idea is to encode a fuzzy system in a neural network, and to apply standard approaches like backpropagation in order to train such a network. This way, *neuro-fuzzy* systems combine the representational advantages of fuzzy systems with the flexibility and adaptivity of neural networks. Interpretations of fuzzy membership include similarity, preference, and uncertainty. A primary motivation was to provide an interface between a numerical scale and a symbolic scale that is usually composed of linguistic terms. Thus, fuzzy sets have the capability to interface quantitative data with qualitative knowledge structures expressed in terms of natural language. In general, due to their closeness to human reasoning, solutions obtained using fuzzy approaches are easy to understand and to apply. This provides the user with comprehensive information and often data summarization for grasping the essence of discovery from a large amount of information in a complex system.

14.8 REVIEW QUESTIONS AND PROBLEMS

1. Find some examples of fuzzy variables in daily life.
2. Show graphically and explain why the law of contradiction is violated in the fuzzy-set theory.
3. The MF of a fuzzy set is defined as

$$\mu_A(x) = \begin{cases} 1 & \text{for} \quad 0 < x < 20 \\ (50 - x)/30 & \text{for} \quad 20 \le x < 50 \\ 0 & \text{for} \quad x \ge 50 \end{cases}$$

   (a) What will be a linguistic description of the fuzzy set A if x is the variable "age" in years?
   (b) Give an analytical description for $\mu_B(x)$ if B is a fuzzy set "age is close to 60 years."
4. Assume you were told that the room temperature is around 70 degrees Fahrenheit. How you would represent this information?
   (a) by a set notation,
   (b) by a fuzzy set notation.
5. Consider the fuzzy sets A, B, and C defined on the interval x = [0, 10] with corresponding μ functions:

$$\mu_A(x) = x/(x+2) \quad \mu_B(x) = 2^{-x} \quad \mu_C(x) = \begin{cases} x^2/24 & \text{for } x \in [0, 4.89] \\ 1 & \text{otherwise} \end{cases}$$

   Determine analytically and graphically:
   (a) A' and B'
   (b) $A \cup C$ and $A \cup B$
   (c) $A \cap C$ and $A \cap B$
   (d) $A \cup B \cup C$
   (e) $A \cap C'$
   (f) Calculate the α-cuts for A, B, and C if α = 0.2, α = 0.5, and α = 1.
6. Consider two fuzzy sets with triangular MFs A(x, 1, 2, 3) and B(x, 2, 2, 4). Find their intersection and union graphically, and express them analytically using the min and max operators.

7. If X = {3, 4, 5} and Y = {3, 4, 5, 6, 7}, and the binary fuzzy relation R = “Y is much greater than X” is defined by the analytical MF

$$\mu_R(X, Y) = \begin{cases} (Y - X)/(X + Y + 2) & \text{if} \quad Y > X \\ 0 & \text{if} \quad Y \le X \end{cases}$$

what will be corresponding relation matrix of R (for all discrete X and Y values)?

8. Apply the extension principle to the fuzzy set

$$A = 0.1/-2 + 0.4/-1 + 0.8/0 + 0.9/1 + 0.3/2$$

where the mapping function $f(x) = x^2 - 3$.

(a) What is the resulting image B where B = f(A)?

(b) Sketch this transformation graphically.

9. Assume that the proposition “if x is A then y is B” is given where A and B are fuzzy sets:

$$A = 0.5/x_1 + 1/x_2 + 0.6/x_3$$

$$B = 1/y_1 + 0.4/y_2$$

Given a fact expressed by the proposition “x is A*,” where

$$A^* = 0.6/x_1 + 0.9/x_2 + 0.7/x_3$$

derive the conclusion in the form “y is B*” using the generalized *modus ponens* inference rule.

10. Solve Problem number 9 by using

$$A = 0.6/x_1 + 1/x_2 + 0.9/x_3$$

$$B = 0.6/y_1 + 1/y_2$$

$$A^* = 0.5/x_1 + 0.9/x_2 + 1/x_3$$

11. The test scores for the three students are given in the following table:

	Math	Physics	Chemistry	Language
Henry	66	91	95	83
Lucy	91	88	80	73
John	80	88	80	78

Find the best student using multifactorial evaluation, if the weight factors for the subjects are given as the vector W = [0.3, 0.2, 0.1, 0.4].

12. Search the Web to find the basic characteristics of publicly available or commercial software tools that are based on fuzzy sets and fuzzy logic. Make a report of your search.

14.9 REFERENCES FOR FURTHER STUDY

Chen, Y., T. Wang, B. Wang, Z. Li, A Survey of Fuzzy Decision Tree Classifier, *Fuzzy Information and Engineering*, Vol. 1, No. 2, 2009, pp. 149–159.

Decision-tree algorithm provides one of the most popular methodologies for symbolic knowledge acquisition. The resulting knowledge, a symbolic decision tree along with a simple inference mechanism, has been praised for comprehensibility. The most comprehensible decision trees have been designed for perfect symbolic data. Over the years, additional methodologies have been investigated and proposed to deal with continuous or multi-valued data, and with missing or noisy features. Recently, with the growing popularity of fuzzy representation, some researchers have proposed to utilize fuzzy representation in decision trees to deal with similar situations. This paper presents a survey of current methods for Fuzzy Decision Tree (FDT) designment and the various existing issues. After considering potential advantages of FDT classifiers over traditional decision-tree classifiers, we discuss the subjects of FDT including attribute selection criteria, inference for decision assignment, and stopping criteria.

Cox, E., *Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration*, Morgan Kaufmann, San Francisco, CA, 2005.

Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration is a handbook for analysts, engineers, and managers involved in developing data-mining models in business and government. As you will discover, fuzzy systems are extraordinarily valuable tools for representing and manipulating all kinds of data, and genetic algorithms and evolutionary programming techniques drawn from biology provide the most effective means for designing and tuning these systems. You do not need a background in fuzzy modeling or genetic algorithms to benefit, for this book provides it, along with detailed instruction in methods that you can immediately put to work in your own projects. The author provides many diverse examples and also an extended example in which evolutionary strategies are used to create a complex scheduling system.

Laurent, A., M. Lesot, eds., *Scalable Fuzzy Algorithms for Data Management and Analysis, Methods and Design*, IGI Global, Hershey, PA, 2010.

The book presents innovative, cutting-edge fuzzy techniques that highlight the relevance of fuzziness for huge data sets in the perspective of scalability issues, from both a theoretical and experimental point of view. It covers a wide scope of research areas including data representation, structuring and querying, as well as information retrieval and data mining. It encompasses different forms of databases, including data warehouses, data cubes, tabular or relational data, and many applications, among which are music warehouses, video mining, bioinformatics, semantic Web and data streams.

Li, H. X., V. C. Yen, *Fuzzy Sets and Fuzzy Decision-Making*, CRC Press, Inc., Boca Raton, 1995.

The book emphasizes the applications of fuzzy-set theory in the field of management science and decision science, introducing and formalizing the concept of fuzzy decision making. Many interesting methods of fuzzy decision making are developed and illustrated with examples.

Pal, S. K., S. Mitra, *Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing*, John Wiley & Sons, Inc., New York, 1999.

The authors consolidate a wealth of information previously scattered in disparate articles, journals, and edited volumes, explaining both the theory of neuro-fuzzy computing and the latest methodologies for performing different pattern-recognition tasks using neuro-fuzzy networks—classification, feature evaluation, rule generation, and knowledge extraction. Special emphasis is given to the integration of neuro-fuzzy methods with rough sets and genetic algorithms to ensure a more efficient recognition system.

Pedrycz, W., F. Gomide, *An Introduction to Fuzzy Sets: Analysis and Design*, The MIT Press, Cambridge, 1998.

The book provides a highly readable, comprehensive, self-contained, updated, and well-organized presentation of the fuzzy-set technology. Both theoretical and practical aspects of the subject are given a coherent and balanced treatment. The reader is introduced to the main computational models, such as fuzzy modeling and rule-based computation, and to the frontiers of the field at the confluence of fuzzy-set technology with other major methodologies of soft computing.