1

Introduction

Pregada fo Natana per totes les dones
que digués la manera segons la qual
per art poguessen atrobar a eleger
la dona qui es millor a abadessa.
(...)
Per a aquella manera dix Natana es
atrobada veritat per la qual veritat
porem atrobar aquella dona qui es pus
cuvinent e mellor a esser abadessa.[1]

Ramon Llull, [234] (p. 80)

Information fusion techniques, in general, and aggregation operators (or aggregation functions), in particular, are extensively used in several fields of human knowledge. They are used to produce the most comprehensive and specific datum about an entity from data supplied by several information sources (or the same source at different periods of time). They are used in systems to reduce some type of noise, increase accuracy, summarize information, extract information, make decisions, and so on. To illustrate this, we consider below some examples in different fields. Some of the typical applications are also included.

Economics: Aggregation techniques are used to define indices about prices such as the *Retail Price Index (RPI)* and, in general, to summarize any kind of economic information. Listings of countries or companies, where individuals are ordered according to their ranking with respect to several criteria, are frequently published in journals and newspapers. Examples are the *Human Development Index (HDI)*, which is an average of the life

[1] Natana was asked by all the sisters to describe the method according to which, with the system, one can find and elect the sister who is suited best to be abbess. (...) "By this method," said Natana, "is found the truth; by this truth we will be able to find the sister who is most suitable and best to be our abbess." Translation from [176].

expectancy index, the educational attainment index, and the adjusted real *gross domestic product (GDP)* per capita.

Biology: Methods to fuse sequences of DNA and RNA are used in several applications. Aggregation operators have also been developed to combine information about taxonomies (classifications of species). More specifically, methods exist to combine dendrograms (tree-like structures) and partitions.

Education: Aggregation operators are extensively used in education for assessing students' knowledge in a given subject or to assign them an overall rating for several subjects. Different methods are used in different countries, according to tradition and to the scale used when giving grades (both numerical and ordinal). Scores for evaluating educational institutions (e.g., universities) are another example of the use of aggregation operators.

Computer Science: Aggregation operators are used for different purposes. On the one hand, we have artificial intelligence applications, which are commented on in more detail below. On the other hand, we have decision making procedures that are applied, for example, to evaluate and select hardware and software.

Within artificial intelligence, information fusion is also widely applied, and its use is rapidly increasing as more complex systems are being developed. For example, its uses in robotics (e.g., fusion of data provided by sensors), vision (e.g., fusion of images), knowledge based systems (e.g., decision making in a multicriteria framework, integration of different kinds of knowledge, and verification of knowledge-based systems correctness) and data mining (e.g., ensemble methods) are well known. Recent advances in multiagent systems extend the range of information fusion applications in systems where an agent needs to consider the behavior of other agents to make decisions on the basis of distributed information.

Although the number of information fusion applications in artificial intelligence is large, it can be said that there are only two ultimate goals. They are (i) to make decisions and (ii) to have a better understanding of the application domain. We describe them in more detail below:

Decision making: This consists either of selecting the best alternative (*alternative selection*) or building one new alternative (or solution) from a set of them (*alternative construction*).

- In *alternative selection*, fusion is used to evaluate the alternatives. A typical situation is one where there is a set of alternatives and each is evaluated against several criteria (this situation corresponds to the multicriteria decision making – MCDM – problem). For example, when a buying agent has received several offers and wants to select the best one, it needs to consider the best price, the best quality, and so on. This situation can be modeled in terms of several preferences (or utility functions) or by using a single but multivalued preference. That is, for

alt	Criteria Satisfaction on: Price	Quality	Comfort
FordT	0.2	0.8	0.3
206	0.7	0.7	0.8
...	...		

$\rightarrow$

alt	Consensus
FordT	0.35
206	0.72
...	...

$\rightarrow$

alt	Ranking
206	0.72
FordT	0.35
...	...

Fig. 1.1. Decision making: (a) multicriteria or multivalued preferences; (b) aggregation of degrees of satisfaction (aggregation of preferences) and construction of the global degree of satisfaction; (c) ranking of the alternatives according to the global degree of satisfaction (preferences)

each offer, we consider the degree of satisfaction in terms of price, quality, and so on. Figure 1.1 illustrates the case in point. The figure includes several criteria $c_1, \ldots, c_N$ for each alternative.

The alternative selection problem is usually solved in a two stage process:

(i) For each decision alternative, aggregate the degrees of satisfaction of all criteria. In this way, we obtain for each alternative a single aggregated value that corresponds to a global degree of satisfaction.
(ii) Rank the alternatives with respect to the global degree of satisfaction.

It is clear that the cornerstone of the process is the aggregation method used in the first stage. Figure 1.1 illustrates the whole process.

Systems modeling group decisions also fit in this class of alternative selection problems. In this case, different experts in a group have different opinions and the goal is to obtain some consensus. This field of study is known as group decision making (GDM).

- In *alternative construction*, fusion corresponds to the whole process of building a new alternative from the original ones. It is important to underline that it is often the case that the alternatives correspond to partial solutions and that different alternatives might be incomparable or mutually incompatible. This process has to consider the importance and the reliability of the alternatives, their constraints, and the approaches used when building them. Algorithms for *plan merging* and *ensemble methods* in machine learning can be studied from this perspective.

  Plan merging consists of integrating partial plans to build a more complex one. In the integration process the preconditions and the effects of each partial plan have to be considered, as they define constraints on the order in which the partial plans can be executed. For

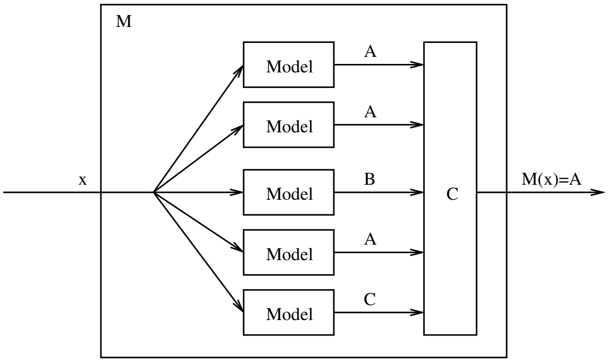

Fig. 1.2. Ensemble methods for classification: $\mathbf{x}$ represents the instance to be classified, and $\mathbb{C}$ represents a method for aggregating partial solutions

example, the plan for tightening a nut cannot be applied after assembling.

Ensemble methods consist of building several models from examples and then combining them to define a new one. The new model is intended to be more reliable and with less error than each of the original ones. Figure 1.2 illustrates this case: in a classification problem, several classifiers or *models* M_i are constructed from a set of examples using, for example, different supervised machine learning techniques. Then, for a particular instance (or situation) $\mathbf{x}$, all models are applied each giving a solution (or class) $M_i(\mathbf{x})$. In the figure, each model leads to A, B, or C. Then, the solution of the whole system that is denoted by $M(\mathbf{x})$ estimates the class of the instance $\mathbf{x}$. This is computed from the classes $M_i(\mathbf{x})$ applying some consensus procedure $\mathbb{C}$. This procedure strongly depends on how $M_i(\mathbf{x})$ are represented. In the problem represented in Figure 1.2, we can use the voting procedure as the consensus procedure $\mathbb{C}$.

Improving the understanding of the application domain: A system solely working with data obtained from a single source of information faces several inconveniences caused by insufficient data quality. In particular, we underline the following difficulties: (i) lack of accuracy of the supplied data due to errors caused by the information source (either intentional or accidental) or due to errors in transmission; (ii) lack of reliability of the sources; (iii) too narrow information supplied in relation to the working domain (the information only describes a part of the application's domain).

To deal with these problems, information fusion techniques can be used. The techniques can increase the reliability of the system, improving their data quality and extending their domain of application. In fact, in some circumstances, such techniques permit the extraction of features that are impossible to perceive from individual sources. Extraction of 3D representation of objects from several images corresponds to this case.

Note that in the setting of improving the quality of the data, information fusion can be applied at the time the system is built or at runtime (for example, by combining the newly acquired information with the previously established one). Knowledge revision can be seen from this perspective.

Although information fusion is a useful tool appropriate for improving the capabilities of intelligent systems, it is important to underline that difficulties arise in their use because such data are frequently not comparable and sometimes inconsistent. Therefore, systems have to embed simple fusion techniques in larger software tools so that results are consistent. These issues are described in more detail in the next section.

1.1 Fusion and Integration

This section defines some of the terms in the field of information fusion and integration.

In Section 1.2 we present a general architecture for information integration based on the processes commonly admitted in multisensor fusion and integration. Information integration is considered here as a general framework that embeds information fusion. This follows the approach in the sensor field, where multisensor fusion and multisensor integration are also differentiated. Additionally, we shall use the term *aggregation operators* to refer to concrete mathematical functions. According to this, we describe the terms *information integration*, *information fusion*, and *aggregation operators* as follows.

Information integration: This corresponds to the use of information from several sources (or from the same source but obtained at different times) to accomplish a particular task.

Information fusion: Information integration requires particular techniques for combining the information. Information fusion is the actual process of combining these different data into one single datum. Therefore, information fusion refers to particular mathematical functions, algorithms, methods, and procedures for data combination. According to this, information fusion is one of the processes embedded in an information integration architecture. In the following, we will use combination as a synonym of fusion.

Aggregation operators: These operators (also referred to as *means* or *mean operators*) correspond to particular mathematical functions used for information fusion. Generally, we consider mathematical functions that combine N values in a given domain D (e.g., N real numbers) and return a value in the same domain (e.g., another real number). Denoting these functions by $\mathbb{C}$ (from *Consensus*), aggregation operators are functions of the form:

Unanimity or idempotency: $\mathbb{C}(a, \ldots, a) = a$ for all a
Monotonicity: $\mathbb{C}(a_1, \ldots, a_N) \geq \mathbb{C}(a'_1, \ldots, a'_N)$ when $a_i \geq a'_i$
Symmetry: For any permutation π on $\{1, \ldots, N\}$ it holds that

$$\mathbb{C}(a_1, \ldots, a_N) = \mathbb{C}(a_{\pi(1)}, \ldots, a_{\pi(N)})$$

Fig. 1.3. Main properties of aggregation operators

$$\mathbb{C} : D^N \to D$$

Usually, operators fuse input values taking into account some information about the sources (data suppliers). That is, operators are parametric so that additional knowledge (background knowledge, following artificial intelligence jargon) on the sources can be considered in the fusion process. We express this by $\mathbb{C}_P$, where P represents the parameters of $\mathbb{C}$.

As an example, we can consider the arithmetic mean as one such aggregation operator:

$$\mathbb{C}(a_1, \ldots, a_N) = \sum_{i=1}^{N} a_i/N$$

This expression does not include any information on the data suppliers. Instead, the weighted mean is another aggregation operator that includes a weight for each data supplier:

$$\mathbb{C}_{\mathbf{p}}(a_1, \ldots, a_N) = \sum_{i=1}^{N} p_i \cdot a_i/N$$

Here, p_i is the weight/relevance for the source supplying datum a_i.

Aggregation operators are usually required to satisfy unanimity (defined in Figure 1.3) and, when D is an ordinal scale, monotonicity. The two properties imply that aggregation operators are functions that yield a value between the minimum and the maximum of the input values. Formally, they are operators $\mathbb{C}$ that satisfy internality:

$$\min_i a_i \leq \mathbb{C}(a_1, \ldots, a_N) \leq \max_i a_i \qquad (1.1)$$

Moreover, in some circumstances symmetry is also required. Here, symmetry stands for the fact that the order of the arguments is not relevant. In other words, there is no source distinguishable.

From this point of view, it is clear that all aggregation operators are information fusion methods. However, only information fusion methods with a *straightforward* mathematical definition are considered here as aggregation operators. Therefore, not all information fusion methods are

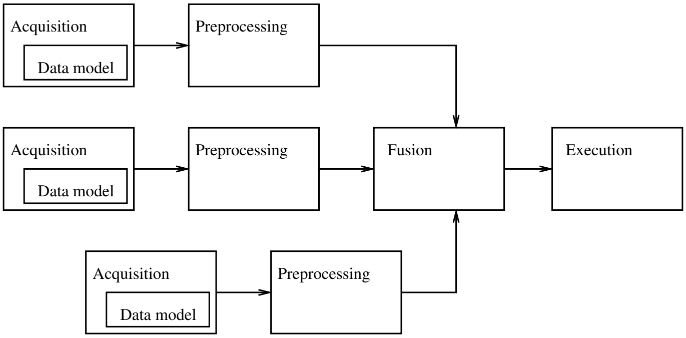

Fig. 1.4. General architecture for data fusion

aggregation operators. In particular, methods with a complex operational definition (e.g., complex computer programs) are not considered in this book as such. Naturally, the division between both terms is rather fuzzy.

1.2 An Architecture for Information Integration

Now, we turn to a general architecture for information integration. This architecture (see Figure 1.4 for a graphical representation) distinguishes the following stages.

1. **Acquisition:** The first stage corresponds to the process of gathering information from the information sources. This stage is also called detection. In order to have good data quality, a good source model is required, that is, a model of the uncertainty and error of the sources, so that it is possible to have a measure of the quality of the information. This measure can then be used in the fusion process so that it takes into account the reliability of the sources. The requirement of a source model is needed when combining sensory information (*sensor model* to determine the reliability of a particular sensor) or human (symbolic) knowledge (to determine, for example, if the supplied information is within the expert's domain of expertise or belongs to some general knowledge).

   Following the analogy with multisensor fusion, acquisition can be passive (when the information recorded is already present in the surroundings of the system) or active (when the information recorded is a consequence of an action initiated by the system).
2. **Preprocessing:** This second stage consists of preparing the data for the fusion process (i.e., of making data computationally appropriate). Several procedures are encompassed in this stage and they range from simple (like noise reduction and sensor recalibration procedures) to complex (edge detection and filtering methods). Procedures are considered preprocessing

as long as they only use the information of a single information source. This stage also includes procedures for making data commensurable and for solving the registration problem. Several pieces of data are commensurable when they refer to the same position in space and instant in time. The registration problem corresponds to the determination of information from each sensor that refers to the same features in the environment.

Aspects of both data commensurability and registration are relevant not only when we are considering numerical data from sensors but also for other kinds of data and applications, e.g., symbolic data in knowledge elicitation. In this latter setting, knowledge either elicited from experts or (automatically) extracted from databases has to be commensurate with other information before being integrated. Otherwise, results will not be meaningful.

3. **Fusion:** Once data are preprocessed and, thus, are commensurable, they can be fused. At this stage, aggregation operators or more complex fusion methods are applied to obtain a new datum. Typically, all input data uses the same representation formalism, which is also used to represent the outcome of the system. For example, the outcome of a set of images is another image. Nevertheless, some systems differ from the approach. For example, different input data use different formalisms. Then, instead of direct fusion, one source can be used to guide or cue the data from other sources. This is referred to as *guiding* or *cueing* and consists of indirect fusion. The case of visual information guiding the operation of a tactile array mounted on the end of a manipulator is an example of this situation.
4. **Execution:** Appropriate procedures are applied using the datum obtained in the fusion stage. Two kinds of procedures can be distinguished: action application and data interpretation. Control systems correspond to the first case. They use the outcome of the fusion process to decide what action to take. Exploratory robots might correspond to the second case, as they will analyze the new data and add them to their knowledge base. This case corresponds to a world model revision because the system modifies the state of its own model for the operating environment. Again, this classification is rather fuzzy, as the analysis of the data can change the behavior of the robot.

Note that all the procedures and functions that participate in this architecture are task-specific and, thus, change according to the application. For example, a decision making process in a multicriteria environment requires first the acquisition of the values for each criteria. Next, the preprocessing stage consists of the normalization of the data or data translation into a uniform space (e.g., the $[0, 1]$ interval). Then, the fusion is applied using a particular aggregation operator and, finally, a decision is made selecting the alternative that is best rated. In contrast, fusion for obstacle detection in a robot navigation system requires different procedures: gathering sensor data, making them commensurable, fusion, and, finally, raising an alarm if an obstacle is found. Nevertheless, although different procedures are used, the stages above apply to all such cases.

1.3 Information Fusion Methods

In the previous section we have focused on fusion processes and their role in an information integration architecture. We now turn to the fusion methods themselves.

Information fusion methods can be studied from different perspectives. In the rest of this section, we describe some of the dimensions in use for classifying them. To some extent, this classification is independent of the type of information source used (sensor or expert) and whether all the information is acquired at the same instant or at different times.

Type of information: Two main categories are distinguished. They correspond to redundant and complementary information.
- Redundant information occurs when several information sources describe the same features in the environment. Differences in the data, expected to be small, are due to the lack of the source's reliability. Redundant data are fused to reduce uncertainty and increase data accuracy.
- Complementary information corresponds to the case of sources describing different features of the environment (different subspaces). Different data describes different characteristics that are not similar. Fusion is applied so that the system model cover all subspaces.

Type of data representation: A basic consideration for any aggregation operator or fusion method is the type of data it is going to fuse. At present, there exists a large number of aggregation operators applicable to a broad range of data representation formalisms. For example, aggregation operators on the following formalisms have been considered in the literature: *numerical data*, *ordinal scales*, *fuzzy sets*, *belief functions*, *dendrograms*, *DNA sequences*, among others. In fact, any kind of data representation formalism is adequate for applying fusion techniques because the plurality rule (mode or voting) can be applied to data of almost any type.

Level of abstraction: Due to the information flow within systems (low-level data is transformed into high-level information), fusion techniques can often be applied at different levels of abstraction. For example, in a multisensor fusion system for tank detection, the following levels can be distinguished: signal, pixel, feature, and symbol. Similarly, in a knowledge elicitation problem using data from multiple experts, fusion can be performed either at the matrix level (directly on the raw data supplied by the expert) or at the similarity level (using similarities extracted from experts' raw data).

Let us consider the expression

$$\mathbb{C}(a_1, a_2, \ldots, a_N) = argmin_c\{\sum_{a_i} d(c, a_i)\},$$

where a_i are numbers in $\mathbb{R}$ and where d is a distance defined over D. Then, the following hold:

1. When $d(a, b) = (a-b)^2$, $\mathbb{C}$ is the arithmetic mean. That is, $\mathbb{C}(a_1, a_2, \ldots, a_N) = \sum_{i=1}^{N} a_i/N$.
2. When $d(a, b) = |a-b|$, $\mathbb{C}$ is the median. The median of $a_1, a_2, \ldots, a_N$ is the element that occupies the central position when the elements a_i are ordered. The median is formally defined in Definition 6.7.
3. When $d(a, b) = 1$ if and only if $a = b$, $\mathbb{C}$ is the plurality rule (mode or voting). That is, $\mathbb{C}(a_1, a_2, \ldots, a_N)$ selects the element in $\mathbb{R}$ that appears most often in $(a_1, a_2, \ldots, a_N)$.

Fig. 1.5. Aggregation as the object that is located at the minimum distance of the objects being aggregated

When several levels can be considered, the selection of the appropriate level depends on the information available. It is usually the case that redundant information is fused at low levels because two pieces of redundant information are usually similar in structure. In contrast, complementary information is usually fused at higher levels of abstraction, as pieces of information are not so similar. For example, in the case of the tank detection system, data from two radars will be fused at the signal level (low level) if both measure the same property at the same time. In contrast, the data from a radar and a radio signal detector should be fused at the symbol level (high level), as in this case the data gathered by the two data suppliers are of a completely different nature and, thus, only the elaborated conclusions (for example, whether data seem to indicate the presence or absence of a tank) can be combined. Nevertheless, there are situations in which two information sources can only be fused at a single level.

1.3.1 Function Construction

A pivotal consideration in any information fusion system is the actual method used for combining information. Its definition is the cornerstone of any integration system. Two methods can be distinguished. They roughly correspond to *a priori* and *a posteriori* analyses of the method's properties.

Definition from properties: The starting point for defining the method is a set of properties considered as a requirement for the method. From these properties the function is derived using mathematical tools. This is the approach used when applying functional equations (see Chapter 3). The definition of aggregation as the object that minimizes a given expression follows the same idea.

This approach is formulated as follows: the aggregation of the values a_1, a_2, ..., $a_N \in D$, denoted by $\mathbb{C}(a_1, a_2, \ldots, a_N)$, is the object c located at the minimum distance of the objects being aggregated. That is,

$$\mathbb{C}(a_1, a_2, \ldots, a_N) = argmin_c\{\sum_{a_i} d(c, a_i)\}, \tag{1.2}$$

where d is a distance defined over D. The approach is valid in any domain D where a distance d is defined. Figure 1.5 gives an example of it when D is the set of real numbers.

Heuristic definition: In this case, the function is selected or defined because it seems to satisfy user requirements or expectations. The function is studied and its properties analyzed later.

An alternative method has also been proposed for function construction. It can be considered as an intermediate approach between a heuristic definition and a definition from properties.

Definition from examples: This manner of definition follows classical statistical estimation theory and supervised machine learning methods. The function is built as an estimator of some available examples. Therefore, the function approximates example outcomes given example inputs. A typical method is to use neural networks for such approximations.

1.4 Goals of Information Fusion

Now that we have introduced information fusion and outlined some of its relevant aspects, we focus on its goals.

We have said that information fusion deals with all the aspects of the fusion process, and its main task is to deal with fusion methods. Due to the development of new representation formalisms, the consideration of new applications, and the growth of computational power, information fusion is a dynamic field, and new methods are constantly being defined. At the same time, existing methods are being analyzed to determine their properties. The two main goals of the field are (i) formalization of aggregation processes and (ii) study of existing methods. The goals are described in more detail below.

Formalization of the aggregation process: This is to find formal descriptions for processes (sometimes, intuitive processes) that are used for decision making and information fusion. Formal descriptions are needed so

Arrow's impossibility theorem applies to aggregation of preferences (over a set of alternatives). It proves that when there are at least three alternatives and at least two preferences, there is no aggregation function that, for all sets of preferences satisfies the following properties:

1. Any preference can be obtained as the result of the function.
2. The function does not imply dictatorship (i.e., the function is not just one of the preferences).
3. The function is monotone, i.e., if one preference is modified so that one alternative is *promoted*, the function should at least avoid *demoting* such alternative.
4. The function satisfies the *independence of irrelevant alternatives*. That is, the final preference of x over y should be independent of preferences for other alternatives.

Fig. 1.6. Arrow's impossibility theorem

that problems can be solved in an effective and sound way. Nevertheless, model building (the procedure of building a formal description) is not an easy task. For help, the development of methodologies for function selection and tools for parameter determination (e.g., algorithms) are required. Moreover, in some situations, we need to consider the definition of new aggregation operators, as existing methods are not appropriate because they do not satisfy the desired properties or, worse, do not fit with the current representation formalism in use. The goal can be decomposed as follows:

1. Function definition: The construction of new functions on the basis of new properties or when considering new knowledge representation formalisms has been studied for a long time. For example, in the framework of aggregation of preferences (or of alternative selection based on preferences), Llull (thirteenth century) and Nicholas of Cusa (Nicholas Cusanus) (fifteenth century) proposed methods that were later rediscovered by Condorcet and Borda (eighteenth century). They are the Condorcet rule (with the Copeland method for solving ties) and the Borda count. A related approach, important in real-world applications, is to study when no function exists that satisfies a set of properties. Arrow's impossibility (or incompatibility) theorem is a result of this kind. We recall that it applies to functions to aggregate preferences and that Arrow proved that there is no aggregation function that satisfies a set of *natural* axioms. The theorem is reproduced in Figure 1.6.

2. Function selection: This corresponds to methods for deciding the most appropriate function in a given situation. At present, this can be done, as pointed out in Section 1.3.1, heuristically, on the basis of properties or from examples.

3. *Parameter determination:* This stands for algorithms and mechanisms for finding the best parameterization of a given aggregation operator. Methods are mainly based on expert interviews or are example based.

Study of existing methods: For most knowledge representation systems there exists a large set of aggregation methods that can be applied. To apply them properly we need to know their intrinsic differences. Three categories can be distinguished in relation to the properties:

1. *Function characterization:* This is to know, on the one hand, which properties a particular operator satisfies and, on the other hand, which operators satisfy a set of properties. Functional equations are basic tools for function characterization.
2. *Determination of function's modeling capabilities:* The selection of an aggregation function corresponds to a tradeoff between expressivity and simplicity. In this respect, we know that aggregation operators can be used to build universal approximators (to approximate an arbitrary function at the desired level of detail). There exist some general models based on quasi-arithmetic means and Choquet integrals. However, to use such general models in practice is a difficult task, because on the one hand they require a large number of parameters and on the other hand they are difficult to interpret. In contrast, the arithmetic mean does not use any parameter, while its modeling capability is very limited (it corresponds to a completely determined hyperplane). In this framework, the determination of a function's modeling capability corresponds to locating it in the broad range of operators between the arithmetic mean and the general model.
3. *Relationship between operators and parameters:* Most aggregation operators are parametric and, therefore, their behavior strongly depends on the parameters. It is important to know how parameters can affect the result. For example, to know whether there exists a parameterization that implies the dictatorship property to one of the information sources (dictatorship can be represented with the weighted mean but not with the OWA operator; see Section 6.1), it is important to know how sensitive the operator is to changes in the data (according to parameterization) or how much the output is changed when the parameters change (needed when parameters are extracted from examples). To help in this analysis, some indices have been defined. Some of them (e.g., *orness*) will be reviewed in Chapter 7.

In this section, we have given a classification of current goals of information fusion and its research. Nevertheless, this classification is not crisp, as there are some research topics that can be found across several different areas. One of them is parameter determination according to the bias-variance trade-off. This is somehow equivalent to the selection of a model that sufficiently fits the data, but does not overfit it. This requires approaches from function selection, parameter determination, and also approaches related to functions' modeling capabilities.

1.5 Bibliographical Notes

1. **Information fusion:** Information fusion and integration is a broad field, with applications in several fields of the human knowledge. Due to this, aside from its pure mathematical research, work on it is published in journals and conferences on a wide range of topics. Our bias, and, therefore, the bibliographical references used and consulted for preparing this book, is towards artificial intelligence, mathematics, economics, remote sensing, and multisensor fusion applications.
2. **Information integration and architectures:** The way we have structured this chapter and our vision of the field are mainly based on sensor fusion and integration. Reference books in this field include [1] and [47]. See also the review paper by Hall and Llinas [179]. The chapter by Luo and Kay [241] in [1] gives a nice state-of-the-art description (from the 1990s) of data fusion and sensor integration. Most of the concepts reviewed can be easily translated to other fields, such as artificial intelligence. Several reference papers on fusion and related issues (e.g., sensor and data fusion, decision making) have been collected by Sadjadi in [345].

   Differences between information integration and information fusion explained in this chapter mainly correspond to the ones in [241], while our definition of information fusion is based on [435] and [166].

   Sensor fusion has devoted much effort to research on architectures. The architecture presented here is based on [47], with elements of [241]. In particular, the definition of preprocessing as "putting the data in a form that is computationally appropriate" is from [47]. Additionally, the difference betwen active and passive acquisition can be found in [193].
3. **Aggregation operators:** There is no standard definition of aggregation operators. For example, Cauchy [66] and more recently Ovchinnikov [306] only require a function returning a value between the minimum and the maximum, while [138] and [449] also require symmetry. In this book, we follow the first approach initially and then add some consideration on the background knowledge later.

   As stated, internality (Equation 1.1) means that an operator leads to a value between the minimum and the maximum. This property is used by Cauchy (1821) in [66]. Ovchinnikov in [306] refers to operators that satisfy this property as compensative functions. [246] refers to them as internal functions.

   Additional references on aggregation operators and related topics are given in the bibliographical notes of other chapters of this book, specially Chapters 4 and 6.

Aggregation operators defined as the minimization of a distance (as in Equation 1.2 in Section 1.3.1) have been extensively used in the literature. For example, Fodor and Roubens in their book [146] (p. 143) use this approach to define the aggregation of relations. Similar results focused on biology can be found in [36] and [87]. They correspond, respectively, to methods for aggregating dendrograms and sequences. See also [141] for a recent application of these aggregation methods to bioinformatics. In this setting, the resulting aggregation function is known as the *median rule*. The examples given in Figure 1.5 are proved in Gini's book (1958) [163]. In particular, the result about the arithmetic mean is proved on p. 168, and the one about the median on p. 176. The property concerning the plurality rule is given on p. 185.

Jackson (1921) [201] includes some results for the same problem when the distance equals $d_p(a,b) = |a-b|^p$ for some $p > 1$. It shows that for $p > 1$ there is a single solution. It also studies the case for $p = 1$, which corresponds to the mode. It shows that there is a unique solution when N is either of the form $N = 2k+1$ or of the form $N = 2k$ with $a_{s(k)} = a_{s(k+1)}$ where s is an order statistics (a permutation such that $a_{s(i)} \leq a_{s(i+1)}$). In the case with $N = 2k$ and $a_{s(k)} \neq a_{s(k+1)}$, any value a in the interval $[a_{s(k)}, a_{s(k+1)}]$ is a valid solution. Nevertheless, the paper also shows that the following holds for the limit of $p \to 1$:

$$\lim_{p \to 1} \arg\min_{c} \left\{ \sum_{i=1}^{N} |c - a_i|^p \right\} = m,$$

where m is characterized by

$$(m - a_{s(1)}) \cdots (m - a_{s(k)}) = (a_{s(k+1)} - m) \cdots (a_{s(N)} - m).$$

From this, it can be shown that for $N = 2$, m should be $m = (a_1 + a_2)/2$ and that for $N = 4$, m corresponds to

$$m = \frac{a_4 a_3 - a_2 a_1}{(a_4 + a_3) - (a_2 + a_1)}.$$

Note that the standard definition of median for $N = 2k$, $(a_{s(k)} + a_{s(k+1)})/2$, does not correspond, in general, to this limit (see Definition 6.7).

The cases with $p = 1$ and $p = 2$ (corresponds to the arithmetic mean) were already studied by different authors. For example, it was known by Laplace [221] (supplement 1812-1818) and Svanberg (attributed) [20] p. 194-195 (1821). The case of $p \to \infty$ corresponds to the midrange of $\{a_1, \ldots, a_N\}$. That is, $(a_{s(1)} + a_{s(N)})/2$. Foster (1922) [149] studied the case of $p \to 0$, showing that it corresponds to the mode. The case of weighted distances was studied in [35] (1938). It leads to the mode ($p \to 0$), weighted median ($p = 1$), weighted mean ($p = 2$), and, again, the midrange for $p \to \infty$.

4. **Applications and examples:** The cited chapter by Luo and Kay [241] describes several systems in some detail. They are examples of sensor fusion. Among them, we underline the example of the tank detection system, where fusion is performed at several levels. This example was outlined in Section 1.3. The other example in the same section on knowledge elicitation is taken from [409].

   Luo and Kay also give an example that corresponds to the indirect fusion (guiding and cueing) described in Section 1.2. It is the description of a robotic object recognition system that uses vision to guide tactile sensing. Other examples of aggregation operators for either numerical or ordinal scales are given in [43]. In particular, [43] includes a description of the Human Development Index and several methods for aggregating grades. Some fusion methods in biology are described in [244], [67], and [86]. [244] deals with fusion of taxonomies ([318] is an application of such aggregation methods for comparing phylogenetic trees), while [67] and [86] deal with fusion of sequences. Methods for the aggregation of partitions, also used to aggregate nonhierarchical classifications in biology, can be found in [143] and [266]. Examples of fusion techniques for computer science can be found in [104] and [105].

   Decision making is described in several books. See [340] for a state-of-the-art (1996) description of the field. Other examples briefly pointed out in this chapter include plan merging and ensemble methods. Methods for plan merging are described in [96, 150]. Ensemble methods are a successful technique applied in machine learning and are nowadays described in most machine-learning books. See [182] and [436].
5. **Goals of information fusion:** Section 1.4 is basically based on our own research. Ramon Llull (thirteenth century) findings on electoral systems can be found in [234] (Chapter XXIV), [176], and also on a Web page [235]. [176] and [235] include English translations as well as transcripts of Llull's original works in either Catalan (for example, the novel Blanquerna [234] written c. 1283 [369]) or Latin (*Artifitium electionis personarum* and *De arte eleccionis*). Llull's election method anticipated Condorcet (eighteenth century) (he uses Copeland's method for solving ties). Nicholas of Cusa (or Cusanus) introduced an alternative method to Llull's in 1431 (in his work *De concordantia catholica*) that corresponds to Borda's account. Ramon Llull and Cusanus were motivated by a need to find a method for honest elections in the Church.

   The papers by McLean [257] and McLean and London [258] are also of interest here. They discuss Ramon Llull's contributions in the context of medieval voting, and the influence of Ramon Llull in Cusanus. Chapter 37 of Book III of *De concordantia Catholica* by Cusanus is reproduced in [257] and [258]. This book was written while Cusanus was attending the Council of Basel (1431-1434). [258] argues that the method proposed by Llull in Blanquerna corresponds to the Borda count. In this respect, we agree with the later interpretation by Hägele and Pukelsheim [176], rather than with the one by McLean and London.

The papers *De arte eleccionis* and *Artifitium electionis personarum* were rediscovered, respectively, by Honecker [191] in 1937 and by Perez Martínez [320] in 1959. The first work was found in the library of the Sankt Nikolaus-Hospital/Cusanusstift in Bernkastel-Kues and seems to have been copied by Cusanus himself (see [176] p. 6).

Arrow's impossibility theorem was given in [23]. For a history of voting procedures, see [426] or [259]. Arrow's theorem is described in several books on preference, choice, and decision. See [332], and also the handbook edited by Arrow, Sen, and Suzumura [24].

The definition of models based on aggregation operators that are universal approximators can be found in [399] and [277, 290, 413]. The former work defines a model based on quasi-weighted means and the latters define models based on Choquet integrals.

The bias-variance tradeoff is described in most machine learning and statistical learning books. See [182] and [296].