CHAPTER 4

Case Studies

4.1 Introduction

FL has been applied in many fields and to many applications. Frequently, one must already be an expert in a particular field to understand the details of the application. While it is tempting to provide the reader with many diverse applications, to demonstrate the wide-range of applicability for type-2 FLSs, I have chosen not to do this. By now, it is already well established that FL is widely applicable, so I don't need to further demonstrate this. Also, I don't want the reader who is unfamiliar with the details of an application to feel left out. So, I have chosen two applications, which I treat as *case studies*, that I believe can be easily understood by all readers. These applications are used throughout the rest of this book, and are *forecasting of time-series* and *knowledge mining using surveys*. Other applications are described in Chapter 14.

4.2 Forecasting of Time-Series

Let $s(k)$ $(k = 1, 2, \ldots, N)$ be a time series, such as daily temperatures of Sante Fe, New Mexico, or hourly measurements of the Dow-Jones stock index. Measured values of $s(k)$ are denoted $x(k)$, where $x(k) = s(k) + n(k)$ and $n(k)$ denotes measurement errors—noise. The problem of forecasting a time-series (i.e., prediction) is:

> Given a window of p past measurements of $s(k)$, namely $x(k-p+1)$, $x(k-p+2), \ldots, x(k)$, determine an estimate of a future value of s, $\hat{s}(k+l)$,

where p and l are fixed positive integers.

Note that if the measurements are noise-free (i.e., perfect), then $x(k-p+1)$, $x(k-p+2)$, ..., $x(k)$ in this problem statement are replaced by $s(k-p+1)$, $s(k-p+2)$, ..., $s(k)$.

Forecasting is a very important problem that appears in many disciplines. Better weather forecasts can, for example, save lives in the event of a catastrophic hurricane; better financial forecasts can improve the return on an investment; etc.

When $l = 1$, we obtain the single-stage forecaster of s, when $l = 2$ we obtain the two-stage forecaster of s, and, in general, for arbitrary values of l, we obtain an l-stage forecaster. For illustrative purposes, we shall focus our attention on the case of $l = 1$.

Suppose that we are given a collection of N data points, $x(1), x(2), \ldots, x(N)$. Then, as is commonly done when neural networks are used to forecast a time-series, we shall partition this data set into two subsets: a *training* data subset with D data points, $x(1), x(2), \ldots, x(D)$, and a *testing* data subset with $N - D$ data points, $x(D+1), x(D+2), \ldots, x(N)$. Because we will use a window of p data points to forecast the next data point, there are at most $D-p$ training pairs, $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(D-p)}$, where

$$\begin{aligned}
\mathbf{x}^{(1)} &= [x(1), x(2), \ldots, x(p), x(p+1)]^T \\
\mathbf{x}^{(2)} &= [x(2), x(3), \ldots, x(p+1), x(p+2)]^T \\
&\cdots \\
\mathbf{x}^{(D-p)} &= [x(D-p), x(D-p+1), \ldots, x(D-1), x(D)]^T
\end{aligned} \tag{4-1}$$

In (4-1), the first p elements of $\mathbf{x}^{(t)}$ are the inputs to the forecaster and the last element of $\mathbf{x}^{(t)}$ is the desired output of the forecaster, i.e.,

$$\mathbf{x}^{(t)} = [p \times 1 \text{ input, desired output}]^{\mathrm{T}} = [x_1^{(t)}, x_2^{(t)}, \ldots, x_p^{(t)}, x_{p+1}^{(t)}]^T \tag{4-2}$$

where $t = 1, 2, \ldots, D-p$. The training data are used in a fuzzy logic system (FLS) forecaster to establish its rules.

There are at least three ways to extract rules from the numerical training data:

1. Let the data establish the centers of the fuzzy sets that appear in the antecedents and consequents of the rules.

2. Pre-specify fuzzy sets for the antecedents and consequents and then associate the data with these fuzzy sets.
3. Establish the architecture of a FLS and use the data to optimize its parameters.

We briefly describe these approaches next.

Because a predicted value of s will depend on p past values of x, there will be p antecedents in each rule. Let these p antecedents be denoted $x_1, x_2, \ldots, x_p$. The interesting feature of time-series forecasting is that, although each rule has p antecedents, these antecedents are all associated with the *same* variable, e.g., daily temperature in Sante Fe, and so is the consequent.

4.2.1 Extracting rules from the data

Method 1: *Let the data establish the centers of the fuzzy sets that appear in the antecedents and consequents of the rules.*

For purposes of single-stage forecasting, here are $D-p$ rules that we can extract from the $D-p$ training pairs, $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(D-p)}$ [Mendel (1995)]:

$$R^1\text{: IF } x_1 \text{ is } F_1^1 \text{ and } x_2 \text{ is } F_2^1 \text{ and } \cdots \text{ and } x_p \text{ is } F_p^1 \text{, THEN } y \text{ is } G^1$$

In this rule, which is obtained from $\mathbf{x}^{(1)}$, F_1^1 is a fuzzy set whose membership function is centered at $x(1)$, F_2^1 is a fuzzy set whose membership function is centered at $x(2)$, ..., F_p^1 is a fuzzy set whose membership function is centered at $x(p)$, and G^1 is a fuzzy set whose membership function is centered at $x(p+1)$.

$$R^2\text{: IF } x_1 \text{ is } F_1^2 \text{ and } x_2 \text{ is } F_2^2 \text{ and } \cdots \text{ and } x_p \text{ is } F_p^2 \text{, THEN } y \text{ is } G^2$$

In this rule, which is obtained from $\mathbf{x}^{(2)}$, F_1^2 is a fuzzy set whose membership function is centered at $x(2)$, F_2^2 is a fuzzy set whose membership function is centered at $x(3)$, ..., F_p^2 is a fuzzy set whose membership function is centered at $x(p+1)$, and G^2 is a fuzzy set whose membership function is centered at $x(p+2)$.

• • • • • • • • • •

$$R^{D-p}: \text{IF } x_1 \text{ is } F_1^{D-p} \text{ and } x_2 \text{ is } F_2^{D-p} \text{ and } \cdots \text{ and } x_p \text{ is } F_p^{D-p}, \text{ THEN } y \text{ is } G^{D-p}$$

In this rule, which is obtained from $\mathbf{x}^{(D-p)}$, F_1^{D-p} is a fuzzy set whose membership function is centered at $x(D - p)$, F_2^{D-p} is a fuzzy set whose membership function is centered at $x(D - p + 1)$, ..., F_p^{D-p} is a fuzzy set whose membership function is centered at $x(D - 1)$, and G^{D-p} is a fuzzy set whose membership function is centered at $x(D)$.

In this first approach to obtaining rules from numerical data, we see that the centers of the antecedent and consequent membership functions are completely determined by the data that are used to create the rules. Usually, all other membership function parameters (e.g., the standard deviation of a Gaussian membership function) are specified ahead of time by the designer.

Method 2: *Pre-specify fuzzy sets for the antecedents and consequents and then associate the data with these fuzzy sets.*

In this second approach ([Wang (1992a, 1994)], [Wang and Mendel (1992c)]), we begin by establishing fuzzy sets for all the antecedents and the consequent. This is done by first establishing domain intervals for all input and output variables. For the example of time-series forecasting, these domain intervals are all the same, because $x_1, x_2, \ldots, x_p$ and y are all sampled values of the measured time series, $x(k)$ (k = 1, 2, ...). Let us assume that, by examining the measured time-series, we establish that $x(k) \in [X^-, X^+]$ for $\forall k$. Next, we divide this domain interval into a pre-specified number of overlapping regions, where the lengths of these overlapping regions can be equal or unequal. Each overlapping region is then labeled and is assigned a membership function. Resolution in forecasting can be controlled by the coarseness of the fuzzy sets that are associated with $x(k)$. Measured values of $x(k)$ are permitted to lie outside of its domain interval, because if $x(k) > X^+$ then $\mu_X(x(k)) = 1$, or if $x(k) < X^-$ then $\mu_X(x(k)) = 0$.

Fuzzy rules are generated from the given data pairs using the following three-step procedure [Wang and Mendel (1992c)]:

1. Determine the degrees (i.e., the membership function values) of the elements of $\mathbf{x}^{(t)}$. As an example, in Figure 4-1 we consider the case when p = 5. We see that $x_1^{(t)}$ has degree 0.45 in $B2$ and 0.75 in $B1$, $x_2^{(t)}$ has degree 0.2 in $S1$ and 0.75 in $S2$, $x_3^{(t)}$ has degree 0.45 in $S2$ and 0.6 in $S3$, $x_4^{(t)}$ has degree 0.4 in $S1$ and 0.75 in CE, $x_5^{(t)}$ has degree 1.0 in $S1$ and 0.2 in CE, and $x_6^{(t)}$ has

degree 0.3 in $B3$ and 0.6 in $B2$.

2. Assign each variable to the region with maximum degree, e.g., $x_1^{(t)}$ is considered to be $B1$, $x_2^{(t)}$ is considered to be $S2$, $x_3^{(t)}$ is considered to be $S3$, $x_4^{(t)}$ is considered to be CE, $x_5^{(t)}$ is considered to be $S1$, and $x_6^{(t)}$ is considered to be $B2$.
3. Obtain one rule from one pair of input–output data, e.g.,

   IF $x_1^{(t)}$ is $B1$ and $x_2^{(t)}$ is $S2$ and $x_3^{(t)}$ is $S3$ and $x_4^{(t)}$ is CE and $x_5^{(t)}$ is $S1$,
   THEN $y^{(t)}$ is $B2$

This three-step procedure is repeated for the $D-p$ training pairs in equation (4-1), i.e., $t = 1, \ldots, D\text{-}p$.

Because there can be lots of data, it is quite likely that there will be some *conflicting rules*, i.e., rules with the same antecedents but different consequents. We resolve this by assigning a degree, $D(R^t)$, to each rule and accept only the rule from a conflict group that has maximum degree, where

$$D(R^t) \equiv \mu_X(x_1^{(t)})\mu_X(x_2^{(t)})\cdots\mu_X(x_p^{(t)})\mu_X(y^{(t)})\ ; \qquad (4\text{-}3)$$

hence, there will be *at most D–p* linguistic rules of the form just obtained for $R^{(t)}$ in step 3. For our example, we find, from step 1 of our three-step procedure, that $D(R^j) = 0.75\times 0.75\times 0.60\times 0.75\times 1.0\times 0.60 = 0.1519$.

A generalized version of this procedure is described in Chapter 5 as the "one-pass method." By "one-pass" we mean that the data are used just one time to obtain all of the rules. Note that, according to this definition, Method 1 can also be referred to as a one-pass method.

Method 3: *Establish the architecture for a FLS and then use the data to optimize its parameters.*

In this third approach, we fix the architecture of the FLS ahead of time, e.g., we fix the number of rules, the number of rule-antecedents, the shapes of the antecedent and consequent membership functions, the inference method, the t-norm, the kind of fuzzification, and the kind of defuzzifier. The resulting FLS has parameters associated with it that have to be specified. These parameters are optimized by using the data. Sometimes the data are only used one time to do this, in which case even this method could be called "one-pass;" however, many times the data are used multiple times to obtain the best possible performance, in which case this is a multiple-pass method.

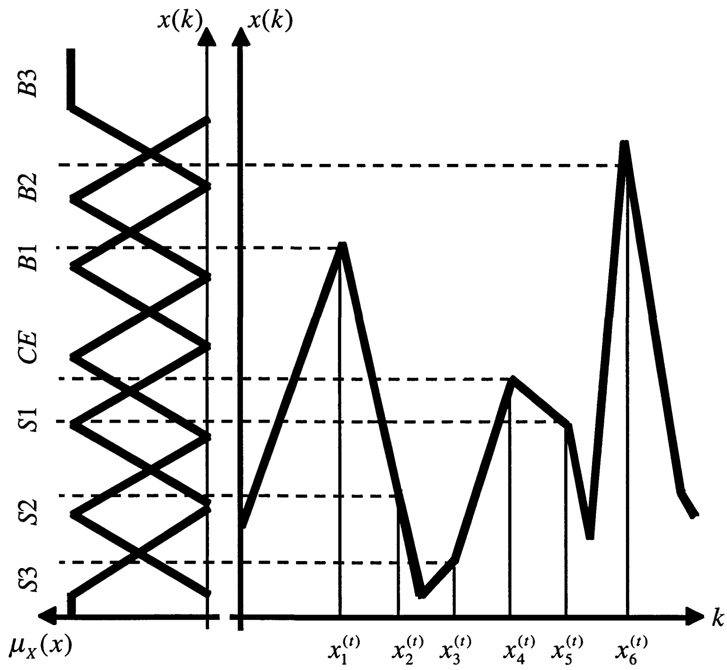

Figure 4-1: Construction to determine degrees of the elements of $\mathbf{x}^{(t)}$ for a representative time-series. The time-series is in the right-hand plot. Membership functions for seven fuzzy sets are in the left-hand plot. Each dashed line projects from a value of the sampled time-series to its intersections with one or two fuzzy sets, which provide the membership function values for that point.

4.2.2 Mackey–Glass chaotic time-series

In the rest of this book (except for Chapters 13 and 14), we shall direct our attention at single-stage forecasting for a specific time-series, one that is chaotic and obtained from the Mackey–Glass equation.

Today, chaos is having an impact on many different fields including physics, biology, chemistry, economics, and medicine (e.g., [Casdagli (1992)], [Farmer (1982)], and [Rasband (1990)]). Very briefly, chaotic behavior can be described as bounded fluctuations of the output of a *non-linear* system with high degree of sensitivity to *initial conditions* [Casdagli (1992)], i.e., trajectories with nearly identical initial conditions can differ a lot from each other. A system exhibiting chaotic dynamics evolves in a *deterministic* manner; however, the correlation of observations from such a system appears to be limited, so the observations appear to be uncorrelated; thus, forecasting for such a system is particularly difficult [Rasband (1990)].

In 1977 Mackey and Glass published an important paper in which they "associate the onset of disease with bifurcations in the dynamics of first-order differential-delay equations which model physiological systems." Equation (4b) of that paper has become known as the *Mackey–Glass equation.* It is a non-linear delay differential equation, one form of which is

$$\frac{ds(t)}{dt} = \frac{0.2s(t-\tau)}{1+s^{10}(t-\tau)} - 0.1s(t) \tag{4-4}$$

For $\tau > 17$ (4-4) is known to exhibit chaos.

To demonstrate the qualitative nature of the Mackey–Glass equation, we display representative portions of the associated Mackey–Glass time series [i.e., the solution of (4-4)] for two values of τ in Figure 4-2 (a) and (b). We also depict the corresponding two-dimensional phase plots in Figure 4-2 (c) and (d). From these plots we are able to distinguish periodic behavior for small values of τ and chaotic behavior for larger values of τ.

The Mackey–Glass time series (for $\tau > 17$) has become one of the benchmark problems for time-series prediction in both the neural network and fuzzy logic fields (e.g., [Lapedus and Farber (1987)], [Moody (1989)], [Moody and Darkin (1989)], [Jones et al. (1990)], [Sanger (1991)], [Wang (1994)], [Hohenson and Mendel (1996)], and [Jang et al. (1997)]).

As we mentioned earlier, all of the single-stage forecasters that we shall design in the rest of this book (except for Chapter 13), using different kinds of FLSs, are based on $D - p$ training pairs. These training pairs are obtained by simulating (4-4) for $\tau = 30$, which we did by converting (4-4) to a discrete-time equation by using Euler's method with a step size equal to 1 [Quinney (1985)]. Because of the 30 time-unit delay, the resulting discrete-time equation requires 31 initial values. These first 31 values of $s(k)$ (i.e., $s(1), s(2), \ldots, s(31)$) were chosen randomly a number of times until a time series was obtained that looked interesting (see Figure 5-11a). Thereafter, the same 31 initial values were used to produce a deterministic (albeit, chaotic) time series that is used in the forecasting studies that are described in Chapters 5, 6, and 10–12.

One of the important features of time-series forecasting is that it can be used to illustrate *all* of the FLSs that are covered in the rest of this book. Table 4-1 summarizes six situations that are covered in later chapters. These situations are distinguished from one another by the natures of the signal, additive measurement noise, training and testing data, and measurements after the design of the forecaster is completed.

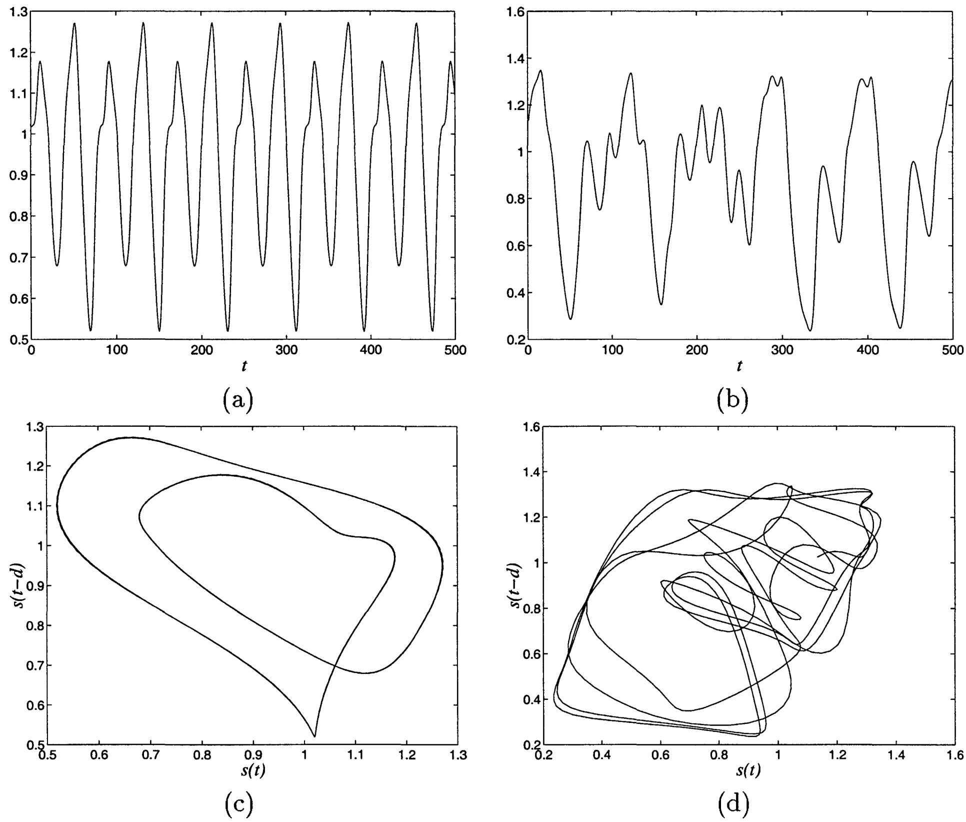

Figure 4-2: (a) and (b) are representative samples of the Mackey–Glass time-series after letting transients relax. (c) and (d) are the corresponding phase plots of the time segments depicted in (a) and (b). Note that "*d*" in $s(t-d)$ on the vertical axis denotes the delay used in the Mackey–Glass equation; it is 13 for (c) and 30 for (d).

❀❀❀❀❀❀❀❀❀❀❀

Table 4-1: Forecasters covered in later chapters of this book. Measurements = signal + noise.

		Distinguishing Features			
Name of Forecaster	**Chapter**	**Signal**	**Measurement Noise**	**Training and Testing Data**	**Measurements After the Design is Completed**
Singleton type-1	5	Deterministic	None	Noise free	Noise free
Non-singleton type-1	6	Deterministic	Stationary	Noisy	Noisy
Singleton type-2	10	Deterministic	Stationary	Noisy	Noise free
Non-singleton type-2 with type-1 inputs	11	Deterministic	Stationary	Noisy	Noisy
Non-singleton type-2 with type-2 inputs	12	Deterministic	Non-stationary	Noisy	Noisy
TSK	13	Random	None	Noise free	Noise free

4.3 Knowledge Mining Using Surveys

Knowledge mining,[1] as used in this book, means extracting information in the form of IF–THEN rules from people. These rules can be modeled using a FLS, which can then be used as a fuzzy logic advisor (FLA) to make decisions or judgments. By "judgment" we mean an assessment of the *level* of a variable of interest. For example, in everyday social interaction, each of us is called upon to make judgments about the meaning of another's behavior (e.g., kindness, generosity, flirtation, harassment). Such judgments are far from trivial, since they often affect the nature and direction of the subsequent social interaction and communications. Although a variety of factors may enter into our decision, behavior (e.g., touching, eye contact) is apt to play a critical role in assessing the level of the variable of interest.

As an engineering example, consider one of the traffic control functions for an asynchronous transfer mode network, called *connection admission control* (CAC).[2] CAC decides whether to accept or reject a telephone call based on the availability of network capacity required to support its quality of service. Here the judgment is associated with the variable CAC. For example, if the total

[1] Another term for this is *knowledge engineering*.
[2] CAC is described in more detail in Section 14.7.

average input rate of real-time voice and video traffic is moderately high and the total average input rate of non-real-time data traffic (e.g., fax) is low, then the confidence of accepting a call is a moderate amount.

4.3.1 Methodology for knowledge mining

In developing a FLA for engineering or social variables, it is useful to adopt the following methodology [Mendel et al. (1999)]:

1. *Identify the behavior of interest.* This step, although obvious, is highly application dependent. For social judgments, we have already mentioned the behaviors of kindness, generosity, flirtation, and harassment; other social variables of interest might be level of violence or amount of sexually explicit material in a television program (leading to, perhaps, a FL V-chip). For engineering judgments, we have mentioned connection admission control; other engineering variables of interest might include toxicity, video quality, sound quality, environmental contamination level, etc.
2. *Determine the indicators of the behavior of interest.* This sometimes requires:

    a. Establishing a list of candidate indicators (e.g., for flirtation [Mendel et al. (1999)], six candidate indicators are touching, eye contact, acting witty, primping, smiling, and complementing).
    b. Conducting a survey in which a representative population is asked to rank-order in importance the indicators on the list of candidate indicators. In some applications it may already be known what the relative importance of the indicators is, in which case a survey is not necessary.
    c. Choosing a meaningful subset of the indicators, because not all of them may be important. In Step 6, where people are asked to provide consequents for a collection of IF–THEN rules by means of a survey, the survey must be kept manageable, because most people do not like to answer lots of questions; hence, it is very important to focus on the truly significant indicators. Factor analysis, from statistics, can be used to help establish the relative significance of indicators.

3. *Establish scales for each indicator and the behavior of interest.* If an indicator is a physically measurable quantity (e.g., temperature, pressure), then the scale is associated with the expected range between the minimum and maximum values for that quantity. On the other hand, many indicators are not measurable by means of instrumentation (e.g., touching, flirtation, harassment, video quality, etc.). Such indicators need to have a scale associated with them, or else it will not be possible to design or activate a FLA. Commonly used scales are 1 through 5, 0 through 5, 0 through 10, etc.

4. *Establish names and interval information for each of the indicator's fuzzy sets and behavior of interest's fuzzy sets*. The issues here are:

   a. What names should be used for the fuzzy sets so that each indicator's scale and the behavior of interest's scale are completely covered by the fuzzy sets?
   b. What are the numerical intervals that a representative group (who will later take the survey in Step 6) associate with the named fuzzy sets?
   c. What is the smallest number of fuzzy sets that should be used for each indicator and behavior of interest?

   Surveys can be used to provide answers to each of these questions. We have already demonstrated in Chapter 2 that words can mean different things to different people; hence, the results of this step's surveys can be used to provide the FOUs for all of the fuzzy sets that will be used in the FLA.
5. *Establish the rules*. Rules are the heart of the FLA; they link the indicators of a behavior of interest to that behavior. The following issues need to be addressed:

   a. How many antecedents will the rules have? As mentioned earlier, people generally do not like to answer complicated questions; so, we advocate using rules that have either one or two antecedents. An interesting (non-engineering) interpretation for a two-antecedent rule is that it provides the correlation effect that exists in the mind of the survey respondent between the two antecedents. Using only one or two antecedents does not mean that a person does not use more than this number of indicators to make a judgment; it means that a person uses the indicators one or two at a time (this should be viewed as a *conjecture*).
   b. How many rule bases need to be established? Each rule base leads to its own FLA. When there is more than one rule base, each of the advisors is a FL *sub-advisor*, and the outputs of these sub-advisors can be combined to create the structure of the overall FLA. If, e.g., we had established that four indicators were equally important for the judgment of flirtation, then there could be up to four single-antecedent rule bases as well as six two-antecedent rule bases. A decision must be made about which of the rule bases would actually be used. This can be done by means of another survey in which the respondents are asked to rank-order the rule bases in order of importance. Later, when (and if) the outputs of the different rule bases are combined, they can be weighted using the results of this step.

6. *Survey people (experts) to provide consequents for the rules*. If, e.g., a single antecedent has five fuzzy sets associated with it, then respondents would be asked five questions. For two-antecedent rules, where each antecedent is again described by five fuzzy sets, there would be 25 questions. The order of the questions should be randomized so that respondents don't correlate their answers from one question to the next. In Step 4 earlier, the names of the consequent fuzzy sets were established. Each rule is associated with a question of the form:

IF antecedent 1 is (state one of antecedent 1's fuzzy sets) and antecedent 2 is (state one of antecedent 2's fuzzy sets), THEN there is _______ of the behavior.

The respondent is asked to choose one of the given names for the consequent's fuzzy sets. The rule base surveys will lead to rule consequent histograms, because everyone will not answer a question the same way.

4.3.2 Survey results

Here we present survey results that will be used in later chapters of this book to design FLAs. We do this for a generic behavior, to illustrate our design procedures, so as not to get lost in the details of a specific social or engineering behavior. Based on the results described in Chapter 2, we used the following five terms for antecedents and consequent: *none to very little*, *some*, *a moderate amount*, *a large amount*, and *a maximum amount*. Table 4-2, which is a repeat of Table 2-3, summarizes the data collected from the Step 4 survey for these labels.

Table 4-2: Processed survey results for labels of fuzzy sets.

		Mean		Standard Deviation	
No.	**Range Label**	**Start (*a*)**	**End (*b*)**	**Start (σ_a)**	**End (σ_b)**
1	*None to very little* (NVL)	0	1.9850	0	0.8104
2	*Some* (S)	2.5433	5.2500	0.9066	1.3693
3	*A moderate amount* (MOA)	3.6433	6.4567	0.8842	0.8557
4	*A large amount* (LA)	6.4833	8.7500	0.7484	0.5981
5	*A maximum amount* (MAA)	8.5500	10	0.7468	0

We limited our FLA to rule bases for one- and two-antecedent rules. In the spirit of generic results, we use x_1 and x_2 to denote the generic antecedents and y to denote the generic consequent for these rules. Tables 4-3 through 4-5 provide the data collected from 47 respondents to the Step 6 surveys. The antecedents for each rule appear in the parentheses after the rule number.

Table 4-3: Histogram of survey responses for single-antecedent rules between indicator x_1 and consequent *y*. Entries denote the number of respondents out of 47 that chose the consequent.

	Consequent				
Rule No.	*None to Very Little (NVL)*	*Some (S)*	*A Moderate Amount (MOA)*	*A Large Amount (LA)*	*A Maximum Amount (MAA)*
1 (NVL)	42	3	2	0	0
2 (S)	33	12	0	2	0
3 (MOA)	12	16	15	3	1
4 (LA)	3	6	11	25	2
5 (MAA)	3	6	8	22	8

Table 4-4: Histogram of survey responses for single-antecedent rules between indicator x_2 and consequent *y*. Entries denote the number of respondents out of 47 that chose the consequent.

		Consequent			
Rule No.	*None to Very Little (NVL)*	*Some (S)*	*A Moderate Amount (MOA)*	*A Large Amount (LA)*	*A Maximum Amount (MAA)*
1 (NVL)	36	7	4	0	0
2 (S)	26	17	4	0	0
3 (MOA)	2	16	27	2	0
4 (LA)	1	3	11	22	10
5 (MAA)	0	3	7	17	20

4.3.3 Methodology for designing a FLA

In Chapter 5 we design a singleton type-1 FLA in which all uncertainties from the surveys, which are summarized in Tables 4-2 through 4-5, are ignored. In Chapter 10, on the other hand, we design a singleton type-2 FLA in which all of the uncertainties from the surveys are accounted for. From the results in Tables 4-2 through 4-5, we see that two sources of uncertainties that were discussed in Chapter 2 are indeed present, namely uncertainties about the words used for the antecedents and consequents (Table 4-2) and uncertainties about the rule consequents (Tables 4-3 through 4-5). Why we will limit our attention just to singleton designs (i.e., to designs in which the FLA is activated by crisp measurements) is clarified next, when we explain how one would make use of a FLA after it is designed.

Table 4-5: Histogram of survey responses for two-antecedent rules between indicators x_1 and x_2 and consequent *y*. Entries denote the number of respondents out of 47 that chose the consequent.

	Consequent				
Rule No.	***None to Very Little (NVL)***	***Some (S)***	***A Moderate Amount (MOA)***	***A Large Amount (LA)***	***A Maximum Amount (MAA)***
1 (NVL/NVL)	38	7	2	0	0
2 (NVL/S)	33	11	3	0	0
3 (NVL/MOA)	6	21	16	4	0
4 (NVL/LA)	0	12	26	8	1
5 (NVL/MAA)	0	9	16	19	3
6 (S/NVL)	31	11	4	1	0
7 (S/S)	17	23	7	0	0
8 (S/MOA)	0	19	19	8	1
9 (S/LA)	1	8	23	13	2
10 (S/MAA)	0	7	17	21	2
11 (MOA/NVL)	7	23	16	1	0
12 (MOA/S)	5	22	20	0	0
13 (MOA/MOA)	2	7	22	15	1
14 (MOA/LA)	1	4	13	17	12
15 (MOA/MAA)	0	4	12	24	7
16 (LA/NVL)	7	13	21	6	0
17 (LA/S)	3	11	23	10	0
18 (LA/MOA)	0	3	18	18	8
19 (LA/LA)	0	1	9	17	20
20 (LA/MAA)	1	2	6	11	27
21 (MAA/NVL)	2	16	18	11	0
22 (MAA/S)	2	9	22	13	1
23 (MAA/MOA)	0	3	15	18	11
24 (MAA/LA)	0	1	7	17	22
25 (MAA/MAA)	0	2	3	12	30

4.3.4 How to use a FLA

Each FLA that we shall design can be referred to as a *consensus* FLA, because it is obtained by using survey results from a population of people. In this section we describe how one can use the resulting FLAs.

Figure 4-3 depicts one way to use a FLA to advise an individual about a social judgment. It assumes that an individual is given the same questionnaire that was used in Step 6 of the knowledge mining process, which led to the consensus FLA. Their completed questionnaire can be interpreted as the individual's FLA, and its output can be plotted on the same plot as the output of the consensus FLA. These outputs can then be compared, and if some or all of the individual's outputs are "far" from those of the consensus FLA, then some action could be taken to sensitize the individual about these differences. Figure 4-4 depicts this for a type-1 consensus FLA, whereas Figure 4-5 depicts this for a type-2 consensus FLA.

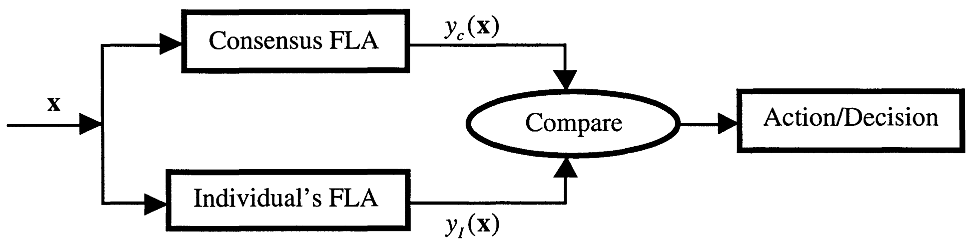

Figure 4-3: One way to use the FLA for a social judgment.

We immediately see a problem with the type-1 comparisons, namely, how "far" must the differences be between the individual FLA and the consensus FLA before some action (e.g., sensitivity training) is taken? This can be difficult to establish when we are comparing two functions, especially since "far" is in itself a fuzzy term.

This problem is handled directly with the type-2 comparisons in Figure 4-5. Note that the individual's FLA is still type-1, and has not changed from Figure 4-4 to 4-5. It is treated as type-1 because the individual takes the survey only one time; hence, there is no uncertainty associated with his or her consequents. The type-2 consensus FLA is represented on Figure 4-5 by two curves, $y_{c2,L}(x)$ and $y_{c2,R}(x)$. These represent the left-hand and right-hand curves, respectively, for the type-reduced sets (which are described in Chapters 9 and 10) of the type-2 consensus FLA. The difference between these curves represents a measure of the uncertainties due to the words used in the surveys as well as the consensus consequents. Observe from Figure 4-5 that the individual's FLA curve falls within the bounds of the type-reduced set; hence, no actions need to be taken. This conclusion is quite different from the one that might have been reached by examining the curves in Figure 4-4 where it appears that there is a significant difference between the individual's behavior level and the consensus FLA's behavior level for larger values of x. How to design the type-1 and type-2 FLAs will be described in Chapters 5 and 10, respectively.

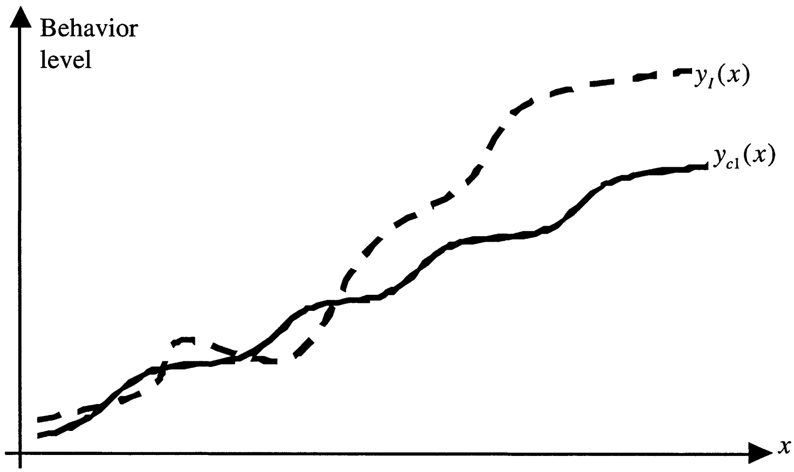

Figure 4-4: Comparison of a type-1 consensus FLA behavior level and an individual's FLA behavior level.

Another way to use a FLA is depicted in Figure 4-6. After the consensus FLA has been designed, it is exposed to a situation, say $x = x'$ (in this discussion **x** is assumed to be a scalar indicator, x), for which it provides the consensus output $y_c(x')$. Then some action or decision occurs. The problem that was associated with the type-1 FLA for a social judgment is the same for an engineering judgment, namely, we would have to take an action or make a decision based only on a point value. This is again resolved by using a type-2 consensus FLA, as depicted in Figure 4-7. Now the region defined by the type-reduced set (i.e., the "uncertain" region) is one where the designer is free to make a decision. For any $y_{c2,L}(x) < f(x) < y_{c2,R}(x)$, using a different decision boundary will lead to different engineering judgment. For example, if we use $y_{c2,L}(x)$ we accept a call when $x = x''$, or, on the other hand, if we use $y_{c2.R}(x)$ we reject the call when

$x = x''$. This kind of soft decision has the potential to be used in network control, signal detection and classification, communication receivers, etc.

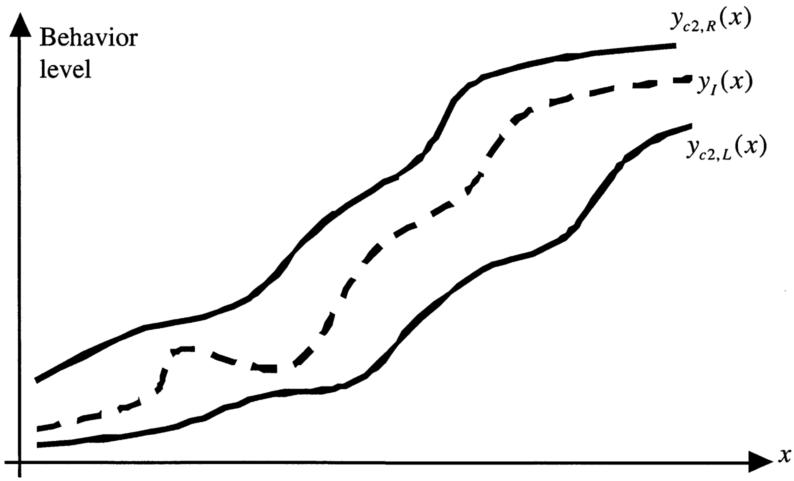

Figure 4-5: Comparison of a type-2 consensus FLA behavior level and an individual's FLA behavior level.

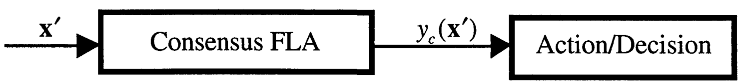

Figure 4-6: A way to use the FLA for an engineering judgment.

EXERCISES

4-1: The way in which we have established the training data to forecast a time-series is to ensure maximum overlap between successive training elements. Many other ways can be created to use the training data that either do not have so much overlap, or do not have any overlap at all between successive training elements.

(a) For the same N data points, create a training set that advances two points to the right, from one element in the training set to the next, instead of just one point to the right [as in (4-1)]. Suppose that the training set is to consist of 50% of all the data. What are the rules for this set of training data? How many rules will there be? What are the testing elements?

(b) For the same N data points, create a training set that has no overlap from one element in the training set to the next. As in part (a), suppose that the training set is to consist of 50% of all the data. What are the rules for this set of training data? How many rules will there be? What are the testing elements?

4-2: Explain some other ways to use a FLA for:

(a) social judgments

(b) engineering judgments

4-3: Suppose that the FLA is comprised of 3 FL sub-advisors. Explain how to use this FLA to make: (a) a social judgment decision, or (b) an engineering judgment decision or action.

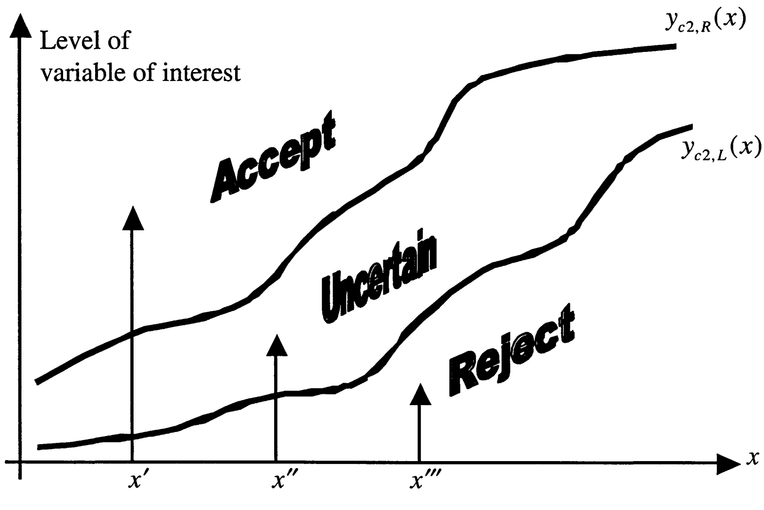

Figure 4-7: A way to use the FLA for an engineering judgment. The judgments are to accept when $x = x'$, reject when $x = x'''$, and must be further clarified when $x = x''$.