Information Fusion and Data Science

Series Editor: Henry Leung

Basel Solaiman
Éloi Bossé

Possibility Theory for the Design of Information Fusion Systems

Springer

Information Fusion and Data Science

Series editor
Henry Leung, University of Calgary, Calgary, AB, Canada

This book series provides a forum to systematically summarize recent developments, discoveries and progress on multi-sensor, multi-source/multi-level data and information fusion along with its connection to data-enabled science. Emphasis is also placed on fundamental theories, algorithms and real-world applications of massive data as well as information processing, analysis, fusion and knowledge generation. The aim of this book series is to provide the most up-to-date research results and tutorial materials on current topics in this growing field as well as to stimulate further research interest by transmitting the knowledge to the next generation of scientists and engineers in the corresponding fields. The target audiences are graduate students, academic scientists as well as researchers in industry and government, related to computational sciences and engineering, complex systems and artificial intelligence. Formats suitable for the series are contributed volumes, monographs and lecture notes.

More information about this series at http://www.springer.com/series/15462

Basel Solaiman • Éloi Bossé

Possibility Theory for the Design of Information Fusion Systems

Springer

Basel Solaiman
Image and Information Processing
Department
IMT Atlantique
Brest, France

Éloi Bossé
Image and Information Processing
Department
IMT Atlantique
Brest, France

ISSN 2510-1528 ISSN 2510-1536 (electronic)
ISBN 978-3-030-32852-8 ISBN 978-3-030-32853-5 (eBook)
https://doi.org/10.1007/978-3-030-32853-5

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Possibility theory is a mathematical theory, coined by L.A. Zadeh in the late 1970s (1978) to deal with vague pieces of information described by means of fuzzy sets and fuzzy logic. Thereafter, Didier Dubois and Henri Prade have been the main founders of this theory to the extent that we have today, a credible alternative to probability theory. A considerable body of literature has flourished around fuzzy sets and possibility theory concepts in a very wide range of applications, from mathematics and logics to advanced engineering methodologies, from medical domain to finance, from human factors to consumer products, and so on. There is a plethora of books and papers describing this rich domain of applications.

The ambition of this book is to address a niche still uncovered by the existing available books: a comprehensive assemblage of the basic concepts, the mathematical developments, and the engineering methodologies to position and exploit possibility theory for the design of computer-based decision-support systems. Usually, decision-support systems comprise three main parts: analysis (analytics), synthesis (information fusion), and prescription (decide and act). Literature shows that possibility theory can be applied to the three parts.

This book consists of nine chapters. The first three chapters discuss the fundamental possibilistic concepts: distribution, necessity measure, possibility measure, joint distribution, and the important concept of conditioning. Chapter 4 examines the concept of similarity that plays an essential role in a wide range of application fields like pattern recognition, reasoning, data and knowledge mining but with respect to what can possibility theory bring to implement that complicated concept. Chapter 5 addresses the links and transformations between the interrelated uncertainty modeling theories. The following next two chapters treat aspects of decision-making through possibilistic and fuzzy integrals, fusion operators, and decision-making criteria in the framework of possibility theory. Chapter 8 presents three low-level complexity applications of possibilistic concepts: (1) on pixel-based image classification, (2) on spatial unmixing, and (3) on image segmentation.

The book is concluded by Chapter 9 on the use of possibility theory in the design of information fusion systems in today's ever-increasing complexity of our real world. Information overload and complexity are core problems to most organizations of today. The advances in networking capabilities have created the conditions of complexity by enabling richer, real-time interactions between and among individuals, objects, systems, and organizations. Fusion of Information and Analytics Technologies (FIAT) are key enablers for the design of current and future decision-support systems to support prognosis, diagnosis, and prescriptive tasks in such complex environments. Hundreds of methods and technologies exist, and several books have been dedicated to either analytics or information fusion so far. This book presents the overall picture in which possibility theory can be of any use.

Brest, France Basel Solaiman
Brest, France Éloi Bossé

Contents

Chapter 1
Introduction to Possibility Theory

1.1 Introduction

The problem of decision-making, arising from everyday practice, belongs to the oldest problem types studied from the seventeenth century, and since then the best reference decision-maker is certainly the human brain. For this reason, modeling of the human brain reasoning operation in decision-making has been a topic of intense studies in many application domains leading to creative methodologies, algorithms, and deductive approaches giving, thus, way to sustainable researches and developments. The major difficulty faced by systems supporting decision-making is due to the fact that we have to deal with imperfect decision-relevant information. At this level, we have to admit a basic assumption that human knowledge, reasoning, and exchanged evidences and information are intrinsically, for the most part, characterized and expressed by a degree of ambiguity and uncertainty rather than in a probabilistic uncertainty manner. In fact, uncertainty and ambiguity capture two rather different types of information imperfections.

Uncertainty is the main cognitive process that makes human free to choose. Its presence (due to lack of knowledge, imperfect, or insufficient information) is the price affecting experts' decision when handling complex systems. It is derived by the nondeterministic membership of a point from the set of decisions containing all possible elementary decisions (called *singletons*, *states of the world*, *basic events*, *decisions*, etc.). The framework of uncertainty modeling has been rooted in probability theory in which the analyst's uncertainty about the integrity of the model is expressed in probabilistic terms. For a long time, probability theory has been considered as the unique normative model to cope with imperfection by presenting a classical well-founded framework manipulating uncertain but precise information. Nevertheless, probability theory, as good as it is, does not remain the best alternative where imprecision is inherent in the studied domain, where available information is simply preferences or ambiguous. In fact, ambiguity is derived from the partial

B. Solaiman, É. Bossé, *Possibility Theory for the Design of Information Fusion Systems*, Information Fusion and Data Science,
https://doi.org/10.1007/978-3-030-32853-5_1

membership of one or several points from the set of decisions. Thereby, over the last five decades, a lot of effort has been put into developing new nonclassical uncertainty theories (fuzzy sets theory, belief functions or evidence theory, imprecise probability theory, and possibility theory). Among the new theories, possibility theory is said to be amenable to the framework for representation of human perceptive uncertainty. This point has been suggested by prominent systems scientists such as Shackle [1] and Cohen [2]. They argue that the traditional approaches for choice modeling using probability theory do not completely represent the true level of uncertainty in people's behavior. Possibility theory deals with uncertainty when the evidence points to a nested set of propositions; and hence, it can deal with propositions that refer to an interval as well as a single value. Possibility theory deals with uncertain information where the available knowledge is expressed in an ambiguous framework and representing our inability to distinguish which of several alternatives is the true one in a particular situation. The word "possibility" can be interpreted in several ways: physical, epistemic, and logical [3].

In the physical sense, "possible" refers to feasibility or easiness. For example, in the sentence "it is possible for a vehicle to hold six passengers," possibility means the physical capacity of the vehicle. In the epistemic sense, "possible" means plausible. Finally, in the logical point of view, possibility provides a means to deal with incompleteness. With a given piece of incomplete information about an event, logical interpretation of possibility gives a degree of confidence in the occurrence of the event as in *it is possible that it will rain tomorrow*.

The concept of possibilities was, first, mooted by the economist G. L.S. Shackle [1, 3] who, unhappy with the use of subjective probability for handling uncertainty, proposed an alternative formalism. This formalism was the calculus of *potential surprise* where uncertainty about an event is characterized by a subjective measure of the degree to which the observer in question would be surprised by its occurrence. Potential surprise is clearly linked to the intuitive notion of possibility. If an event is entirely possible, then there is no surprise attached to its occurrence. If an event is wholly impossible, or is believed to be so, then if it occurs, it will be accompanied by the maximum degree of surprise.

Nevertheless, possibility theory was later introduced again by L. Zadeh [4], who related possibility theory to fuzzy sets theory. Consider the following example to show the scope of application of possibility theory. Let Ω denote a collection of individuals within which we are looking for precisely one and only one person but we don't know the person who we are searching for looks like. This uncertain environment, or situation, constitutes the global framework of application of possibility theory (which is exactly the same as the framework of other uncertain information processing theories like probability theory and belief functions theory). Now, imagine the available knowledge, called *evidence*, to identify our individual is given as a *fuzzy evidence* such as the individual we are looking for is *young*. Faced with this situation, our human reasoning will:

(i) "Indirectly" attribute, to each individual in Ω, a membership degree, or a belongingness value, on the scale from 0 to 1, to the fuzzy evidence *young*

(ii) Project this membership degree into a kind of mental possibilistic space of representation, where each individual in Ω has a *possibility degree of being* the one we are looking for.

This operation is referred to as the projection of the available evidences into a *possibilistic knowledge representation* form. In possibility theory, the result of this projection is called a *possibility distribution* defined on the set Ω.

In the case where the available evidence is not fully reliable, the obtained possibility distribution has to be "adjusted" taking into account this important knowledge of what is called information sources reliability. This possibility distribution adjustment is called possibility distributions *discounting*. If several sources of knowledge are available (the individual we are looking is *young*, *tall*, etc.), then and after the first step of transforming these sources into possibilistic knowledge representation form, the resulting possibility distributions have to be "merged" together, using adequate *possibilistic fusion operations*, in order to "resume" our global state of knowledge into a single possibility distribution over Ω. An important question is related to the decision-making process. In fact, as possibility theory operates in uncertain environments, the ultimate expected output is, therefore, to identify one and only one individual within the set Ω. At this level, possibility theory offers some interesting decision-making tools like the *possibility measure* (producing a degree of possibility that the individual we are looking for lies within a subset $A \subseteq \Omega$ of individuals) and the *certainty measure* (producing a *degree of trust* that the individual we are looking for lies within a subset $A \subseteq \Omega$ of individuals).

This simple example of searching for an individual within a set can be extended to extremely important categories of engineering problems: pattern recognition, investment risk evaluation, classification, estimation, automatic target recognition and tracking, etc. Before going in-depth through different concepts of possibility theory, it is important to precisely define the concept of "information," hereafter called *information element*, as well as different forms of information imperfections and make a brief visit to some existing theories allowing to process imperfect information.

1.2 Information Concept

1.2.1 Information Element Definition

One of the historical barriers to technology transfer in information processing systems has been the lack of a unifying terminology. *Information*, by contrast, is a most heterogeneous term. What is information? While there exists a conventional, commonsense hierarchy of information ranging from data (usually raw) to information (processed data) to knowledge (synthesized information), even these are not precise distinctions. The dividing lines are blurry and partly subjective. Recall that the success of an information processing system is strongly related to the way its basic components are defined and to the quality of their associated knowledge as well as to the knowledge produced by the processing system.

Nevertheless, there is a relatively small body of literature in the information processing community that addresses this topic; moreover, there is no precise definition of what *information* is and what defines information characteristics. A clear statement of *what is information* and *what is informative* can lead to a strong qualitative understanding of the fundamental nature of information. A definition of information should capture the essential nature of the information and should allow frameworks, theories, and results to be transferred across disciplinary boundaries [5, 6]. According to Losee [5], information can be defined in terms of a process, or a function, in the following way: "Information is the value currently attached or instantiated to a characteristic or a variable returned by a function, or a process. The value returned by a function is informative about the function's argument, or about the function, or about both." This definition has the merit of positioning information as a relational concept linking data sets [7]. Asking for the meaning, the word "information" seems trivial since we use this word so frequently. Consider a simple example. What does it mean if we are given the value $x = 39$? From a mathematical point of view, it is just a positive integer; it is simply *data*. If $x = 39$ °C, then this becomes more "informative" since x certainly refers to the temperature (i.e., x denotes an element from a set of meaningful semantics). Nevertheless, this is still not enough informative because we still don't know the temperature of what! Now, if we are given that this concerns the temperature of Paul, then we possess a real information element concerning the temperature of Paul, and we can even start "reasoning" on this information by saying that Paul has a fever and must be examined by a physician. In other words, *observing a data value from a given set isn't enough to make the observed data an informative act* [7]. We need the context in which that information has been obtained. Having real information means that we know the concern of the information and how the content outcome is obtained. This leads to the following pragmatic definition of information [7, 8] which extends the definition proposed by Losee [5] by including a formal structure associated with important information characteristics:

Definition An information element is *a functional relation between two data sets: definition and content sets, through an informative function*[1]. Here, the "informative function" has to be understood in its cyber-physical sense.[2]

Therefore, the main components of an information element are (Fig. 1.1):

1. A ***definition set***: representing the potential information input elements

[1]Here, "function" has to be taken not in its formal mathematical sense but rather as *an activity or purpose natural to or intended for a person or thing*.

[2]Cyber-physical means a mechanism (or a machine) that is controlled or monitored by computer-based algorithms. See https://en.wikipedia.org/wiki/Cyber-physical_system for a definition of a cyber-physical system.

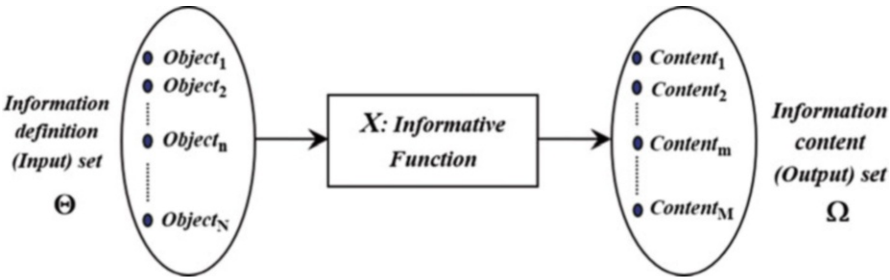

Fig. 1.1 Basic information element structure

2. A ***content set***: encoding the possible knowledge produced by the information such as measurements or estimations of physical parameters, decisions, hypothesis, etc.
3. An ***input-output relational function***: producing the mathematical or physical model representation associating the input elements with the produced information contents

An information element is always informative about something, being a component of the output or result of the informative relation.[3] In fact, having the sole information content isn't enough to make an input object/event as an informative act; the information meaning as an entity (*definition set*, *informative relation*, *content set*) must be perceived to make the "information" informative. This aspect has already been pointed by Stonier [9]: "we must not confuse the content reading and/or interpretation of information with information itself."

An information element is called ***exhaustive*** if and only if the content set contains all possible outcomes produced by the informative relation. This property, related to the information content set, is also called the ***closed world assumption***. Otherwise, the information is said to be operating under the ***open world assumption***.

Another important property is also used to characterize an information element: exclusivity. In fact, an information element is called ***exclusive*** if and only if two different information contents cannot be simultaneously produced as the outcome of the informative function.

Consider, for instance, the case of a digital image. In this case, if we consider the digital image as a 2D array representing gray levels, then what is considered is *not* information but just a kind of *abstract data* (i.e., observed pixel gray levels). This abstract data becomes a full information element when it is associated with the basic objects we are imaging as well as with the physical model leading to obtain the

[3]A *relation* between two sets is a collection of ordered pairs containing one object from each set. If the object is from the first set and the object is from the second set, then the objects are said to be related if the ordered pair is in the relation. A *function* is a type of relation. But, a relation is allowed to have the object in the first set to be related to more than one object in the second set. So a relation may not be represented by a function machine, because, given the object to the input of the machine, the machine couldn't spit out a unique output object that is paired to.

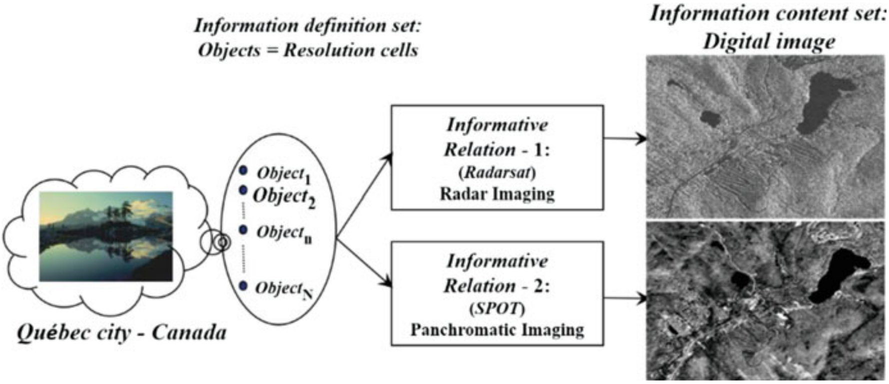

Fig. 1.2 Remote sensing examples of information using two informative functions (active radar imaging and passive panchromatic imaging functions)

digital image (i.e., the physical process used by the imaging sensor). In Fig. 1.2, image information elements of the Québec City are given using two imaging modalities (i.e., two distinct informative physical functions: *radar imaging* and *optical panchromatic imaging*).

The knowledge of the informative function is crucial for all information processing tasks that some call as "intelligent" tasks like scene interpretation, information fusion, data/knowledge mining, etc. For example, in remote sensing, let us imagine that a resolution cell (i.e., input object) is "observed" as having a zero gray level (i.e., information content). The interpretation of this cell in terms of its thematic contents (i.e., giving a semantic meaning to the observed gray level) cannot be conducted if the *physical model* of the sensor is not known. If the used sensor is an imaging radar, then, the resolution cell contents can be interpreted as being a flat surface or may correspond to a shadow area. On the other hand, if the imaging sensor corresponds to a given spectral band in multispectral imaging, then the resolution cell thematic content corresponds to the content absorbing the emitted electromagnetic signals in the considered spectral band. From this example we can easily understand that the adjunction of the physical model (i.e., informative function) and the resolution (i.e., input element) makes the essential difference between data and information.

If we go further in this explanation, we can easily understand the difference between "generic" data processing techniques and what can be called "knowledge-based" processing techniques. To illustrate this idea, we can simply imagine the difference between classical image filtering techniques and filtering techniques adapted to speckled images (radar, sonar, ultrasound, etc.). In fact, the knowledge incorporated in speckled images filtering techniques corresponds to the mathematical modeling of the physical model exploited by the imaging sensors (i.e., a multiplicative Rayleigh noise). Moreover, the explicit positioning of the informative function within the basic information element structure clarifies the concept of ***information partiality*** or ***incompleteness*** (i.e., the information does not capture all relevant aspects of a phenomenon, an entity, or an input object). In fact, considering a sensor-based information element and since a sensor exploits a specific physical process in order to extract one or several "facets" of the observed objects, then the huge informative aspects of an observed object will be restricted to those acquired through the "physical" window of the considered sensor. For instance, the spectral signature of objects is only measured in few small spectral bands in a multispectral imaging system. In radar imaging systems, this partial sensor vision is related to the frequency, polarization, and geometrical acquisition configuration used by the sensor. As a direct consequence, and to overcome information incompleteness, the use and the development of information fusion systems are very desirable and even become crucial.

It is important to notice that this information element definition and structure are extremely general and can be applied to all types of encountered information elements: sensor issued information, data transformation information, feature extraction information, decisional information, etc.

1.2.2 Intrinsic Information Imperfection Types

Information imperfection usually arises at the early stages of the development of information processing systems, since it pervades the description of the domain or the real-world situation. It can be considered as the multifaceted concept characterizing the fact that a considered information element lacks fulfilling a predefined targeted objective. Most efforts for handling imperfect information have focused on "modeling" imperfections and on "processing" imperfect information elements through mature mathematical theories and approaches.

Quality of information (QoI) [10] provides the foundation and the reasoning framework for the conception, design development, operation of information processing, and fusion systems. Considerable research on studying and classifying various quality aspects into broad categories has been conducted. Wang and Strong [11] have classified QoI into four major categories: *intrinsic*, *contextual*, *representational*, and *accessibility*.

In the framework of information processing systems, QoI is addressed at only two levels: intrinsic and contextual levels. The intrinsic level concerns information characterization in terms of imperfection nature, interpretation, and modeling while considering the information element as an "independent" entity out of the global fusion context, whereas the contextual level concerns the information characterization in terms of its impact, completeness, relevance, conflict, redundancy, etc. within the global fusion context.

From an intrinsic point of view, various sources of imperfections are encountered ranging from the early input definition to the content outcome of the information element (including the nature of the informative function, or relation, as well as the available external sources of knowledge used by the information element).

Consider the information element $I = (\Theta, X, \Omega)$ of Fig. 1.1, where Θ (resp. Ω) denotes the definition (resp. content) set and X the informative relation. Imperfection modeling is studied in terms of impact of different information imperfection sources on the information content outcome. Three major intrinsic imperfection types are considered: *uncertainty*, *imprecision*, and *ambiguity*.

I. ***Uncertainty*****:** Most information processing systems efforts are concerned with adequate modeling of information uncertainty, which is the result of noisy, imprecise, erroneous or ill-suited to the problem data, ambiguous observations, and incomplete and poorly defined *prior* knowledge [12].

Assume the two following conditions hold:

1. The informative relation, X, is a *punctual outcome relation* (i.e., one information content, x_{True} from Ω is produced by X).
2. The set of information contents Ω is exhaustive and exclusive (i.e., the information content produced by X is unique and certainly belongs to Ω).

The information is said to be affected by uncertainty, if and only if the true information content x_{True} is *unknown* with certainty. The main objective of uncertainty imperfection modeling and processing is to represent and to deal with this lack of knowledge and to "determine" with total certainty the unknown true content outcome of the considered information (the class of an observed object, the decision to consider, etc.). Two major approaches are used to model and to process uncertainty type of imperfection: *probabilistic* and *evidential* approaches. Given that uncertainty affects punctual outcome informative functions and relations, both probabilistic and evidential approaches consider the *total certainty* as having a global measure of unity. Depending on the available knowledge concerning the true information content, x_{True}, both approaches differ in the way this global measure is distributed into different elements from Ω.

The probabilistic approach makes a "punctual certainty distribution" of the unity total certainty on different information contents (i.e., each information content, $x \in \Omega$, called a singleton, captures a partial amount of certainty $\Pr\{x\}$ where different $\Pr\{x\}$ add to one). $\Pr\{x\}$ is interpreted as the probability that x is the true information content. An information element $I = (\Theta, X, \Omega)$ affected by uncertainty type of imperfection and for which a probability distribution of uncertainty is available is called a *probabilistic information*.

The evidential approach (based on belief functions theory [13]) constitutes an excellent alternative to the probabilistic approach when the available knowledge does not allow making a punctual certainty distribution. In fact, evidential approach is based on making a "subset certainty distribution" of the unit total certainty on different subsets of Ω (i.e., each information content subset, $A \subseteq \Omega$, captures a partial amount of certainty $m(A)$ with different $m(A)$ add to one). $m(A)$ is interpreted as the mass of belief that the true information formation content x_{True} is in A. An information element $I = (\Theta, X, \Omega)$ affected by uncertainty type of imperfection and for which a mass distribution of uncertainty is available is called an *evidential information*.

Notice that the probabilistic approach can be considered as a special case of the evidential approach. Relaxing the punctual certainty distribution constraint, by the evidential approach, gives a "practical" dimension and a "facility" of knowledge representation. Nevertheless, this relaxation reduces the precision and the quality of information processing results when compared to the probabilistic approach.

II. ***Imprecision*:** Imprecision type of imperfection is an issue pertaining to the intrinsic quality of information. It refers to the case where the available knowledge about the true information content is available as a subset Ω_1 of the information content set Ω (i.e., $x_{\text{True}} \in \Omega_1 \subseteq \Omega$). In this case, the information $I = (\Theta, X, \Omega)$ is called an *imprecise information*. Therefore, imprecise information involves the lack of precise knowledge of the information content and, thus, should not be considered as erroneous. In the decision-making domain, imprecision represents the uncertainty as "a state of mind" of an agent which does not possess the needed information or knowledge to make a precise decision; the agent is in a state of uncertainty: "I'm not sure that this object is a table" [14]. Special kinds of imprecise information include:

- *Disjunctive* information content subset (e.g., John's age is either 31 or 32, the class of the object is either C_1 or C_2, etc.).
- *Negative* information content subset (e.g., John's age is not 30, etc.).
- *Range* information content subset (e.g., John's age is between 30 and 35, or John's age is over 30).
- *Error Margins* content subset (e.g., measured missile range is 100 ± 5 Km).

The two "boundary" kinds of imprecision are *precise* information (i.e., $\Omega_1 = \{x_{\text{True}}\}$), and *null* (also called missing data or total ignorance) information (i.e., Ω_1 encompasses the entire set of possible information content Ω). Notice that imprecise information is generally associated with an accuracy measurement quantifying the closeness of agreement between the information produced outcome (i.e., Ω_1) and the true information content, x_{True}. From an information processing point of view, an imprecise information is considered as a special case of evidential information where the total certainty is attributed to the subset Ω_1 (i.e., $m(\Omega_1) = 1$, and $m(A) = 0$ for all $A \neq \Omega_1$).

Remarks

- The aim of information processing approaches when dealing with both forms of imperfection (imprecision and/or uncertainty) is mainly to determine the unique true information content x_{True} with the highest precision and certainty degrees.
- Imprecision is often confused with uncertainty because both imperfection types are related to the same root (i.e., originated by punctual informative function where the "unique" true content is unknown: precisely, case of imprecision; or certainly, case of uncertainty). Also, both imprecision and uncertainty can be present at the same time, and one can cause the other. It is important to be able to tell the difference between these two antagonistic concepts, even if they can be included in a broader meaning for uncertainty (knowing that $x_{\text{True}} \in \Omega_1 \subseteq \Omega$ does not imply the precise and certain knowledge of x_{True}). To illustrate the

difference and potential "mixture" between imprecision and uncertainty, consider the following two situations:

1. Paul has at least two children and I'm sure about it.
2. Paul has three children but I'm not sure about it.

In the first information, the number of children is imprecise but certain; whereas, in the second information, the number of children is precise but uncertain [15].

III. ***Ambiguity*****:** Literally, information is said to be ambiguous if it is unclear what the information refers to the fact that it can be interpreted in several ways or that its truth or validity is not totally verified. From an informational point of view and assuming that the information content set Ω is exhaustive, then two types of ambiguity are encountered:

- *Non-specificity*: i.e., multiple content outcomes are produced simultaneously by the informative function.
- *Partial truth*: i.e., the information content is partially produced.

L. Zadeh [16] has proposed to model this imperfection type as a fuzzy set defined on the information content set, where each content outcome x is associated with a *membership value* $\mu(x)\in[0,1]$ representing the "strength" or the "truth" of production of the outcome x by the informative function: $\mu(x) = 0$ means that x is not produced (obtained or concerned); and $\mu(x) = 1$ means that x is fully produced.

An information element $I = (\Theta, X, \Omega)$ affected by the ambiguity type of imperfection and for which a membership function $\mu(.)$ is available is called an *ambiguous (or fuzzy) information*. Notice that in this case, it is nonsense to determine single information content. Therefore, the major objective of the application of fuzzy concepts is to combine and to conduct the fusion of multiple ambiguous information elements.

1.3 Possibilistic Information Concept

A particular situation, of high importance, where "hybrid" forms of imperfections is frequently encountered. This concerns the case where the information element is affected by the uncertainty imperfection type (i.e., having a punctual informative function where the true output x_{True} is unknown with certainty), but the available knowledge about x_{True} is "weaker" than probabilities (subjective knowledge, ambiguous, etc.). In this case, each content outcome x is associated with a *possibility value* $\pi(x)\in[0,1]$ representing the possibility strength that the outcome x to be the unique true information content. This type of information imperfection is called *epistemic uncertainty*, and the associated information element is referred as a *possibilistic information* [4]. A practical feature of possibility theory that is worth emphasizing is its interest for modeling uncertainty as well as preferences. This theory is detailed in the next chapters.

References

1. G.L.S. Shackle, *Decision Order and Time in Human Affairs* (Cambridge University Press, New York, 2010)
2. L.J. Cohen, *The Probable and the Provable* (Clarendon Press, Oxford, 1977)
3. E.N. Zalta, U. Nodelman, C. Allen, J. Perry, *Stanford Encyclopedia of Philosophy* (Stanford University. The Metaphysics Research Lab, Stanford, 2003)
4. L. Zadeh, Fuzzy sets as the basis for a theory of possibility. Fuzzy Sets Syst. **1**, 3–28 (1978)
5. R.M. Losee, A discipline independent definition of information. J. Am. Soc. Inf. Sci. **48**, 254–269 (1997)
6. R.M. Losee, *The Science of Information: Measurement and Applications* (Academic Press, San Diego, 1990)
7. B. Solaiman, Information fusion concepts: From information elements definition to the application of fusion approaches, in *Proceedings SPIE Volume 4385 - Sensor Fusion: Architectures, Algorithms, and Applications*, (2001), pp. 205–212
8. É. Bossé, A. Guitouni, P. Valin, An essay to characterize information fusion systems, in *Proceedings of the 9th International Conference on Information Fusion (FUSION2006)*, (Firenze, 2006)
9. T. Stonier, *Information and the Internal Structure of the Universe* (Springer, London, 1990)
10. É. Bossé, G. Rogova (eds.), *Information Quality in Information Fusion and Decision Making*Information Fusion and Data Science (Springer Nature, Basel, 2019)
11. R.Y. Wang, D.M. Strong, Beyond accuracy: What data quality means to data consumers. J. Manag. Inf. Syst. **12**, 5–34 (1996)
12. A. Appriou, Situation assessment based on spatially ambiguous multisensor measurements. Int. J. Intell. Syst. **16**, 1135–1166 (2001)
13. G. Shafer, *A Mathematical Theory of Evidence* (Princeton University Press, London, 1976)
14. A.-L. Jousselme, P. Maupin, É. Bossé, Uncertainty in a situation analysis perspective, in *Proceedings of the Sixth International Conference of Information Fusion*, (2003)
15. P. Smets, Imperfect information: Imprecision and uncertainty, in *Uncertainty Management in Information Systems*, (Springer, New York, 1997), pp. 225–254
16. L.A. Zadeh, Fuzzy sets. Inf. Control. **8**, 338–353 (1965)

Chapter 2
Fundamental Possibilistic Concepts

2.1 Introduction

In the real-world problems, it is impossible to avoid uncertainties. Uncertainty sources are diverse: incomplete domain knowledge, noisy and conflicting data, incomplete information, linguistic imprecise and ambiguous knowledge, etc. Up to middle of the twentieth century, most theoretical advances were devoted to the theory of probabilities. The second half of the twentieth century was very prolific for the development of new theories dealing with uncertainties [1, 2]. One of these theories is possibility theory [3, 4] that was introduced to allow reasoning to be carried out in the framework of a vague knowledge about the uncertainties. In fact, possibility theory can be described as a collection of techniques centered on the concept of a possibility distribution used for the representation and manipulation of the ambiguous or vague knowledge about the encountered uncertainty. In this chapter, the fundamental concept of possibility distributions is detailed in terms of its definition, its informative facets, and its different distribution models. Two important concepts are also detailed: the discounting concept (allowing to adjust a possibility distribution in order to take into consideration some external reliability knowledge) and the extension principle that allows to compute the new possibility distribution resulting from the projection of the set of alternatives using a deterministic extension projection function. Different operators allowing the merging of several possibility distributions are then detailed. Two set measures allowing to characterize subsets occurrence (i.e., possibility and necessity measures) are defined and their different characteristics are expressed. An important issue detailed within this chapter is related to subnormal possibility distributions where the available ambiguous knowledge is inconsistent.

B. Solaiman, É. Bossé, *Possibility Theory for the Design of Information Fusion Systems*, Information Fusion and Data Science,
https://doi.org/10.1007/978-3-030-32853-5_2

2.2 Possibility Distributions Concept

In a similar way to all theories dealing with uncertainty, through possibility theory, the estimation of the tendency, or likelihood, of the occurrence of an elementary event (i.e., singleton) is represented by a *possibility distribution* ($\pi - d$) depicting our state of knowledge of what is plausible from what is less plausible. The $\pi - d$ shows the graded partial belief of the occurrence of an elementary event ranging from "0" (for impossible ones) to "1" (for absolutely possible ones). Namely, let Ω denote a finite set of *mutually exclusive* alternatives that are of concern to us (diagnosis, hypothesis, classes, decisions, etc.). This means that in any situation, one and only one of these alternatives, called the *true alternative*, occurs at a time, and that Ω contains all potential alternatives that may occur. Ω is called *reference set*, *universe of discourses*, *frame of discernment*, *decision set*, *set of alternatives*, *set of states of the world*, etc. In this document, Ω will be called the *set of alternatives*. Each element x from Ω is called: *elementary alternative*, *basic hypothesis*, *elementary decision*, *state of the world*, *singleton*, etc. In this document, x will be simply called an *alternative*.

2.2.1 Defining a Possibility Distribution

Let Ω denote a finite set of *mutually exclusive* alternatives, where the unique occurring alternative, i.e., the true alternative, is unknown. This, in fact, resumes the uncertainty type of information imperfection where we face the problem of "discovering" or identifying the identity of the true occurring alternative. A *possibility distribution* π, defined on the set of alternatives Ω, is a point-wise mapping from the set Ω into the unit interval, i.e.,

$$\begin{array}{rcl} \pi : \Omega & \rightarrow & [0, 1] \\ x & \rightarrow & Poss\{x\} = \pi(x) \end{array}$$

The value $\pi(x)$ is interpreted as being our degree of belief or as representing a *flexible restriction* (i.e., a constraint) of the value of x on the set of alternatives Ω. It shows the graded partial belief of the occurrence of different alternatives with the following conventions:

- $\pi(x) = 1$ means that the alternative x is believed to be fully possible (i.e., the occurrence of x is totally compatible with the knowledge available about Ω).
- $\pi(x) = 0$ means that the alternative x is believed to be fully impossible to be the true alternative (i.e., the occurrence of x is totally incompatible, or in a total contradiction, with the knowledge available about Ω).
- $\pi(x) = p \in]0,1[$ indicates that the alternative x is considered as having a partial possibility to degree p of being the true occurring alternative.

- $\pi(x_1) > \pi(x_2)$ means that x_1 is a *preferred alternative* to x_2 for being the true alternative.

  Note that:

- Most of the authors "impose" a normalization condition to the possibility distribution meaning that at least one alternative should be fully possible (i.e., $\exists x_0 \in \Omega$: $\pi(x_0) = 1$). In this case, the possibility distribution is referred to as being *normal*. Otherwise, the possibility distribution is called *subnormal*, i.e., $\forall x \in \Omega$: $\pi(x) < 1$ (subnormal distributions will be discussed in detail later in this chapter).
- The unit interval [0, 1] of π may be replaced by any linearly ordered, possibly finite scale.
- As possibility distributions are defined over basic alternatives (also called *singletons*) and not on the events (i.e., not on subsets of Ω), they are called *point functions*.
- A possibility distribution, π, could be viewed as describing possible values that could be assigned to some unknown variable X taking values in the finite set of alternatives $\Omega = \{x_1, \ldots, x_N\}$ and assuming that X represents a possibilistic information. The unique true alternative is known through the possibility distribution π which acts as an elastic constraint on the alternatives that can be assigned to X. Therefore, $\pi(x)$ represents to what extent it is possible that x is the true alternative: $\pi(x) = Poss\{X = x_{True}\}$.
- (X, π) is called a *possibilistic variable*.

The two extreme forms of knowledge, i.e., the complete knowledge and the total ignorance, are simply modeled by the two following possibility distributions:

- Complete knowledge, CK, (i.e., the occurring true alternative x_{True} is known):

$$\pi_{\mathrm{CK}}(x_{True}) = 1 \text{ and } \pi_{\mathrm{CK}}(x) = 0, \quad \forall x \neq x_{True}$$

- Total ignorance, TI, (i.e., total lake of knowledge concerning the occurring alternative):

$$\pi_{\mathrm{TI}}(x) = 1, \quad \forall x \in \Omega.$$

A possibility distribution π defined on the set of alternatives Ω such that $\pi(x) > 0$ for all $x \in \Omega$ is called *nondogmatic* as it does not definitely exclude any alternative x from Ω (all $x \in \Omega$ are considered as being possible alternatives). The *height* of a possibility distribution π, denoted by $h(\pi)$, is the highest possibility value taken by different alternatives:

$$h(\pi) = max_{x \in \Omega}\{\pi(x)\}$$

The *core* of a possibility distribution π, denoted by $Core(\pi)$, is defined as the subset of fully possible alternatives:

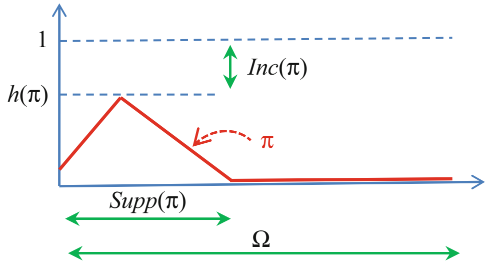

Fig. 2.1 Informative facets embedded in a possibility distribution

$$Core(\pi) = \{x : x \in \Omega, \pi(x) = 1\}$$

In general, a possibility distribution π defined on a set of alternatives Ω vehicles three main informative knowledge facets (Fig. 2.1).

1. *The domain of possible*, $Supp(\pi)$, is also called the support of the possibility distribution. In fact, the set of alternatives Ω is partitioned into two domains: the domain of possible containing alternatives having a possibility degree different from zero (i.e., $Supp(\pi) = \{x \in \Omega$ such that $\pi(x) > 0\})$ and the domain of impossible $\Omega / Supp(\pi) = \{x \in \Omega$ such that $\pi(x) = 0\}$.
2. *Possibilistic ordering*: The second informative source of knowledge is related to relative possibility degrees attributed to different alternatives. In fact, considering two alternatives x_1, $x_2 \in \Omega$ for which the possibilistic source of information has attributed the two possibility degrees $\pi(x_1)$ and $\pi(x_2)$. Besides the importance of the values of the possibility degrees attributed to both alternatives, the fact that $\pi(x_1) > \pi(x_2)$, for instance, encapsulates a relevant informative aspect for which the possibilistic source of knowledge "considers" the occurrence of the alternative x_1 as more credible than the alternative x_2.
3. *Inconsistency*: The third informative source of knowledge encapsulated in a possibility distribution is its degree of inconsistency defined as:

$$Inc(\pi) = 1 - h(\pi)$$

$(Inc(\pi) \in [0,1])$, where $h(\pi)$ indicates the height of the possibility distribution (i.e., the highest possibility degree). In fact, this element reflects the degree to which the possibility distribution can confirm, or not, if at least one of the alternatives is fully possible to occur. Unfortunately, this important informative knowledge source element simply "disappears" when a possibility distribution is normalized (forcing, thus, to have at least one alternative to be fully possible).

2.2.2 Possibility Distribution Models

Different types of encountered possibility distributions are detailed in this section. Each type is assumed to model a given form of the available knowledge about the identity of the true, unique but unknown alternative from Ω.

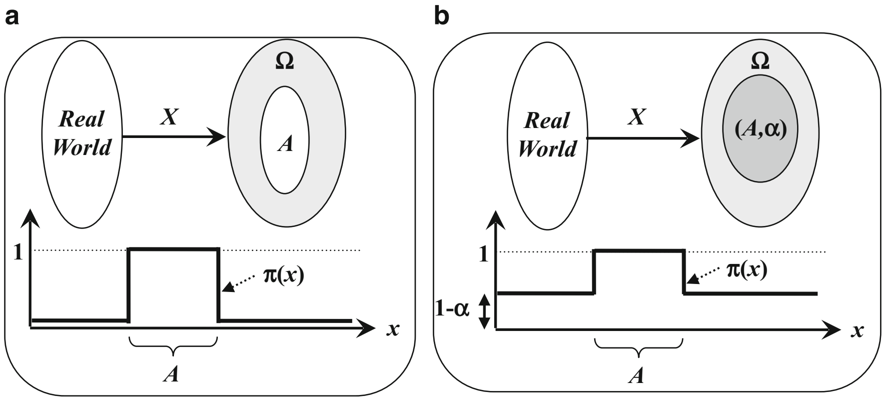

Fig. 2.2 Some types of imprecise type possibility distribution functions. (**a**) Imprecise information, (**b**) α-Certain imprecise information

(A) Imprecise type possibility distribution *Imprecise information* on the set of alternatives Ω, assumed to be exhaustive and the alternatives are mutually exclusive (i.e., one and only one alternative occurs at time), is defined as an uncertain information for which the knowledge about the true alternative is expressed as a subset $A \subseteq \Omega$. This type of information is frequently encountered and is more natural than giving a "point" alternative. Some examples of imprecise information are encountered when an expert claims that the true value x lies within the interval $x_0 \pm \Delta x$ or when a physician declares that the patient's illness is certainly one out of a subset of mutually exclusive illnesses.

The following distribution, called *imprecise information possibility distribution*, allows representing easily this type of information (Fig. 2.2a):

$$\begin{aligned} \pi : \Omega &\rightarrow [0,1] \\ x &\rightarrow \pi(x) = \mathbf{1I}(A) = \begin{cases} 1 & \text{if} \quad x \in A \\ 0 & \text{if} \quad x \notin A \end{cases} \end{aligned}$$

where **1I** (A) denotes the classical algebraic characteristic function of the subset A. Nevertheless, imprecise information possibility distributions are too restrictive. In fact, claiming that $\pi(x) = 0$ for some $x \notin A$ means that x is fully impossible to occur. This is too strong for the expert who is then tempted to give a wide uninformative interval support of π. It is worthwhile to notice that this type of possibility distributions assumes binary values and doesn't offer the possibility to express partial degrees of possibility.

(B) α-Certain imprecise type possibility distribution An *imprecise information with a certainty factor α*, (also called α-Certain imprecise information) is an imprecise information where the available knowledge about the true alternative is expressed as a subset $A \subseteq \Omega$ associated with a certainty level of trust $\alpha \in [0,1]$ concerning the occurrence of A, for instance, the information delivered by a physician declaring that the patient's illness is certainly one out of the two mutually exclusive illnesses $\{H_1, H_2\}$ and that he is 90% sure. In this case, $A = \{H_1, H_2\}$ with a certainty factor $\alpha = 0.9$. The possibility distribution modeling this type of information has been suggested in [5] as follows (Fig. 2.2b):

$$\begin{aligned} \pi : \Omega \quad &\rightarrow \quad [0,1] \\ x \quad &\rightarrow \quad \pi(x) = max\{\mathbf{1I}(A), 1-\alpha\} \end{aligned}$$

which expresses that the "remaining trust," i.e., $1 - \alpha$, will be considered as the possibility degree that the true alternative lies outside A.

(C) Epistemic type possibility distribution The objective of a possibilistic assessment is to produce a measure of the degree to which the available/acquired knowledge supports each alternative in the set of alternatives Ω. The result of this assessment is a possibility distribution. In several cases, the available knowledge about the true alternative is given as a constraint defined in terms of a "fuzzy concept" defined on Ω. It is important to notice that the concept of possibility distributions is closely related to that of fuzzy sets. Let Ω denote an exhaustive set of *mutually exclusive* alternatives, on which a fuzzy evidence A is defined, and denote $\mu_A(x)$ the corresponding membership function:

$$\begin{aligned} \mu_A : \Omega \quad &\rightarrow \quad [0,1] \\ x \quad &\rightarrow \quad \mu_A(x) \end{aligned}$$

Assuming that A is the available knowledge about the true occurring alternative, the question is then: "How can we obtain a possibility distribution on the set of alternatives?"

L. Zadeh [3] has formulated the so called *possibility postulate* which may be considered as the basis for a possibilistic interpretation of the fuzzy evidence:

Possibility postulate In the absence of any information regarding the true alternative from Ω than that conveyed by the fuzzy evidence A, then, $\mu_A(.)$ induces a possibility distribution $\pi_A(x)$ which equates the possibility for an alternative $x \in \Omega$, be the true one, to the grade of membership $\mu_A(.)$.

Example Assume that we are to assess the possibility of the occurrence of a number from 0 to 10 given that the available evidence is that the number is *small* (Fig. 2.3).

Since we know that the compatibility of the concept *Small* with the number "4" is 0.6, $\mu_{Small}(4) = 0.6$, we conclude that the occurrence possibility degree of 4 is considered as being 0.6. By a similar argument, the possibility of 0, 1, 2, and 3 is 1 and of the numbers greater than 5 is 0. Notice that the membership function, $\mu_A(x)$, is viewed as computing a degree of assurance, certainty, or possibility that an alternative $x \in \Omega$ satisfies the property of being a member of the fuzzy set (or the ambiguous piece of information) defined by μ_A. Therefore, if we were to select a

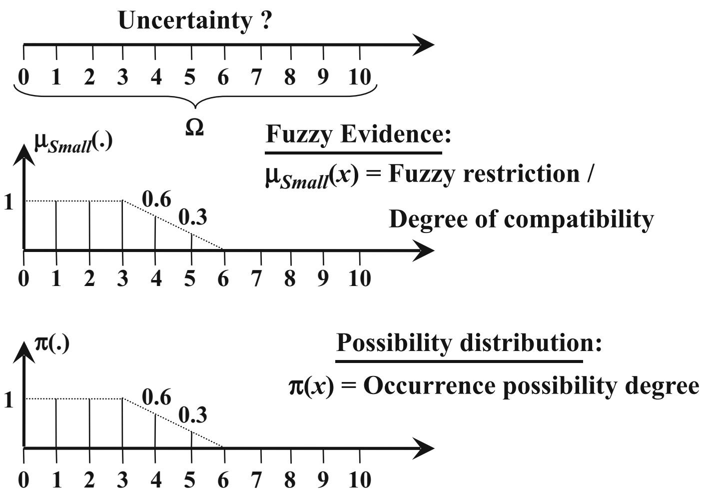

Fig. 2.3 Epistemic possibility distribution assessment

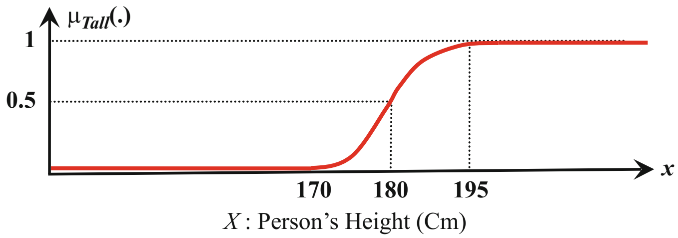

Fig. 2.4 Possibility distribution function associated with *Tall*

given $x \in \Omega$ as being the true alternative, we would be only certain to a degree $\mu_A(x)$ that x satisfies the aforementioned ambiguous information. For instance, if our goal is to know the exact height of Peter, and we only know that Peter is *Tall*, then the possible values of Peter's height are restricted by the possibility distribution associated with the fuzzy set *Tall* and defined by (Fig. 2.4):

$$\pi_{Peter\text{'s } height}(x) =_{\text{def}} \mu_{Tall}(x)$$

It is crucial to understand that the values duplication (i.e., $\pi_{Peter\text{'s } height}(x) = \mu_{Tall}(x)$) is to be positioned at the numerical and not at the semantical level. In fact, $\mu_{Tall}(x)$ reflects the compatibility of one "feature" of persons (here, the size) with the fuzzy property *Tall*. This compatibility value is extended (or projected) to become representing the possibility degree for a given size to be the true one for a given person. An interesting example in mammographic image interpretation can be

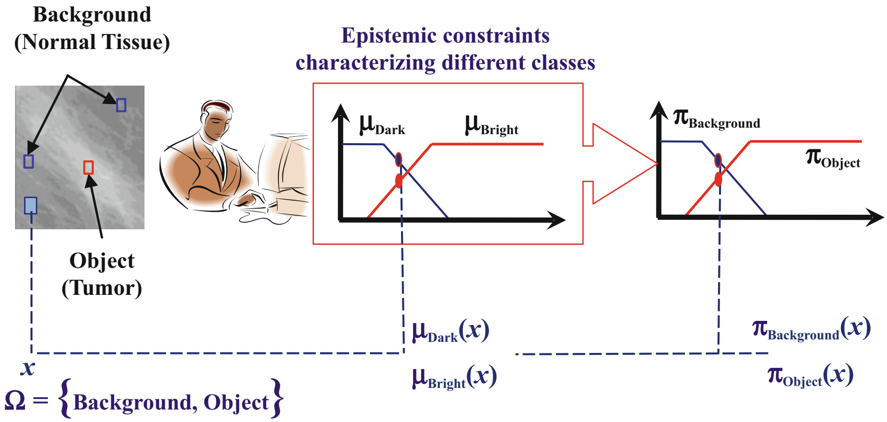

Fig. 2.5 Possibility distribution functions associated with normal tissue and tumor classes in mammography

considered (Fig. 2.5). In this example, physicians express their knowledge describing normal and tumor tissues by characterizing these tissues as being "observed" as *dark* and *bright* pixels in mammographic images.

Notice that this description represents two fuzzy sets (*dark* and *bright*) defined on the observed gray levels definition set. The membership functions of these fuzzy sets constitute the epistemic constraints allowing to define the possibility degrees for an observed pixel to belong to a normal or to a tumor tissue.

It is worthwhile to mention that all standard types defining membership function and representing the fuzzy constraints (i.e., triangular, trapezoidal, Gaussian, singleton-based, piecewise linear, etc.) can be applied for the definition of the epistemic type possibility distributions.

(D) Qualitative possibility distributions Experts frequently meet difficulties in providing precise numerical values of possibility degrees. It seems, thus, more natural for them to give an "order relation" between different alternatives of the universe Ω.

Consider a finite universe of alternatives $\Omega = \{x_1, \ldots, x_N\}$ and an ordered scale $\mathcal{L} = \{a_0 = 1, a_1, \ldots, a_L, a_{L+1} = 0\}$ such that $a_0 = 1 > a_1 > \cdots > a_L > a_{L+1} = 0$. A qualitative possibility distribution is defined as a function associating to each alternative (from the universe Ω) an element $a \in \mathcal{L}$ (from the ordered scale $\mathcal{L}$) enabling, thus, to express that some alternatives are "more possible" than others (without referring to any numerical value, and this, makes the difference with other quantitative possibility distributions setting. In other words, assigning "qualitative" values a_k as possibility values (i.e., $\pi(x_n) = a_k$) implies alternatives ranking importance representation rather than pure numerical possibilistic degrees. However, we can derive an infinity of quantitative possibility distributions from a qualitative one.