Chapter - III

SOFT COMPUTING TECHNIQUES
CHAPTER – III

SOFT COMPUTING

3.1 Introduction

Soft computing constitutes a collection of disciplines, which include fuzzy logic (FL), neural networks (NNs), genetic algorithms (GAs), and probabilistic reasoning (PR). It is fast emerging as a tool to help computer-based intelligent systems mimic the ability of the human mind to employ modes of reasoning that are approximate rather than exact. The basic thesis of soft computing is that precision and certainty carry a cost and that intelligent systems should exploit, wherever possible, the tolerance for imprecision and uncertainty. In other words, it provides the foundation for the conception and design of high MIQ (Machine Intelligent Quotient) systems, and therefore forms the basis of future generation computing systems. Considerable success has been achieved in the application of this principle. We are now at a juncture where the ideas implicit in soft computing will begin to have significant impact in many other domains of application. A large number of researchers all over the world are engaged in developing soft-computing methodologies for designing intelligent systems since last one decade. In large measure, fuzzy logic, neurocomputing, and probabilistic reasoning are complementary, not competitive. In many cases a problem can be solved most effectively by using FL, NN and PR in combination rather than exclusively. A striking example of a particularly effective combination is neurofuzzy systems. Such systems are becoming increasingly visible as consumer products ranging from air conditioners and washing machines to photocopiers and camcorders. Less visible but perhaps even more important are neurofuzzy systems in industrial applications. The significant point is that in both consumer products and industrial systems, the employment of soft computing techniques leads to systems, which have high MIQ.

It is a fact that we live in a world that is pervasively imprecise, uncertain, and hard to be categorical about. It is also a fact that precision and certainty carry a cost. Consider the case of travelling salesman problem, which is frequently used as a test bed for assessing the effectiveness of various methods of solution. The important point about this problem is the steep rise in computing as a function of precision of solution. A more familiar example that
illustrates the point is the problem of parking a car. We find it relatively easy to park a car because the final position of the car is not specified precisely. If it were, the difficulty of parking the car would increase geometrically with the increase in precision, and eventually parking would become impossible. These and many similar examples lead to the basic premise and the guiding principle of soft computing [142].

The basic premises of soft computing are

- The real world is pervasively imprecise and uncertain.
- Precision and certainty carry cost.

The guiding principle of soft computing is

- Exploit the tolerance for imprecision, uncertainty, and partial truth to achieve tractability, robustness, and low solution cost.

### 3.2 Fuzzy Set Theory

Mathematics is the mother of all sciences. Over the centuries, mathematicians and scientists have introduced many novel theories, and some of these theories have found novel applications. The theory of fuzzy sets is one such mathematical theory.

Professor Zadeh [143] coined the word Fuzzy Sets. Fuzzy sets deal with sets of objects or phenomena which are vague and do not have sharp boundaries whereas the traditional set theory models the world as black or white and makes no provision for sets of gray. This two-valued logic has proved very effective and successful in solving well-defined problems, which are characterized by the precise description of the system being dealt with in quantitative form. However, a class of problems exists that does not lend itself readily to this approach. These problems are typically complex or ill structured in nature and are often left to human beings to deal with rather than being automated. The concepts are no longer clear cut like true or false but are relatively vague, for example, more or less true but most likely false. In the 1920s, a mathematician Lukasiewicz [144] challenged this premise and proposed that a gradation may exist between these two extremes. About ten years later, a physicist Black [145] introduced the concept of vagueness, that a set may contain elements that are partly in and partly out of the set. In 1965, Lotfi Zadeh of the University of
California at Berkley [143] came up with a formal methodology for handling sets of this type proposed by Black and called these fuzzy sets. The calculus of fuzzy sets is a very promising tool for dealing with cognitive uncertainty. Indeed, the applications of these fuzzy sets, which once were thought to be dull and dry, can be found in many scientific and scholarly works. It is true that Boole [146] introduced the beautiful notion of binary sets, which is the foundation of modern digital computer but Boolean logic is unable to model the human cognition and thinking process.

Lotfi Zadeh introduced the theory of fuzzy set as an extension to traditional set theory and developed the corresponding fuzzy logic to manipulate the fuzzy sets. A fuzzy set allows for the degree of membership of an item in a set to be any real number between zero and one. This allows human observations, expressions and expertise to be more closely modelled. Since its introduction fuzzy set theory has attracted the attention of many researchers in mathematical and engineering fields as well as in computer science and has become established. Currently a large number of successful applications of fuzzy logic to many real world control problems have been reported. In fact these applications have surpassed the expectations of the pioneers in this area. It was originally thought that most of the applications of fuzzy logic would be in those knowledge-based systems in which the resident information is both imprecise and uncertain. Contrary to this expectation, most of the present applications of fuzzy logic have taken place in systems that have imprecision but not uncertainty.

In order to clarify the basic difference between fuzzy logic and fuzzy set theory Professor Lotfi A. Zadeh, the father of fuzzy set theory, mentioned it in the following words:

"With regard to fuzzy logic, there is an issue of semantics that is in need of clarification. Specifically, it is frequently not recognized that the term fuzzy logic is actually used in two different senses. In narrow sense, fuzzy logic (FLn) is a logical system—an extension of multi-valued logic that is intended to serve as a logic of approximate reasoning. In a wider sense, fuzzy logic (FLw) is more or less synonymous with fuzzy set theory; that is, the theory of classes with unsharp boundaries. In this perspective, FL = FLw, and FLn is merely a branch of FL. Today the term fuzzy logic is used predominantly in its wider sense. It is in this sense that any field X can be "fuzzified"—and hence generalized—by replacing the concept of a crisp set in X by a fuzzy set. In application to basic fields such as set theory,
arithmetic, topology, graph theory, probability theory, and logic, fuzzification leads to fuzzy set theory, fuzzy arithmetic, fuzzy topology, fuzzy graph theory, and fuzzy logic in its narrow sense. Similarly, in application to applied fields like neurocomputing, stability theory, pattern recognition, and mathematical programming, fuzzification leads to fuzzy neurocomputing, fuzzy stability theory, fuzzy pattern recognition, and fuzzy mathematical programming [147].”

3.2.1 Membership Function

The fuzzy set theory is a generalization of the set theory and provides a means for the representation of imprecision and vagueness. Fuzzy set theory becomes identical with it in the limiting case where the properties being dealt with are ‘crisp’. A set can be described either by naming all its members (the list method) or by specifying some well-defined properties satisfied by the members of the set (the rule method). The list method, however, can be used only for finite sets [148]. The set $A$ whose members are $a_1, a_2, ..., a_n$ is usually written as

$$A = \{a_1, a_2, ..., a_n\}, \quad (3.1)$$

and the set $B$ whose members satisfy the properties $P_1, P_2, ..., P_n$ is usually written as

$$B = \{b \mid B \text{ has properties } P_1, P_2, ..., P_n \}, \quad (3.2)$$

where the symbol $\mid$ denotes the phrase “such that.” Each member in the universe is either in the set, or not. The process by which individuals from the universal set $U$ are determined to be either members or non-members of a set can be defined by a characteristic, or discriminate, function. For a given set $A$, this function assigns a value $\mu_A(u)$, where $u \in U$ such that

$$\mu_A(u) = \begin{cases} 1 & \text{if and only if } u \in A \\ 0 & \text{if and only if } u \notin A \end{cases} \quad (3.3)$$

Thus, the function maps elements of the universal set to the set containing 0 and 1. This can be indicated by

$$\mu_A(u) = U \rightarrow \{0, 1\} \quad (3.4)$$
The number of elements that belong to a set $A$ is called the *cardinality* of the set and is denoted by $|A|$. 

Fuzzy sets are based on the idea of extending the range of the characteristic function so that it covers the real numbers in the interval $[0, 1]$. The membership value assigned to an element in the universe is no longer restricted to just two possibilities, but can be 0, 1, or any value in-between [149]. Such a function is called a *membership function* and the set defined by it a *fuzzy set*.

Each fuzzy set, $A$, is defined in terms of a relevant universal set $U$ by a membership function, denoted as $\mu_A(u)$, where $u \in U$. This function assigns to each element $u$ of $U$ a member, in the closed interval $[0,1]$, that characterize the degree of membership of $u$ in $A$. That is, the membership function can take all values between zero and one including the discrete values of 0 and 1. More formally, membership functions are functions of the form

$$\mu_A(u) : U \rightarrow [0,1] \quad (3.5)$$

In defining a membership function, the universal set $U$ is always assumed to be a classical set. Given a fuzzy set $A$, which is a subset of the universal set, $U$, the support of $A$ denoted by $\text{Supp} (A)$, is an ordinary set defined by

$$\text{Supp} (A) = \{u \in U \mid \mu_A(u) > 0\} \quad (3.6)$$

A fuzzy set can also be written as

$$A = \{\mu_A(u) / u \in \text{supp}(A)\} \quad (3.7)$$

which means that only those fuzzy elements whose membership function is greater than zero contribute to fuzzy set $A$.

### 3.2.2 Linguistic Variables

A concept in fuzzy logic that plays a key role in exploiting the tolerance for imprecision is the linguistic variable. A linguistic variable, as its name suggests, is a variable whose values are words or sentences in a natural or synthetic language. For example, *age* is a linguistic variable if its *linguistic values* are *young, old, middle-aged, very old, not very young*, and so on. A linguistic variable is interpreted as a label of a fuzzy set that is characterised by a *membership function*, as illustrated in Figure 3.1. Thus, if $u$ is a numerical age, say 53, then
\( \mu_{\text{middle-age}}(53) \) is the grade of membership of 53 in middle-aged. Subjectively, \( \mu_{\text{middle-age}}(u) \) is interpreted as the degree to which \( u \) fits the perception of middle-aged in a specified context.

In a general setting, a linguistic variable, \( V \), can be viewed as a micro-language with context-free grammar and attributed-grammar semantics. The context-free grammar defines the legal values of \( V \). For example, in the case of age, the legal values are young, not young, not very young, quite old, middle-aged, and so on. The attributed-grammar semantics provides a mechanism for computing the membership function of any value of \( V \) from the knowledge of the membership function of the so-called primary terms — young and old, for example. A primary term plays the role of a generator whose meaning (its membership function) must be calibrated in context. For example, the meaning of not very young might be computed as

\[
\mu_{\text{not very young}}(u) = 1 - (\mu_{\text{young}}(u))^2
\]  

where very plays the role of an intensifier and young is a primary term whose membership function is specified in context. Most current applications of fuzzy logic employ a simple framework of membership functions. Specifically, the membership functions are assumed to be triangular or trapezoidal, and the number of linguistic values is usually in the range of three to seven. The triangular linguistic values are usually ranging from negative large, negative medium, and negative small (NL, NM, and NS) to zero (Z) to positive small, positive medium, and positive large (PS, PM, and PL) as illustrated in Figure 3.2.

Figure 3.1 Interpretation of middle-aged as a linguistic value

In a general setting, a linguistic variable, \( V \), can be viewed as a micro-language with context-free grammar and attributed-grammar semantics. The context-free grammar defines the legal values of \( V \). For example, in the case of age, the legal values are young, not young, not very young, quite old, middle-aged, and so on. The attributed-grammar semantics provides a mechanism for computing the membership function of any value of \( V \) from the knowledge of the membership function of the so-called primary terms — young and old, for example. A primary term plays the role of a generator whose meaning (its membership function) must be calibrated in context. For example, the meaning of not very young might be computed as

\[
\mu_{\text{not very young}}(u) = 1 - (\mu_{\text{young}}(u))^2
\]  

where very plays the role of an intensifier and young is a primary term whose membership function is specified in context. Most current applications of fuzzy logic employ a simple framework of membership functions. Specifically, the membership functions are assumed to be triangular or trapezoidal, and the number of linguistic values is usually in the range of three to seven. The triangular linguistic values are usually ranging from negative large, negative medium, and negative small (NL, NM, and NS) to zero (Z) to positive small, positive medium, and positive large (PS, PM, and PL) as illustrated in Figure 3.2.
3.2.3 Granulation

In a related sense, the use of words may be viewed as a form of fuzzy quantization or more generally as granulation, as Figure 3.3 shows [142]. Basically, granulation involves a replacement of a constraint of the form

\[ X - a \]

with a constraint of the form

\[ X \text{ is } A \]

where \( A \) is a fuzzy subset of \( U \), the universe of \( X \). For example,

\[ X - 2 \]

might be replaced with

\[ X \text{ is small} \]

In fuzzy logic, \( X \text{ is } a \) is interpreted as a characterization of the possible values of \( X \), with \( A \) representing a possibility distribution. Thus, the possibility that \( X \) can take a value \( u \) is given by

\[ \text{Poss}\{X = u}\} = \mu_A(u) \]

(3.9)

It is in this sense that \( X \text{ is } \mu_A(u) \), with possibility interpreted as ease of attainment or assignment, may be interpreted as an elastic constraint on \( X \).
3.2.4 Fuzzy Set Operations

The use of fuzzy sets provides a basis for the systematic manipulation of vague and imprecise concepts using fuzzy set operations performed by manipulating the membership functions. Set operation on fuzzy sets $A$ and $B$ can be represented using their membership functions $\mu_A(u)$ and $\mu_B(u)$. Here we summarize only the standard fuzzy operations, by far the most common operations in practical applications of fuzzy set theory. These operations are as follows:

Set union

The union of two fuzzy sets $A$ and $B$ with membership functions $\mu_A(u)$ and $\mu_B(u)$ is the fuzzy set whose membership function $A \cup B$ is given by:

$$A \cup B \Leftrightarrow \{u \mid (u \in A \lor u \in B) \land \mu (A \cup B)(u) = \max (\mu_A(u), \mu_B(u))\} \quad (3.10)$$

Set intersection

The intersection of two fuzzy sets $A$ and $B$ with membership functions $\mu_A(u)$ and $\mu_B(u)$ is the fuzzy set whose membership function $A \cap B$ is given by:

$$A \cap B \Leftrightarrow \{u \mid (u \in A \land u \in B) \land \mu (A \cap B)(u) = \min (\mu_A(u), \mu_B(u))\} \quad (3.11)$$
Set Complement

The complement of a (normalized) fuzzy set \( A \) with membership function \( \mu_A(u) \) is defined as the fuzzy set on the same universe with the membership function \( \mu_{\neg A}(u) \):

\[
\neg A = \{ u \mid (u \notin A \land \mu_{\neg A}(u) = 1 - \mu_A(u))/u \}
\]  

Set equality

Two fuzzy sets \( A \) and \( B \) are equal if they are defined on the same universe and the membership function is the same for both, that is,

\[
A = B \iff \{ u \mid (u \in A \land u \in B) \land \mu_A(u) = \mu_B(u) \}
\]  

Set containment

\[
A \subseteq B \iff \{ u \mid \forall u (u \in A \rightarrow u \in B) \land \mu_A(u) \leq \mu_B(u) \}
\]  

Set Concentration (CON)

\[
\mu_{CON(A)}(u) = \{ u \mid (u \in A \land \mu_{CON(A)}(u) = (\mu_A(u))^2) \}
\]  

Set Dilation (DIL)

\[
\mu_{DIL(A)}(u) = \{ u \mid (u \in A \land \mu_{DIL(A)}(u) = (\mu_A(u))^{1/2}) \}
\]  

These operators such as concentration, dilation, and complementation are usually used to represent linguistic hedges that act as modifiers to linguistic variables represented in fuzzy sets. For example, the concentration operator can be used to approximate the effect of the linguistic modifier 'very'. That is, \( very(A) = Con(A) \), where \( A \) is a linguistic variable. The concentration operation causes small changes in magnitude for the high membership degrees and broader changes in magnitude for the low membership degrees. The effect of dilation is the opposite of the effect of the concentration operation.

Similarity Relations

Similarity relations are useful for describing how similar two elements from the same domain are, as the name implies. Given two elements, \( x \) and \( y \), a similarity relation, \( S(x, y) \), for given domain \( D_i \), maps these two elements into an element in the unit interval \([0,1]\). The
more similar two elements are, the higher the similarity value. If the two elements are the same, that is, if we compare an element with itself, the similarity is 1 (the highest similarity value). The similarity relation is the basis of the similarity-based fuzzy relational database model. More formally, a similarity relation can be defined as follows:

A similarity relation is a mapping, \( S : D_i \times D_i \rightarrow [0,1] \), such that for \( x, y, z \in D_i \), the following rules hold:

\[
\begin{align*}
S(x, x) &= 1 \quad \text{(reflexivity)}, \\
S(x, y) &= S(y, x) \quad \text{(symmetry)}, \\
S(x, z) &\geq \max_{y \in D_i} (\min (S(x, y), S(y, z))) \quad \text{(max-min transitivity)}.
\end{align*}
\]

Size of Fuzzy set

In classical set theory, the size of a set is often taken as number of elements in the set. The size of a fuzzy set is defined as the sum of the grades of membership of its elements.

3.3 Applications of Fuzzy Logic

Fuzzy set theory provides us with a respectable inventory of theoretical tools for dealing with concepts expressed in natural language. These tools enable us to represent linguistic concepts and to manipulate them in a great variety of ways for various purposes; they enable us to express and deal with various relations, functions, and equations that involve linguistic concepts; and they allow us to fuzzify any desired area of classical mathematics to facilitate emerging applications. The past few years have witnessed a rapid growth in the number and variety of applications of fuzzy logic. The applications range from consumer products such as cameras, camcorders, washing machines, and microwave ovens to industrial process control, medical instrumentation, decision support systems, and portfolio selection.

Electrical engineering was the first engineering discipline within which the utility of fuzzy sets and fuzzy logic was recognized in the form of fuzzy controller [150]. Fuzzy controllers, contrary to classical controllers, are capable of utilizing knowledge elicited from human operators. This is crucial in control problems for which it is difficult or even impossible to construct precise mathematical models, or for which the acquired models are
difficult or expensive. These difficulties may result from inherent nonlinearities, the time-varying nature of the processes to be controlled, large unpredictable environmental disturbances, degrading sensors or other difficulties in obtaining precise and reliable measurements, and a host of other factors. It has been observed that experienced human operators are generally able to perform well under these circumstances [148].

Among other engineering disciplines, the utility of fuzzy set theory was recognized surprisingly early in civil engineering. Some initial ideas regarding the application of fuzzy sets in civil engineering emerged in the early 1970s and were endorsed by the civil engineering community quite enthusiastically [151]. One important category of problems in civil engineering for which fuzzy set theory has already proven useful consists of problems of assessing or evaluating existing constructions [152]. Typical examples of these problems are the assessment of fatigue in metal structure, the assessment of quality of highway pavements, and the assessment of damage in buildings after an earthquake.

It was realized around the mid-1980s, primarily in the context of mechanical engineering design [153], that fuzzy set theory is suited to facilitate the whole design process, including the early stages. The basic idea is that fuzzy sets allow the designer to describe the designed artefact as approximately as desired at the early stages of the design process. This approximate description in terms of imprecise input parameters is employed to calculate the corresponding approximate characterization of relevant output parameters. The latter are compared with given performance criteria and information obtained by this comparison is then utilized to determine appropriate values of input parameters.

Since the mid-1980s, the interest in fuzzy set theory has also been visible in industrial engineering. This interest is primarily connected with the use of fuzzy controllers in manufacturing, fuzzy expert systems for various special areas of industrial engineering, and virtually all types of fuzzy decision making, as well as fuzzy linear programming [154].

Fuzzy set theory is also becoming important in computer engineering and knowledge engineering. Its role in computer engineering primarily involves the design of specialized hardware for fuzzy logic developed by Yamakawa [155]. It consists of units that implement individual inference rules (rule units), units that implement the Max operation needed for aggregating inferences made by multiple inference rules (Max units), and defuzzification
units. Each rule unit is capable of implementing one fuzzy inference rule with three
antecedents or less. One of seven linguistic labels can be chosen for each antecedent and
each consequent: NL (negative large), NM (negative medium), NS (negative small), AZ
(average zero), PS (positive small), PM (positive medium), or PL (positive large).
Fuzzy sets representing these linguistic labels for antecedents may be expressed by
membership functions of either triangular or trapezoidal shapes. The triangles or trapezoids
can be positioned as desired by choosing appropriate values of voltage in the circuits, and
their left and right slopes can be adjusted as desired by choosing appropriate resistance
values of variable resistors in the circuits. For consequents, the shapes of membership
functions may be arbitrary. Its role in knowledge engineering involves knowledge
acquisition, knowledge representation, and human-machine interaction.

Imprecision and uncertainty play a large role in the field of medicine. The field has,
for this reason, become one of the most fruitful and active areas of application for the theory
of fuzzy sets and the theory of evidence. Overviews of these applications are presented by
Adlassnig [156] and Gupta, Martin-Clouaire, and Nikiforuk [157]. The physician generally
gathers knowledge about the patient from the past history, physical examination, laboratory
test results, and other investigative procedures such as X-ray and ultrasonics. The
knowledge provided by each of these sources carries with it varying degrees of uncertainty.

Applications of fuzzy sets within the field of decision-making consist of
‘fuzzifications’ of classical theories of decision-making [158, 159]. While decision making
under conditions of risk and uncertainty have been modelled by probabilistic decision
theories and by game theories. Fuzzy decision theories deal with vagueness or fuzziness
inherent in subjective or imprecise determinations of preferences, constraints, and goals.

Applications of the mathematics of uncertainty and information within the field of
computer science have been quite extensive, particularly in those endeavours concerned
with the storage and manipulation of knowledge in a manner compatible with human
thinking [160]. This includes the construction of database and information storage and
retrieval systems as well as the design of computerized expert systems [161, 162].

An expert system, as the name suggests, models the reasoning process of a human
expert within a specific domain of knowledge in order to make the experience,
understanding, and problem-solving capabilities of the expert available to the non-expert for purposes of consultation, diagnosis, learning, decision support, or research. Most expert systems consist of a domain-specific knowledge base and problem-solving or reasoning algorithms known as an inference engine. The facts, relations, judgements, opinions, and "rules of thumb" contained within the expert knowledge base usually manifest varying degrees of imprecision and uncertainty. Thus the management of uncertainty in the design of expert systems is of key importance for the successful modelling of the reasoning process, and the utility of fuzzy set theory and the theory of evidence for this purpose has been and continues to be extensively studied [163, 164].

The utility of fuzzy set theory in pattern recognition and cluster analysis was already recognized in the mid-1960s, and the literature dealing with fuzzy pattern recognition and fuzzy clustering is now quite extensive [165, 166].

There are some additional application areas of fuzzy set theory where its potential has not yet been fully utilized. One area in which fuzzy set theory has a great potential is psychology [167]. Psychology is not only a field in which it is reasonable to anticipate profound applications of fuzzy set theory, but also one that is very important to the development of fuzzy set theory itself. In particular, the area of psycholinguistics is essential for studying the connection between the human use of linguistic terms in different contexts with the associated fuzzy sets and operations on fuzzy sets. To understand this connection is necessary for properly incorporating the notion of a context into the formalism of fuzzy set theory. This will undoubtedly help us, in the long run, to better understand human communication and design machines capable of communicating with people in a human-friendly way with natural language.

Applications of fuzzy set theory in natural sciences are relatively scarce [148]. This is due to the fact that classical methods based on crisp sets and additive measures have worked quite well in these areas and, consequently, there have been no pressing needs to replace them with the more realistic methods based on fuzzy set theory. One exception is quantum mechanics, where the need for non-classical methods is acute. Based on some preliminary investigation, it is reasonable to expect that fuzzy set theory will play profound roles in quantum mechanics in the near future. Other areas of physics are likely to be affected by fuzzy set theory. Areas of physics in which utility of fuzzy set theory has already
been demonstrated include non-equilibrium thermodynamics and dimensional analysis. Fuzzy set theory is also likely to have a strong impact upon chemistry in near future.

Applications in biology have not been visible so far, which is somewhat surprising since the potential is enormous. Fuzzy sets will undoubtedly play an important role in narrowing down the large gap that currently exists between theoretical and experimental biology. It is reasonable to expect that some applications of fuzzy sets in biology will be profound. In the related area of ecology, a few applications have already been explored.

The science of meteorology deals with a complex system, which, in truth, encompasses the entire planet. For this reason, meteorological descriptions as well as forecasts have always relied on the kind of robust summary offered by vague linguistic terms such as hot weather, draught, or low pressure. Applications of fuzzy set theory to meteorology, therefore, constitute attempts to deal with the complexity of the study by taking advantage of the representation of vagueness offered in this mathematical formalism. Three different applications of fuzzy set theory within the field of meteorology are presented by Cao and Chen [168].

We may continue to speculate about other ways in which fuzzy thinking in its various forms is likely to affect our lives in the future. We may speculate about its effects on education, social policy, law, art, and so on.

3.4 Uncertainty and Fuzzy Set Theory

The concept of information is intimately connected with the concept of uncertainty. The most fundamental aspect of this connection is that uncertainty involved in any problem-solving situation is a result of some information deficiency. Information in reference to the modelling of the real systems may be incomplete, imprecise, fragmentary, not fully reliable, vague, contradictory, or deficient in some other way. In general, these various information deficiencies may result in different types of uncertainty [169].

3.4.1 Types of Uncertainty

In literature, the word uncertain has a broad semantic content. For example, Webster's New Twentieth Century Dictionary gives six clusters of meanings for this term:
1. Not certainly known, questionable, problematical;
2. Vague, not definite or determined;
3. Doubtful, not having certain knowledge, not sure;
4. Ambiguous;
5. Not steady or constant, varying;
6. Liable to change or vary, not dependable or reliable.

On the basis of these meanings, Klir [148] divides uncertainty into two categories: vagueness and ambiguity. In general, vagueness is associated with the difficulty of making sharp or precise distinctions in the world; that is, some domain of interest is vague if it cannot be delimited by sharp boundaries. Ambiguity, on the other hand, is associated with one-to-many relations, that is, situations in which the choice between two or more alternatives is left unspecified.

Each of these two distinct forms of uncertainty—vagueness and ambiguity—is connected with a set of kindred concepts. Some of the concepts connected with vagueness are fuzziness, haziness, cloudiness, uncleanness, indistinctiveness, and sharplessness; some of the concepts connected with ambiguity are nonspecificity, one-to-many relation, variety, generality, diversity, and divergence.

3.4.2 Measures of Uncertainty

It is well known that a measure of uncertainty can also be used for measuring information. That is, the amount of uncertainty regarding some situation represents the total amount of potential information in this situation. According to this view, the reduction of uncertainty by a certain amount indicates the gain of an equal amount of information. A measure of uncertainty, when adopted as a measure of information, does not include semantic and pragmatic aspects of information. As such, it is not adequate for dealing with information in human communication. However, when we are dealing with structural (syntactic) aspects of systems, such a measure is not only adequate but highly desirable. It can be used for measuring the degree of constraint among variables of interest, thus comprising a powerful tool for dealing with systems problems such as systems modelling, analysis, or design.

In general, uncertainty has been expressed in the following two alternative ways:
1. Given a universal set $U$, a fuzzy subset $A$ of $U$ is defined by a function

$$
\mu_A(u) = U \rightarrow [0, 1] \tag{3.17}
$$

such that $\mu_A(u)$ expresses the grade of membership of $u$ in $A$, i.e., the degree of compatibility of $x$ with the concept represented by the fuzzy set $A$. Clearly $\mu_A(u) = 0$ means that $u$ is definitely not a member of $A$, and $\mu_A(u) = 1$ means that $u$ is definitely a member of $A$; if either $\mu_A(u) = 0$ or $\mu_A(u) = 1$ for each $u \in U$, $A$ is a crisp set.

2. Given a universal set $U$ and the set of all its crisp subsets (power set) $\wp(U)$, a function

$$
g : \wp(U) \rightarrow [0, 1] \tag{3.18}
$$

is defined such that $g(A)$ expresses the uncertainty of whether a particular arbitrary element of $U$, which is not a priori located in any of the sets of $\wp(U)$, belongs to a particular crisp set $A$. In order to qualify for this purpose, function $g$ must satisfy the following requirements:

- (g1) $g(\emptyset) = 0$ and $g(U) = 1$ (boundary conditions);
- (g2) for all $A, B \in \wp(U)$, if $A \subseteq B$, then $g(A) \leq g(B)$ (monotonicity);
- (g3) for all $A_i \in \wp(U)$, $i \in N$, if $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ or $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$, then $\lim_{i \to \infty} g(A_i) = g(\lim_{i \to \infty} A_i)$ (continuity).

Alternative I, expressed by function (3.17), was introduced by Zadeh in 1965 [143]. It allows us to define subsets of $U$ that do not have sharp boundaries. We thus deal with the power set $\wp(U)$ of all fuzzy subsets of $U$. Fuzzy sets in $\wp(U)$ can be operated upon by a variety of operators of fuzzy complementation, intersection, union, etc. [170, 171]. Each set of operator forms a special theory of fuzzy sets.

Alternative II, expressed by function (3.18), was proposed by Sugeno in 1977 [172]. It is based on defining a generalized measure, called a fuzzy measure, on the power set $\wp(U)$ of all crisp subsets of $U$. When requirements (g1-g3) are augmented with some additional requirements, various special theories are obtained.
The concept of fuzzy set, defined by (3.17), provides a basic mathematical framework for dealing with vagueness. The need for such a framework was successfully argued by Max Black as early as 1937 in a paper that perhaps contains the best discussion of the concept of vagueness ever written [145].

The concept of a fuzzy measure, defined by (3.18), provides a general framework for dealing with ambiguity. That is, function (3.18) specifies a set of alternative subsets of $U$ that are associated with any given element of $U$ to various degrees according to the available evidence. Three types of ambiguity are defined within this framework. The first is connected with the size of the subsets that are designated by function (3.18) as prospective locations of the element. The larger the subsets, the less specific the characterization. This type of ambiguity has the meaning of nonspecificity in evidence, as expressed by function (3.18).

The second type of ambiguity is exhibited by disjoint subsets of $U$ that are designated by function (3.18) as prospective locations of the element of concern. In this case, evidence focusing on one subset conflicts with evidence focusing on the other subsets. This type of ambiguity thus has the meaning of conflict or dissonance in evidence.

The third type of ambiguity is associated with the numbers of subsets of $U$ that are designated by function (3.18) and that does not overlap totally or overlap partially. The multitude of partially or totally conflicting evidence is a source of confusion. This type of ambiguity therefore characterizes confusion in evidence.

It follows from this preliminary discussion that the two alternative ways of developing fuzzy set theory reflect the two fundamentally different types of uncertainty—vagueness and ambiguity. Consequently each of them is a distinct framework within which appropriate measures of the corresponding type of uncertainty must be formulated.

3.4.3 Techniques for Measuring Uncertainty

It is probably impossible to capture all of the semantics of real world information. Frequently, the observation of and the knowledge about the real world are deficient and as a consequence its modelling and hence its representation is imperfect in some way. Prior to the entry of fuzzy set theory into our mathematical repertory, the only well-developed mathematical apparatus for dealing with uncertainty was probability theory. Although useful
and successful in many applications, probability theory is, in fact, appropriate for dealing with only a very special type of uncertainty. Its limitations have increasingly been recognized [173, 174]. Several mathematical models of uncertainty that depart from probability approach have been proposed during the past three decades. The main ones are Shafer's Evidence theory and Zadeh's possibility theory. There have also been attempts to handle the problems of incomplete information using classical logic. Many default reasoning logics have been proposed and the study of non-monotonic logic is currently gaining much attention. There is a basic difference in the type of incompleteness that is represented by the numeric and non-numeric approaches.

(A) Numeric Approaches

(i) Probability Theory

For a long time this was the only theory employed for handling uncertainty. The main idea can be described as follows:

Let \( P \) be a finite set of propositions such that:

If \( p \in P \), then \( \neg p \in P \) and

If \( p \in P \), \( q \in P \), \( p \land q = F \), then \( p \lor q \in P \)

A probability measure \( M \) is a function from \( P \) to \([0, 1]\) such that

\[ M(F) = 0 \], where \( F \) is the ever false proposition (\( F \in P \)).

\[ M(T) = 1 \], where \( T \) is the ever false proposition (\( T \in P \)).

\( p \in P \), \( q \in P \), then \( M(p \lor q) = M(p) + M(q) \).

These axioms have the following noticeable consequences:

\[ \forall p \in P, M(p) + M(\neg p) = 1 \] and

If \( p \) entails \( q \) (i.e. \( p \rightarrow q = T \)) then \( M(q) \geq M(p) \).

Cheeseman in his paper [175] argues that specifying two numbers, that is, probability and its standard deviation solves the problem. Ignorance can be represented by a
probability number along with a large standard deviation whereas a knowledge packed situation can be represented by a probability number and a very small standard deviation.

Historically, many different concepts of probability have been presented. The two main surviving ones are (1) \( \text{probability}_1 = \text{degree of confirmation} \); (2) \( \text{probability}_2 = \text{relative frequency in the long run} \). As stated by Carnap [176] these two are different explanations for representing different concepts. Confusion has been caused by the use of same term \( \text{probability} \) for both of them. The probability number only reflects an observer’s knowledge about the world. Confusion arises when efforts are made to provide a frequency ratio interpretation for all probability-based knowledge. For performing inference and knowledge representation it is the concept of \( \text{probability}_1 \) that is relevant. Cox [177] has also shown that there is a field of probabilistic inference, which lies outside the range of a theory of probability based on the frequency interpretation.

(ii) Evidence Theory

Shafer introduced this theory in the seventies as an extension to probability theory [178]. If \( q \) is some variable, let \( \Theta \) represents the set of all possible values of \( q \). Then the propositions may be represented as ‘The true value of \( q \) is in \( T \) ’ where \( T \) is a subset of \( \Theta \). This is abbreviated as ‘\( q \) is in \( T \) ’. Thus, the propositions are represented by subsets of the set \( \Theta \). The variable \( q \) can be an arbitrary parameter taking numeric or non-numeric values. The set \( \Theta \) is called the frame of discernment for variable \( q \). \( P(\Theta) \), the power set of \( \Theta \), may be viewed as the set of events, each corresponding to a proposition. The model employs the familiar idea of using a number between zero and one to indicate the degree of support provided to a proposition by the body of available evidence. The numerical function is defined as follows:

Let \( m \) be a function from \( P(\Theta) \) to \( [0, 1] \) such that

\[
m(\emptyset) = 0 \quad \text{and} \quad \sum_{A \subseteq \Theta} m(A) = 1
\]  

(3.19)

\( m \) is called the basic probability assignment and represents a body of evidence regarding the value of \( q \). Unlike many other ideas of the past, this theory does not focus on the act of judgement by which the numbers for the function \( m \) are determined. Instead, it focuses on the combination of degrees of belief based on various distinct bodies of evidence.
Qualitative interpretation of values for the function $m$ gives a good insight into the utility of this theory. The function values such that

$$m(G) = 1 \text{ and } \forall A \neq G, \ m(A) = 0, \ A \subseteq \emptyset$$

(3.20)

represent a body of evidence that is certain. It states that $q$ is in $G$, exactly and certainly. This certain body of evidence does not necessarily lead to precise specification of the value of $q$, unless $G$ is a singleton of $\emptyset$. The quantity $m(A)$ measures the belief that one commits exactly to $A$, and not the total belief committed to elements and subsets within $A$. To obtain the measure of the total belief committed to $A$, one must add to $m(A)$ the quantities $m(B)$ for all proper subsets $B$ of $A$.

$$Bel(A) = \sum_{B \subseteq A} m(B)$$

(3.21)

A subset $A$ of a frame $\emptyset$ is called a focal element of a belief function $Bel$ over $\emptyset$ if $m(A) > 0$. The union of all the focal elements of a belief function is called its core. If $C$ is the core of a belief function $Bel$ over $\emptyset$, then a subset $B$ of $\emptyset$ satisfies $Bel(B) = 1$ if and only if $C \subseteq B$. This function $Bel$ is referred to as Credibility Function represented as $Cr(A)$. By duality, a plausibility function $Pl$ is defined from $Cr$ as

$$\forall A \subseteq \emptyset, \ Pl(A) = 1 - Cr(\neg A)$$

(3.22)

Writing it in terms of function $m$ we get,

$$pl(A) = \sum_{B \subseteq \emptyset} m(B) - \sum_{B \subseteq (\emptyset - A)} m(B) = \sum_{A \subseteq \emptyset, B} m(B)$$

(3.23)

A situation of total ignorance can be represented by an $m$ function such that $m(\emptyset) = 1$ and $\forall A \neq \emptyset, \ A \subset \emptyset, \ m(A) = 0$. In this case we will have $Cr(p) = Cr(\neg p) = 0$ and $Pl(p) = Pl(\neg p) = 1$ for all propositions $p$ corresponding to subsets of $\emptyset$.

In addition to $Cr$ and $Pl$, some other functions of $m$ may be defined which reflect various properties of the evidence. One such function is called the Commonality Number for a subset and is defined as follows:

$$Q(A) = \sum_{B \subseteq \emptyset, A \subseteq B} m(B)$$

(3.24)
Shafer has shown that the basic probability assignment that produces a given belief function is unique and can be recovered from the belief function. The same is true for commonality numbers. Thus, $Cr$, $Pl$, and $Q$ are different macro properties derived from a basic probability mass assignment to a frame of discernment.

One advantage of this approach over a probability framework is that explicit representation of ignorance is possible. The major deficiency is that while it provides a method for representing uncertain information, very little has been done to perform reasoning or inference with information represented in terms of belief functions. Assessment of quantities has been the major drawback for any numerical approach to handle uncertainty. This problem appears when the $m$ function for a frame of discernment must be evaluated from the raw evidence. This evidence in most situations is qualitative in nature and converting it into quantitative parameters is only subjective. Also, updating of the $m$ values in the light of new evidence poses another problem. In the case of probability theory a background provided by prior probabilities can be updated by new information that is available in terms of conditional probabilities. No such formalism exists in the framework of Evidence theory. Also, the new evidence may require creation or deletion of focal elements or readjustment of $m$ values for the same core of evidence. There is a lack of experimental results relating to evidence theory.

(iii) Possibility Theory

There are two periods in the development of fuzzy set theory. The first one (1965-1975) has put emphasis on the notion of a fuzzy class as a class without definite boundaries. The main applications of this idea were fuzzy clustering and classification, a smooth interface between numerical and symbolic knowledge, and the use of interpolative reasoning as done in fuzzy control. From 1975 on, Zadeh developed the idea of a fuzzy set as an elastic constraint on possible situations, parameter values, etc. This research culminated in the late seventies and early eighties by the introduction of possibility theory.

A special branch of evidence theory that deals only with bodies of evidence whose focal elements are nested is referred to as possibility theory. Special counterparts of belief measures and plausibility measures in possibility theory are called necessity measures and possibility measures, respectively. However, the theory of possibility was independently
proposed by Zadeh as a development of Fuzzy Set Theory, for representing vagueness inherent in some linguistic terms [179].

Let $F$ be a fuzzy subset of a universe of discourse $U$, characterized by the membership function $\mu_F$. The grade of membership $\mu_F(u)$ can be interpreted as the compatibility of $u$ with the concept labelled $F$. Let $X$ be a variable taking values in $U$. Then, $F$ acts as a fuzzy restriction, $R(X)$, on the values $X$ can take. The proposition \textit{‘$X$ is $F$’}, which translates into $R(X) = F$, associates a possibility distribution $\Pi_X$ with $X$ which is postulated to be equal to $R(X)$, i.e., $\Pi_X = R(X)$ Possibility distribution function $\pi_x$ is numerically equal to the membership function of $F$, i.e., $\pi_x = \mu_F$. If $U$ is the exhaustive set of the possible values of $X$, at least one value must be completely possible. Then the attached possibility distribution is said to be normalized.

The concept of possibility theory has been built upon fuzzy set theory and is well suited for representing the imprecision of vague linguistic predicates like ‘young’, ‘tall’ etc. The vague predicate ‘young’ induces a fuzzy set and the corresponding possibility distribution. From a semantic point of view, the values restricted by a possibility distribution are more or less all the eligible values for the variable. Cheeseman in [175] has argued that possibility can be viewed as the probability of the proposition ‘event $x$ is possible’ instead of the proposition ‘occurrence of event $x$’. Another useful concept handled well by Possibility theory is that of imprecise quantification. These ideas can then be related to linguistic concepts like most, some, few, majority, etc.

(B) Non-numeric methods

The types of uncertainty and incompleteness that are handled by numeric and non-numeric approaches are qualitatively very different. Numeric methods can carry the uncertainty through the processes of combination of evidence and inference. In non-numeric methods, the uncertainty must be resolved by making assumptions before combination or inferencing can be performed, but the results can be revised when new information is received. The qualitative difference between numeric and non-numeric approaches is that in the former, each premise may be partially believed, but inferences may have a high degree of confidence, whereas in the latter, each premise is fully believed or disbelieved but one’s confidence in the inferences will depend on the amount and type of underlying assumptions.
The decision arrived at by the proof procedure does not say anything about the amount of uncertainty attached to the decision. The non-numeric approach models the incompleteness of information but fails to represent the ignorance about the value of a variable, which is done very well by the method of belief functions in evidence theory. Some of the important non-numeric methods are discussed below:

(i) **Non-monotonic Logic**

It is widely acknowledged that classical logic stems from old attempts at developing a formal model of human reasoning. These attempts mainly come from philosophers. The first half of the century has witnessed significant progress in classical logic as a tool for founding mathematical reasoning. In contrast, the last 30 years, and the emergence of Artificial Intelligence, have pointed out the deficiencies of classical logic as a tool for modelling commonsense reasoning. When inferring from incomplete, uncertain or contradictory information, man does not follow the strict rules of classical logic. Most of the recent research in the domain of problem solving has been restricted to systems based on logic or more specifically, the first order predicate logic (FOPL). In some of the systems that have been attempted, incomplete information is represented as disjunctions of the several possibilities.

Classical symbolic logic lacks tools for describing how to revise a formal theory to deal with inconsistencies caused by new information. The non-monotonic Logics provide a theoretical framework for such updating of world models. This process is well handled by the Truth Maintenance System [TMS] of Doyle [180]. The TMS (also known as belief revision and revision maintenance systems) is a program, which maintains the consistency of the knowledge being used by the problem solver. It maintains it in the form of a tree where a node represents a hypothesis and its children represent various justifications. Each justification may consist of many other hypotheses or assumptions. When some new information causes a contradiction, TMS revises the assumptions to restore the consistency and informs the reasoner about the revisions.

(ii) **Theory of Endorsements**

Another non-numeric approach that has been recently presented is Cohen's Theory of Endorsements [181]. The model has been developed and tested around the program
SOLOMAN, which does decision making for portfolio investment. One stated goal of the effort was to completely avoid the use of any numbers in the representation of uncertainty. According to the author, the strength of evidence in any situation is a summary of several factors that pertain to certainty. The numerical approaches somehow summarize all the pro and con evidence into one single number. Thus, the summary representation is inadequate. The crux of the approach is therefore centred on the idea of dealing with the reasons for believing or disbelieving a hypothesis. The structures of such reasons are called endorsements; reasons for believing being positive endorsements and the reasons for disbelieving, the negative endorsements.

The system uses first order predicate calculus for representing the endorsements and works with tasks that are generated to satisfy goals in accordance with the backward chaining control strategy. Another kind of task that the system handles is a resolution task. It also maintains a database of propositions or statements that are believed. In a rule based inference systems, the conclusion of a rule is asserted if the condition of the rule is satisfied.

(iii) Certainty Factor

Many early systems such as CASNET, INTERNIST and MYCIN worked well to handle uncertainty but had only a weak theoretical basis. In MYCIN, the concept of certainty factors (CFs) was used. In this scheme facts and rules are assigned a number in the interval [0, 1]. The certainties of two premises \( a \) and \( b \) are transmitted to their conjunction by the fuzzy operator \( \text{CF} (a \& b) = \min [\text{CF} (a), \text{CF} (b)] \). With two rules, ‘if \( a \) then \( c \)’ and ‘if \( b \) then \( c \)’, the certainties combine with the function \( 1 - \text{CF} (c) = [1 - \text{CF} (a)] [1 - \text{CF} (b)] \). Also this rule is associative. Such multivalued logics have been presented in the literature in recent years as ‘fuzzy’ logics. Comparing the CF’s approach with evidence theory, we see that the former seems to deal with single proposition in isolation. Evidence theory very efficiently handles this problem by confining all competing hypotheses to a single frame of discernment. The CF’s approach can be viewed as a special case of belief function where each frame has only one focal point consisting of a singleton. A different approach was used in PROPECTOR, an expert system for aiding in geological exploration. The uncertainty calculus of this system is loosely related to probability theory.
3.5 Neural Networks

The processing of information in a computer is done principally in a different way than in human or animal brains. Whereas the computer is strongly influenced by the von Neuman architecture [182], the brain is a gigantic parallel processing system [183]. The processing speed of the brain as compared to that of a computer is something like the speed of a snail compared to that of a car. The superiority of a computer in processing numbers and doing mathematical calculations is obvious. On the other hand there are activities of the brain, such as the pattern recognition, perception, and the motor control, which are many times faster than the fastest digital computer in existence today. Fifth and sixth generation computers have been built based on advance in speed, miniaturization, and parallelization of sophisticated hardware, but the results thus far have largely fallen short of expectations.

The reasons for these failures can be traced to a basic misunderstanding of brain function, which can be attributed to the imagination of an eminent physiologist, Warren McCulloch [184]. In his early work with Walter Pitts [185], he proposed that the action potential, which was then seen as the universal means of communication in the brain, served as a digital token or symbol for the performance of logical processing by networks of neurons that operated in accordance with the rules of Boolean algebra.

This was a misconception of the function of neurons, which are not digital switches, and of the brain, which is not a logical device. However, enormous utility of the conception as applied in artificial intelligence quickly overshadowed its erroneous nature as applied in biology. Computers were developed as devices for the manipulation of symbols according to specified rules. By means of brilliant innovation of von Neumann both the rules and the symbols to be processed could be read into the machines. Artificial intelligence went one step further by taking the rules not from brain function but from the overt behaviour of experts as observed by other experts. Thus replacing studies of the brain by studies in psychology and industry [186].

3.5.1 Historical Perspective

The filed of neural network is not new. Possibly, the first reference to neural network theory was that by McCulloch and Pitts [185]. They modelled a biological neuron based on the simplified concept that "the amount of activity at any given point in the brain is the sum of
all the weighted signals of other points entering into the given point." Replacing the point in the brain with neuron, neural activity was modelled as the sum of weighted signals to neuron. The McCulloch and Pitts neural model resembles what is known as a binary logic device. The McCulloch and Pitts model is also called as Perceptron.

The next major development after the McCulloch and Pitts neural model was proposed, occurred in 1949 when D.O. Hebb [187] proposed a learning mechanism for the brain that became the starting point for artificial neural network learning (training) algorithms. He postulated that as brain learns, it changes its connectivity patterns. More specifically, his learning hypothesis is as follows: "When the axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that the A's efficiency, as one of the cells firing cell B, is increased." Hebb further proposed that if one cell repeatedly assists in firing another, the knobs of the synapse, are the junction, between the cells would grow so as to increase the area of contact.

This idea of learning mechanism was first incorporated in artificial neural network by E. Rosenblatt in 1958 [188]. He combined the simple McCulloch and Pitts model with the adjustable synaptic weights based on Hebbian learning hypothesis to form the first artificial neural network with the capability to learn. The delta rule or the least mean squares (LMS) learning algorithm was developed by Widrow and Hoff in 1960 [189]. This model was called ADALINE for ADAptive LInear NEuron. This learning algorithm first introduced the concept of supervised learning using a teacher, which guides the learning process. It is the recent generalization of this learning rule into the backpropagation algorithm that has led to the resurgence in biologically based neural network research today. This states that if there is a difference between the actual output pattern and the desired output pattern during training then the weights are changed to reduce the difference. The amount of change of weights is equal to the error of outputs times the values of the inputs, times the learning rate. Many networks use some variation of this formula for training.

In 1969 research in the field of artificial neural networks suffered a serious setback. Minsky and Papert published a book called Perceptrons [190] in which they proposed that single layer neural networks have limitations in their abilities to process data, and are capable of any mapping that is linearly separable. They pointed out, carefully applying
mathematical techniques that the logical exclusive OR (XOR) function could not be realized by Perceptrons. Further, Minsky and Papert argued that research into multi-layer neural network could be unproductive. Due to this pessimistic view of Minsky and Papert, the field of artificial neural networks entered into an almost total eclipse for nearly two decades. Fortunately, Minsky and Papert's judgement has been disproved; all non-linear separable problems can be solved by multi-layer perceptron networks. Nevertheless, a few dedicated researchers such as Kohonen, Grosberg, Anderson, Hopfield continued their efforts.

A renaissance in the field of neural networks started in 1982 with the publication of the dynamic neural architecture by Hopfield [191]. This was followed by the landmark publication "Parallel Distributed Processing" by McClelland and Rumelhart [192] who introduced into the backpropagation learning technique for multi-layer neural networks.

In 1988 Linsker described a new principle for self-organization in a perceptual network [193]. The principle is designed to preserve maximum information about input activity patterns, subject to such constraints as synaptic connections and synapse dynamic range. In 1989, Mead's book, Analog VLSI and Neural Systems, was published [194]. In the early 1990s, Vapnik et al. invented a computationally powerful class of supervised learning networks, called support vector machines, for solving pattern recognition, regression, and density estimation problem [195]. It is now well established that Chaos constitutes a key aspect of physical phenomena. According to Freeman [196], patterns of neural activity are not imposed from outside the brain; rather they are constructed from within.

Perhaps more than any other publication, the 1982 paper by Hopfield and the 1986 two-volume book by Rumelhart and McLelland were the most influential publications responsible for the resurgence of interest in neural networks in the 1980s. Neural networks have certainly come a long way from the early days of McCulloch and Pitts. Indeed, they have established themselves as an interdisciplinary subject with deep roots in the neurosciences, psychology, mathematics, the physical sciences, and engineering. Needless to say, they are here to stay, and will continue to grow in theory, design, and applications.

3.5.2 Definition

There is no universally accepted definition of an ANN. Some of the important definitions are as follows:
According to DARPA Neural Network Study [197] "...a neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements or nodes."

According to Haykin [198] "A neural network is massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects: (i) Knowledge is acquired by the network through a learning process and (ii) Interneuron connection strengths known as synaptic weights are used to store the knowledge."

According to Fausett [199] "An Artificial neural network is an information processing system that has certain performance characteristics in common with biological neural networks."

According to Nigrin [200] "A neural network is a circuit composed of a very large number of simple processing elements that are neurally based. Each element operates only on local information. Furthermore each element operates asynchronously, thus there is no overall system clock."

According to Zurada [201] "Artificial neural systems, or neural networks, are physical cellular systems which can acquire, store, and utilize experiential knowledge."

In general, a common definition of neural network may be given as "A neural network is a network of many simple processors ("units"), each having a small amount of local memory. The units are connected by communication channels ("connections"), which usually carry numeric data, encoded by any of various means. The units operate only on their local data and on the inputs they receive via the connections the restriction to local operations is often relaxed during training."

### 3.5.3 Models of a Neuron

A neuron is an information-processing unit that is fundamental to the operation of a neural network. The block diagram of Figure 3.4 shows the model of a neuron, which forms the basis for designing artificial neural networks. Three basic elements of neuron model are [198]:

73
1. A set of *synapses* or *connecting links*, each of which is characterized by a *weight* or *strength* of its own. Specifically, a signal $x_j$ at the input of synapse $j$ connected to neuron $k$ is multiplied by the synaptic weight $w_{kj}$. The first subscript refers to the neuron in question and the second subscript refers to the input end of the synapse to which the weight refers. Unlike a synapse in the brain, the synaptic weight of an artificial neuron may lie in a range that includes negative as well as positive values.

2. An adder for summing the input signals, weighted by the respective synapses of neuron.

3. An activation function for limiting the amplitude of the output of a neuron. The activation function is also referred to as a squashing function in that it squashes (limits) the permissible amplitude range of the output signal to some finite value.

Typically, the normalized amplitude range of the output of a neuron is written in unit interval $[0, 1]$ or alternatively $[-1, 1]$. The neuron model of Figure 3.4 also includes an externally applied bias, denoted by $b_k$. The bias $b_k$ has the effect of increasing or lowering the net input of the activation function, depending on whether it is positive or negative, respectively. In mathematical terms, we may describe a neuron $k$ by writing the following pair of equations:

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \quad (3.25)$$

and

$$y_k = \phi(u_k + b_k) \quad (3.26)$$

where $x_1, \ldots, x_m$ are the input signals; $w_{k1}, \ldots, w_{km}$ are the synaptic weights of neuron $k$; $u_k$ is the linear combiner output due to the input signals; $b_k$ is the bias; $\phi(\cdot)$ is the activation function.
function; and \( y_k \) is the output signal of the neuron. The bias \( b_k \) is used for affine transformation to the output \( u_k \) of linear combiner in the model of Figure 3.4, as shown by

\[ v_k = u_k + b_k \]  \hspace{1cm} (3.27)

In particular, depending on whether the bias \( b_k \) is positive or negative, the relationship between the induced local field or activation potential \( v_k \) of neuron \( k \) and the linear combiner output \( u_k \) is modified in the manner illustrated in Figure 3.5. The bias \( b_k \) is an external parameter of artificial neuron \( k \). Equivalently, we may formulate the combination of Eqs. (3.25) to (3.27) as follows:

\[ v_k = \sum_{j=0}^{m} w_{kj} x_j \]  \hspace{1cm} (3.28)

and

\[ y_k = \phi(v_k) \]  \hspace{1cm} (3.29)

In Eq. (3.28) we have added a new synapse. Its input is

\[ x_0 = +1 \]  \hspace{1cm} (3.30)

and its weight is,

\[ w_{k0} = b_k \]  \hspace{1cm} (3.31)

![Figure 3.5 Affine transformation produced by the presence of a bias](image)

We may therefore reformulate the model of neuron \( k \) as in Figure 3.6. In this figure, the effect of the bias is accounted for by doing two things: (1) adding a new input signal fixed at +1, and (2) adding a new synaptic weight equal to the bias \( b_k \).
3.5.4 Types of Activation Function

The activation function denoted by \( \phi(v) \), defines the output of a neuron in terms of the induced local field \( v \). Three basic types of activation functions are [198]:

1. **Threshold Function.** For this type of function, described in Figure 3.7 (a), we have

\[
\phi(v) = \begin{cases} 
1 & \text{if } v \geq 0 \\
0 & \text{if } v < 0 
\end{cases} \quad (3.32)
\]

In engineering literature, this form of a threshold function is commonly referred to as a Heaviside function. Correspondingly, the output of a neuron \( k \) employing such a threshold function is expressed as

\[
y_k = \begin{cases} 
1 & \text{if } v_k \geq 0 \\
0 & \text{if } v_k < 0 
\end{cases} \quad (3.33)
\]

where \( v_k \) the induced local field of the neuron; that is,

\[
v_k = \sum_{j=1}^{m} w_{kj} x_j + b_k \quad (3.34)
\]

such a neuron is referred to in the literature as the **McCulloch-Pitts model**, in recognition of the pioneering work done by McCulloch and Pitts [145]. In this model, the output of a neuron takes on the value of 1 if the induced local field of that neuron is nonnegative, and 0 otherwise. This statement describes the **all-or-none property** of the McCulloch-Pitts model.
2. **Piecewise-Linear Function.** For the piecewise-linear function described in Figure 3.7(b) we have

\[
\phi(v) = \begin{cases} 
1, & v \geq +1/2 \\
0, & v \leq -1/2 \\
v, & +1/2 > v > -1/2 
\end{cases}
\]  

(3.35)

where the amplification factor inside the linear region of operation is assumed to be unity. This form of activation function may be viewed as an approximation to a non-linear amplifier. The following situations may be special forms of the piecewise-linear function:

- A linear combiner arises if the linear region of operation is maintained without running into saturation.

- The piecewise-linear function reduces to a **threshold function** if the amplification factor of the linear region is made infinitely large.

3. **Sigmoid Function.** The sigmoid function is the most common form of activation function used in the construction of artificial neural networks. It is defined as a strictly increasing function that exhibits a graceful balance between linear and non-linear behaviour [202]. An example of the sigmoid function is the **logistic function** [203], defined by

\[
\phi(v) = \frac{1}{1 + \exp(-av)}
\]  

(3.36)

where \(a\) is the **slope parameter** of the sigmoid function. By varying the parameter \(a\), we obtain sigmoid functions of different slopes, as illustrated in Figure 3.7(c). In fact, the slope at the origin equals \(a/4\). In the limit, as the slope parameter approaches infinity, the sigmoid function becomes simply a threshold function. Whereas a threshold function assumes the value of 0 or 1, a sigmoid function assumes a continuous range of values from 0 to 1. Also the sigmoid function is differentiable, whereas the threshold function is not. Differentiability is an important feature of neural network theory.

The activation function defined in Eqs. (3.32), (3.35), and (3.36) range from 0 to +1. It is sometimes desirable to have the activation function range from −1 to +1, in which case the activation function assumes an anti-symmetric form with respect to the origin; that is, the
activation function is an odd function of the induced local field. Specifically, the threshold function of Eq. (3.32) is now defined as

![Activation Function Diagrams](image)

Figure 3.7 (a) Threshold function. (b) Piecewise-linear function. (c) Sigmoid function for varying slope parameter $a$. 

78
\[
\phi(v) = \begin{cases} 
1 & \text{if } v > 0 \\
0 & \text{if } v = 0 \\
-1 & \text{if } v < 0
\end{cases}
\] (3.37)

which is commonly referred to as the \textit{signum function}. For the corresponding form of a sigmoid function we may use the \textit{hyperbolic tangent function}, defined by

\[
\phi(v) = \tanh(v)
\] (3.38)

### 3.5.5 Artificial Neural Network

Artificial neural networks have been developed as generalizations of mathematical models of human cognition or neural biology, based on the assumptions that:

(i) Information processing occurs at many simple elements called neurons.

(ii) Signals are passed between neurons over connection links.

(iii) Each connection link has an associated weight, which, in a typical neural net, multiplies the signal transmitted.

(iv) Each neuron applies an active activation function (usually nonlinear) to its net input (sum of weighted input signals) to determine its output signal.

A neural net consists of a large number of simple processing elements called neurons, units, cells, or nodes. Each neuron is connected to other neurons by means of directed communication links, each with an associated weight. The weights represent information being used by the net to solve a problem. Each neuron has an internal state, called its \textit{activation} or \textit{activity level}, which is a function of the inputs it has received. Typically, a neuron sends its activation as a signal to several other neurons.

A neural network is characterized by (1) its pattern of connection between the neurons (called its \textit{architecture}), (2) its method of determining the weights on the connections (called its \textit{training}, or \textit{learning}, algorithm), and (3) its \textit{activation function}.

The McCulloch-Pitts model of a Neuron is mathematically described as:

\[
O_i = a(f(B_i, W_{i1} * I_1, W_{i2} * I_2...))
\] (3.39)
where \( a \) is the activation function, \( W_1, W_2, \ldots \) are weights incorporated with inputs \( I_1, I_2, \ldots \), \( B_i \) is the bias weight or threshold of the neuron and \( f \) is the function estimating the net input.

### 3.5.6 Biological Neural Networks

There is a close analogy between the structure of a biological neuron (i.e., a brain or nerve cell) and artificial neuron. In fact, the structure of an individual neuron varies much less from species to species than does the organization of the system.

A biological neuron has three types of components that are of particular interest in understanding an artificial neuron: its dendrites, soma, and axon. The many dendrites receive signals from other neurons. The signals are electric impulses that are transmitted across a synaptic gap by means of a chemical process. The action of the chemical transmitter modifies the incoming signal (by scaling the frequency of the signals that are received) in a manner similar to the action of the weights in an artificial neural network [199].

The soma, or cell body, sums the incoming signals. When sufficient input is received, the cell fires; that is, it transmits a signal over its axon to other cells. It is often supposed that a cell either fires or does not at any instant of time, so that transmitted signals can be treated as binary. However, the frequency of firing varies and can be viewed as a signal of either greater or lesser magnitude. This corresponds to looking at discrete time steps and summing all activity (signals received or signals sent) at a particular point of time.

The transmission of the signal from a particular neuron is accomplished by an action potential resulting from differential concentration of ions on either side of the neuron’s axon sheath (the brain’s “white matter”). The ions most directly involved are potassium, sodium, and chloride.

A generic biological neuron is illustrated in Figure 3.8, together with axons from two other neurons (from which the illustrated neuron could receive signals) and dendrites for two other neurons (to which the original neuron would send signals).

Several key features of the processing elements of artificial neural networks are suggested by the properties of biological neurons, viz., that:
Figure 3.8 Biological neuron

1. The processing element receives many signals.
2. Signals may be modified by a weight at the receiving synapse.
3. The processing element sums the weighted inputs.
4. Under appropriate circumstances (sufficient input), neuron transmits a single output.
5. The output from a particular neuron may go to many other neurons.

Other features of artificial neural networks suggested by biological neurons are:

6. Information processing is local
7. Memory is distributed:
   a. Long-term memory resides in the neurons’ synapses or weights.
   b. Short-term memory corresponds to the signals sent by the neurons.
8. A synapse’s strength may be modified by experience.
9. Neurotransmitters for synapses may be excitatory or inhibitory.

Yet another important characteristic that artificial neural networks, share with biological neural systems is *fault tolerance*. Biological neural systems are fault tolerant in two respects. First, we are able to recognize many input signals that are somewhat different from any signal we have seen before. An example of this is our ability to recognize a person in a picture we have not seen before or to recognize a person after a long period of time.

Second we are able to tolerate damage to the neural system itself. Humans are born with as many as 100 billion neurons. Most of these are in the brain, and most are not replaced when they die. In spite of our continuous loss of neurons, we continue to learn.
Even in cases of traumatic neural loss, other neurons can sometimes be trained to take over the functions of the damaged cell. In a similar manner, artificial neural networks can be designed to be insensitive to small damage to the network, and the network can be retrained in cases of significant damage (e.g., loss of data and some connections).

3.5.7 Applications

The study of neural networks is an extremely interdisciplinary field, both in its development and its application. Some of the areas in which neural networks are currently being applied are discussed below:

There are many applications of neural networks in the general areas of signal processing. One of the first commercial applications was (and still is) to suppress noise on a telephone line. The neural net used for this purpose is a form of ADALINE. The need for adaptive echo cancelers has become more pressing with the development of transcontinental satellite links for long-distance telephone circuits [204]. One of the most common examples of application of neural net in control problem is ‘truck backer-upper’ [205]. This is used to provide steering directions to a trailer truck attempting to back up to a loading dock. Information is available describing the position of the cab of the truck, the position of the rear of the trailer, the (fixed) position of the loading dock, and the angles that the truck and the trailer make with the loading dock. The neural net is able to learn how to steer the truck in order for the trailer to reach the dock, starting with the truck and trailer in any initial configuration that allows enough clearance for a solution to be possible.

Many interesting problems fall into the general area of pattern recognition. One specific area in which many neural network applications have been developed is the automatic recognition of handwritten characters (digits or letters). The large variation in sizes, positions, and styles of writing make this a difficult problem for traditional techniques. General purpose multilayer neural nets, such as the backpropagation net have been used for recognizing handwritten zip codes [206]. One of many examples of the application of neural networks to medicine was developed by Anderson et al. in the mid-1980s. It has been called the ‘Instant Physician’ [207]. The idea behind this application is to train an auto-associative memory neural network to store a large number of medical records, each of which includes information on symptoms, diagnosis, and treatment for a particular case. After training, the
net can be presented with input consisting of a set of symptoms; it will then find the full stored pattern that represents the 'best' diagnosis and treatment.

Learning to read English text aloud is a difficult task, because the correct phonetic pronunciation of a letter depends on the context in which the letter appears. A traditional approach to the problem would typically involve constructing a set of rules for the standard pronunciation of various groups of letters, together with a look-up table for the exceptions. One of the most widely known examples of a neural network approach to the problem of speech production is NETtalk [208], a multilayer neural net. In contrast to the need to construct rules and look-up tables for the exceptions, NETtalk’s only requirement is a set of examples of the written input, together with the correct pronunciation for it. The written input includes both the letter that is currently being spoken and three letters before and after it. Additional symbols are used to indicate the end of a word or punctuation. The net is trained using the 1,000 most common English words. After training, the net can read new words with very few errors.

Progress is being made in the difficult area of speaker-independent recognition of speech. A number of useful systems now have a limited vocabulary or grammar or require retraining for different speakers. Several types of neural networks have used for speech recognition, including multilayer nets or multilayer nets with recurrent connections, Lippmann [209] summarizes the characteristics of many of these nets. One net that is of particular interest, both because of its level of development toward a practical system and because of its design, was developed by Kohonen [210] using the self-organizing map. He calls his net a ‘phonetic typewriter’.

Neural networks are being applied in a number of business settings. The mortgage assessment work by Nestor, Inc. is a good example of neural applications in business. Although it may be thought that the rules, which form the basis for mortgage underwriting are well understood, it is difficult to specify completely the process by which experts make decisions in marginal cases. In addition, there is a large financial reward for even a small reduction in the number of mortgage that becomes delinquent. The basic idea behind the neural network approach to mortgage risk assessment is to use past experience to train the net to provide more consistent and reliable evaluation of mortgage applications [199].
3.6 Fuzzy Neural Systems

Artificial neural networks (ANNs) are well known massively parallel computing models which have exhibited excellent behaviour in the resolution of complex artificial intelligence problems. However, many researchers refuse to use them because of their shortcoming of being "black boxes." The word black box is used here in the sense that it is a difficult task to determine how an ANN makes a particular decision. This is a significant weakness because it is hard to trust the reliability of networks addressing real-world problems without the ability to produce comprehensive decisions.

On the other hand, fuzzy rule-based systems (FRBSs), developed using logic, have become a field of active research during the last few years. These algorithms have proved their strengths in tasks such as the control of complex systems, producing fuzzy control. But fuzzy set theory also provides an excellent way of modelling knowledge.

The relation between both worlds (ANNs and FRBSs) has been extensively studied. Indeed, there is a close relation since equivalence results have been obtained. However, all of these results are approximate. Benitez et al. [211] further established not just the equivalence but equality between ANNs and FRBSs that use comprehensive fuzzy rules. This connection yields two immediate and important conclusions. First, we can apply what has been discovered for one of the modes to the other. Second, we can translate the knowledge embedded in the neural network into a more cognitive acceptable language, fuzzy rules. In other words, we obtain an understandable interpretation of neural nets.

Fuzzy systems and neural networks contain their own advantages and drawbacks. One area of combining them, popularly known as fuzzy neural networks, seeks maximization of desirable properties and the reduction of disadvantages in both the systems. Several methods of fusion from fuzzy systems and neural networks are reported in literature [212—217]. Different learning strategies are used in those applications e.g. unsupervised learning [218], supervised learning [192] and differential competitive learning [216].

3.6.1 Fuzzy Rule-Based Systems

Rules, in general, represent in a natural way causality relationships between inputs and outputs of a system, corresponding to the usual linguistic construction "IF a set of conditions
is satisfied, THEN a set of consequences is inferred." Fuzzy logic [219, 220] provides a natural tool to model and process uncertainty, hence, fuzzy rules have the additional advantage over classical production rules of allowing a suitable management of vague and uncertain knowledge. They represent knowledge using linguistic labels instead of numeric values, thus, they are more understandable for humans and may be easily interpreted.

Systems using fuzzy rules are termed FRBSs [221, 222]. These systems map $n$-dimensional spaces into $m$-dimensional spaces. As depicted in Figure 3.9, they are composed of four parts: fuzzifier, knowledge base, inference engine, and defuzzifier. The fuzzifier converts real valued inputs into fuzzy values. The knowledge base includes fuzzy rule base and database. Fuzzy rules have the form

$$R_i: \text{If } x_1 \text{ is } A_1^i \text{ and } x_2 \text{ is } A_2^i \text{ and } \ldots \text{ and } x_n \text{ is } A_n^i, \text{ then } y \text{ is } B_i$$

where $x_1, \ldots, x_n$ are the inputs, $y$ is the output and $A_1^i, \ldots, A_n^i,$ and $B_i$ are linguistic labels. Membership functions of these linguistic terms are contained in the database. The inference engine calculates fuzzy output from fuzzy inputs using fuzzy implication function and finally the defuzzifier yields a real-value output from the inferred fuzzy output.

![Figure 3.9 Structure of fuzzy rule based systems](image)

3.6.2 Fuzzy systems as feedforward neural networks

Fuzzy systems can be mapped into feedforward type of neural networks. Those systems are denoted as fuzzy neural networks. The neurons used in many fuzzy neural networks have a slightly different structure. The activation $a$ is restricted to be either linear or sigmoidal. For
the function $f$ multiplication and the soft versions of minimum [219] and maximum can also be used instead of the summation.

Soft minimum (Min) and soft maximum (Max) can be described as:

$$\text{Min}_K(I_1...I_n) = \frac{\sum_{i=1}^{n} I_i e^{-KJ_i}}{\sum_{i=1}^{n} e^{-KJ_i}}$$

(3.40)

$$\text{Max}_K(I_1...I_n) = \frac{\sum_{i=1}^{n} I_i e^{-KJ_i}}{\sum_{i=1}^{n} e^{-KJ_i}}$$

(3.41)

where $K$ is a variable, that can be virtually increased to infinity, together with the network reaching the convergence.

**Rule Neurons**

The premise of a fuzzy rule can be implemented as a rule neuron. Three types of rules are considered for implementation.

- Simple rules with premises containing a single fuzzy variable.
- Conjunctive rules with many fuzzy variables in premises.
- Disjunctive rules with many fuzzy variables in premises.

The implementation of Min-Max, Product-Sum or Lukasiewicz T-norms and T-Conorms are straightforward using the neuron model used. In our system, conjunction and disjunction operators are implemented as Min and Max, respectively; e.g. the premise of a conjunctive rule with two inputs can be implemented with a single neuron using the parameters: $a = \text{linear}, B_i = 0, f = \text{Min}$, and weights $W_{11} = W_{12} = 1$.

**Antecedent membership functions**

Virtually any membership function can be obtained using a multi-layer perceptron network that has to be separately trained. But using a lesser number of neurons, and exploiting the possibilities of shifting, scaling and reflecting the sigmoidal transfer function, a satisfactory
solution can be reached, without elaborate training. Considering three possible adjectives: low \((L)\), medium \((M)\), and high \((H)\), the formation of membership function is illustrated (Figure 3.10). Two sigmoidal functions are useful in creation of membership functions:

![Image of membership functions]

**Figure 3.10 Antecedent membership functions**

\[
a_1[I, C, \alpha] = \frac{1}{1 + e^{-C(I-\alpha)}} \quad \text{(3.42)}
\]

\[
a_2[I, C, \alpha] = \frac{1}{1 + e^{C(I-\alpha)}} \quad \text{(3.43)}
\]

The sigmoid (4) is the mirror reflection of the sigmoid (5) on the \(Y\)-axis. Gain \(C\) is a positive variable, used to change the steepness of the sigmoid curve (e.g. if \(C\) goes to infinity, the sigmoid curve tends to be the step function), and \(\alpha\) is a positive variable employed in shifting the sigmoids:

- **Low:** \(L = a_2[I, C_L, C_L, \alpha_L]\) \quad \text{(3.44)}
- **High:** \(H = a_1[I, C_H, C_H, \alpha_H]\) \quad \text{(3.45)}
- **Medium:** medium can be realized in different ways. One way is to use two sigmoid neurons from both types. A third linear neuron with \(f = \text{Min}\) is connected to the two sigmoid neurons by fixed connection weights of unity.

\[
M_1 = a_1[I, C_{M1}, C_{M1}, \alpha_{M1}] \quad \text{(3.46)}
\]

\[
M_2 = a_2[I, C_{M2}, C_{M2}, \alpha_{M2}] \quad \text{(3.47)}
\]

\[M = \text{Min} \{M_1, M_2\} \quad \text{(3.48)}
\]
Another way of implementing medium is to subtract one shifted sigmoid neuron from another shifted sigmoid neuron using a third linear neuron:

\[ M_1 = a_1 [I_b, C_{M1}, \alpha_{M1}] \]  
\[ M_2 = a_2 [I_b, C_{M2}, \alpha_{M1}] \]  
\[ M = \sum \{M_1, -M_2\} \]

The value of \( \alpha_{M1} < \alpha_{M2} \) in both cases.

**Consequent membership functions**

Let \( U \) denote the finite set of possible normalized output values from the rule inference of a fuzzy system:

\[ U = \{U_1, U_2, ..., U_n\} \tag{3.52} \]

where \( \forall i: 0 \leq U_i \leq 1 \). The fuzzy output must be defuzzified to get the crisp output \( U_{out} \).

The set of maximum membership values \( \mu \) (from the envelop shown in Figure 3.11) corresponding to the set \( U \) can be denoted as

\[ \mu = \{\mu_1, \mu_2, ..., \mu_n\} \tag{3.53} \]

Figure 3.11 shows the creation of consequent membership values using two layers of linear neurons having \( f = \text{Maximum} \). Each neuron in the first layer denotes an output membership function. Both layers are connected with weights corresponding to maximum possible membership values of the low, medium and high adjectives. Whenever a rule strength reaches the highest possible value (in normalized case 1), the fuzzy output corresponds to consequent membership curves shown in-dashed lines. Lower rule strengths represent the values of an appropriately scaled curve (e.g. solid curve\( \mu_i \)). The curve \( \mu_i^{a} \), is the modified fuzzy output, where \( \alpha_i \in \alpha_1 \) and \( \alpha_1 \) is the set variable confidence measure. This can be considered as an addition of nonlinear noise to the fuzzy output. Consequent membership functions can be tuned by adjusting the weights connecting the layers in away similar to the tuning of antecedent membership functions.
Defuzzification

Standard defuzzification methods such as centre of gravity (COG) can be implemented as neural networks [219],

\[
U_{\text{out}}^{\text{COG}} = \frac{\sum_{i=1}^{n} \mu_i U_i}{\sum_{i=1}^{n} \mu_i} \tag{3.54}
\]

The easiest way of implementing application oriented defuzzification is to use a trainable perceptron neural network. All the rule strengths are weighted and added to the output neuron with sigmoidal activation function. This is also called black box defuzzification in [215].