Part VI

FUZZY SYSTEMS

Two-valued, or Boolean, logic is a well-defined and used theory. Boolean logic is especially important for implementation in computing systems where information, or knowledge about a problem, is binary encoded. Boolean logic also played an important role in the development of the first AI reasoning systems, especially the inference engine of expert systems [315]. For such knowledge representation and reasoning systems, propositional and first-order predicate calculus are extensively used as representation language [539, 629]. Associated with Boolean logic is the traditional two-valued set theory, where an element either belongs to a set or not. That is, set membership is precise. Coupled with Boolean knowledge, two-valued set theory enabled the development of exact reasoning systems.

While some successes have been achieved using two-valued logic and sets, it is not possible to solve all problems by mapping the domain into two-valued variables. Most real-world problems are characterized by the ability of a representation language (or logic) to process incomplete, imprecise, vague or uncertain information. While two-valued logic and set theory fail in such environments, fuzzy logic and fuzzy sets give the formal tools to reason about such uncertain information. With fuzzy logic, domains are characterized by linguistic terms, rather than by numbers. For example, in the phrases *''it is partly cloudy''*, or *''Stephan is very tall''*, both *partly* and *very* are linguistic terms describing the *magnitude* of the fuzzy (or linguistic) variables *cloudy* and *tall*. The human brain has the ability to understand these terms, and infer from them that it will most probably not rain, and that Stephan might just be a good basket ball player (note, again, the fuzzy terms!). However, how do we use two-valued logic to represent these phrases?

Together with fuzzy logic, fuzzy set theory provides the tools to develop software products that model human reasoning (also referred to as approximate reasoning). In fuzzy sets, an element belongs to a set to a degree, indicating the certainty (or uncertainty) of membership.

The development of logic has a long and rich history, in which major philosophers played a role. The foundations of two-valued logic stemmed from the efforts of Aristotle (and other philosophers of that time), resulting in the so-called *Laws of Thought* [440]. The first version of these laws was proposed around 400 B.C., namely the *Law of the Excluded Middle*. This law states that every proposition must have only one of two outcomes: either *true* or *false*. Even in that time, immediate objections were given with examples of propositions that could be true, and simultaneously not true.

It was another great philosopher, Plato, who laid the foundations of what is today referred to as fuzzy logic. It was, however, only in the 1900s that Lejewski and Lukasiewicz [514] proposed the first alternative to the Aristotelian two-valued logic. Three-valued logic has a third value which is assigned a numeric value between *true* and *false*. Lukasiewicz later extended this to four-valued and five-valued logic. It was only recently, in 1965, that Lotfi Zadeh [944] produced the foundations of infinite-valued logic with his mathematics of fuzzy set theory.

Following the work of Zadeh, much research has been done in the theory of fuzzy systems, with applications in control, information systems, pattern recognition and decision support. Some successful real-world applications include automatic control of dam gates for hydroelectric-powerplants, camera aiming, compensation against vibrations in camcorders, cruise-control for automobiles, controlling air-conditioning systems, document archiving systems, optimized planning of bus time-tables, and many more. While fuzzy sets and logic have been used to solve real-world problems, they were also combined with other CI paradigms to form hybrid systems, for example, fuzzy neural networks and fuzzy genetic algorithms [957].

A different set theoretic approach which also uses the concept of membership functions, namely rough sets (introduced by Pawlak in 1982 [668]), is sometimes confused with fuzzy sets. While both fuzzy sets and rough sets make use of membership functions, rough sets differ in the sense that a lower and upper approximation to the rough set is determined. The lower approximation consists of all elements that belong with full certainty to the corresponding set, while the upper approximation consists of elements that may possibly belong to the set. Rough sets are frequently used in machine learning as classifier, where they are used to find the smallest number of features to discern between classes [600]. Rough sets are also used for extracting knowledge from incomplete data [600, 683]. Hybrid approaches that employ both fuzzy and rough sets have also been developed [843].

The remainder of this Part is organized as follows: Chapter 20 discusses fuzzy sets, while fuzzy logic and reasoning are covered in Chapter 21. A short overview of fuzzy controllers is given in Chapter 22. The Part is concluded with an overview of rough set theory in Chapter 23.

Chapter 20

Fuzzy Sets

Consider the problem of designing a set of all tall people, and assigning all the people you know to this set. Consider classical set theory where an element is either a member of the set or not. Suppose all tall people are described as those with height greater than 1.75m. Then, clearly a person of height 1.78m will be an element of the set *tall*, and someone with height 1.5m will not belong to the set of tall people. But, the same will apply to someone of height 1.73m, which implies that someone who falls only 2cm short is not considered as being tall. Also, using two-valued set theory, there is no distinction among members of the set of tall people. For example, someone of height 1.78m and one of height 2.1m belongs equally to the set! Thus, no semantics are included in the description of membership.

The alternative, fuzzy sets, has no problem with this situation. In this case all the people you know will be members of the set *tall*, but to different degrees. For example, a person of height 2.1m may be a member of the set to degree 0.95, while someone of length 1.7m may belong to the set with degree 0.4.

Fuzzy sets are an extension of crisp (two-valued) sets to handle the concept of *partial truth*, which enables the modeling of the uncertainties of natural language. The vagueness in natural language is further emphasized by linguistic terms used to describe objects or situations. For example, the phrase *when it is very cloudy, it will most probably rain*, has the linguistic terms *very* and *most probably* – which are understood by the human brain. Fuzzy sets, together with fuzzy reasoning systems, give the tools to also write software, which enables computing systems to understand such vague terms, and to reason with these terms.

This chapter formally introduces fuzzy sets. Section 20.1 defines fuzzy sets, while membership functions are discussed in Section 20.2. Operators that can be applied to fuzzy sets are covered in Section 20.3. Characteristics of fuzzy sets are summarized in Section 20.4. The chapter is concluded with a discussion of the differences between fuzziness and probability in Section 20.5.

Computational Intelligence: An Introduction, Second Edition A.P. Engelbrecht

20.1 Formal Definitions

Different to classical sets, elements of a fuzzy set have membership degrees to that set. The degree of membership to a fuzzy set indicates the certainty (or uncertainty) that the element belongs to that set. Formally defined, suppose X is the domain, or universe of discourse, and $x \in X$ is a specific element of the domain X. Then, the fuzzy set A is characterized by a membership mapping function [944]

$$\mu_A : X \to [0, 1] \tag{20.1}$$

Therefore, for all $x \in X$, $\mu_A(x)$ indicates the certainty to which element x belongs to fuzzy set A. For two-valued sets, $\mu_A(x)$ is either 0 or 1.

Fuzzy sets can be defined for discrete (finite) or continuous (infinite) domains. The notation used to denote fuzzy sets differ based on the type of domain over which that set is defined. In the case of a discrete domain X, the fuzzy set can either be expressed in the form of an n_x-dimensional vector or using the sum notation. If $X = \{x_1, x_2, \cdots, x_{n_x}\}$, then, using set notation,

$$A = \{(\mu_A(x_i)/x_i) | x_i \in X, i = 1, \cdots, n_x\} \tag{20.2}$$

Using sum notation,

$$A = \mu_A(x_1)/x_1 + \mu_A(x_2)/x_2 + \cdots + \mu_A(x_{n_x})/x_{n_x} = \sum_{i=1}^{n_x} \mu_A(x_i)/x_i \tag{20.3}$$

where the sum should not be confused with algebraic summation. The use of sum notation above simply serves as an indication that A is a set of ordered pairs. A continuous fuzzy set, A, is denoted as

$$A = \int_X \mu(x)/x \tag{20.4}$$

Again, the integral notation should not be algebraically interpreted.

20.2 Membership Functions

The membership function is the essence of fuzzy sets. A membership function, also referred to as the characteristic function of the fuzzy set, defines the fuzzy set. The function is used to associate a degree of membership of each of the elements of the domain to the corresponding fuzzy set. Two-valued sets are also characterized by a membership function. For example, consider the domain X of all floating-point numbers in the range $[0, 100]$. Define the crisp set $A \subset X$ of all floating-point numbers in the range $[10, 50]$. Then, the membership function for the crisp set A is represented in Figure 20.1. All $x \in [10, 50]$ have $\mu_A(x) = 1$, while all other floating-point numbers have $\mu_A(x) = 0$.

Membership functions for fuzzy sets can be of any shape or type as determined by experts in the domain over which the sets are defined. While designers of fuzzy sets

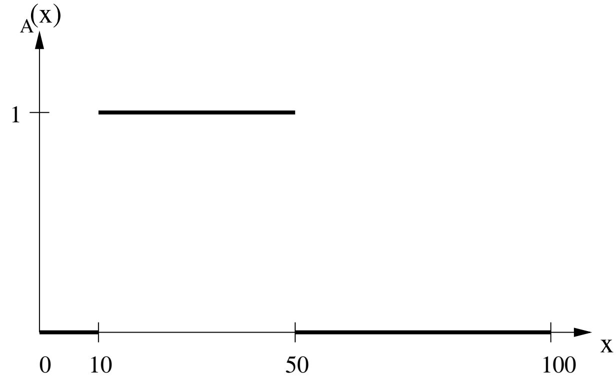


Figure 20.1 Illustration of Membership Function for Two-Valued Sets

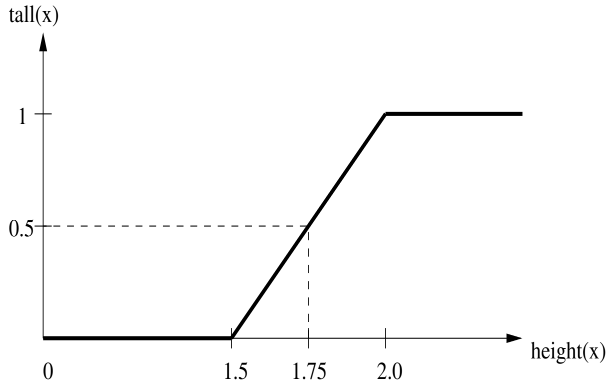


Figure 20.2 Illustration of *tall* Membership Function

have much freedom in selecting appropriate membership functions, these functions must satisfy the following constraints:

- A membership function must be bounded from below by 0 and from above by 1.
- The range of a membership function must therefore be $[0, 1]$.
- For each $x \in X$, $\mu_A(x)$ must be unique. That is, the same element cannot map to different degrees of membership for the same fuzzy set.

Returning to the *tall* fuzzy set, a possible membership function can be defined as (also illustrated in Figure 20.2)

$$tall(x) = \begin{cases} 0 & \text{if } length(x) < 1.5m \\ (length(x) - 1.5m) \times 2.0m & \text{if } 1.5m \leq length(x) \leq 2.0m \\ 1 & \text{if } length(x) > 2.0m \end{cases} \tag{20.5}$$

Now, assume that a person has a length of 1.75m, then $\mu_A(1.75) = 0.5$.

While the *tall* membership function above used a discrete step function, more complex discrete and continuous functions can be used, for example:

- **Triangular** functions (refer to Figure 20.3(a)), defined as

$$\mu_A(x) = \begin{cases} 0 & \text{if } x \leq \alpha_{min} \\ \frac{x-\alpha_{min}}{\beta-\alpha_{min}} & \text{if } x \in (\alpha_{min}, \beta] \\ \frac{\alpha_{max}-x}{\alpha_{max}-\beta} & \text{if } x \in (\beta, \alpha_{max}) \\ 0 & \text{if } x \geq \alpha_{max} \end{cases} \tag{20.6}$$

- **Trapezoidal** functions (refer to Figure 20.3(b)), defined as

$$\mu_A(x) = \begin{cases} 0 & \text{if } x \leq \alpha_{min} \\ \frac{x-\alpha_{min}}{\beta_1-\alpha_{min}} & \text{if } x \in [\alpha_{min}, \beta_1) \\ \frac{\alpha_{max}-x}{\alpha_{max}-\beta_2} & \text{if } x \in (\beta_2, \alpha_{max}) \\ 0 & \text{if } x \geq \alpha_{max} \end{cases} \tag{20.7}$$

- **Γ-membership** functions, defined as

$$\mu_A x = \begin{cases} 0 & \text{if } x \leq \alpha \\ 1 - e^{-\gamma(x-\alpha)^2} & \text{if } x > \alpha \end{cases} \tag{20.8}$$

- **S-membership** functions, defined as

$$\mu_A(x) = \begin{cases} 0 & \text{if } x \leq \alpha_{min} \\ 2\left(\frac{x-\alpha_{min}}{\alpha_{max}-\alpha_{min}}\right)^2 & \text{if } x \in (\alpha_{min}, \beta] \\ 1 - 2\left(\frac{x-\alpha_{max}}{\alpha_{max}-\alpha_{min}}\right)^2 & \text{if } x \in (\beta, \alpha_{max}) \\ 1 & \text{if } x \geq \alpha_{max} \end{cases} \tag{20.9}$$

- **Logistic** function (refer to Figure 20.3(c)), defined as

$$\mu_A(x) = \frac{1}{1 + e^{-\gamma x}} \tag{20.10}$$

- **Exponential-like** function, defined as

$$\mu_A(x) = \frac{1}{1 + \gamma(x-\beta)^2} \tag{20.11}$$

  with $\gamma > 1$.

- **Gaussian** function (refer to Figure 20.3(d)), defined as

$$\mu_A(x) = e^{-\gamma(x-\beta)^2} \tag{20.12}$$

It is the task of the human expert of the domain to define the function that captures the characteristics of the fuzzy set.

20.3 Fuzzy Operators

As for crisp sets, relations and operators are defined for fuzzy sets. Each of these relations and operators are defined below. For this purpose let X be the domain, or universe, and A and B are fuzzy sets defined over the domain X.

Equality of fuzzy sets: For two-valued sets, sets are equal if the two sets have exactly the same elements. For fuzzy sets, however, equality cannot be concluded if the two sets have the same elements. The degree of membership of elements to the sets must also be equal. That is, the membership functions of the two sets must be the same.

Therefore, two fuzzy sets A and B are equal if and only if the sets have the same domain, and $\mu_A(x) = \mu_B(x)$ for all $x \in X$. That is, $A = B$.

Containment of fuzzy sets: For two-valued sets, $A \subset B$ if all the elements of A are also elements of B. For fuzzy sets, this definition is not complete, and the degrees of membership of elements to the sets have to be considered.

Fuzzy set A is a subset of fuzzy set B if and only if $\mu_A(x) \leq \mu_B(x)$ for all $x \in X$. That is, $A \subset B$.

Figure 20.4 shows two membership functions for which $A \subset B$.

Complement of a fuzzy set (NOT): The complement of a two-valued set is simply the set containing the entire domain without the elements of that set. For fuzzy sets, the complement of the set A consists of all the elements of set A, but the membership degrees differ. Let $\overline{A}$ denote the complement of set A. Then, for all $x \in X$, $\mu_{\overline{A}}(x) = 1 - \mu_A(x)$. It also follows that $A \cap \overline{A} \neq \emptyset$ and $A \cup \overline{A} \neq X$.

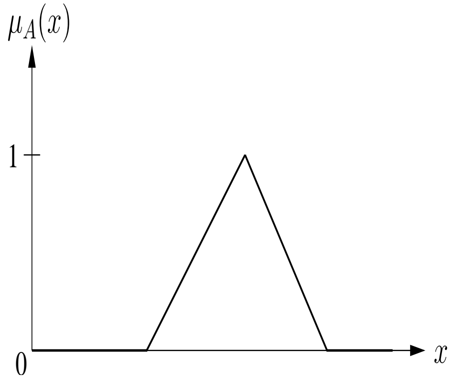
Intersection of fuzzy sets (AND): The intersection of two-valued sets is the set of elements occurring in both sets. Operators that implement intersection are referred to as t-norms. The result of a t-norm is a set that contain all the elements of the two fuzzy sets, but with degree of membership that depends on the specific t-norm. A number of t-norms have been used, of which the min-operator and the product operator are the most popular. If A and B are two fuzzy sets, then

- **Min-operator**: $\mu_{A\cap B}(x) = \min\{\mu_A(x), \mu_B(x)\}, \ \forall x \in X$
- **Product operator**: $\mu_{A\cap B}(x) = \mu_A(x)\mu_B(x), \ \forall x \in X$

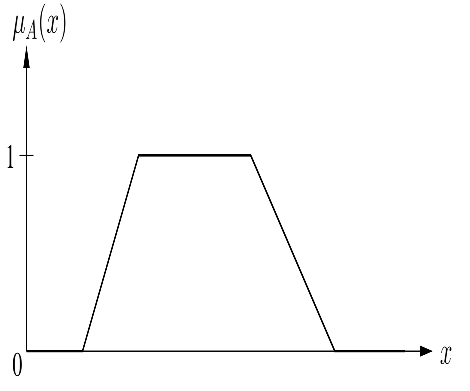
The difference between the two operations should be noted. Taking the product of membership degrees is a much stronger operator than taking the minimum, resulting in lower membership degrees for the intersection. It should also be noted that the ultimate result of a series of intersections approaches 0.0, even if the degrees of memberships to the original sets are high.

Other t-norms are [676],

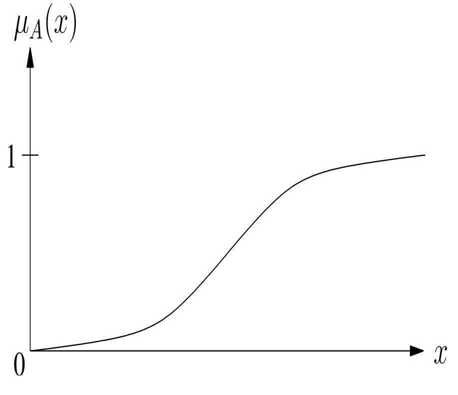
- $\mu_{A\cap B}(x) = \frac{1}{1+\sqrt[p]{\left(\frac{1-\mu_A(x)}{p}\right)^p+\left(\frac{1-\mu_B(x)}{p}\right)^p}}$, for $p > 0$.
- $\mu_{A\cap B}(x) = \max\{0, (1+p)(\mu_A(x)+\mu_B(x)-1) - p\mu_A(x)\mu_B(x)\}$, for $p \geq -1$.



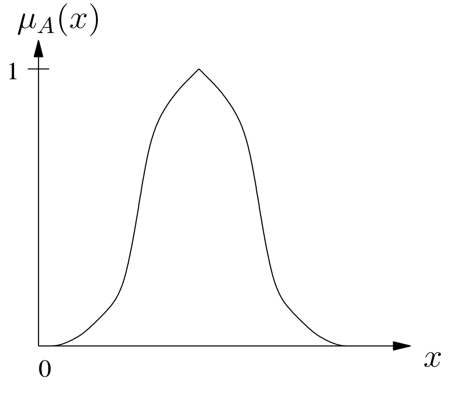
(a) Triangular Function



(b) Trapezoidal Function



(c) Logistic Function



(d) Gaussian Function

Figure 20.3 Example Membership Functions for Fuzzy Sets

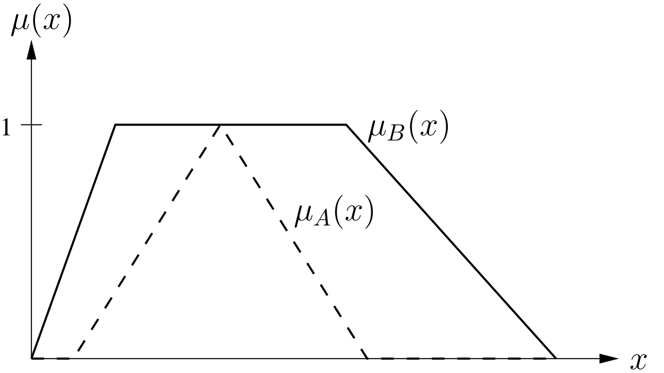


Figure 20.4 Illustration of Fuzzy Set Containment

- $\mu_{A\cap B}(x) = \frac{\mu_A(x)\mu_B(x)}{p+(1-p)(\mu_A(x)+\mu_B(x)-\mu_A(x)\mu_B(x))}$, for $p > 0$.
- $\mu_{A\cap B}(x) = \frac{1}{\sqrt[p]{\frac{1}{(\mu_A(x))^p}+\frac{1}{(\mu_B(x))^p}-1}}$
- $\mu_{A\cap B}(x) = \frac{\mu_A(x)\mu_B(x)}{\max\{\mu_A(x)\mu_B(x),p\}}$, for $p \in [0,1]$.
- $\mu_{A\cap B}(x) = p \times \min\{\mu_A(x), \mu_B(x)\} + (1-p) \times \frac{1}{2}(\mu_A(x) + \mu_B(x))$, where $p \in [0,1]$ [930].

Union of fuzzy sets (OR): The union of two-valued sets contains the elements of all of the sets. The same is true for fuzzy sets, but with membership degrees that depend on the specific uninion operator used. These operators are referred to as s-norms, of which the max-operator and summation operator are most frequently used:

- **Max-operator**: $\mu_{A\cup B}(x) = \max\{\mu_A(x), \mu_B(x)\}$, $\forall x \in X$, or
- **Summation operator**: $\mu_{A\cup B}(x) = \mu_A(x)+\mu_B(x)-\mu_A(x)\mu_B(x)$, $\forall x \in X$

Again, careful consideration must be given to the differences between the two approaches above. In the limit, a series of unions will have a result that approximates 1.0, even though membership degrees are low for the original sets!

Other s-norms are [676],

- $\mu_{A\cup B}(x) = \frac{1}{1+\sqrt[p]{\left(\frac{\mu_A(x)}{1-\mu_A(x)}\right)^p+\left(\frac{\mu_B(x)}{1-\mu_B(x)}\right)^p}}$, for $p > 0$.
- $\mu_{A\cup B}(x) = \min\{1, \mu_A(x) + \mu_B(x) + p\mu_A(x)\mu_B(x)\}$, for $p \geq 0$.
- $\mu_{A\cup B}(x) = \frac{\mu_A(x)+\mu_B(x)-\mu_A(x)\mu_B(x)-(1-p)\mu_A(x)\mu_B(x)}{1-(1-p)\mu_A(x)\mu_B(x)}$, for $p \geq 0$.
- $\mu_{A\cup B}(x) = 1 - \frac{1}{\sqrt[p]{\frac{1}{(1-\mu_A(x))^p}+\frac{1}{(1-\mu_B(x))^p}-1}}$
- $1 - \frac{(1-\mu_A(x))(1-\mu_B(x))}{\max\{(1-\mu_A(x)),(1-\mu_B(x)),p\}}$ for $p \in [0,1]$.
- $\mu_{A\cup B}(x) = p \times \max\{\mu_A(x), \mu_B(x)\} + (1-p) \times \frac{1}{2}(\mu_A(x) + \mu_B(x))$, where $p \in [0,1]$ [930].

Operations on two-valued sets are easily visualized using Venn-diagrams. For fuzzy sets the effects of operations can be illustrated by graphing the resulting membership function, as illustrated in Figure 20.5. For the illustration in Figure 20.5, assume the fuzzy sets A, defined as floating point numbers between $[50, 80]$, and B, defined as numbers *about* 40 (refer to Figure 20.5(a) for definitions of the membership functions). The complement of set A is illustrated in Figure 20.5(b), the intersection of the two sets are given in Figure 20.5(c) (assuming the *min* operator), and the union in Figure 20.5(d) (assuming the *max* operator).

20.4 Fuzzy Set Characteristics

As discussed previously, fuzzy sets are described by membership functions. In this section, characteristics of membership functions are overviewed. These characteristics include normality, height, support, core, cut, unimodality, and cardinality.

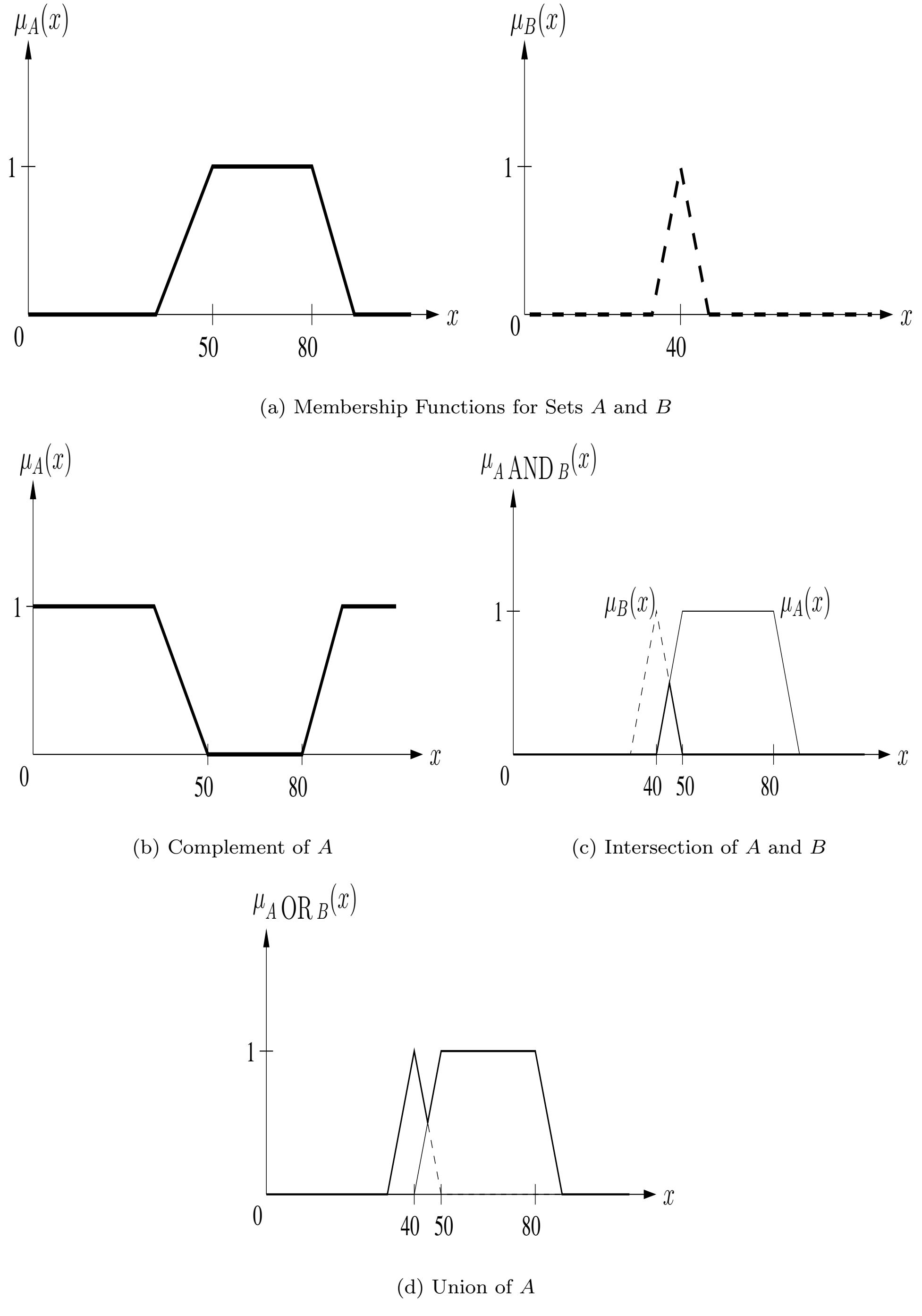


Figure 20.5 Illustration of Fuzzy Operators

Normality: A fuzzy set A is normal if that set has an element that belongs to the set with degree 1. That is,

$$\exists x \in A \bullet \mu_A(x) = 1 \tag{20.13}$$

then A is normal, otherwise, A is subnormal. Normality can alternatively be defined as

$$\sup_x \mu_A(x) = 1 \tag{20.14}$$

Height: The height of a fuzzy set is defined as the supremum of the membership function, i.e.

$$height(A) = \sup_x \mu_A(x) \tag{20.15}$$

Support: The support of fuzzy set A is the set of all elements in the universe of discourse, X, that belongs to A with non-zero membership. That is,

$$support(A) = \{x \in X | \mu_A(x) > 0\} \tag{20.16}$$

Core: The core of fuzzy set A is the set of all elements in the domain that belongs to A with membership degree 1. That is,

$$core(A) = \{x \in X | \mu_A(x) = 1\} \tag{20.17}$$

α-cut: The set of elements of A with membership degree greater than α is referred to as the α-cut of A:

$$A_\alpha = \{x \in X | \mu_A(x) \geq \alpha\} \tag{20.18}$$

Unimodality: A fuzzy set is unimodal if its membership function is a unimodal function, i.e. the function has just one maximum.

Cardinality: The cardinality of two-valued sets is simply the number of elements within the sets. This is not the same for fuzzy sets. The cardinality of fuzzy set A, for a finite domain, X, is defined as

$$card(A) = \sum_{x \in X} \mu_A(x) \tag{20.19}$$

and for an infinite domain,

$$card(A) = \int_{x \in X} \mu_A(x) dx \tag{20.20}$$

For example, if $X = \{a, b, c, d\}$, and $A = 0.3/a + 0.9/b + 0.1/c + 0.7/d$, then $card(A) = 0.3 + 0.9 + 0.1 + 0.7 = 2.0$.

Normalization: A fuzzy set is normalized by dividing the membership function by the height of the fuzzy set. That is,

$$normalized(A) = \frac{\mu_A(x)}{height(x)} \tag{20.21}$$

Other characteristics of fuzzy sets (i.e. concentration, dilation, contrast intensification, fuzzification) are described in Section 21.1.1.

The properties of fuzzy sets are very similar to that of two-valued sets, however, there are some differences. Fuzzy sets follow, similar to two-valued sets, the commutative, associative, distributive, transitive and idempotency properties. One of the major differences is in the properties of the cardinality of fuzzy sets, as listed below:

- $card(A) + card(B) = card(A \cap B) + card(A \cup B)$
- $card(A) + card(\overline{A}) = card(X)$

where A and B are fuzzy sets, and X is the universe of discourse.

20.5 Fuzziness and Probability

There is often confusion between the concepts of fuzziness and probability. It is important that the similarities and differences between these two terms are understood. Both terms refer to degrees of certainty (or uncertainty) of events occurring. But that is where the similarities stop. Degrees of certainty as given by statistical probability are only meaningful before the associated event occurs. After that event, the probability no longer applies, since the outcome of the event is known. For example, before flipping a fair coin, there is a 50% probability that heads will be on top, and a 50% probability that it will be tails. After the event of flipping the coin, there is no uncertainty as to whether heads or tails are on top, and for that event the degree of certainty no longer applies. In contrast, membership to fuzzy sets is still relevant after an event occurred. For example, consider the fuzzy set of tall people, with Peter belonging to that set with degree 0.9. Suppose the event to execute is to determine if Peter is good at basketball. Given some membership function, the outcome of the event is a degree of membership to the set of good basketball players. After the event occurred, Peter still belongs to the set of tall people with degree 0.9.

Furthermore, probability assumes independence among events, while fuzziness is not based on this assumption. Also, probability assumes a closed world model where everything is known, and where probability is based on frequency measures of occurring events. That is, probabilities are estimated based on a repetition of a finite number of experiments carried out in a stationary environment. The probability of an event A is thus estimated as

$$Prob(A) = \lim_{n \to \infty} \frac{n_A}{n} \tag{20.22}$$

where n_A is the number of experiments for which event A occurred, and n is the total number of experiments. Fuzziness does not assume everything to be known, and is based on descriptive measures of the domain (in terms of membership functions), instead of subjective frequency measures.

Fuzziness is not probability, and probability is not fuzziness. Probability and fuzzy sets can, however, be used in a symbiotic way to express the probability of a fuzzy event.

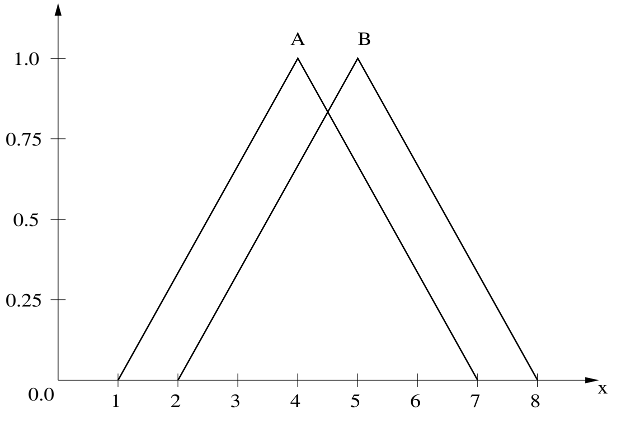


Figure 20.6 Membership Functions for Assignments 1 and 2

20.6 Assignments

1. Perform intersection and union for the fuzzy sets in Figure 20.5 using the t-norms and s-norms defined in Section 20.3.
2. Give the height, support, core and normalization of the fuzzy sets in Figure 20.5.
3. Consider the two fuzzy sets:

$$\begin{aligned}\text{long pencils} &= \{pencil1/0.1, pencil2/0.2, pencil3/0.4, pencil4/0.6,\\ &\quad pencil5/0.8, pencil6/1.0\}\\ \text{medium pencils} &= \{pencil1/1.0, pencil2/0.6, pencil3/0.4, pencil4/0.3,\\ &\quad pencil5/0.1\}\end{aligned}$$

   (a) Determine the union of the two sets.
   (b) Determine the intersection of the two sets.
4. What is the difference between the membership function of an ordinary set and a fuzzy set?
5. Consider the membership functions of two fuzzy sets, A and B, as given in Figure 20.6.
   (a) Draw the membership function for the fuzzy set $C = A \cap \overline{B}$, using the min-operator.
   (b) Compute $\mu_C(5)$.
   (c) Is C normal? Justify your answer.
6. Consider the fuzzy sets A and B such that $core(A) \cap core(B) = \emptyset$. Is fuzzy set $C = A \cap B$ normal? Justify your answer.
7. Show that the min-operator is
   (a) commutative
   (b) idempotent
   (c) transitive

Chapter 21

Fuzzy Logic and Reasoning

The previous chapter discussed theoretical aspects of fuzzy sets. The fuzzy set operators allow rudimentary reasoning about facts. For example, consider the three fuzzy sets *tall*, *good_athlete* and *good_basketball_player*. Now assume

$$\mu_{tall}(Peter) = 0.9 \text{ and } \mu_{good\_athlete}(Peter) = 0.8$$

$$\mu_{tall}(Carl) = 0.9 \text{ and } \mu_{good\_athlete}(Carl) = 0.5$$

If it is known that a good basketball player is tall and is a good athlete, then which one of Peter or Carl will be the better basketball player? Through application of the intersection operator,

$$\mu_{good\_basketball\_player}(Peter) = \min\{0.9, 0.8\} = 0.8$$

$$\mu_{good\_basketball\_player}(Carl) = \min\{0.9, 0.5\} = 0.5$$

Using the standard fuzzy set operators, it is possible to determine that Peter will be better at the sport than Carl.

The example above is a very simplistic situation. For most real-world problems, the sought outcome is a function of a number of complex events or scenarios. For example, actions made by a controller are determined by a set of fuzzy if-then rules. The if-then rules describe situations that can occur, with a corresponding action that the controller should execute. It is, however, possible that more than one situation, as described by if-then rules, are simultaneously active, with different actions. The problem is to determine the best action to take. A mechanism is therefore needed to infer an action from a set of activated situations.

What is necessary is a formal logic system that can be used to reason about uncertainties in order to derive at plausible actions. Fuzzy logic [945] is such a system, which together with an inferencing system form a tool for approximate reasoning. Section 21.1 provides a short definition of fuzzy logic and a short overview of its main concepts. The process of fuzzy inferencing is described in Section 21.2.

21.1 Fuzzy Logic

Zadeh [947] defines fuzzy logic (FL) as a logical system, which is an extension of multi-valued logic that is intended to serve as a logic for approximate reasoning. The two

Computational Intelligence: An Introduction, Second Edition A.P. Engelbrecht

most important concepts within FL is that of a linguistic variable and the fuzzy if-then rule. These concepts are discussed in the next subsections.

21.1.1 Linguistics Variables and Hedges

Lotfi Zadeh [946] introduced the concept of linguistic variable (or fuzzy variable) in 1973, which allows computation with words in stead of numbers. Linguistic variables are variables with values that are words or sentences from natural language. For example, referring again to the set of tall people, *tall* is a linguistic variable. Sensory inputs are linguistic variables, or nouns in a natural language, for example, temperature, pressure, displacement, etc. Linguistic variables (and hedges, explained below) allow the translation of natural language into logical, or numerical statements, which provide the tools for approximate reasoning (refer to Section 21.2).

Linguistic variables can be divided into different categories:

- Quantification variables, e.g. all, most, many, none, etc.
- Usuality variables, e.g. sometimes, frequently, always, seldom, etc.
- Likelihood variables, e.g. possible, likely, certain, etc.

In natural language, nouns are frequently combined with adjectives for quantifications of these nouns. For example, in the phrase *very tall*, the noun *tall* is quantified by the adjective *very*, indicating a person who is "taller" than tall. In fuzzy systems theory, these adjectives are referred to as hedges. A hedge serves as a modifier of fuzzy values. In other words, the hedge *very* changes the membership of elements of the set *tall* to different membership values in the set *very_tall*. Hedges are implemented through subjective definitions of mathematical functions, to transform membership values in a systematic manner.

To illustrate the implementation of hedges, consider again the set of tall people, and assume the membership function μ_{tall} characterizes the degree of membership of elements to the set *tall*. Our task is to create a new set, *very_tall* of people that are very tall. In this case, the hedge *very* can be implemented as the square function. That is, $\mu_{very\_tall}(x) = \mu_{tall}(x)^2$. Hence, if Peter belongs to the set *tall* with certainty 0.9, then he also belongs to the set *very_tall* with certainty 0.81. This makes sense according to our natural understanding of the phrase *very tall*: Degree of membership to the set *very_tall* should be less than membership to the set *tall*. Alternatively, consider the set *sort_of_tall* to represent all people that are sort of tall, i.e. people that are shorter than tall. In this case, the hedge *sort of* can be implemented as the square root function, $\mu_{sort\_of\_tall}(x) = \sqrt{\mu_{tall}(x)}$. So, if Peter belongs to the set *tall* with degree 0.81, he belongs to the set *sort_of_tall* with degree 0.9.

Different kinds of hedges can be defined, as listed below:

- **Concentration hedges** (e.g. *very*), where the membership values get relatively smaller. That is, the membership values get more concentrated around points with higher membership degrees. Concentration hedges can be defined, in general terms, as

$$\mu_{A'}(x) = \mu_A(x)^p, \quad for\ p > 1 \tag{21.1}$$

where $A^{'}$ is the concentration of set A.

- **Dilation hedges** (e.g. *somewhat, sort of, generally*), where membership values increases. Dilation hedges are defined, in general, as

$$\mu_{A'}(x) = \mu_A(x)^{1/p} \quad for\ p > 1 \tag{21.2}$$

- **Contrast intensification hedges** (e.g. *extremely*), where memberships lower than 1/2 are diminished, but memberships larger than 1/2 are elevated. This hedge is defined as,

$$\mu_{A'}(x) = \begin{cases} 2^{p-1}\mu_A(x)^p & \text{if } \mu_A(x) \leq 0.5 \\ 1 - 2^{p-1}(1-\mu_A(x))^p & \text{if } \mu_A(x) > 0.5 \end{cases} \tag{21.3}$$

  which intensifies contrast.

- **Vague hedges** (e.g. *seldom*), are opposite to contrast intensification hedges, having membership values altered using

$$\mu_{A'}(x) = \begin{cases} \sqrt{\mu_A(x)/2} & \text{if } \mu_A(x) \leq 0.5 \\ 1 - \sqrt{(1-\mu_A(x))/2} & \text{if } \mu_A(x) > 0.5 \end{cases} \tag{21.4}$$

  Vague hedges introduce more "fuzziness" into the set.

- **Probabilistic hedges**, which express probabilities, e.g. likely, not very likely, probably, etc.

21.1.2 Fuzzy Rules

For fuzzy systems in general, the dynamic behavior of that system is characterized by a set of linguistic fuzzy rules. These rules are based on the knowledge and experience of a human expert within that domain. Fuzzy rules are of the general form

$$\text{if antecedent(s) then consequent(s)} \tag{21.5}$$

The antecedent and consequent of a fuzzy rule are propositions containing linguistic variables. In general, a fuzzy rule is expressed as

$$\text{if } A \text{ is } a \text{ and } B \text{ is } b \text{ then } C \text{ is } c \tag{21.6}$$

where A and B are fuzzy sets with universe of discourse X_1, and C is a fuzzy set with universe of discourse X_2. Therefore, the antecedent of a rule form a combination of fuzzy sets through application of the logic operators (i.e. complement, intersection, union). The consequent part of a rule is usually a single fuzzy set, with a corresponding membership function. Multiple fuzzy sets can also occur within the consequent, in which case they are combined using the logic operators.

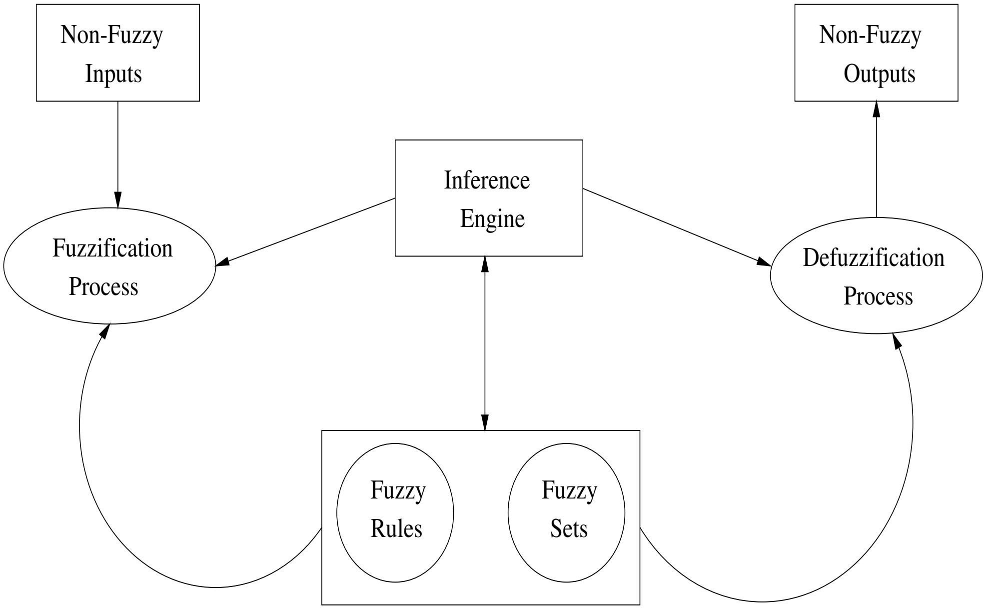


Figure 21.1 Fuzzy Rule-Based Reasoning System

Together, the fuzzy sets and fuzzy rules form the knowledge base of a fuzzy rule-based reasoning system. In addition to the knowledge base, a fuzzy reasoning system consists of three other components, each performing a specific task in the reasoning process, i.e. fuzzification, inferencing and defuzzification. The different components of a fuzzy rule based system are illustrated in Figure 21.1.

The question is now what sense can be made from a single fuzzy rule. For example, for the rule,

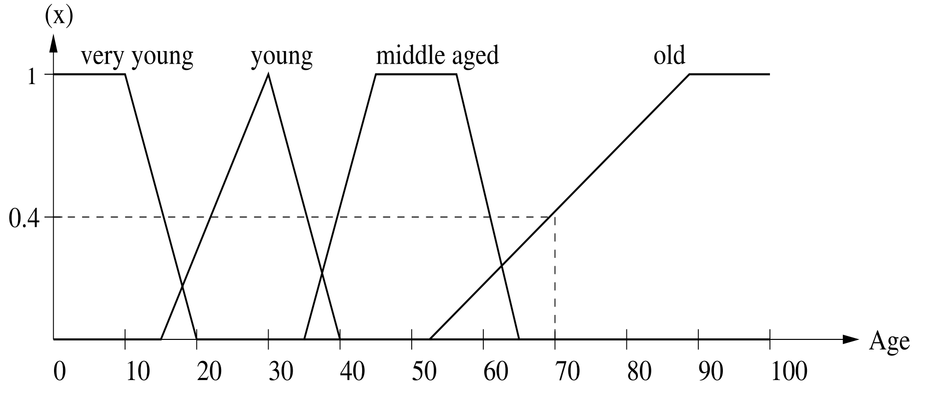
$$\text{if } Age \text{ is } Old \text{ the } Speed \text{ is } Slow \tag{21.7}$$

what can be said about *Speed* if *Age* has the value of 70? Given the membership functions for *Age* and *Speed* in Figure 21.2(a), then find $\mu_{Old}(70)$, which is 0.4. For linguistic variable *Speed* find the intersection of the horizontal line with membership function *Slow*. This gives the shaded area in Figure 21.2(b). A defuzzification operator (refer to Section 21.2) is then used to find the center of gravity of the shaded area, which gives the value of $Speed = 3$.

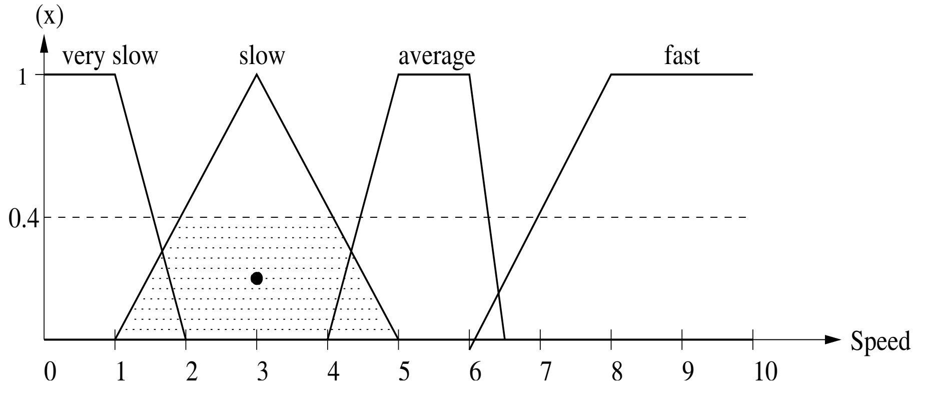
This simple example leads to the next question: How is a plausible outcome determined if a system has more than one rule? This question is answered in the next section.

21.2 Fuzzy Inferencing

Together, the fuzzy sets and fuzzy rules form the knowledge base of a fuzzy rule-based reasoning system. In addition to the knowledge base, a fuzzy reasoning system consists



(a) *Age* Membership Functions



(b) *Speed* Membership Functions

Figure 21.2 Interpreting a Fuzzy Rule

of three other components, each performing a specific task in the reasoning process, i.e. fuzzification, inferencing, and defuzzification (refer to Figure 21.1).

The remainder of this section is organized as follows: fuzzification is discussed in Section 21.2.1, fuzzy inferencing in Section 21.2.2, and defuzzification in Section 21.2.3.

21.2.1 Fuzzification

The antecedents of the fuzzy rules form the fuzzy "input space," while the consequents form the fuzzy "output space". The input space is defined by the combination of input fuzzy sets, while the output space is defined by the combination of output sets. The fuzzification process is concerned with finding a fuzzy representation of non-fuzzy input values. This is achieved through application of the membership functions associated with each fuzzy set in the rule input space. That is, input values from the universe of discourse are assigned membership values to fuzzy sets.

For illustration purposes, assume the fuzzy sets A and B, and assume the corresponding membership functions have been defined already. Let X denote the universe of discourse for both fuzzy sets. The fuzzification process receives the elements $a, b \in X$, and produces the membership degrees $\mu_A(a), \mu_A(b), \mu_B(a)$ and $\mu_B(b)$.

21.2.2 Inferencing

The task of the inferencing process is to map the fuzzified inputs (as received from the fuzzification process) to the rule base, and to produce a fuzzified output for each rule. That is, for the consequents in the rule output space, a degree of membership to the output sets are determined based on the degrees of membership in the input sets and the relationships between the input sets. The relationships between input sets are defined by the logic operators that combine the sets in the antecedent. The output fuzzy sets in the consequent are then combined to form one overall membership function for the output of the rule.

Assume input fuzzy sets A and B with universe of discourse X_1 and the output fuzzy set C with X_2 as universe of discourse. Consider the rule

$$\text{if } A \text{ is } a \text{ and } B \text{ is } b \text{ then } C \text{ is } c \tag{21.8}$$

From the fuzzification process, the inference engine knows $\mu_A(a)$ and $\mu_B(b)$. The first step of the inferencing process is then to calculate the firing strength of each rule in the rule base. This is achieved through combination of the antecedent sets using the operators discussed in Section 20.3. For the example above, assuming the *min*-operator, the firing strength is

$$\min\{\mu_A(a), \mu_B(b)\} \tag{21.9}$$

For each rule k, the firing strength α_k is thus computed.

The next step is to accumulate all activated outcomes. During this step, one single fuzzy value is determined for each $c_i \in C$. Usually, the final fuzzy value, β_i, associated with each outcome c_i is computed using the *max*-operator, i.e.

$$\beta_i = \max_{\forall k}\{\alpha_{k_i}\} \tag{21.10}$$

where α_{k_i} is the firing strength of rule k which has outcome c_i.

The end result of the inferencing process is a series of fuzzified output values. Rules that are not activated have a zero firing strength.

Rules can be weighted *a priori* with a factor (in the range [0,1]), representing the degree of confidence in that rule. These rule confidence degrees are determined by the human expert during the design process.

21.2.3 Defuzzification

The firing strengths of rules represent the degree of membership to the sets in the consequent of the corresponding rule. Given a set of activated rules and their corresponding firing strengths, the task of the defuzzification process is to convert the output of the fuzzy rules into a scalar, or non-fuzzy value.

For the sake of the argument, suppose the following hedges are defined for linguistic variable C (refer to Figure 21.3(a) for the definition of the membership functions): large decrease (LD), slight increase (SI), no change (NC), slight increase (SI), and large increase (LI). Assume three rules with the following C membership values: $\mu_{LI} = 0.8, \mu_{SI} = 0.6$ and $\mu_{NC} = 0.3$.

Several inference methods exist to find an approximate scalar value to represent the action to be taken:

- The **max-min method**: The rule with the largest firing strength is selected, and it is determined which consequent membership function is activated. The centroid of the area under that function is calculated and the horizontal coordinate of that centroid is taken as the output of the controller. For our example, the largest firing strength is 0.8, which corresponds to the *large_increase* membership function. Figure 21.3(b) illustrates the calculation of the output.
- The **averaging method**: For this approach, the average rule firing strength is calculated, and each membership function is clipped at the average. The centroid of the composite area is calculated and its horizontal coordinate is used as output of the controller. All rules therefore play a role in determining the action of the controller. Refer to Figure 21.3(c) for an illustration of the averaging method.
- The **root-sum-square method**: Each membership function is scaled such that the peak of the function is equal to the maximum firing strength that corresponds to that function. The centroid of the composite area under the scaled functions are computed and its horizontal coordinate is taken as output (refer to Figure 21.3(d)).
- The **clipped center of gravity method**: For this approach, each membership function is clipped at the corresponding rule firing strengths. The centroid of the composite area is calculated and the horizontal coordinate is used as the output of the controller. This approach to centroid calculation is illustrated in Figure 21.3(e).

The calculation of the centroid of the trapezoidal areas depends on whether the domain of the functions is discrete or continuous. For a discrete domain of a finite number of values, n_x, the output of the defuzzification process is calculated as ($\sum$ has its algebraic meaning)

$$output = \frac{\sum_{i=1}^{n_x} x_i \mu_C(x_i)}{\sum_{i=1}^{n_x} \mu_C(x_i)} \tag{21.11}$$

In the case of a continuous domain ($\int$ has its algebraic meaning),

$$output = \frac{\int_{x \in X} x\mu(x)dx}{\int_{x \in X} \mu(x)dx} \tag{21.12}$$

where X is the universe of discourse.

21.3 Assignments

1. For the two pencil fuzzy sets in the assignments of Chapter 20, define a hedge for the set *very long pencils*, and give the resulting set.
2. Consider the following rule base:

   if x is Small then y is Big

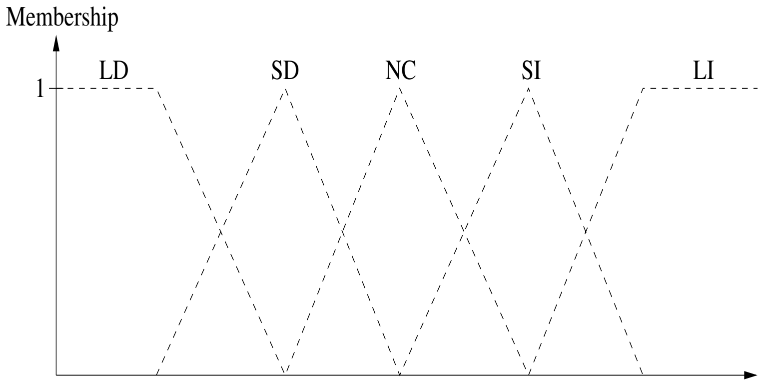
   if x is Medium then y is Small

   if x is Big then y is Medium

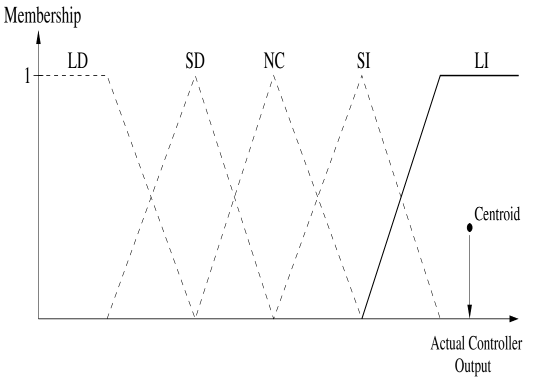
   Given the membership functions illustrated in Figure 21.4, answer the following questions:

   (a) Using the clipped center of gravity method, draw the composite function for which the centroid needs to be calculated, for $x = 2$.

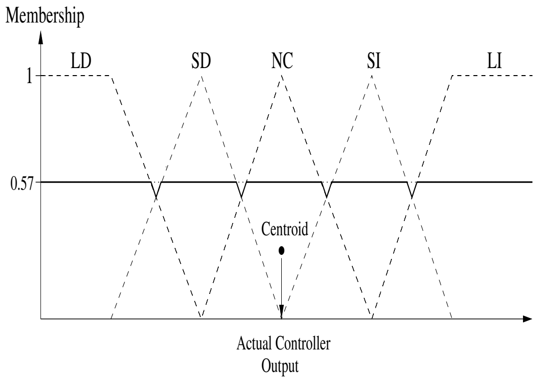
   (b) Compute the defuzzified output on the discrete domain, $Y = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$
3. Repeat the assignment above for the root-sum-square method.
4. Develop a set of fuzzy rules and membership functions to adapt the values of w, c_1 and c_2 of a *gbest* PSO.
5. Show how a fuzzy system can be used to adapt the learning rate of a FFNN trained using gradient descent.



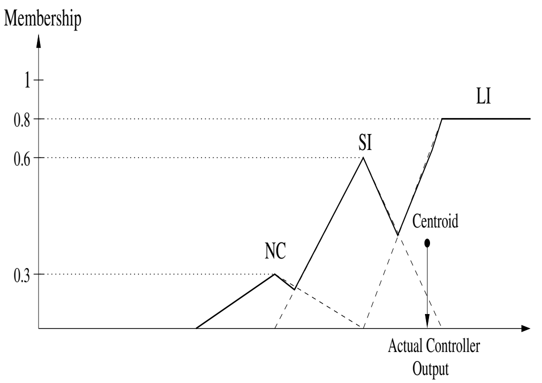
(a) Output Membership Functions



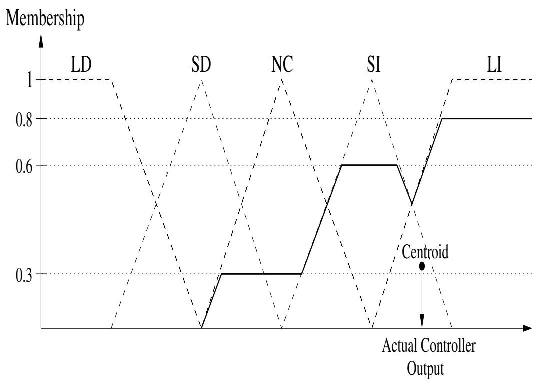
(b) Max-Min Method



(c) Averaging Method

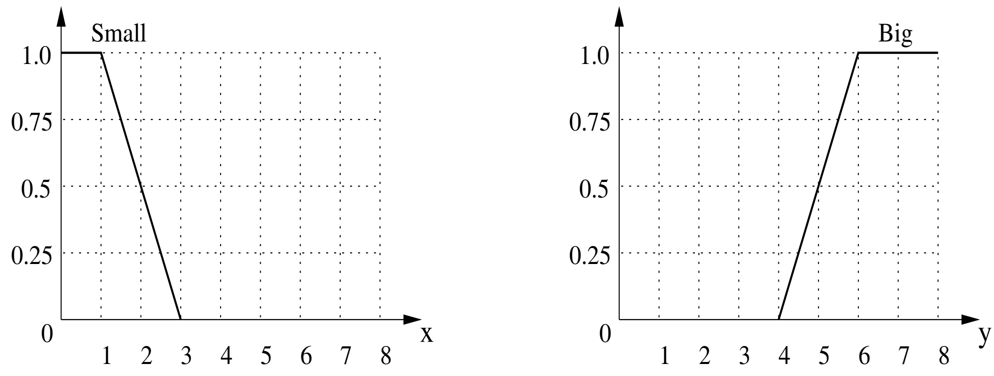


(d) Root-Sum-Square Method

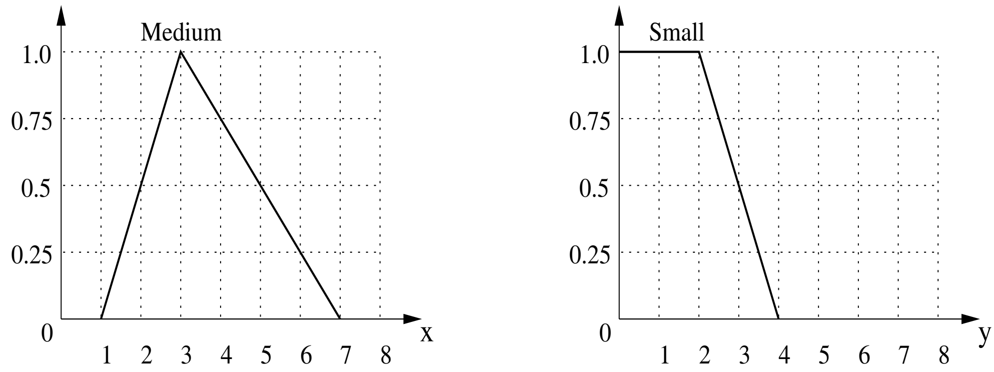


(e) Clipped Center of Gravity Method

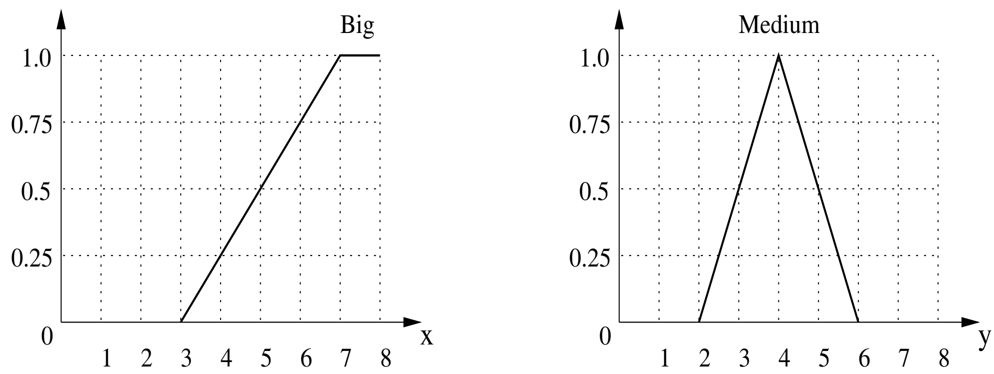
Figure 21.3 Defuzzification Methods for Centroid Calculation



(a) if x is Small then y is Big



(b) if x is Medium then y is Small



(c) if x is Big then y is Medium

Figure 21.4 Membership Functions for Assignments 2 and 3