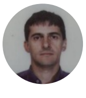
Choice under imprecision: a simple possibility-theory-based technique illustrated with a problem of weed identification

ARTICLE *in* COMPUTERS AND ELECTRONICS IN AGRICULTURE · AUGUST 1989

Impact Factor: 1.76 · DOI: 10.1016/0168-1699(89)90028-8



Roger Martin-Clouaire
French National Institute for …

Computers and Electronics in Agriculture, 4 (1989) 103-120
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

Choice under Imprecision: A Simple Possibility-Theory-Based Technique illustrated with a Problem of Weed Identification

JACQUES BADIA and ROGER MARTIN-CLOUAIRE

Station de Biométrie et Intelligence Artificielle, Institut National de la Recherche Agronomique (INRA), B.P. 27, 31326 Castanet-Tolosan Cédex (France)

(Accepted 10 March 1989)

ABSTRACT

Badia, J. and Martin-Clouaire, R., 1989. Choice under imprecision: a simple possibility-theory-based technique illustrated with a problem of weed identification. *Comput. Electron. Agric.*, 4: 103-120.

Choosing among a predefined set of alternatives is a very common problem in many domains of human activity. This paper presents a simple tool for:
- describing a collection of objects (or alternatives) along a given set of attributes in a situation where the available knowledge is pervaded by imprecision; and
- retrieving from among them those being compatible (to some extent) with some data (possibly also imprecise) gathered along the same set of attributes and concerning an unknown object to be identified.

Our approach advocates the use of possibility theory, which is a relatively recent and very powerful tool for dealing with the vagueness and uncertainty of available knowledge. The proposed technique has been implemented as an autonomous module and incorporated into a microcomputer system specifically designed for the problem of weed identification in cultures located in continental France. In this paper, the weed-identification application serves the basic purpose of illustration, and the method is described in a way which preserves its generality. Background material on possibility theory and a short discussion of the differences from probability theory are provided.

INTRODUCTION

The wide variety of problems commonly encountered by decision-makers includes planning, allocating resources, and assessing risk. Basically, a decision problem is characterized as selecting one (or a small subset) from several options so as to maximize some function of many possible variables, attributes or criteria.

The decision-making problem addressed in this paper consists of identifying

0168-1699/89/$03.50

an unknown object described by some data gathered along a given set of attributes among a collection of objects, knowing their descriptions along the same attributes. In other words, starting from a matrix of objects (or options or alternatives) against attributes representing knowledge of a collection of objects, the problem is to find which objects best correspond to an unknown one characterized by the values assessed for each attribute. In this paper, a single decision-maker situation is assumed.

This problem basically involves a matching assessment and bears some resemblance[1] to the one of selecting objects that satisfy some properties (patterns) among a data base of objects, each described along a given set of attributes also used in specifying the desired properties. To give an expertise-free example of such a selection problem, suppose that the set of objects is a collection of second-hand cars and that the data base (knowledge matrix) gives for each the values taken by such attributes as age, cost, mileage, speed and fuel consumption. What we are interested in is to find one or several cars satisfying a particular request, such as being 10 years old, costing $1500, having done 90 000 km, able to reach 110 km/h and having a consumption of 8 liters per 100 km.

Stated thus, the problem does not look very difficult to solve provided the information used in the matrix and in the query is of good quality (i.e. precise and certain). Unfortunately, an inherent characteristic of knowledge available to humans is that it is imperfect, in the sense of being incomplete, uncertain, imprecise, possibly inconsistent, or otherwise not totally suited to the judgmental task at hand. In this paper, we address the case where the attribute values used in the matrix and those specified for the unknown object(s) to be identified are imprecise or vague.

More specifically, to elaborate on our previous example, what is more interesting and considered here is to be able to retrieve cars satisfying a softer description, such as being around 10 years old and rather cheap, displaying less than 100 000 km, not too limited in speed and having a moderate fuel consumption. Moreover, due to limitation in information-gathering, some matrix cells may also be imprecise or vague. For instance, the mileage of a particular car (say Car-1) may be reported as 'roughly between 85 000 and 110 000 km' if for some reason the mileometer cannot be trusted and its consumption is likely to be known as 'around 10 liters per 100 km'.

Consequently, the problem tackled here is basically one of representation

[1]Selecting and identifying are slightly different indeed. For instance, in the second-hand-cars example the selection process requires that, for each car, we find out whether the data describing it in the knowledge matrix entail the desired properties expressed in the query. On the other hand, with the same knowledge matrix, an identification investigation would serve to determine which cars are compatible with the data observed on an unknown car (to be identified). In this case, the basic task to be accomplished consists of evaluating whether the observed data of the unknown car entail the specific characteristics of a car described in the data base.

and processing of imprecision and uncertainty. To deal with the kind of imperfect information encountered in this problem, it is our belief that the classical techniques of probability theory cannot be used, because imprecision is different from randomness and subjective judgement does not fit well into a probabilistic framework. Rather, we advocate the use of a particular mathematical model based on the theory of fuzzy sets, also called possibility theory (Zadeh, 1978). An attribute value is imprecise when the described parameter (e.g. the age of a given car) is not specified through a single value as, for instance, in 'the car is more than 10 years old'. Indeed, in this example, the attribute value is actually a set. However, in order to deal with the extreme case of imprecision known as fuzziness or vagueness (e.g. 'the car is around 10 years old') we also need to represent classes that do not have crisp boundaries, i.e. classes in which there is no sharp transition from membership to non-membership. Such classes have been named fuzzy sets by Zadeh (1965). For example, the class of old cars is a fuzzy set. So are the classes of objects characterized by such commonly used adjectives as large, small, substantial, important, which are sometimes called 'fuzzy predicates'.

In the next section, we introduce some basic material about possibility theory. A subsequent section shows how it can be used in a simple way to identify a poorly described object from a set of poorly defined alternatives. The technique developed for this general problem is presented through an application to weed identification, the project having motivated the undertaking of the present work. Finally, some concluding remarks discuss the scope and utility of the approach in general and point out directions toward which the presented model could be extended.

BACKGROUND ON POSSIBILITY THEORY

Our intention in this section is merely to lay bare enough of the mathematical tools of possibility theory to enable readers to understand the application made to our selection (or choice) problem. We first give some informal definitions of uncertainty and imprecision. Basic mathematics of possibility theory are then introduced and a short discussion about the differences from probability theory is also provided. More on possibility theory may be found in, for instance, Dubois and Prade (1988).

Imprecision versus uncertainty

Before proceeding it is crucial to distinguish between imprecision and uncertainty, although these words have already been used and partially defined in the introduction. A proposition (or equivalently, an event) is imprecise when it is not elementary, but composed of a cluster of elementary propositions. For instance, 'the age of the car is somewhere in [9, 11]' is an imprecise proposition that can be seen as the disjunction of the three elementary propositions: 'the age of the car is 9 years', 'the age of the car is 10 years', 'the age of the car is 11 years'. An imprecise proposition is said to be vague or fuzzy if the set of elementary propositions it refers to has no precise boundaries as 'the age of the car is roughly 10 years'. An uncertain proposition is one the truth of which is not completely known. Truth is a notion that always refers to what is actually known of reality (i.e. the actual or assumed state of facts), and depends on the individual assessing it. A proposition may be both imprecise and uncertain. The uncertainty of a non-vague proposition can be expressed by assigning a numerical grade of confidence to the assumption that the proposition is true and one grade to the assumption it is false. Non-vague propositions can only be either true or false. Vague propositions are special in the sense that they can be neither true nor false but something in between. For instance, some individual (having his own discernment of what 'young men' means) could say that 'Jean is young' is rather true knowing that Jean is actually 30. Uncertainty and imprecision are closely linked notions since the more imprecise a statement the more certain it can be.

Possibility and necessity measures

Consider the set $\mathbb{P}$ of events associated to a body of uncertain knowledge. Each event can be seen as a subset of a reference set Ω called the event always true or the universe. A particular event is the impossible one denoted by $\emptyset$. To deal with uncertainty and imprecision we need a mathematical function measuring the confidence in the occurrence of any particular event (i.e. subset of Ω). A possibility measure Π is one such measure defined as a function from $\mathbb{P}$ to $[0, 1]$ satisfying:

$$
\begin{aligned}
&\text{(a)} \quad \Pi(\emptyset)=0 \\
&\text{(b)} \quad \Pi(\Omega)=1 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (1) \\
&\text{(c)} \quad \Pi(A \cup B)=\max[\Pi(A), \Pi(B)] \quad \text{for any A and B in } \mathbb{P}
\end{aligned}
$$

A worth-noticing consequence of (1) is that:

$$\max[\Pi(A), \Pi(\neg A)]=1 \quad \text{for any A in } \mathbb{P} \tag{2}$$

where $\neg A$ is the complement of A (i.e. the opposite event of A). Formula (2) ensures that, for two opposite events, at least one has a possibility equal to 1 (i.e. is completely possible); however, the fact that an event has a possibility equal to 1 does not prevent the opposite event having a non-zero possibility or, even, a possibility equal to 1. This property mirrors the least-commitment attitude characterizing a common-sense judgement of possibility. Saying that A and $\neg A$ are both completely possible corresponds to a situation of total ignorance where the occurrence of A is not more surprising than the one of $\neg A$. Another way to justify the word 'possibility' as a name for such a confidence measure is to think in terms of physical possibility; the axiom (c) tells nothing other than that to realize $A \cup B$ it is sufficient to realize the easiest of A or B.

Generally, a possibility measure Π can be built from a so-called possibility distribution π, which is a function from Ω to $[0, 1]$, in the following way:

$$\Pi(A) = \sup_{u \in A}[\pi(u)] \quad \text{for any A in } \mathbb{P} \tag{3}$$

Note that $\Pi(\{u\}) = \pi(u)$, where $\{u\}$ denotes a singleton; π can be viewed as a fuzzy restriction on the possible values of a variable, say X, which takes its value in Ω. Then, the possibility measure Π defined by formula (3) enables us to compute the possibility that X lies in a given subset A of Ω, or, in other words, the possibility that the proposition 'X is in A' is true knowing that the value of X is one of the more or less possible values described by π. Since $\Pi(\Omega) = 1$ we must have $\sup_{u \in \Omega}[\pi(u)] = 1$; a possibility distribution such that at least one element u° such that $\pi(u^\circ) = 1$ exists is said to be normalized.

In possibility theory, another measure of confidence is also used, called necessity or certainty measure (denoted by N), and can be built from Π according to the duality expressed by:

$$N(A) = 1 - \Pi(\neg A) \quad \text{for any A in } \mathbb{P} \tag{4}$$

A necessity (or certainty) measure is such that $N(\varnothing) = 0$ and $N(\Omega) = 1$ but satisfies a dual axiom of (c) as follows:

$$N(A \cap B) = \min[N(A), N(B)] \quad \text{for any A and B in } \mathbb{P} \tag{5}$$

Consequently, we have:

$$\min[N(A), N(\neg A)] = 0 \quad \text{for any A in } \mathbb{P} \tag{6}$$

$$N(A) = \inf_{u \notin A}[1 - \pi(u)] \quad \text{for any A in } \mathbb{P} \tag{7}$$

Formula (4) expresses that the necessity of an event corresponds to the impossibility of the opposite event, which is the usual relationship between possibility and necessity in modal logic. Formula (6) entails that, among two opposite events, at most one has a non-zero necessity (i.e. two opposite propositions cannot be certain simultaneously).

Fuzzy sets

A fuzzy set F on the universe Ω is defined by a membership function μ_F which is a function from Ω to $[0, 1]$. For any u in Ω, $\mu_F(u)$ must be interpreted as the membership degree of u in F and expresses the compatibility of u with the concept conveyed by F. Vague categories may then be represented on an objective universe that may be numerical (e.g. Ω is set of ages of cars) or not (e.g. Ω is set of cars) depending on the need. See Fig. 1 for an example involving

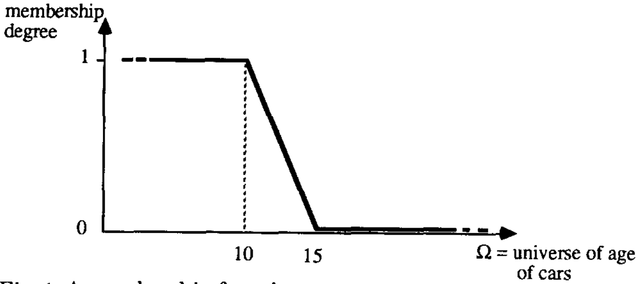


Fig. 1. A membership function.

the fuzzy set of ages compatible (to a degree) with the concept 'not old' applied to cars. Ω could have been taken as a discrete set instead of a continuum. If Ω is the set of real numbers then the fuzzy sets on Ω express fuzzy quantities.

A possibility distribution can always be seen as the membership function of a fuzzy set, but the converse is not true. The subtle difference between them is explained in the next section. The notions of union, intersection and complementation have been extended to fuzzy sets. The most commonly used definitions of these operations are:

$$\mu_{F\cup G}(u)=\max[\mu_F(u),\mu_G(u)]$$
$$\mu_{F\cap G}(u)=\min[\mu_F(u),\mu_G(u)] \tag{8}$$
$$\mu_{\neg F}(u)=1-\mu_F(u) \quad \text{for any } u \text{ in } \Omega$$

Formulae (3) and (7) define the possibility and the necessity of crisp events (i.e. represented by non-fuzzy sets). They can be extended to fuzzy events (i.e. represented by fuzzy sets) in the following way:

$$\Pi(A)=\sup_{u\in\Omega}\min[\mu_A(u),\pi(u)] \tag{9}$$
$$N(A)=\inf_{u\in\Omega}\max[\mu_A(u),1-\pi(u)] \tag{10}$$

where μ_A is the membership function of the fuzzy set A. Formulae (1), (4) and (5) still stand in the case of fuzzy events, but neither (2) nor (6) are guaranteed to hold.

Main differences with probability theory

When the available knowledge is in the form of frequencies of elementary events, the measure of confidence, P, that can be constructed naturally satisfies the well-known additivity axiom:

$$P(A\cup B)=P(A)+P(B) \quad \text{for any A and B that are disjoint} \tag{11}$$

Then P is nothing but a probability measure and (11) is the probabilistic counterpart of (1) and (5). Possibility, necessity and probability measures can all be computed from distributions on elements of Ω. Normalization conditions exist for both kinds of distributions involved: $\sum_{u \in \Omega} p(u) = 1$ in the case of a discrete probability distribution p and $\sup_{u \in \Omega}[\pi(u)] = 1$ with a possibility distribution π. Nevertheless, probability and possibility theories depart from each other in terms of what they can model of the real world. The major difference appears clearly in the contrast between the property expressed by formula (2) and the probability counterpart which is given by $P(A) + P(\neg A) = 1$. In other words, the probability of an event determines completely the probability of the opposite event, whereas the possibility of an event is poorly related to the possibility of the opposite one.

A possibility-based representation of uncertainty is appropriate when evidence in favor of the truth of a proposition is not directly linked to evidence in favor of the opposite, which is especially the case when uncertainty is subjective. Information humans commonly have to process is often of this kind. Faced with subjective uncertainty, some theorists have introduced an interpretation in terms of subjective probabilities. However, the relevance of this modeling of subjective judgements by probabilities is questionable on several points including limited expressive power, too strong underlying assumptions regarding the available evidence, and assessment difficulties (Dubois and Prade, 1989). A deep discussion of these issues is beyond the scope of this paper; however, pointing out the limitation of probability to model ignorance (even partial) helps in understanding why other theoretical frameworks have been created.

In probability theory, total incapacity in deciding in favor of the event A or the opposite $\neg A$ is translated into $P(A) = P(\neg A) = 0.5$. This is unsatisfactory because, first, it is then impossible to distinguish the case of total ignorance from the case of total uncertainty where A and $\neg A$ have effectively as much chance to occur and, second, it leads to paradoxical results as exemplified next. Consider $\Omega = \{$brown shirt, yellow shirt, no shirt$\}$ which contains the only three possible elementary events qualifying a murderer to be identified. The knowledge of a non-witness is vacuous by definition. In probability this situation translates (via the maximum-entropy principle) into the assignment of the same probability degree 1/3 to each elementary event. A consequence of this is, for instance, that it is twice as probable that the unknown person had a shirt rather than no shirt (i.e. P(shirt) = (P(brown shirt) + P(yellow shirt) = 2/3), which seems a very strong statement in a situation of total ignorance!

The inability of probability theory to represent partial ignorance which often pervades what is still valuable knowledge has been noticed by several practitioners in different domains of application, and has led to independent developments of new numerical approaches to uncertainty. Besides possibility theory, there is currently a great deal of research done mainly in the artificial intelligence context on the so-called Dempster–Shafer theory of evidence (Shafer, 1976) which offers, as a by-product, a common framework for comparing and understanding the different kinds of uncertain information (probabilistic, possibilistic and others).

In summary, what is important to keep in mind is that a single number (e.g. a probability) is not always sufficient to assess uncertainty. This is why a possibility assessment is used together with a necessity one. Ignorance about an event A is expressed by $\Pi(A)=1$ and $N(A)=0$; total uncertainty is equivalent to $\Pi(A)=N(A)=0.5$.

WEED IDENTIFICATION IN THE POSSIBILISTIC SETTING

The research institute the authors belong to has set up the project of constructing a large weed-control computer system designed for crop farmers of continental France. Ultimately, this system should concern about 450 species belonging to the dicotyledonous weeds of the French flora and about 50 species of various families including gramineas, rushes and sedges. So far a prototype, called MALHERB (Badia, 1988), restricted to the task of weed identification, has been built.

Given an unknown weed to be identified, and depending on what information can be obtained by observing it, the determination may be done in a one or two-step process. In the first, an identification based on morphological characteristics of the leaves, stem and roots is performed. If this step provides only a set of possible weeds (instead of a single one) as a result, then the system goes through the second step which pushes the identification further (for pruning the set of possible alternatives obtained through the first step, which are typically less than ten in number) by taking into account other characteristics concerning the environment where the unknown plant has been found. It is in this last step that possibility theory is used, as discussed in the sequel.

Assume we have a collection of weeds (typically less than ten, as we said) known in terms of the value taken by each of them along the four environmental attributes which are: climatic zone, issuing crop, acid/basic state of soil, and texture of soil. Now, given a yet unidentified weed observed along the same attributes, we want to find which known weeds most closely resemble it. Solving this problem incorporates three different tasks: representing the available knowledge, assessing to what extent the observation on an attribute satisfies what is known about this attribute for a given weed, and aggregating for any weed the above assessments made for each attribute.

These three tasks are now considered in turn. Finally we discuss the best way to use the results provided by our technique after these three tasks have been performed. The presentation is made in a way preserving the generality of the approach.

Knowledge representation

The available knowledge about a set of R alternatives (weeds) is supposed to be represented in a table having C columns, each corresponding to an attribute, and R rows. Each attribute is single-valued; it can only take one value at a time (though there may be several candidates) as opposed to a multivalued attribute as, for instance, 'time of chemical cleaning on the field' could be since its value is likely to be a set of dates. An element K_{ij} at the ith row and the jth column stands for the class of possible values that can be taken by the jth attribute of the ith alternative. Usually, such a class is fuzzy rather than crisp and it can be represented by a membership function μ_{ij} on the domain Ω_j (or reference set) of the concerned attribute. Then[2], $\mu_{ij}(u)=1$ means that u certainly belongs to the set, whereas $\mu_{ij}(u)=0$ signifies that it certainly does not belong to it. A further hypothesis in the model is that the attributes are not interactive, which means that the value taken by any of them does not depend on the values taken by the others.

Each fuzzy set μ_{ij} is a piece of agricultural botanic knowledge that corresponds to the question: "what are the values that the jth attribute may take if the weed under consideration is the ith one." Each attribute takes value on a discrete domain. As far as notation is concerned, given a domain $\{a, b, c, d, e\}$, a fuzzy set F may be defined by $\{(a, 0), (b, 1), (c, 0.4), (d, 0.2), (e, 0.5)\}$ with the understanding that, for instance, the pair $(c, 0.4)$ signifies that the grade of membership of c in F is 0.4 (i.e. $\mu_F(c)=0.4$) and that a definitely does not belong to F (i.e. $\mu_F(a)=0$). Pairs of the kind $(a, 0)$ are usually omitted in the definition of a fuzzy set so that F may have been given by $\{(b, 1), (c, 0.4), (d, 0.2), (e, 0.5)\}$. All the fuzzy sets considered in this application are assumed normalized (i.e. for any ill-bounded class there is always at least one element that certainly belongs to it). The respective domains of the attributes we are dealing with in the weed identification problem are given in Table 1. Each box corresponds to one attribute and each element of a domain has a symbolic name (shorter than its real one and used henceforth for convenience) given on the left-hand side.

Besides having to represent the expert knowledge conveyed by the fuzzy sets of values that each attribute may take for each alternative (see Table 2 where there are three alternatives), we have to represent the data collected by a user who wants to identify which alternatives best match her/his observations which are made with respect to the same attributes. Since any observation may be imprecise or vague, the user's data (denoted by $(O_1, ..., O_C)$) are best represented by a collection of possibility distributions π_j, $j=1, ..., C$ on the corre-

[2]In some cases, though not encountered in our application, μ_{ij} may be a fuzzy predicate (e.g. old) or fuzzy concept. Then, the membership degree $\mu_{ij}(u)$ conveys the compatibility of u with the concept under consideration.

TABLE 1

Attribute domains

Attribute 1: Climatic type	Attribute 2: Issuing crop
a_1 Mediterranean	a_2 fall colza
b_1 sub-Mediterranean	b_2 fall cereals or fall protein-rich crop
c_1 oceanic	c_2 spring cereals or spring protein-rich crop
d_1 temperate	d_2 potatoes
e_1 others	e_2 beet
	f_2 sunflower or soya
	g_2 corn or grain sorghum
	h_2 vineyard or Mediterranean non-irrigated orchard
	i_2 temperate or Mediterranean irrigated orchard
	j_2 market-gardening
Attribute 3: Acid/basic state of soil	**Attribute 4: Texture of soil**
a_3 basic	a_4 clay
b_3 neutral	b_4 compound: clay to silt
c_3 acid	c_4 compound: silt to sand
	d_4 clay and sand
	e_4 sand

TABLE 2

Fuzzy sets representing botanic knowledge about three weeds

Weed	Attribute			
	Climatic type	Issuing crop	Acid/basic state of soil	Texture of soil
Rumex crispus	$\{(a_1, 1), (b_1, 1), (c_1, 1), (d_1, 1), (e_1, 1)\}$	$\{a_2, 0.8), (b_2, 0.8), (c_2, 0.8), d_2, 0.8), (e_2, 0.8), (f_2, 0.8), (g_2, 0.8), (h_2, 0.8), (i_2, 1)\}$	$\{(a_3, 1), (b_3, 0.8), (c_3, 0.8)\}$	$\{(a_4, 0.8), (b_4, 1), (c_4, 1), (d_4, 0.8)\}$
Rumex obtusifolius	$\{(a_1, 1), (b_1, 1), (c_1, 1), (d_1, 1), (e_1, 1)\}$	$\{(a_2, 0.8), (b_2, 0.8), (c_2, 0.8), (d_2, 0.8), (e_2, 0.8), (f_2, 0.8), (g_2, 0.8), (h_2, 0.8), (i_2, 1)\}$	$\{(a_3, 0.8), (b_3, 0.8), (c_3, 1)\}$	$\{(a_4, 0.2), (b_4, 0.8 (c_4, 0.8), (d_4, 1)\}$
Rumex pulcher	$\{(a_1, 1), (b_1, 1), (c_1, 0.2)\}$	$\{(h_2, 1), (i_2, 0.8)\}$	$\{(a_3, 1), (b_3, 1), (c_3, 1)\}$	$\{(a_4, 1), (b_4, 1), (c_4, 1), (d_4, 1)\}$

sponding domains; $\pi_j(u)$ is the grade of possibility that u is indeed the value of the jth attribute. Considering the observation on the jth attribute, $\pi_j(u)=1$ means that u is completely possible (but not that it is certain that u is the value of the attribute except if $\pi_j(v)=0$ for any v different of u), while $\pi_j(u)=0$ means that u is totally impossible as a value of the jth attribute. We assume that for any attribute there is always at least one completely possible value (i.e. the involved possibility distributions are normalized). Thus, collected data take the form shown in Table 3. Note that the datum associated to the third attribute is precise whereas the others are imprecise.

The semantic difference between μ_{ij}, which represents an ill-bounded class, and π_j, which represents an ill-known value, is a good example of the subtle distinction between a fuzzy set and a possibility distribution that was alluded to earlier. A possibility distribution contains the idea that any two values of the domain are mutually exclusive (i.e. the attribute has a unique value); a fuzzy set does not. Nevertheless, since a possibility distribution restricts the set of values a single-valued variable may take it also implicitly defines the membership function of the fuzzy set of possible values.

In practical terms, the identification of a membership function or a possibility distribution can be relatively qualitative since possibility theory is not very sensitive to slight variations in evaluating the involved degrees. Actually, what is important is the order induced by these degrees on the set of reference and not the exact degree associated to any particular element. This nice feature comes from the fact that possibility theory essentially uses maximum and minimum operators so that numbers are processed in a way emphasizing their ordinal properties rather than their cardinal virtues. Fundamentally, what matters is to elicit the set of completely possible values (i.e. having a possibility equal to 1) and those completely impossible (i.e. having a possibility equal to 0); the remaining part of the reference set corresponds to gradual transition between 1 and 0 or 0 and 1. From a pragmatic point of view, this transition can often be assumed linear when the reference set is ordered, as shown in Fig. 1. When the domain is discrete, membership may be assessed on a small and finite set of degrees. In our application, we use a scale of five degrees: $\{0, 0.2, 0.5, 0{,}8, 1\}$. Identification of membership functions can also be done from statistical data, but this topic is beyond the scope of this paper. See Dubois and Prade (1988) for details.

TABLE 3

Possibility distributions representing data observed on an unknown plant

Climatic type	Issuing crop	Acid/basic state of soil	Texture of soil
$\{(c_1, 0.8), (d_1, 0.8), (e_1, 1)\}$	$\{(d_2, 1), (e_2, 0.8)\}$	$\{(c_3, 1)\}$	$\{(c_4, 1), (d_4, 0.5)\}$

Assessing the matching between a datum and a piece of knowledge

We are interested in evaluating whether the data observed on an unknown object correspond to the ith alternative. The sub-problem considered here is to assess to what extent an observation Q_j represented by a possibility distribution π_j matches (i.e. is compatible with) the corresponding item of knowledge, K_{ij} expressed via the membership function μ_{ij}. As already explained, due to the imperfect quality of the available information it is preferable to make this assessment in terms of two numbers: $\Pi(K_{ij}; O_j)$ and $\mathrm{N}(K_{ij}; O_j)$ representing the possibility of matching and the necessity of matching, respectively. These numbers are obtained as indicated by formulae (9) and (10):

$$\Pi(K_{ij}; O_j) = \sup_{u\in\Omega_j} \min[\mu_{ij}(u), \pi_j(u)] \tag{12}$$

$$\mathrm{N}(K_{ij}; O_j) = \inf_{u\in\Omega_j} \max[\mu_{ij}(u), 1-\pi_j(u)] \tag{13}$$

To be more specific, the degree $\Pi(K_{ij};O_j)$ estimates to what extend it is possible that both K_{ij} and O_j refer to the same value, which means that $\Pi(K_{ij}; O_j)$ is a degree of overlapping of the fuzzy set of values represented by μ_{ij} and the fuzzy set of possible values resulting from the imprecise observation. The degree $\mathrm{N}(K_{ij}; O_j)$ estimates to what extent it is necessary (i.e. certain) that the value to which O_j refers is among the ones compatible with those in K_{ij}; in other words, $\mathrm{N}(K_{ij}; O_j)$ is a degree of inclusion of the fuzzy set of possible values associated with the observation in the fuzzy set expressed by μ_{ij}. It can be checked that the possibility degree is always greater than or equal to the degree of necessity.

In the example provided by Table 2 and Table 3, we have:

- for the first weed (i.e. *R. crispus*)
  $\Pi(K_{11}; O_1) = 1$ and $\mathrm{N}(K_{11}; O_1) = 1$
  $\Pi(K_{12}; O_2) = 0.8$ and $\mathrm{N}(K_{12}; O_2) = 0.8$
  $\Pi(K_{13}; O_3) = 0.8$ and $\mathrm{N}(K_{13}; O_3) = 0.8$
  $\Pi(K_{14}; O_4) = 1$ and $\mathrm{N}(K_{14}; O_4) = 0.8$
- for the second weed (i.e. *R. obtusifolius*)
  $\Pi(K_{21}; O_1) = 1$ and $\mathrm{N}(K_{21}; O_1) = 1$
  $\Pi(K_{22}; O_2) = 0.8$ and $\mathrm{N}(K_{22}; O_2) = 0.8$
  $\Pi(K_{23}; O_3) = 1$ and $\mathrm{N}(K_{23}; O_3) = 1$
  $\Pi(K_{24}; O_4) = 0.8$ and $\mathrm{N}(K_{24}; O_4) = 0.8$
- for the third weed (i.e. *R. pulcher*)
  $\Pi(K_{31}; O_1) = 0.2$ and $\mathrm{N}(K_{31}; O_1) = 0$
  $\Pi(K_{32}; O_2) = 0$ and $\mathrm{N}(K_{32}; O_2) = 0$
  $\Pi(K_{33}; O_3) = 1$ and $\mathrm{N}(K_{33}; O_3) = 1$
  $\Pi(K_{34}; O_4) = 1$ and $\mathrm{N}(K_{34}; O_4) = 1$

The following properties shed some more light on the meaning of the possibility and necessity assessments characterizing the matching. For this purpose, it is useful to define the support s(F) and the core c(F) of a fuzzy set F. The support of F is the set of values that somewhat belong to F, that is $s(F) = \{u \mid \mu_F(u) > 0\}$. The core is the set of values that certainly belong to F, i.e. $c(F) = \{u \mid \mu_F(u) = 1\}$.

- If O_j is precise, which means $\pi_j(u^\circ) = 1$ and $\pi_j(u) = 0$ for any $u \neq u^\circ$, then we have $\Pi(K_{ij}; O_j) = N(K_{ij}; O_j) = \mu_{ij}(u^\circ)$.
- $\Pi(K_{ij}; O_j) = 0$ if and only if $s(K_{ij}) \cap s(O_j) = \varnothing$ (i.e. the matching surely fails if K_{ij} and O_j do not overlap at all).
- $\Pi(K_{ij}; O_j) = 1$ and only if $c(K_{ij}) \cap c(O_j) \neq \varnothing$ (i.e. the matching is completely possible as long as there is at least one value fully belonging to both K_{ij} and O_j).
- $N(K_{ij}; O_j) = 1$ if and only if $c(K_{ij}) \supseteq s(O_j)$ (i.e. the matching necessarily succeeds when all somewhat possible values of O_j belong to those surely in K_{ij}).
- If O_j and K_{ij} are non-fuzzy (i.e. represented by crisp sets) then $\Pi(K_{ij}; O_j)$ and $N(K_{ij}; O_j)$ take value in $\{0, 1\}$ rather than $[0, 1]$ (i.e. classical set theory is recovered as a special case).
- Assume two sources O_j^1 and O_j^2, respectively represented by π_j^1 and π_j^2, are available for an observation and are such that $\pi_j^1(u) \leq \pi_j^2(u)$ for any u in Ω_j (i.e. the second is vaguer than the first) then we have $\Pi(K_{ij}; O_j^1) \leq \Pi(K_{ij}; O_j^2)$ but $N(K_{ij}; O_j^1) \geq N(K_{ij}; O_j^2)$ which are natural properties when keeping in mind that possibility and necessity grades are degrees of intersection and inclusion, respectively.

Aggregating elementary assessments

For a given alternative (the *i*th one for instance) we can perform the above measurements of matching for all attributes. Each such measurement is called an elementary assessment for matching. For simplicity, let us denote the degrees $\Pi(K_{ij}; O_j)$ and $N(K_{ij}; O_j)$ by Π_{ij} and N_{ij} respectively. In order to estimate the global matching between data and the *i*th alternative we have to aggregate or combine the pairs (Π_{ij}, N_{ij}). There are basically three attitudes in front of aggregation: conjuctive, disjunctive and trade-off (i.e. compromise). The choice of the right one is application-dependent; it is determined by what a row in the matrix is supposed to express (i.e. a conjunction, a disjunction or an intermediary combination of elementary requirements that an unknown object must satisfy in order to be identified). In our application, the aggregation is consistent with the conjunctive combination but we have also conducted experiences with a trade-off combination. A conjunctive aggregation signifies that the data is said to match the alternative if each observation O_j matches the corresponding K_{ij}. Contrastingly, in a trade-off combination the global matching may still be satisfactory even though one (or few) observation(s) may not fit the description of the *i*th alternative provided the others are in agreement with it along the remaining attributes. The key question to ask in order to choose between the two kinds of combination is: Can good fits on some attributes compensate poor fits on others? If the answer is positive then a trade-off combination is appropriate; if not, the conjunctive one is better.

There are several ways to perform a conjunctive aggregation. In our application, we use the minimum operator to combine separately the degrees of possibility and the degrees of necessity. As given by formula (14) and (15) we obtain two numbers, denoted by UGM^c and LUP^c, standing for upper and lower global matching index respectively (the superscript c means conjuctive):

$$\text{UGM}^c = \min_{j=1,\ldots C} [\Pi_{ij}] \tag{14}$$

$$\text{LGM}^c = \min_{j=1,\ldots,C} [\text{N}_{ij}] \tag{15}$$

In the example constituted by Table 2 and Table 3 it follows that:

UGM^c (*R. crispus*) $=0.8$	LGM^c (*R. crispus*) $=0.8$
UGM^c (*R. obtusifolius*) $=0.8$	LGM^c (*R. obtusifolius*) $=0.8$
UGM^c (*R. pulcher*) $=0$	LGM^c (*R. pulcher*) $=0$

The minimum operator could have been replaced by the product, for instance, but the results could be slightly different because the choice of an operator has semantic implications. See Dubois and Prade (1989) for a discussion of the different operators. Choosing the minimum operator offers the advantage that UGM^c and LGM^c are still possibility and necessity degrees, respectively (on the Cartesian product of the domains).

The trade-off aggregation is performed by averaging over the elementary assessments and leads similar to a pair (UGM^t, LUP^t) of upper and lower global matching indexes (the superscript means trade-off) defined by:

$$\text{UGM}^t = (\textstyle\sum_{j=1,\ldots,C} \Pi_{ij})/C \tag{16}$$

$$\text{LGM}^t = (\textstyle\sum_{j=1,\ldots,C} \text{N}_{ij})/C \tag{17}$$

If the example constituted by Table 2 and Table 3 is considered under the trade-off combination interpretation we obtain:

UGM^t (*R. crispus*) $=0.9$	LGM^t (*R. crispus*) $=0.85$
UGM^t (*R. obtusifolius*) $=0.9$	LGM^t (*R. obtusifolius*) $=0.9$
UGM^t (*R. pulcher*) $=0.55$	LGM^t (*R. pulcher*) $=0.5$

However, though not yet tested in our application, more complex aggregation schemes may be necessary in order to model the different levels of importance that may be attached to the attributes. As shown in Dubois et al. (1988) the conjuctive aggregation formulas can be extended into:

$$\text{UGM}^c = \min_{j=1,\ldots,C} \max [\Pi_{ij}, 1-w_{ij}] \tag{18}$$

$$\text{LGM}^c = \min_{j=1,\ldots,C} \max [\text{N}_{ij}, 1-w_{ij}] \tag{19}$$

where the importance levels are represented by the weights w_{ij}, $j=1, \ldots, C$, expressed on the $[0, 1]$-scale. The greater w_{ij} the greater the importance of the jth attribute for the ith alternative. We also assume that $\max_{j=1,\ldots,C}[w_{ij}]=1$ (normalization), i.e. the most important attributes are graded by 1. Note that if all the w_{ij} are equal to 1 (i.e. equal importance), then the formulae (14) and (15) are recovered. When $w_{ij}=0$ the elementary matching degrees concerning the jth attribute are not taken into account.

Modeling importance in the trade-off aggregation can be obtained by extending the formulas (16) and (17) into:

$$\text{UGM}^{\text{t}} = \textstyle\sum_{j=1,\ldots,C} r_{ij}\Pi_{ij} \tag{20}$$

$$\text{LGM}^{\text{t}} = \textstyle\sum_{j=1,\ldots,C} r_{ij}\text{N}_{ij} \tag{21}$$

where the r_{ij} are weights in $[0, 1]$ such that $\sum_{j=1,\ldots,C} r_{ij}=1$ (normalization), expressing the relative importance of the attributes. Equal importance translates into $r_{ij}=1/C$ for all j; formulas (16) and (17) are then recovered.

Final interpretation and discussion

Having computed upper and lower global matching indexes for each alternative still does not completely solve the problem at hand. Indeed, there is a final interpretation step to go through.

The proposed technique helps in ruling out some alternatives and can be used to rank the remaining ones that are compatible to some extent with the observed data. The ruled-out alternatives are those having an upper global matching index less than some threshold (i.e. 0.5 for instance with the conjunctive combination). In our example, the weed *R. pulcher* is eliminated for this reason. The ranking of alternatives is based on the so-called Pareto ordering that compares pairs of the form (UGM, LGM). By definition, (UGM(1), LGM(1)) is greater than (UGM(2), LGM(2)) if and only if UGM(1) > UGM(2) and LGM(1) ≥ LGM(2) or UGM(1) ≥ UGM(2) and LGM(1) > LGM(2). However, situations where UGM(1) > UGM(2) and LGM(1) < LGM(2) may be encountered. The Pareto ordering is only partial but the lower global matching index may be considered as more important than the other one since it comes out of the combination of degrees of certainty.

In addition, there are cases where the assessments of the matching between the data and two particular weeds are providing similar results though one is more informative than the other. Indeed, assume that the fuzzy sets defining the possible values that an alternative may take are exactly the whole domains (i.e. the set of all conceivable values) of the corresponding attributes. Then, it is sure beforehand that any unknown object will perfectly match the description of this alternative. Thus, it is more informative to ascertain a matching with an alternative that is defined with more specificity (i.e. precision). This notion of 'informativeness' can be taken into account in the system by using a so-called measure of specificity (Yager, 1982).

However, the available information may be such that nothing at all can be said about the matching between the data and a given alternative. This situation corresponds to the case where the upper global matching index is equal to 1 and the lower global matching index is equal to 0, and may be caused by the fact that the observed data is too imprecise with respect to the available knowledge about the alternative under consideration. Thus, to the question about the goodness of fit between the data and a given alternative our technique may answer "I don't know", which means no artifact is introduced to go around such an indeterminate outcome and, consequently, arbitrary commitment is avoided.

In general, for some reasons the user may question the validity of a datum (observed on the unknown object) corresponding to a particular attribute. He can then discard this attribute and recompute the assessment of the global matching by taking into account the remaining attributes only.

CONCLUDING REMARKS

The aim of this paper was to show the relevance of possibility theory to some decision-making problem where imprecision and uncertainty are dominant aspects of the available information. Possibility theory provides a rigorous methodology in which incompleteness in the available information is translated into possibility distributions on suitable semantic scales. Its advantages lie in, first, its ability to represent shades of ignorance in a natural way and, second, the simplicity of the calculations it involves so that its computer implementation does not suffer from inefficiency. The concept of possibility distribution is particularly appropriate to represent subjective information and beliefs, based on a limited number of observations. However, as alluded to earlier, statistical information, especially when imprecise outcomes are observed, can be represented not only by histograms but also by possibility distributions (since what is somewhat probable must be somewhat possible and what is somewhat necessary must be somewhat probable). Such possibility distributions (Dubois and Prade, 1988) are approximations of reality and consequently are not as semantically rich as the ones in the histogram form. Nevertheless, the use of such information in the possibility distribution form may be easier to process and may be used together with other possibilistic items of information.

We have illustrated its use in a simple technique applied to weed identification. So far, results obtained in this application are satisfactory and further tests will be conducted in order to refine the definition of the fuzzy sets constituting the knowledge base and the use of importance assessment attached to the attributes. Directions toward which extensions are possible while staying in the same framework include the following points that, for most of them, have already been explored in the context of data base research (Testemale, 1986):

- the introduction of weight of importance that can vary along the values composing the domain in order to represent, for instance, that the age of the car is not important only within a range;
- the addition of some inferential capabilities to deduce or induce the possible values of an attribute from what is known about the other attributes and the relationship between them;
- the ability to process more sophisticated queries; and
- the ability to cope with multivalued attributes.

At the root of the presented technique is a fuzzy-pattern matching procedure (Dubois et al., 1988) that has also been used in rule-based systems (Martin-Clouaire and Prade, 1986). Indeed, in such systems knowledge is under the form 'IF ⟨situation⟩ THEN ⟨action⟩' where ⟨situation⟩ is a logical combination of propositions describing an abstract situation and ⟨action⟩ tells what to do when the actual data correspond to ⟨situation⟩. The ⟨situation⟩ part and the data matched against it may be vague or uncertain. It is then evident that the evaluation of the extent to which data fit the pattern described in ⟨situation⟩ is a problem similar to the one considered in this paper.

It is our belief that agriculture offers many instances of problems analogous to the one addressed in the application part of this paper and, more gradually, many problems where the complexity is due to imprecision and uncertainty in knowledge. Thus, managing such information in a proper way is an important issue and we urge being aware of new theoretical developments pursuing this goal.

REFERENCES

Badia, J., 1988. MALHERB: identification des mauvaises herbes des cultures. Technical report, Station de Biométrie et Intelligence Artificielle, INRA, Castanet-Tolosan, 31 pp.

Dubois, D. and Prade, H. (with the collaboration of H. Farreny, R. Martin-Clouaire, C. Testemale), 1988. Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum, New York, 263 pp.

Dubois, D. and Prade, H., 1989. Representation and combination of uncertainty with belief functions and possibility measures. Comput. Intell., 4 (in press).

Dubois, D., Prade, H. and Testemale, C., 1988. Weighted fuzzy pattern matching. Int. Fuzzy Sets Syst., 28: 1-19.

Martin-Clouaire, R. and Prade, H., 1986. SPII-1: a simple inference engine capable of accommodating both imprecision and uncertainty. In: G. Mitra (Editor), Computer Assisted Decision Making. North-Holland, Amsterdam, pp. 117-131.

Shafer, G.A., 1976. A Mathematical Theory of Evidence. Princeton University Press, Princeton, NJ, 297 pp.

Testemale, C., 1986. A database system dealing with incomplete or uncertain information and vague queries. In: H. Prade and C. Negoita (Editors), Fuzzy Logic in Knowledge Engineering. TÜV Rheinland, Köln, pp. 29-45.

Yager, R.R., 1982. Measuring tranquility and anxiety in decision-making: an application of fuzzy sets. Int. J. Man-Mach. Stud., 8: 139–146.
Zadeh, L.A., 1965. Fuzzy sets. Inf. Control, 8: 338–353.
Zadeh, L.A., 1978. Fuzzy sets as a basis for a theory of possibility. Int. J. Fuzzy Sets Syst., 1: 3–28.