4 Fuzzy expert systems

In which we present fuzzy set theory, consider how to build fuzzy expert systems and illustrate the theory through an example.

4.1 Introduction, or what is fuzzy thinking?

Experts usually rely on **common sense** when they solve problems. They also use vague and ambiguous terms. For example, an expert might say, 'Though the power transformer is **slightly** overloaded, I can keep this load for **a while**'. Other experts have no difficulties with understanding and interpreting this statement because they have the background to hearing problems described like this. However, a knowledge engineer would have difficulties providing a computer with the same level of understanding. How can we represent expert knowledge that uses vague and ambiguous terms in a computer? Can it be done at all?

This chapter attempts to answer these questions by exploring the **fuzzy set theory** (or **fuzzy logic**). We review the philosophical ideas behind fuzzy logic, study its apparatus and then consider how fuzzy logic is used in fuzzy expert systems.

Let us begin with a trivial, but still basic and essential, statement: fuzzy logic is not logic that is fuzzy, but logic that is used to describe fuzziness. Fuzzy logic is the theory of fuzzy sets, sets that calibrate vagueness. Fuzzy logic is based on the idea that all things admit of degrees. Temperature, height, speed, distance, beauty – all come on a sliding scale. The motor is running **really hot**. Tom is a **very tall** guy. Electric cars are **not very fast**. **High**-performance drives require **very rapid** dynamics and **precise** regulation. Hobart is **quite a short** distance from Melbourne. Sydney is a **beautiful** city. Such a sliding scale often makes it impossible to distinguish members of a class from non-members. When does a hill become a mountain?

Boolean or conventional logic uses sharp distinctions. It forces us to draw lines between members of a class and non-members. It makes us draw lines in the sand. For instance, we may say, 'The maximum range of an electric vehicle is short', regarding a range of 300 km or less as short, and a range greater than 300 km as long. By this standard, any electric vehicle that can cover a distance of 301 km (or 300 km and 500 m or even 300 km and 1 m) would be described as long-range. Similarly, we say Tom is tall because his height is 181 cm. If we drew a line at 180 cm, we would find that David, who is 179 cm, is small. Is David really a small man or have we just drawn an arbitrary line in the sand? Fuzzy logic makes it possible to avoid such absurdities.

Fuzzy logic reflects how people think. It attempts to model our sense of words, our decision making and our common sense. As a result, it is leading to new, more human, intelligent systems.

Fuzzy, or multi-valued logic was introduced in the 1930s by Jan Lukasiewicz, a Polish logician and philosopher (Lukasiewicz, 1930). He studied the mathematical representation of fuzziness based on such terms as **tall**, **old** and **hot**. While classical logic operates with only two values 1 (true) and 0 (false), Lukasiewicz introduced logic that extended the range of truth values to all real numbers in the interval between 0 and 1. He used a number in this interval to represent the **possibility** that a given statement was true or false. For example, the possibility that a man 181 cm tall is really tall might be set to a value of 0.86. It is **likely** that the man is tall. This work led to an inexact reasoning technique often called **possibility theory**.

Later, in 1937, Max Black, a philosopher, published a paper called 'Vagueness: an exercise in logical analysis' (Black, 1937). In this paper, he argued that a continuum implies degrees. Imagine, he said, a line of countless 'chairs'. At one end is a Chippendale. Next to it is a near-Chippendale, in fact indistinguishable from the first item. Succeeding 'chairs' are less and less chair-like, until the line ends with a log. When does a **chair** become a **log**? The concept **chair** does not permit us to draw a clear line distinguishing **chair** from **not-chair**. Max Black also stated that if a continuum is discrete, a number can be allocated to each element. This number will indicate a degree. But the question is degree of what. Black used the number to show the percentage of people who would call an element in a line of 'chairs' a **chair**; in other words, he accepted vagueness as a matter of probability.

However, Black's most important contribution was in the paper's appendix. Here he defined the first simple fuzzy set and outlined the basic ideas of fuzzy set operations.

In 1965 Lotfi Zadeh, Professor and Head of the Electrical Engineering Department at the University of California at Berkeley, published his famous paper 'Fuzzy sets'. In fact, Zadeh rediscovered fuzziness, identified and explored it, and promoted and fought for it.

Zadeh extended the work on possibility theory into a formal system of mathematical logic, and even more importantly, he introduced a new concept for applying natural language terms. This new logic for representing and manipulating fuzzy terms was called fuzzy logic, and Zadeh became the Master of fuzzy logic.

Why fuzzy?

As Zadeh said, the term is concrete, immediate and descriptive; we all know what it means. However, many people in the West were repelled by the word fuzzy, because it is usually used in a negative sense.

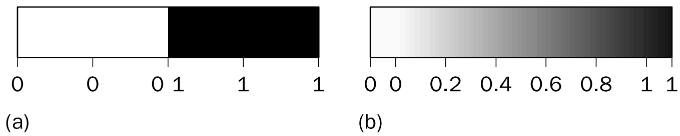


Figure 4.1 Range of logical values in Boolean and fuzzy logic: (a) Boolean logic; (b) multi-valued logic

Why logic?

Fuzziness rests on fuzzy set theory, and fuzzy logic is just a small part of that theory. However, Zadeh used the term fuzzy logic in a broader sense (Zadeh, 1965):

> Fuzzy logic is determined as a set of mathematical principles for knowledge representation based on degrees of membership rather than on crisp membership of classical binary logic.

Unlike two-valued Boolean logic, fuzzy logic is **multi-valued**. It deals with **degrees of membership** and **degrees of truth**. Fuzzy logic uses the continuum of logical values between 0 (completely false) and 1 (completely true). Instead of just black and white, it employs the spectrum of colours, accepting that things can be partly true and partly false at the same time. As can be seen in Figure 4.1, fuzzy logic adds a range of logical values to Boolean logic. Classical binary logic now can be considered as a special case of multi-valued fuzzy logic.

4.2 Fuzzy sets

The concept of a **set** is fundamental to mathematics. However, our own language is the supreme expression of sets. For example, *car* indicates the set of cars. When we say *a car*, we mean one out of the set of cars.

Let X be a classical (**crisp**) set and x an element. Then the element x either belongs to X ($x \in X$) or does not belong to X ($x \notin X$). That is, classical set theory imposes a sharp boundary on this set and gives each member of the set the value of 1, and all members that are not within the set a value of 0. This is known as the **principle of dichotomy**. Let us now dispute this principle.

Consider the following classical paradoxes of logic.

1 Pythagorean School (400 BC):
Question: Does the Cretan philosopher tell the truth when he asserts that 'All Cretans always lie'?
Boolean logic: This assertion contains a contradiction.
Fuzzy logic: The philosopher does and does not tell the truth!

2 Russell's Paradox:
The barber of a village gives a hair cut only to those who do not cut their hair themselves.

Question: Who cuts the barber's hair?
Boolean logic: This assertion contains a contradiction.
Fuzzy logic: The barber cuts and doesn't cut his own hair!

Crisp set theory is governed by a logic that uses one of only two values: true or false. This logic cannot represent vague concepts, and therefore fails to give the answers on the paradoxes. The basic idea of the fuzzy set theory is that an element belongs to a fuzzy set with a certain degree of membership. Thus, a proposition is not either true or false, but may be partly true (or partly false) to any degree. This degree is usually taken as a real number in the interval [0,1].

The classical example in the fuzzy set theory is *tall men*. The elements of the fuzzy set 'tall men' are all men, but their degrees of membership depend on their height, as shown in Table 4.1. Suppose, for example, Mark at 205 cm tall is given a degree of 1, and Peter at 152 cm is given a degree of 0. All men of intermediate height have intermediate degrees. They are partly tall. Obviously, different people may have different views as to whether a given man should be considered as tall. However, our candidates for *tall men* could have the memberships presented in Table 4.1.

It can be seen that the crisp set asks the question, 'Is the man tall?' and draws a line at, say, 180 cm. *Tall men* are above this height and *not tall men* below. In contrast, the fuzzy set asks, 'How tall is the man?' The answer is the partial membership in the fuzzy set, for example, Tom is 0.82 tall.

A fuzzy set is capable of providing a graceful transition across a boundary, as shown in Figure 4.2.

We might consider a few other sets such as 'very short men', 'short men', 'average men' and 'very tall men'.

In Figure 4.2 the horizontal axis represents the **universe of discourse** – the range of all possible values applicable to a chosen variable. In our case, the variable is the human height. According to this representation, the universe of men's heights consists of all tall men. However, there is often room for

Table 4.1 Degree of membership of 'tall men'

		Degree of membership	
Name	**Height, cm**	**Crisp**	**Fuzzy**
Chris	208	1	1.00
Mark	205	1	1.00
John	198	1	0.98
Tom	181	1	0.82
David	179	0	0.78
Mike	172	0	0.24
Bob	167	0	0.15
Steven	158	0	0.06
Bill	155	0	0.01
Peter	152	0	0.00

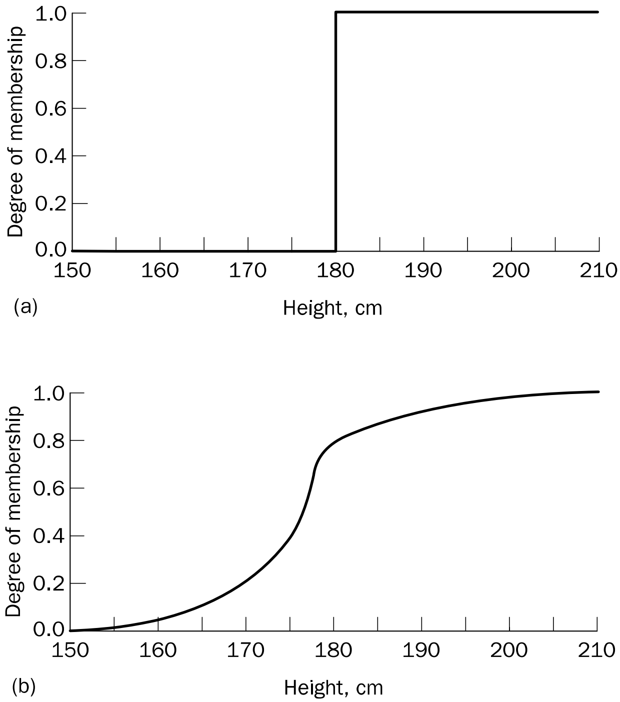


Figure 4.2 Crisp (a) and fuzzy (b) sets of 'tall men'

discretion, since the context of the universe may vary. For example, the set of 'tall men' might be part of the universe of human heights or mammal heights, or even all animal heights.

The vertical axis in Figure 4.2 represents the membership value of the fuzzy set. In our case, the fuzzy set of 'tall men' maps height values into corresponding membership values. As can be seen from Figure 4.2, David who is 179 cm tall, which is just 2 cm less than Tom, no longer suddenly becomes a *not tall* (or *short*) man (as he would in crisp sets). Now David and other men are gradually removed from the set of 'tall men' according to the decrease of their heights.

What is a fuzzy set?

A fuzzy set can be simply defined as a set with fuzzy boundaries.

Let X be the universe of discourse and its elements be denoted as x. In classical set theory, crisp set A of X is defined as function $f_A(x)$ called the **characteristic function** of A

$$f_A(x) : X \to 0, 1, \tag{4.1}$$

where

$$f_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases}$$

This set maps universe X to a set of two elements. For any element x of universe X, characteristic function $f_A(x)$ is equal to 1 if x is an element of set A, and is equal to 0 if x is not an element of A.

In the fuzzy theory, fuzzy set A of universe X is defined by function $\mu_A(x)$ called the **membership function** of set A

$$\mu_A(x) : X \rightarrow [0, 1], \tag{4.2}$$

where

$\mu_A(x) = 1$ if x is totally in A;

$\mu_A(x) = 0$ if x is not in A;

$0 < \mu_A(x) < 1$ if x is partly in A.

This set allows a continuum of possible choices. For any element x of universe X, membership function $\mu_A(x)$ equals the degree to which x is an element of set A. This degree, a value between 0 and 1, represents the **degree of membership**, also called **membership value**, of element x in set A.

How to represent a fuzzy set in a computer?

The membership function must be determined first. A number of methods learned from knowledge acquisition can be applied here. For example, one of the most practical approaches for forming fuzzy sets relies on the knowledge of a single expert. The expert is asked for his or her opinion whether various elements belong to a given set. Another useful approach is to acquire knowledge from multiple experts. A new technique to form fuzzy sets was recently introduced. It is based on artificial neural networks, which learn available system operation data and then derive the fuzzy sets automatically.

Now we return to our 'tall men' example. After acquiring the knowledge for men's heights, we could produce a fuzzy set of *tall men*. In a similar manner, we could obtain fuzzy sets of *short* and *average* men. These sets are shown in Figure 4.3, along with crisp sets. The universe of discourse – the men's heights – consists of three sets: *short*, *average* and *tall men*. In fuzzy logic, as you can see, a man who is 184 cm tall is a member of the *average men* set with a degree of membership of 0.1, and at the same time, he is also a member of the *tall men* set with a degree of 0.4. This means that a man of 184 cm tall has partial membership in multiple sets.

Now assume that universe of discourse X, also called the **reference super set**, is a crisp set containing five elements $X = \{x_1, x_2, x_3, x_4, x_5\}$. Let A be a crisp subset of X and assume that A consists of only two elements, $A = \{x_2, x_3\}$. Subset A can now be described by $A = \{(x_1, 0), (x_2, 1), (x_3, 1), (x_4, 0), (x_5, 0)\}$, i.e. as a set of pairs $\{(x_i, \mu_A(x_i))\}$, where $\mu_A(x_i)$ is the membership function of element x_i in the subset A.

The question is whether $\mu_A(x)$ can take only two values, either 0 or 1, or any value between 0 and 1. It was also the basic question in fuzzy sets examined by Lotfi Zadeh in 1965 (Zadeh, 1965).

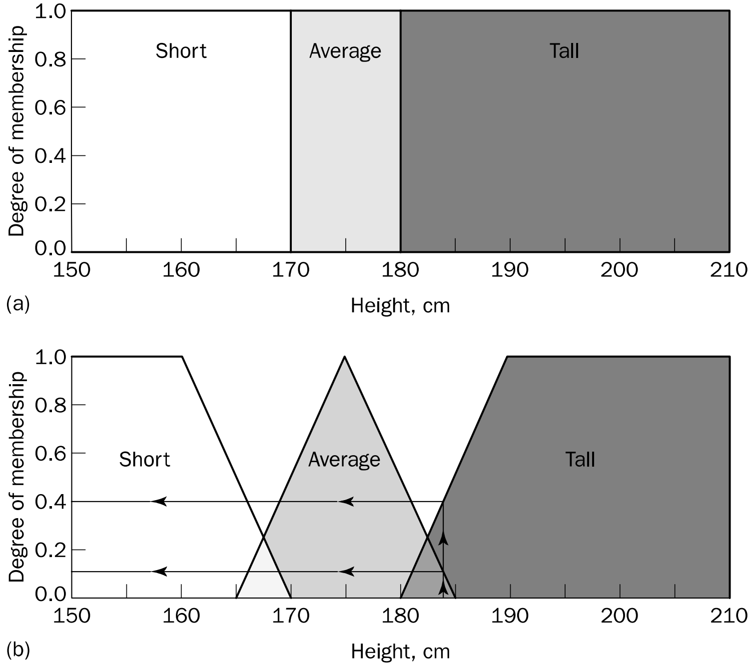


Figure 4.3 Crisp (a) and fuzzy (b) sets of short, average and tall men

If X is the reference super set and A is a subset of X, then A is said to be a fuzzy subset of X if, and only if,

$$A = \{(x, \mu_A(x)\} \qquad x \in X, \mu_A(x) : X \to [0, 1] \tag{4.3}$$

In a special case, when $X \to \{0, 1\}$ is used instead of $X \to [0, 1]$, the fuzzy subset A becomes the crisp subset A.

Fuzzy and crisp sets can be also presented as shown in Figure 4.4.

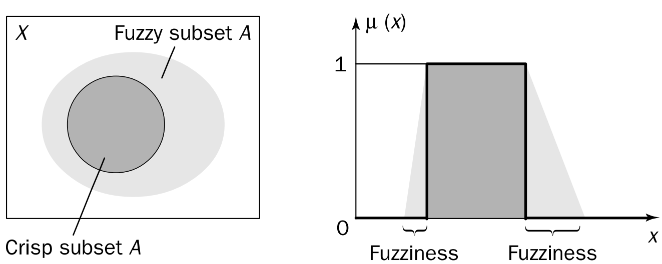


Figure 4.4 Representation of crisp and fuzzy subset of X

Fuzzy subset A of the finite reference super set X can be expressed as,

$$A = \{(x_1, \mu_A(x_1))\}, \{(x_2, \mu_A(x_2))\}, \ldots, \{(x_n, \mu_A(x_n))\} \tag{4.4}$$

However, it is more convenient to represent A as,

$$A = \{\mu_A(x_1)/x_1\}, \{\mu_A(x_2)/x_2\}, \ldots, \{\mu_A(x_n)/x_n\}, \tag{4.5}$$

where the separating symbol / is used to associate the membership value with its coordinate on the horizontal axis.

To represent a continuous fuzzy set in a computer, we need to express it as a function and then to map the elements of the set to their degree of membership. Typical functions that can be used are sigmoid, gaussian and pi. These functions can represent the real data in fuzzy sets, but they also increase the time of computation. Therefore, in practice, most applications use **linear fit functions** similar to those shown in Figure 4.3. For example, the fuzzy set of *tall men* in Figure 4.3 can be represented as a **fit-vector**,

tall men = (0/180, 0.5/185, 1/190) or
tall men = (0/180, 1/190)

Fuzzy sets of *short* and *average men* can be also represented in a similar manner,

short men = (1/160, 0.5/165, 0/170) or
short men = (1/160, 0/170)

average men = (0/165, 1/175, 0/185)

4.3 Linguistic variables and hedges

At the root of fuzzy set theory lies the idea of linguistic variables. A linguistic variable is a fuzzy variable. For example, the statement 'John is tall' implies that the linguistic variable *John* takes the linguistic value *tall*. In fuzzy expert systems, linguistic variables are used in fuzzy rules. For example,

IF wind is strong
THEN sailing is good

IF project_duration is long
THEN completion_risk is high

IF speed is slow
THEN stopping_distance is short

The range of possible values of a linguistic variable represents the universe of discourse of that variable. For example, the universe of discourse of the linguistic variable *speed* might have the range between 0 and 220 km per hour and may include such fuzzy subsets as *very slow*, *slow*, *medium*, *fast*, and *very fast*. Each fuzzy subset also represents a linguistic value of the corresponding linguistic variable.

A linguistic variable carries with it the concept of fuzzy set qualifiers, called **hedges**. Hedges are terms that modify the shape of fuzzy sets. They include adverbs such as *very*, *somewhat*, *quite*, *more or less* and *slightly*. Hedges can modify verbs, adjectives, adverbs or even whole sentences. They are used as

- All-purpose modifiers, such as *very*, *quite* or *extremely*.
- Truth-values, such as *quite true* or *mostly false*.
- Probabilities, such as *likely* or *not very likely*.
- Quantifiers, such as *most*, *several* or *few*.
- Possibilities, such as *almost impossible* or *quite possible*.

Hedges act as operations themselves. For instance, *very* performs concentration and creates a new subset. From the set of *tall men*, it derives the subset of *very tall men*. *Extremely* serves the same purpose to a greater extent.

An operation opposite to concentration is dilation. It expands the set. *More or less* performs dilation; for example, the set of *more or less tall men* is broader than the set of *tall men*.

Hedges are useful as operations, but they can also break down continuums into fuzzy intervals. For example, the following hedges could be used to describe temperature: *very cold*, *moderately cold*, *slightly cold*, *neutral*, *slightly hot*, *moderately hot* and *very hot*. Obviously these fuzzy sets overlap. Hedges help to reflect human thinking, since people usually cannot distinguish between *slightly hot* and *moderately hot*.

Figure 4.5 illustrates an application of hedges. The fuzzy sets shown previously in Figure 4.3 are now modified mathematically by the hedge *very*. Consider, for example, a man who is 185 cm tall. He is a member of the *tall men* set with a degree of membership of 0.5. However, he is also a member of the set of *very tall men* with a degree of 0.15, which is *fairly* reasonable.

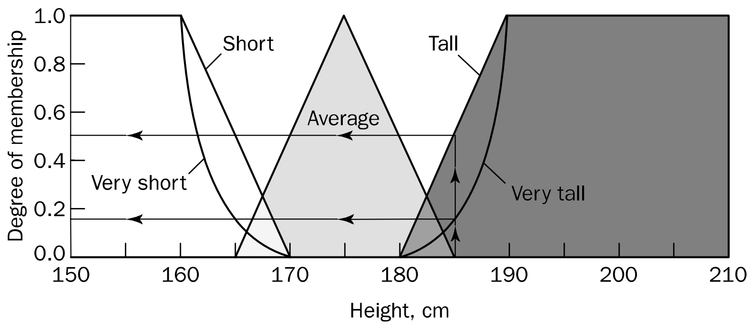


Figure 4.5 Fuzzy sets with *very* hedge

Let us now consider the hedges often used in practical applications.

- *Very*, the operation of concentration, as we mentioned above, narrows a set down and thus reduces the degree of membership of fuzzy elements. This operation can be given as a mathematical square:

$$\mu_A^{very}(x) = [\mu_A(x)]^2 \tag{4.6}$$

Hence, if Tom has a 0.86 membership in the set of *tall men*, he will have a 0.7396 membership in the set of *very tall men*.

- *Extremely* serves the same purpose as *very*, but does it to a greater extent. This operation can be performed by raising $\mu_A(x)$ to the third power:

$$\mu_A^{extremely}(x) = [\mu_A(x)]^3 \tag{4.7}$$

If Tom has a 0.86 membership in the set of *tall men*, he will have a 0.7396 membership in the set of *very tall men* and 0.6361 membership in the set of *extremely tall men*.

- *Very very* is just an extension of concentration. It can be given as a square of the operation of concentration:

$$\mu_A^{very\,very}(x) = [\mu_A^{very}(x)]^2 = [\mu_A(x)]^4 \tag{4.8}$$

For example, Tom, with a 0.86 membership in the *tall men set* and a 0.7396 membership in the *very tall men set*, will have a membership of 0.5470 in the set of *very very tall men*.

- *More or less*, the operation of dilation, expands a set and thus increases the degree of membership of fuzzy elements. This operation is presented as:

$$\mu_A^{more\,or\,less}(x) = \sqrt{\mu_A(x)} \tag{4.9}$$

Hence, if Tom has a 0.86 membership in the set of *tall men*, he will have a 0.9274 membership in the set of *more or less tall men*.

- *Indeed*, the operation of intensification, intensifies the meaning of the whole sentence. It can be done by increasing the degree of membership above 0.5 and decreasing those below 0.5. The hedge *indeed* may be given by either:

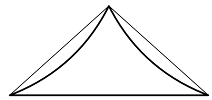
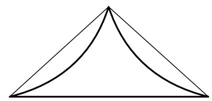
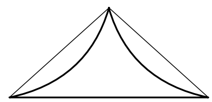
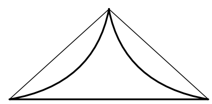
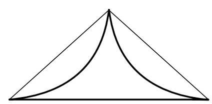
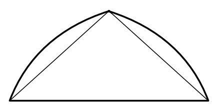
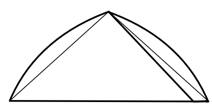
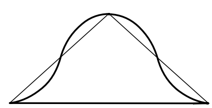
$$\mu_A^{indeed}(x) = 2[\mu_A(x)]^2 \qquad \text{if } 0 \leqslant \mu_A(x) \leqslant 0.5 \tag{4.10}$$

or

$$\mu_A^{indeed}(x) = 1 - 2[1 - \mu_A(x)]^2 \qquad \text{if } 0.5 < \mu_A(x) \leqslant 1 \tag{4.11}$$

If Tom has a 0.86 membership in the set of *tall men*, he can have a 0.9608 membership in the set of *indeed tall men*. In contrast, Mike, who has a 0.24 membership in *tall men* set, will have a 0.1152 membership in the *indeed tall men* set.

Table 4.2 Representation of hedges in fuzzy logic

Hedge	Mathematical expression	Graphical representation
A little	$[\mu_A(x)]^{1.3}$	
Slightly	$[\mu_A(x)]^{1.7}$	
Very	$[\mu_A(x)]^{2}$	
Extremely	$[\mu_A(x)]^{3}$	
Very very	$[\mu_A(x)]^{4}$	
More or less	$\sqrt{\mu_A(x)}$	
Somewhat	$\sqrt{\mu_A(x)}$	
Indeed	$2[\mu_A(x)]^2$ if $0 \leqslant \mu_A \leqslant 0.5$ $1 - 2[1 - \mu_A(x)]^2$ if $0.5 < \mu_A \leqslant 1$	

Mathematical and graphical representation of hedges are summarised in Table 4.2.

4.4 Operations of fuzzy sets

The classical set theory developed in the late 19th century by Georg Cantor describes how crisp sets can interact. These interactions are called **operations**.

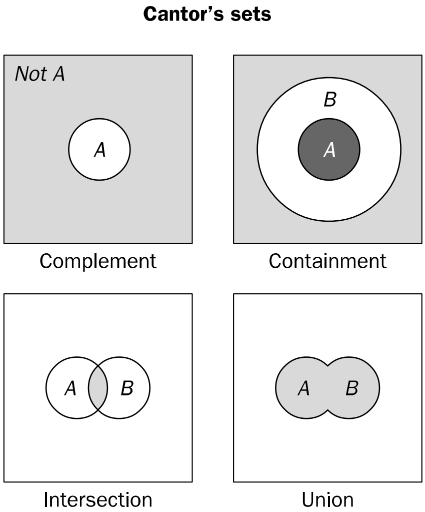


Figure 4.6 Operations on classical sets

We look at four of them: complement, containment, intersection and union. These operations are presented graphically in Figure 4.6. Let us compare operations of classical and fuzzy sets.

Complement

- Crisp sets: Who does not belong to the set?
- Fuzzy sets: How much do elements not belong to the set?

The complement of a set is an opposite of this set. For example, if we have the set of *tall men*, its complement is the set of *NOT tall men*. When we remove the tall men set from the universe of discourse, we obtain the complement. If A is the fuzzy set, its complement $\neg A$ can be found as follows:

$$\mu_{\neg A}(x) = 1 - \mu_A(x) \tag{4.12}$$

For example, if we have a fuzzy set of *tall men*, we can easily obtain the fuzzy set of *NOT tall men*:

$$\begin{aligned} tall\ men &= (0/180, 0.25/182.5, 0.5/185, 0.75/187.5, 1/190) \\ NOT\ \ tall\ men &= (1/180, 0.75/182.5, 0.5/185, 0.25/187.5, 0/190) \end{aligned}$$

Containment

- Crisp sets: Which sets belong to which other sets?
- Fuzzy sets: Which sets belong to other sets?

Similar to a Chinese box or Russian doll, a set can contain other sets. The smaller set is called the **subset**. For example, the set of *tall men* contains all tall men. Therefore, *very tall men* is a subset of *tall men*. However, the *tall men* set is just a subset of the set of *men*. In crisp sets, all elements of a subset entirely belong to a larger set and their membership values are equal to 1. In fuzzy sets, however, each element can belong less to the subset than to the larger set. Elements of the fuzzy subset have smaller memberships in it than in the larger set.

$$\begin{aligned} \textit{tall men} &= (0/180, 0.25/182.5, 0.50/185, 0.75/187.5, 1/190) \\ \textit{very tall men} &= (0/180, 0.06/182.5, 0.25/185, 0.56/187.5, 1/190) \end{aligned}$$

Intersection

- Crisp sets: Which element belongs to both sets?
- Fuzzy sets: How much of the element is in both sets?

In classical set theory, an intersection between two sets contains the elements shared by these sets. If we have, for example, the set of *tall men* and the set of *fat men*, the intersection is the area where these sets overlap, i.e. Tom is in the intersection only if he is tall AND fat. In fuzzy sets, however, an element may partly belong to both sets with different memberships. Thus, a fuzzy intersection is the lower membership in both sets of each element.

The fuzzy operation for creating the intersection of two fuzzy sets A and B on universe of discourse X can be obtained as:

$$\mu_{A\cap B}(x) = min\,[\mu_A(x), \mu_B(x)] = \mu_A(x) \cap \mu_B(x), \quad \text{where } x \in X \tag{4.13}$$

Consider, for example, the fuzzy sets of *tall* and *average men*:

$$\begin{aligned} \textit{tall men} &= (0/165, 0/175, 0.0/180, 0.25/182.5, 0.5/185, 1/190) \\ \textit{average men} &= (0/165, 1/175, 0.5/180, 0.25/182.5, 0.0/185, 0/190) \end{aligned}$$

According to Eq. (4.13), the intersection of these two sets is

$$\textit{tall men} \cap \textit{average men} = (0/165, 0/175, 0/180, 0.25/182.5, 0/185, 0/190)$$

or

$$\textit{tall men} \cap \textit{average men} = (0/180, 0.25/182.5, 0/185)$$

This solution is represented graphically in Figure 4.3.

Union

- Crisp sets: Which element belongs to either set?
- Fuzzy sets: How much of the element is in either set?

The union of two crisp sets consists of every element that falls into either set. For example, the union of *tall men* and *fat men* contains all men who are tall OR fat, i.e. Tom is in the union since he is tall, and it does not matter whether he is fat or not. In fuzzy sets, the union is the reverse of the intersection. That is, the union is the largest membership value of the element in either set.

The fuzzy operation for forming the union of two fuzzy sets A and B on universe X can be given as:

$$\mu_{A\cup B}(x) = max\,[\mu_A(x), \mu_B(x)] = \mu_A(x) \cup \mu_B(x), \quad \text{where } x \in X \tag{4.14}$$

Consider again the fuzzy sets of *tall* and *average* men:

$$\begin{aligned} \textit{tall men} &= (0/165, 0/175, 0.0/180, 0.25/182.5, 0.5/185, 1/190) \\ \textit{average men} &= (0/165, 1/175, 0.5/180, 0.25/182.5, 0.0/185, 0/190) \end{aligned}$$

According to Eq. (4.14), the union of these two sets is

$$\textit{tall men} \cup \textit{average men} = (0/165, 1/175, 0.5/180, 0.25/182.5, 0.5/185, 1/190)$$

Diagrams for fuzzy set operations are shown in Figure 4.7.

Crisp and fuzzy sets have the same properties; crisp sets can be considered as just a special case of fuzzy sets. Frequently used properties of fuzzy sets are described below.

Commutativity

$$\begin{aligned} A \cup B &= B \cup A \\ A \cap B &= B \cap A \end{aligned}$$

Example:

tall men OR *short men* = *short men* OR *tall men*
tall men AND *short men* = *short men* AND *tall men*

Associativity

$$\begin{aligned} A \cup (B \cup C) &= (A \cup B) \cup C \\ A \cap (B \cap C) &= (A \cap B) \cap C \end{aligned}$$

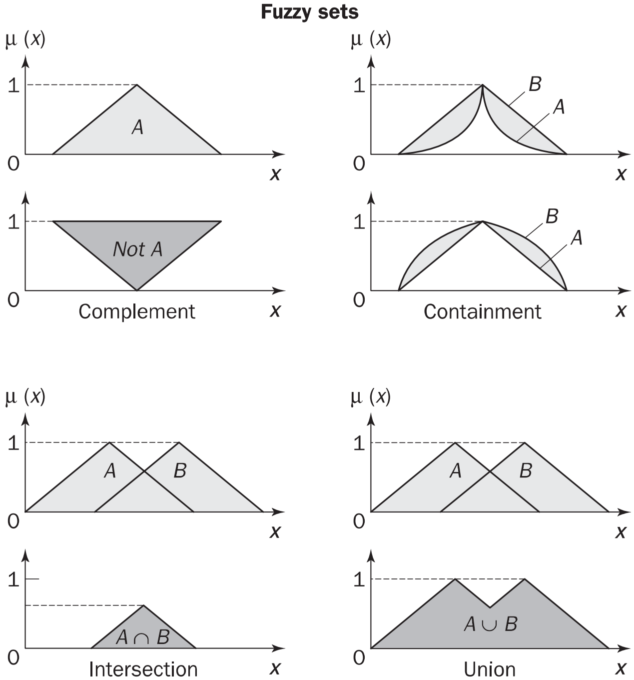


Figure 4.7 Operations of fuzzy sets

Example:

tall men OR (*short men* OR *average men*) = (*tall men* OR *short men*) OR *average men*
tall men AND (*short men* AND *average men*) = (*tall men* AND *short men*) AND *average men*

Distributivity

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

Example:

tall men OR (*short men* AND *average men*) = (*tall men* OR *short men*) AND (*tall men* OR *average men*)
tall men AND (*short men* OR *average men*) = (*tall men* AND *short men*) OR (*tall men* AND *average men*)

Idempotency

$$A \cup A = A$$
$$A \cap A = A$$

Example:

tall men OR *tall men* = *tall men*
tall men AND *tall men* = *tall men*

Identity

$A \cup \emptyset = A$
$A \cap X = A$
$A \cap \emptyset = \emptyset$
$A \cup X = X$

Example:

tall men OR *undefined* = *tall men*
tall men AND *unknown* = *tall men*
tall men AND *undefined* = *undefined*
tall men OR *unknown* = *unknown*

where *undefined* is an empty (**null**) set, the set having all degree of memberships equal to 0, and *unknown* is a set having all degree of memberships equal to 1.

Involution

$\neg(\neg A) = A$

Example:

NOT (NOT *tall men*) = *tall men*

Transitivity

If $(A \subset B) \cap (B \subset C)$ then $A \subset C$
Every set contains the subsets of its subsets.

Example:

IF (*extremely tall men* ⊂ *very tall men*) AND (*very tall men* ⊂ *tall men*)
THEN (*extremely tall men* ⊂ *tall men*)

De Morgan's Laws

$\neg(A \cap B) = \neg A \cup \neg B$
$\neg(A \cup B) = \neg A \cap \neg B$

Example:

NOT (*tall men* AND *short men*) = NOT *tall men* OR NOT *short men*
NOT (*tall men* OR *short men*) = NOT *tall men* AND NOT *short men*

Using fuzzy set operations, their properties and hedges, we can easily obtain a variety of fuzzy sets from the existing ones. For example, if we have fuzzy set A of *tall men* and fuzzy set B of *short men*, we can derive fuzzy set C of *not very tall men and not very short men* or even set D of *not very very tall and not very very short men* from the following operations:

$$\mu_C(x) = [1 - \mu_A(x)^2] \cap [1 - (\mu_B(x)^2]$$
$$\mu_D(x) = [1 - \mu_A(x)^4] \cap [1 - (\mu_B(x)^4]$$

Generally, we apply fuzzy operations and hedges to obtain fuzzy sets which can represent linguistic descriptions of our natural language.

4.5 Fuzzy rules

In 1973, Lotfi Zadeh published his second most influential paper (Zadeh, 1973). This paper outlined a new approach to analysis of complex systems, in which Zadeh suggested capturing human knowledge in fuzzy rules.

What is a fuzzy rule?

A fuzzy rule can be defined as a conditional statement in the form:

IF x is A
THEN y is B

where x and y are linguistic variables; and A and B are linguistic values determined by fuzzy sets on the universe of discourses X and Y, respectively.

What is the difference between classical and fuzzy rules?

A classical IF-THEN rule uses binary logic, for example,

Rule: 1
IF speed is > 100
THEN stopping_distance is long

Rule: 2
IF speed is < 40
THEN stopping_distance is short

The variable speed can have any numerical value between 0 and 220 km/h, but the linguistic variable *stopping_distance* can take either value *long* or *short*. In other words, classical rules are expressed in the black-and-white language of Boolean logic. However, we can also represent the stopping distance rules in a fuzzy form:

Rule: 1
IF speed is fast
THEN stopping_distance is long

Rule: 2
IF speed is slow
THEN stopping_distance is short

Here the linguistic variable *speed* also has the range (the universe of discourse) between 0 and 220 km/h, but this range includes fuzzy sets, such as *slow*, *medium* and *fast*. The universe of discourse of the linguistic variable *stopping_distance* can be between 0 and 300 m and may include such fuzzy sets as *short*, *medium* and *long*. Thus fuzzy rules relate to fuzzy sets.

Fuzzy expert systems merge the rules and consequently cut the number of rules by at least 90 per cent.

How to reason with fuzzy rules?

Fuzzy reasoning includes two distinct parts: evaluating the rule antecedent (the IF part of the rule) and *implication* or applying the result to the consequent (the THEN part of the rule).

In classical rule-based systems, if the rule antecedent is true, then the consequent is also true. In fuzzy systems, where the antecedent is a fuzzy statement, all rules fire to some extent, or in other words they fire partially. If the antecedent is true to some degree of membership, then the consequent is also true to that same degree.

Consider, for example, two fuzzy sets, 'tall men' and 'heavy men' represented in Figure 4.8.

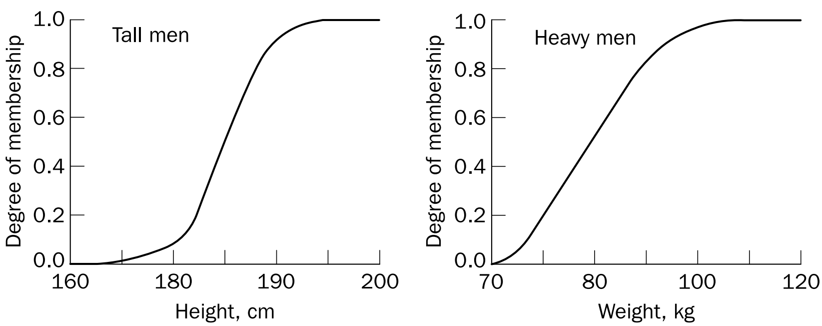


Figure 4.8 Fuzzy sets of *tall* and *heavy* men

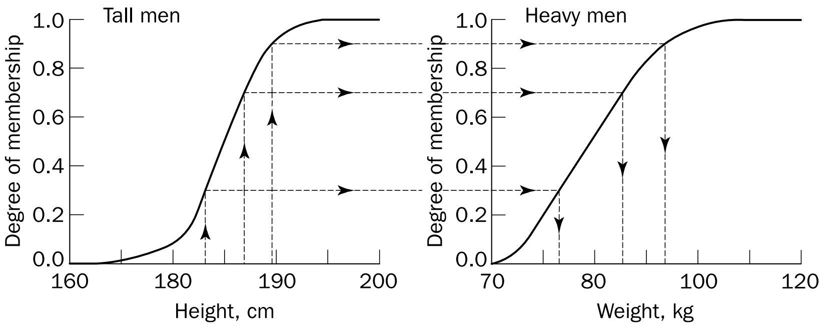


Figure 4.9 Monotonic selection of values for man weight

These fuzzy sets provide the basis for a weight estimation model. The model is based on a relationship between a man's height and his weight, which is expressed as a single fuzzy rule:

IF height is *tall*
THEN weight is *heavy*

The value of the output or a truth membership grade of the rule consequent can be estimated directly from a corresponding truth membership grade in the antecedent (Cox, 1999). This form of fuzzy inference uses a method called **monotonic selection**. Figure 4.9 shows how various values of men's weight are derived from different values for men's height.

Can the antecedent of a fuzzy rule have multiple parts?

As a production rule, a fuzzy rule can have multiple antecedents, for example:

IF project_duration is long
AND project_staffing is large
AND project_funding is inadequate
THEN risk is high

IF service is excellent
OR food is delicious
THEN tip is generous

All parts of the antecedent are calculated simultaneously and resolved in a single number, using fuzzy set operations considered in the previous section.

Can the consequent of a fuzzy rule have multiple parts?

The consequent of a fuzzy rule can also include multiple parts, for instance:

IF temperature is hot
THEN hot_water is reduced;
cold_water is increased

In this case, all parts of the consequent are affected equally by the antecedent.

In general, a fuzzy expert system incorporates not one but several rules that describe expert knowledge and play off one another. The output of each rule is a fuzzy set, but usually we need to obtain a single number representing the expert system output. In other words, we want to get a precise solution, not a fuzzy one.

How are all these output fuzzy sets combined and transformed into a single number?

To obtain a single crisp solution for the output variable, a fuzzy expert system first aggregates all output fuzzy sets into a single output fuzzy set, and then defuzzifies the resulting fuzzy set into a single number. In the next section we will see how the whole process works from the beginning to the end.

4.6 Fuzzy inference

Fuzzy inference can be defined as a process of mapping from a given input to an output, using the theory of fuzzy sets.

4.6.1 Mamdani-style inference

The most commonly used fuzzy inference technique is the so-called Mamdani method. In 1975, Professor Ebrahim Mamdani of London University built one of the first fuzzy systems to control a steam engine and boiler combination (Mamdani and Assilian, 1975). He applied a set of fuzzy rules supplied by experienced human operators.

The Mamdani-style fuzzy inference process is performed in four steps: fuzzification of the input variables, rule evaluation, aggregation of the rule outputs, and finally defuzzification.

To see how everything fits together, we examine a simple two-input one-output problem that includes three rules:

Rule: 1		Rule: 1	
IF	x is $A3$	IF	*project_funding* is *adequate*
OR	y is $B1$	OR	*project_staffing* is *small*
THEN	z is $C1$	THEN	*risk* is *low*
Rule: 2		Rule: 2	
IF	x is $A2$	IF	*project_funding* is *marginal*
AND	y is $B2$	AND	*project_staffing* is *large*
THEN	z is $C2$	THEN	*risk* is *normal*
Rule: 3		Rule: 3	
IF	x is $A1$	IF	*project_funding* is *inadequate*
THEN	z is $C3$	THEN	*risk* is *high*

where x, y and z (*project funding*, *project staffing* and *risk*) are linguistic variables; $A1$, $A2$ and $A3$ (*inadequate*, *marginal* and *adequate*) are linguistic values determined by fuzzy sets on universe of discourse X (*project funding*); $B1$ and $B2$ (*small* and *large*) are linguistic values determined by fuzzy sets on universe of discourse Y (*project staffing*); $C1$, $C2$ and $C3$ (*low*, *normal* and *high*) are linguistic values determined by fuzzy sets on universe of discourse Z (*risk*).

The basic structure of Mamdani-style fuzzy inference for our problem is shown in Figure 4.10.

Step 1: ***Fuzzification***

The first step is to take the crisp inputs, $x1$ and $y1$ (*project funding* and *project staffing*), and determine the degree to which these inputs belong to each of the appropriate fuzzy sets.

What is a crisp input and how is it determined?

The crisp input is always a numerical value limited to the universe of discourse. In our case, values of $x1$ and $y1$ are limited to the universe of discourses X and Y, respectively. The ranges of the universe of discourses can be determined by expert judgements. For instance, if we need to examine the risk involved in developing the 'fuzzy' project, we can ask the expert to give numbers between 0 and 100 per cent that represent the project funding and the project staffing, respectively. In other words, the expert is required to answer to what extent the project funding and the project staffing are really adequate. Of course, various fuzzy systems use a variety of different crisp inputs. While some of the inputs can be measured directly (height, weight, speed, distance, temperature, pressure etc.), some of them can be based only on expert estimate.

Once the crisp inputs, $x1$ and $y1$, are obtained, they are fuzzified against the appropriate linguistic fuzzy sets. The crisp input $x1$ (project funding rated by the expert as 35 per cent) corresponds to the membership functions $A1$ and $A2$ (*inadequate* and *marginal*) to the degrees of 0.5 and 0.2, respectively, and the crisp input $y1$ (project staffing rated as 60 per cent) maps the membership functions $B1$ and $B2$ (*small* and *large*) to the degrees of 0.1 and 0.7, respectively. In this manner, each input is fuzzified over all the membership functions used by the fuzzy rules.

Step 2: ***Rule evaluation***

The second step is to take the fuzzified inputs, $\mu_{(x=A1)} = 0.5$, $\mu_{(x=A2)} = 0.2$, $\mu_{(y=B1)} = 0.1$ and $\mu_{(y=B2)} = 0.7$, and apply them to the antecedents of the fuzzy rules. If a given fuzzy rule has multiple antecedents, the fuzzy operator (AND or OR) is used to obtain a single number that represents the result of the antecedent evaluation. This number (the truth value) is then applied to the consequent membership function.

To evaluate the disjunction of the rule antecedents, we use the OR fuzzy operation. Typically, fuzzy expert systems make use of the classical fuzzy operation *union* (4.14) shown in Figure 4.10 (Rule 1):

$$\mu_{A \cup B}(x) = max\,[\mu_A(x), \mu_B(x)]$$

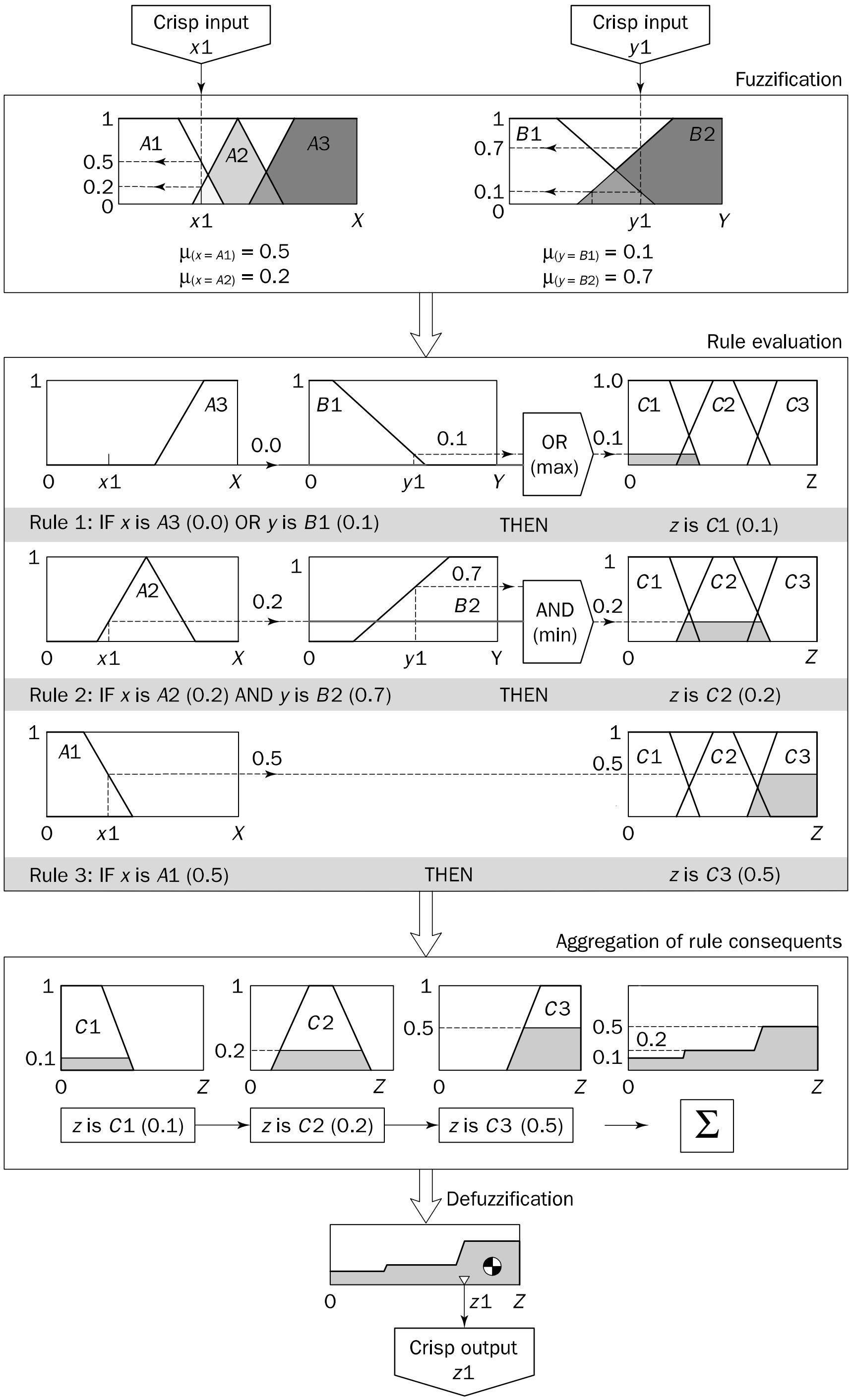


Figure 4.10 The basic structure of Mamdani-style fuzzy inference

However, the OR operation can be easily customised if necessary. For example, the MATLAB Fuzzy Logic Toolbox has two built-in OR methods: *max* and the probabilistic OR method, *probor*. The probabilistic OR, also known as the **algebraic sum**, is calculated as:

$$\mu_{A \cup B}(x) = probor\,[\mu_A(x), \mu_B(x)] = \mu_A(x) + \mu_B(x) - \mu_A(x) \times \mu_B(x) \quad (4.15)$$

Similarly, in order to evaluate the conjunction of the rule antecedents, we apply the AND fuzzy operation *intersection* (4.13) also shown in Figure 4.10 (Rule 2):

$$\mu_{A \cap B}(x) = min\,[\mu_A(x), \mu_B(x)]$$

The Fuzzy Logic Toolbox also supports two AND methods: *min* and the product, *prod*. The product is calculated as:

$$\mu_{A \cap B}(x) = prod\,[\mu_A(x), \mu_B(x)] = \mu_A(x) \times \mu_B(x) \quad (4.16)$$

Do different methods of the fuzzy operations produce different results?

Fuzzy researchers have proposed and applied several approaches to execute AND and OR fuzzy operators (Cox, 1999) and, of course, different methods may lead to different results. Most fuzzy packages also allow us to customise the AND and OR fuzzy operations and a user is required to make the choice.

Let us examine our rules again.

Rule: 1
IF x is $A3$ (0.0)
OR y is $B1$ (0.1)
THEN z is $C1$ (0.1)

$$\mu_{C1}(z) = max\,[\mu_{A3}(x), \mu_{B1}(y)] = max\,[0.0, 0.1] = 0.1$$

or

$$\mu_{C1}(z) = probor\,[\mu_{A3}(x), \mu_{B1}(y)] = 0.0 + 0.1 - 0.0 \times 0.1 = 0.1$$

Rule: 2
IF x is $A2$ (0.2)
AND y is $B2$ (0.7)
THEN z is $C2$ (0.2)

$$\mu_{C2}(z) = min\,[\mu_{A2}(x), \mu_{B2}(y)] = min\,[0.2, 0.7] = 0.2$$

or

$$\mu_{C2}(z) = prod\,[\mu_{A2}(x), \mu_{B2}(y)] = 0.2 \times 0.7 = 0.14$$

Thus, Rule 2 can be also represented as shown in Figure 4.11.

Now the result of the antecedent evaluation can be applied to the membership function of the consequent. In other words, the consequent membership function is **clipped** or **scaled** to the level of the truth value of the rule antecedent.

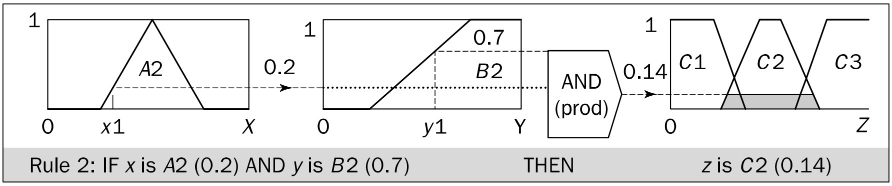


Figure 4.11 The AND *product* fuzzy operation

What do we mean by 'clipped or scaled'?

The most common method of correlating the rule consequent with the truth value of the rule antecedent is to simply cut the consequent membership function at the level of the antecedent truth. This method is called **clipping** or **correlation minimum**. Since the top of the membership function is sliced, the clipped fuzzy set loses some information. However, clipping is still often preferred because it involves less complex and faster mathematics, and generates an aggregated output surface that is easier to defuzzify.

While clipping is a frequently used method, **scaling** or **correlation product** offers a better approach for preserving the original shape of the fuzzy set. The original membership function of the rule consequent is adjusted by multiplying all its membership degrees by the truth value of the rule antecedent. This method, which generally loses less information, can be very useful in fuzzy expert systems.

Clipped and scaled membership functions are illustrated in Figure 4.12.

Step 3: ***Aggregation of the rule outputs***

Aggregation is the process of unification of the outputs of all rules. In other words, we take the membership functions of all rule consequents previously clipped or scaled and combine them into a single fuzzy set. Thus, the input of the aggregation process is the list of clipped or scaled consequent membership functions, and the output is one fuzzy set for each output variable. Figure 4.10 shows how the output of each rule is aggregated into a single fuzzy set for the overall fuzzy output.

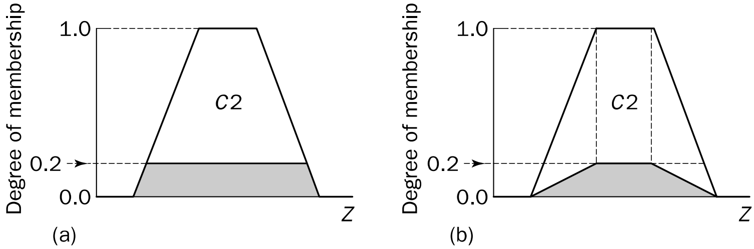


Figure 4.12 Clipped (a) and scaled (b) membership functions

Step 4: ***Defuzzification***

The last step in the fuzzy inference process is defuzzification. Fuzziness helps us to evaluate the rules, but the final output of a fuzzy system has to be a crisp number. The input for the defuzzification process is the aggregate output fuzzy set and the output is a single number.

How do we defuzzify the aggregate fuzzy set?

There are several defuzzification methods (Cox, 1999), but probably the most popular one is the **centroid technique**. It finds the point where a vertical line would slice the aggregate set into two equal masses. Mathematically this **centre of gravity** (COG) can be expressed as

$$\text{COG} = \frac{\displaystyle\int_a^b \mu_A(x)x\,dx}{\displaystyle\int_a^b \mu_A(x)\,dx} \tag{4.17}$$

As Figure 4.13 shows, a centroid defuzzification method finds a point representing the centre of gravity of the fuzzy set, *A*, on the interval, *ab*.

In theory, the COG is calculated over a continuum of points in the aggregate output membership function, but in practice, a reasonable estimate can be obtained by calculating it over a sample of points, as shown in Figure 4.13. In this case, the following formula is applied:

$$\text{COG} = \frac{\displaystyle\sum_{x=a}^{b} \mu_A(x)x}{\displaystyle\sum_{x=a}^{b} \mu_A(x)} \tag{4.18}$$

Let us now calculate the centre of gravity for our problem. The solution is presented in Figure 4.14.

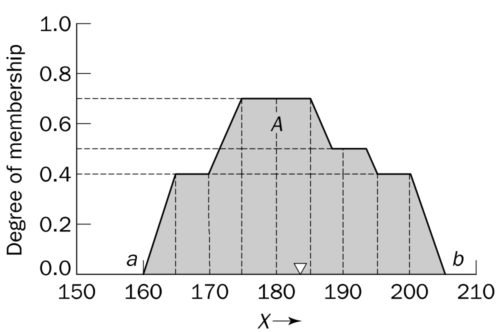


Figure 4.13 The centroid method of defuzzification

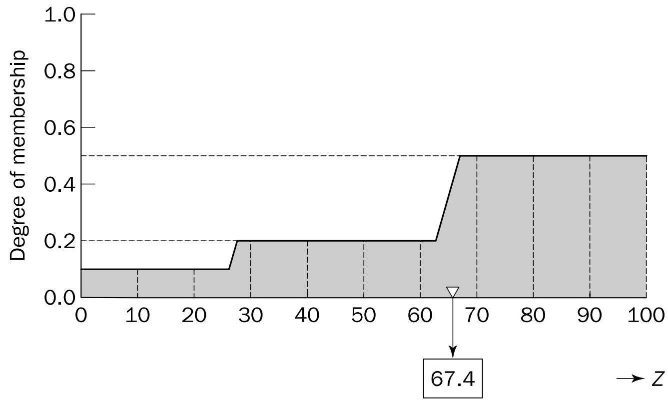


Figure 4.14 Defuzzifying the solution variable's fuzzy set

$$\begin{aligned}\text{COG} &= \frac{(0+10+20)\times 0.1+(30+40+50+60)\times 0.2+(70+80+90+100)\times 0.5}{0.1+0.1+0.1+0.2+0.2+0.2+0.2+0.5+0.5+0.5+0.5}\\ &= 67.4\end{aligned}$$

Thus, the result of defuzzification, crisp output $z1$, is 67.4. It means, for instance, that the risk involved in our 'fuzzy' project is 67.4 per cent.

4.6.2 Sugeno-style inference

Mamdani-style inference, as we have just seen, requires us to find the centroid of a two-dimensional shape by integrating across a continuously varying function. In general, this process is not computationally efficient.

Could we shorten the time of fuzzy inference?

We can use a single spike, a **singleton**, as the membership function of the rule consequent. This method was first introduced by Michio Sugeno, the 'Zadeh of Japan', in 1985 (Sugeno, 1985). A singleton, or more precisely a fuzzy singleton, is a fuzzy set with a membership function that is unity at a single particular point on the universe of discourse and zero everywhere else.

Sugeno-style fuzzy inference is very similar to the Mamdani method. Sugeno changed only a rule consequent. Instead of a fuzzy set, he used a mathematical function of the input variable. The format of the Sugeno-style fuzzy rule is

IF x is A
AND y is B
THEN z is $f(x, y)$

where x, y and z are linguistic variables; A and B are fuzzy sets on universe of discourses X and Y, respectively; and $f(x, y)$ is a mathematical function.

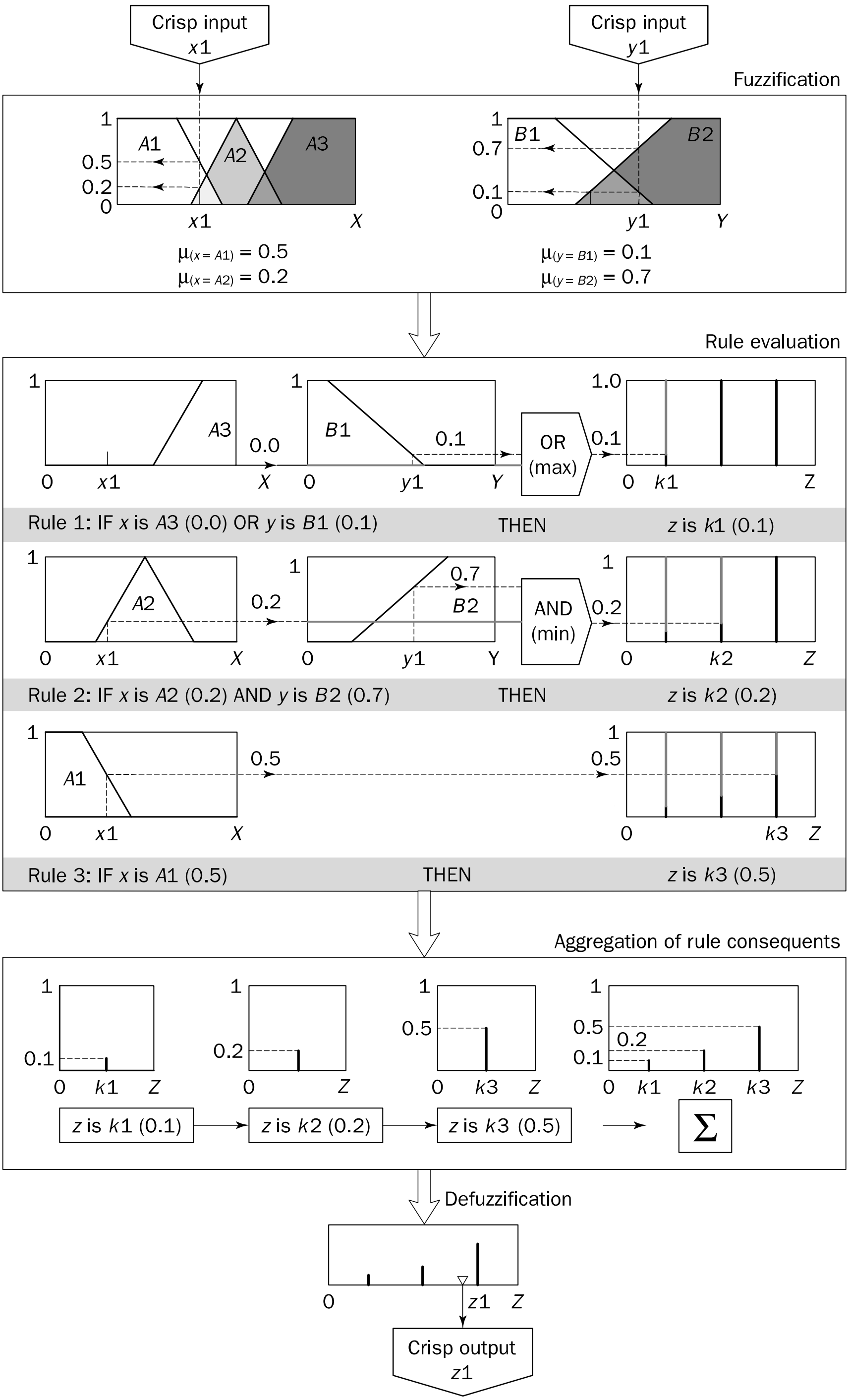


Figure 4.15 The basic structure of Sugeno-style fuzzy inference

The most commonly used **zero-order Sugeno fuzzy model** applies fuzzy rules in the following form:

IF x is A
AND y is B
THEN z is k

where k is a constant.

In this case, the output of each fuzzy rule is constant. In other words, all consequent membership functions are represented by singleton spikes. Figure 4.15 shows the fuzzy inference process for a zero-order Sugeno model. Let us compare Figure 4.15 with Figure 4.10. The similarity of Sugeno and Mamdani methods is quite noticeable. The only distinction is that rule consequents are singletons in Sugeno's method.

How is the result, crisp output, obtained?

As you can see from Figure 4.15, the aggregation operation simply includes all the singletons. Now we can find the **weighted average** (WA) of these singletons:

$$\mathrm{WA} = \frac{\mu(k1) \times k1 + \mu(k2) \times k2 + \mu(k3) \times k3}{\mu(k1) + \mu(k2) + \mu(k3)} = \frac{0.1 \times 20 + 0.2 \times 50 + 0.5 \times 80}{0.1 + 0.2 + 0.5} = 65$$

Thus, a zero-order Sugeno system might be sufficient for our problem's needs. Fortunately, singleton output functions satisfy the requirements of a given problem quite often.

How do we make a decision on which method to apply – Mamdani or Sugeno?

The Mamdani method is widely accepted for capturing expert knowledge. It allows us to describe the expertise in more intuitive, more human-like manner. However, Mamdani-type fuzzy inference entails a substantial computational burden. On the other hand, the Sugeno method is computationally effective and works well with optimisation and adaptive techniques, which makes it very attractive in control problems, particularly for dynamic nonlinear systems.

4.7 Building a fuzzy expert system

To illustrate the design of a fuzzy expert system, we will consider a problem of operating a service centre of spare parts (Turksen *et al.*, 1992).

A service centre keeps spare parts and repairs failed ones. A customer brings a failed item and receives a spare of the same type. Failed parts are repaired, placed on the shelf, and thus become spares. If the required spare is available on the shelf, the customer takes it and leaves the service centre. However, if there is no spare on the shelf, the customer has to wait until the needed item becomes available. The objective here is to advise a manager of the service centre on certain decision policies to keep the customers satisfied.