Chapter 1

Introduction

The traditional aggregation tool in information treatment is the weighted average, or more general, the weighted sum. That is, if the numerical information received from diverse information sources $x_1, x_2, \cdots, x_n$ are $f(x_1), f(x_2), \cdots, f(x_n)$ respectively, then the synthetic amount, weighted sum y, of the information is calculated by

$$y = w_1 f(x_1) + w_2 f(x_2) + \cdots + w_n f(x_n), \tag{1.1}$$

where $w_1, w_2, \cdots, w_n$ are the weights of $x_1, x_2, \cdots, x_n$, respectively. When $0 \le w_i \le 1$ for $i = 1, 2, \cdots, n$ and $\sum_{i=1}^{n} w_i = 1$, the weighted sum shown in (1.1) is called the weighted average. In databases, these information sources $x_1, x_2, \cdots, x_n$ are regarded as attributes and $f(x_1), f(x_2), \cdots, f(x_n)$ are their observations (or say, their records), respectively. An observation can be considered as a function defined on the finite set consisting of these involved information sources. Thus, the weighted sum, essentially, is the Lebesgue integral defined on the set of information sources and is a linear aggregation model. The linear models have been widely applied in information fusion and data mining, such as in multiregression, multi-objective decision making, classification, clustering, Principal Components Analysis (PCA), and so on. However, using linear methods need a basic assumption that there is no interaction among the contributions from individual attributes towards a certain target, such as the objective attribute in regression problems or the classifying attribute in classification problems. This interaction is totally

different from the correlationship in statistics. The latter is used to describe the relation between the appearing values of two considered attributes and is not related to any target attribute.

To describe the interaction among contributions from attributes towards a certain target, the concept of nonadditive set functions, such as λ-measures (called λ-fuzzy measure during the seventies and eighties of the last century), belief measures, possibility measures, monotone measures, and efficiency measures have been introduced. The systematic investigation on nonadditive set functions started thirty five years ago. At that time, they were called *fuzzy measures*. Noticeably, the traditional aggregation tool, the weighted sum, fails when the above-mentioned interaction cannot be ignored and some new types of integrals, such as the Choquet integral, the upper integral and the lower integral, should be adopted. In general, these integrals are nonlinear and are generalizations of the classical Lebesgue integral in the sense that they coincide with the Lebesgue integral when the involved nonadditive measure is simply additive. The fuzzy integral, which was introduced in 1974, is also a special type of nonlinear integrals with respect to so-called fuzzy measures. Since the fuzzy integral adopts the maximum and minimum operators, but not the common addition and the common multiplication, most people do not prefer to use the fuzzy integral in real problems. Currently, the most common nonlinear integral in use is the Choquet integral. It has been widely applied in information fusion and data mining, such as the nonlinear multiregressions and the nonlinear classifications, successfully. However, the corresponding algorithms are relatively complex. Only the traditional algebraic methods are not sufficient to solve most data mining problems based on nonlinear integrals. Some newly introduced soft computing techniques, such as the genetic algorithm and the pseudo gradient search, which are presented in Chapter 7 of this monograph, must be adopted.

In most real problems, there are only finitely many variables. For example, in any real database, there are only finitely many attributes. So, the part of fundamental theory in this monograph is focused on the discussion of the nonadditive set functions and the relevant nonlinear integrals defined on a finite universal set. The readers who are interested in the convergence theorems of the function sequences and integral

sequences with respect to nonadditive set functions may refer to monographs *Fuzzy Measure Theory* (Plenum press, New York, 1992) and *Generalized Measure Theory* (Springer-verlag, New York, 2008).

The current monograph consists of eleven chapters, After the Introduction, Chapters 2 to 5 devote to the fundamental theory on sets, fuzzy sets, set functions, and integrals. Chapters 6 to 11 discuss the applications of the nonlinear integrals in information fusion and data mining, as well as the relevant soft computing techniques. The relation among these chapters is illustrated in Figure 1.1.

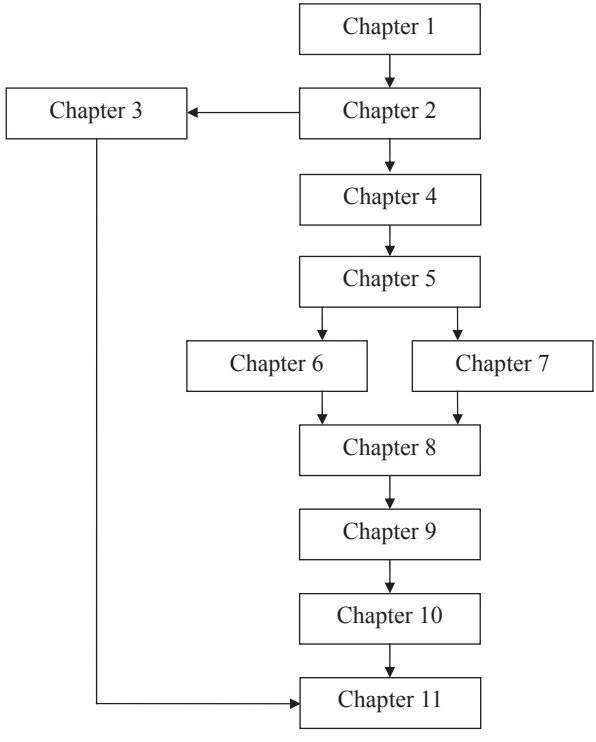

Fig. 1.1 The relation among chapters.