J Intell Inf Syst (2006) 0: 1–21
DOI 10.1007/s10844-006-0976-x

Fuzzy methods for case-based recommendation and decision support

Didier Dubois · Eyke Hüllermeier · Henri Prade

Received: 22 September 2004 / Revised: 28 February 2005 /
Accepted: 15 March 2005

Abstract The paper proposes two case-based methods for recommending decisions to users on the basis of information stored in a database. In both approaches, fuzzy sets and related (approximate) reasoning techniques are used for modeling user preferences and decision principles in a flexible manner. The first approach, case-based decision making, can principally be seen as a case-based counterpart to classical decision principles well-known from statistical decision theory. The second approach, called case-based elicitation, combines aspects from flexible querying of databases and case-based prediction. Roughly, imagine a user who aims at choosing an optimal alternative among a given set of options. The preferences with respect to these alternatives are formalized in terms of flexible constraints, the expression of which refers to cases stored in a database. As both types of decision support might provide useful tools for recommender systems, we also place the methods in a broader context and discuss the role of fuzzy set theory in some related fields.

Q1 **Keywords**

1. Introduction

So-called recommender systems (Kautz, 1998) (see Resnik & Varian (1997) for a concise introduction) have been a focus of research in recent years. Roughly speaking, such systems assist people without sufficient personal experience in making choices between alternatives. For example, such systems may use a database about user preferences in order to predict additional topics or products a (new) user might like (Breese, Heckerman, & Kadie, 1998). So-called collaborative filtering

D. Dubois · H. Prade
IRIT– Institut de Recherche en Informatique de Toulouse, France

E. Hüllermeier (✉)
Department of Computer Science, University of Magdeburg, Germany
e-mail: eyke@informatik.uni-marburg.de

(Goldberg, Nichols, Oki, & Terry, 1992) allows for personalization in e-commerce by exploiting similarities and dissimilarities among customers' preferences (Lin, Alvarez, & Ruiz, 2002).

Recommender systems employ methods from various fields including, among others, database technology and case-based reasoning. In both fields, fuzzy set theory turned out as an appropriate theoretical foundation for several kinds of extensions and further developments. This is mainly due to some distinctive features of fuzzy sets, notably the ability to handle concepts such as similarity, preference and uncertainty in a unifying way, the possibility of representing *graded* degrees of truth or acceptance, and the capability of *approximate* and *plausible* reasoning. These are valuable features when retrieving objects stored in a database, extrapolating from experience given in the form of stored cases or, more generally, when making recommendations to users.

Our focus in this paper is on *case-based decision support*, which can serve as a basis for a special type of recommender system: Suppose that a user having to make a decision has access to cases stored in a database, and that the user's preferences for alternative decisions depend, in a way to be specified, on these cases. In other words, the user does not (at least not only) have, say, *absolute* preferences. Rather, the preferences are expressed in a *relative* way with respect to the stored cases. Given this type of preferences, the system should suggest one or several among the potential alternatives on the basis of the information provided by these cases.

Section 2 recalls some basic concepts of fuzzy sets. To put the contents of the current paper into context, Section 3 provides a brief overview of applications of fuzzy sets in case-based decision support and related fields. Two different types of case-based decision support are then discussed in detail in Sections 4 and 5, respectively. Section 4 extends work that has been presented in Dubois et al. (1997). The approach proposed in Section 5 is a generalization of case-based prediction as introduced in Dubois, Hüllermeier, & Prade (2002).

2. Background on fuzzy sets

A fuzzy subset of a set $\mathcal{D}$ is identified by a so-called membership function (often denoted μ), which is a generalization of the characteristic function $I_A(\cdot)$ of an ordinary set $A \subseteq \mathcal{D}$ (Zadeh, 1965). For each element $x \in \mathcal{D}$, this function specifies the degree of membership of x in the fuzzy set. Usually, membership degrees are taken from the unit interval $[0, 1]$, i.e., a membership function is a mapping $\mathcal{D} \to [0, 1]$. We shall use the same notation for ordinary sets and fuzzy sets. Moreover, we shall not distinguish between a fuzzy set and its membership function, that is, $A(x)$ denotes the degree of membership of the element x in the fuzzy set A.

Fuzzy sets formalize the idea of *graded membership*, i.e., the idea that an element can belong "more or less" to a set. Consequently, a fuzzy set can have "non-sharp" boundaries. Consider the set of large apartments as an example. Is it reasonable to say that $120\,m^2$ is large and $119\,m^2$ is not large? In fact, any sharp boundary of the set of large apartments will appear rather arbitrary. Modeling the concept "large apartment" as a fuzzy set A, it becomes possible to express, for example, that a size of $120\,m^2$ is completely in accordance with this concept ($A(120) = 1$), $80\,m^2$ is "more or less" large ($A(80) = 1/2$, say), and $40\,m^2$ is definitely not large ($A(40) = 0$).

Table 1 Fuzzy set-based methods in case-based reasoning and database technology

	Input (apart from the case base)	Output
Flexible querying	Constraints on properties of a desired object	An ordered list of stored objects (partly) satisfying the constraints
Querying by example	One or several (fictitious) ideal objects + similarity measures	An ordered list of stored objects similar to the ideal object
Case-based prediction	A partially specified case + similarity measures	Predictions for the remaining attribute values
Case-based decision	A new decision problem (stored cases are triples problem/decision/utility)	A recommended decision
Case-based elicitation	A partially specified case + constraints on missing attributes	Recommendations on the values of the remaining attributes

A fuzzy set can have different semantic interpretations. Particularly, a fuzzy set can express three types of cognitive concepts which are of major importance in artificial intelligence, namely *uncertainty*, *similarity*, and *preference* (Dubois & Prade, 1997b). To exemplify, consider the fuzzy set A of apartments with "ideal size," which might be formalized by the mapping $A : x \mapsto \max\{0, 1 - |x - 120|/40\}$ where x is the size in square meters.[1]

Given (imprecise/uncertain) information in the form of a linguistic statement L, saying that a certain apartment has ideal size, $A(x)$ is considered as the *possibility* that the real size of the apartment is x. Formally, the fuzzy set A induces a possibility distribution $\pi(\cdot)$ by virtue of the identity $\pi(x) \equiv A(x)$ (Zadeh, 1978). Note, however, that $\pi(\cdot)$ and A are semantically different concepts, and that x plays different roles, respectively, on the left and right hand side. To emphasize this, one might write $\pi(x|L) \equiv A(L|x)$: Given the imprecise information L about x_0 (the true size), the possibility that $x_0 = x$ is evaluated by the degree to which the fuzzy concept (modeled by) A is compatible with x.

A membership degree $A(x)$ can also be considered as the *similarity* to the prototype of an apartment with ideal size (or, more generally, as the maximal similarity to any member of a set of prototypes) (Ruspini, 1991; Cross & Sudkamp, 2002). In our example, this is an apartment of $120\,m^2$. Another apartment of, say, $100\,m^2$ is similar to this prototype to the degree $A(100) = 1/2$.

In connection with *preference* modeling, a fuzzy set is considered as a *flexible constraint* (Bellmann & Zadeh, 1970; Dubois, Fargier, & Prade, 1994, 1996a,b). In our example, $A(x)$ specifies the degree of satisfaction caused by an apartment of size x: A size of $x = 120$ is fully satisfactory ($A(x) = 1$), whereas a size of $x = 100$ is more or less acceptable, namely to the degree $1/2$. This example already reveals some advantages of flexible constraints over classical (hard) constraints. Particularly, a graded transition between satisfaction and violation of a constraint is often more natural than an abrupt transition caused by a hard threshold. Moreover, flexible constraints allow for different types of *compensation* if an object is assessed on the basis of multiple criteria. For instance, an apartment which is slightly too small might still appear attractive if its location is really excellent.

[1] This example shows that a fuzzy set is generally context-dependent.

3. Flexible case-based recommendation: an overview

The methods discussed in subsequent sections are closely related to some other applications in case-based reasoning and database technology, where fuzzy sets and fuzzy relations serve a similar purpose. In order to get a general idea of the relationship between these methods, this section provides a very brief overview of such applications.

Table 1 gives a summary, in which the methods are classified according to the type of input provided by the user and the output produced by the system. Roughly speaking, the input either consists of a case or of a kind of constraint that specifies a particular property of a case. Likewise, the output is either a complete case (resp. a set of cases), or a property (attribute) of a case. According to this latter distinction, flexible querying and querying by example can be seen as *case retrieval* methods (the output of which is one or several cases), whereas the other three techniques can be considered as a kind of *case completion* method (a partially specified case is completed or supplemented with additional information in a way or another).

3.1. Flexible querying of databases

Extensions of database systems that make the querying of databases more flexible have become a focus of research in recent years (e.g., Larsen et al. 2001). In this connection, fuzzy set-based techniques are of central importance. The basic idea is to specify a query to a database in terms of a flexible constraint. Example: “Find a reasonably large apartment with at least four rooms which is not too expensive.” The cases that satisfy or, say, match this constraint at least to a certain extent can be retrieved from the database. More specifically, they can be presented to the user in descending order of matching degrees (Bosc & Pivert, 1992, 1995).

Apart from fuzzy sets modeling the preference for single attributes – also called *preference profiles* in this context – important tools include so-called combination and weighting operators. The former are used for combining the individual degrees of satisfaction into one overall evaluation. The latter type of operator serves for weighting the (relative) importance of individual attributes. (For instance, the price of an apartment might be more important than its size).

3.2. Querying by example

It is often not a simple matter for a user to express his preferences in terms of (flexible) constraints on properties (attributes) of a desired object. In such cases, it might be more convenient for him to specify or select some prototypical examples (and maybe counter-examples) of what he is looking for. Then, cases are retrieved from the database which do resemble these examples (and do not resemble the counter-examples). In this connection, the similarity-based semantics of fuzzy sets comes into play (Dubois, Prade, & Sèdes, 2001).

Again, the basic scheme can be generalized in various directions. For example, so-called fuzzy quantifiers (to which we shall return in Section 4.4 below) can be employed for relaxing the constraint that an acceptable object must be similar to *all* of the examples (and to *none* of the counter-examples).

3.3. Case-based prediction

Case-based prediction, as formalized within the framework of fuzzy sets and possibility theory in Dubois et al., (1998, 2002), is in line with instance-based learning (IBL) (Aha, Kibler, & Albert, 1991): A (query) case is partially specified in terms of some of its attribute-values. On the basis of the similarity-guided inference principle underlying instance-based learning, predictions for the remaining attributes are then derived from the (completely specified) cases stored in the database.

More precisely, a possibilistic version of the aforementioned principle is proposed which takes the heuristic and hence uncertain character of similarity-based inference into account: *The more similar the known attributes of a given case to the known attributes of the query, the more possibly the remaining attributes are similar as well*. Predictions of the unknown attribute values derived on the basis of this principle are expressed in terms of possibility distributions. That is, each potential attribute value is assigned a certain degree of possibility.

3.4. Case-based decision making

The idea of case-based decision making – to be detailed in Section 4 – has recently been proposed as an alternative to expected utility theory (Gilboa & Schmeidler, 1995). It combines concepts and principles from both decision theory and instance-based learning. Loosely speaking, a case-based decision maker learns by storing already experienced decision problems, along with a rating of the results. Whenever a new problem needs to be solved, possible actions are assessed on the basis of experience from similar situations in which these actions have already been applied.

3.5. Case-based elicitation

A second type of case-based decision – to be introduced in Section 5 – is closely related to case-based prediction and flexible querying. Like in case-based prediction, the output of the system is an attribute value – the user's decision – completing the case under consideration.[2] Like in flexible querying, the user can impose restrictions on the output in the form of flexible constraints. Now, however, these restrictions are *implicit* in the sense that they are instantiated through the cases stored in the database. In order to distinguish this type of decision making from the aforementioned one, we shall call it *case-based elicitation* (optimal choices are elicited against the background of the given constraints and the stored cases).

4. Case-based decision making

This section starts with a brief review of a model originally introduced by Gilboa and Schmeidler (1995), referred to as *case-based decision theory* (CBDT) by the authors. Then, a fuzzy-set based counterpart to this model will be introduced. For

[2] More generally, the partially specified case can be seen as given information, and the attribute value as a suggested alternative.

this approach, we shall furthermore propose a generalization which applies the idea of *fuzzy quantification* to the case-based evaluation of potential decisions.

4.1. The basic model

In a nutshell, the CBDT setup can be characterized as follows (Gilboa & Schmeidler, 1995): Let $\mathcal{P}$ and $\mathcal{A}$ be (finite) sets of problems and acts, respectively, and denote by $\mathcal{R}$ a set of results or outcomes. Choosing act $a \in \mathcal{A}$ for solving problem $p \in \mathcal{P}$ leads to the outcome $r = r(p,a) \in \mathcal{R}$. A utility function $u : \mathcal{R} \to \mathcal{U}$ (resp. $u : \mathcal{P} \times \mathcal{A} \to \mathcal{U}$) assigns utility values to such outcomes; the utility scale $\mathcal{U}$ is taken as the set of real numbers. Let $\sigma P : \mathcal{P} \times \mathcal{P} \to [0,1]$ be a similarity measure quantifying the similarity of problems. Suppose the decision maker to have a (finite) memory $\mathcal{M} = \{(p_1,a_1,r_1)\ldots(p_n,a_n,r_n)\}$ of cases at his disposal, where $(p_k,a_k) \in \mathcal{P} \times \mathcal{A}$, $r_k = r(p_k,a_k)$, $1 \leq k \leq n$, and suppose furthermore that he has to choose an act for a new problem $p_0 \in \mathcal{P}$. If a certain act $a \in \mathcal{A}$ has not been applied to the problem p_0 so far (i.e., there is no case $(p_0,a,r) \in \mathcal{M}$) the decision maker will generally be uncertain about the result $r(p_0,a)$ and, hence, about the utility $u(p_0,a) = u(r(p_0,a))$. According to the assumption underlying the paradigm of CBDT he then evaluates an act based on its performance when dealing with similar problems in the past, as represented by (parts of) the memory $\mathcal{M}$. More precisely, according to Gilboa and Schmeidler (1995) the decision maker chooses an act that maximizes a linear combination of the benefits experienced so far:

$$V(a) = V_{p_0,\mathcal{M}}(a) \overset{\mathrm{df}}{=} \sum_{(p,a,r)\in\mathcal{M}} \sigma_{\mathcal{P}}(p,p_0) \cdot u(r). \tag{1}$$

The summation over an empty set yields the “default value” 0 which plays the role of an “aspiration level.” Despite the formal resemblance between (1) and the well-known expected utility formula one should not ignore some substantial differences between CBDT and expected utility theory (EUT). This concerns not only the conceptual level but also mathematical aspects. Particularly, it should be noted that the similarity weights in (1) do not necessarily sum to 1. Consequently, (1) must not be interpreted as an estimation of the utility $u(r(p0,a))$. Theoretical details of CBDT including an axiomatic characterization as well as a generalization of decision principle (1) are presented in Gilboa and Schmeidler (1995).

4.2. Fuzzy modeling of case-based decisions

Case-based decision making has been realized in Dubois et al. (1998) as a kind of similarity-based approximate reasoning. This approach is in line with methods of qualitative decision theory. In fact, the assumption that uncertainty and preference can be quantified by means of, respectively, a precise probability measure and a cardinal (numeric) utility function (as it is assumed in classical decision theory as well as in Gilboa and Schmeidler's approach) does often appear unrealistic. As opposed to (1), the approach discussed in this section only assumes an *ordinal* setting for modeling decision problems, i.e., ordinal scales such as

$$\{\texttt{very bad}, \texttt{bad}, \texttt{moderate}, \texttt{good}, \texttt{excellent}\} \quad \text{and} \quad \{\texttt{low}, \texttt{middle}, \texttt{high}\}$$

for assessing, respectively, preference and similarity. This interpretation should be kept in mind, especially since (without loss of generality) both scales will subsequently be taken as subsets of the unit interval (e.g., $\{0, 1/4, 1/2, 3/4, 1\}$ and $\{0, 1/2, 1\}$).[3]

4.2.1. Basic measures for act evaluation

Given a memory $\mathcal{M}$ of cases (p_k, a_k, r_k) and a new problem p_0, the following (estimated) utility value is assigned to an act $a \in \mathcal{A}$:

$$V^{\downarrow}_{p_0,\mathcal{M}}(a) \stackrel{\text{df}}{=} \min_{(p,a,r)\in\mathcal{M}} \sigma_{\mathcal{P}}(p, p_0) \rightsquigarrow u(r), \qquad (2)$$

where $\rightsquigarrow$ is a multiple-valued implication operator (a mapping $[0,1] \times [0,1] \to [0,1]$ which is non-increasing in the first and non-decreasing in the second argument). This valuation supports the idea of finding an act a which has always resulted in good outcomes for problems similar to the current problem p_0. Indeed, (2) can be considered as a generalized truth degree of the claim that "whenever a has been applied to a problem p similar to p_0, the corresponding outcome has had a high utility."

An essential idea behind (2) is to avoid the accumulation and compensation effect of the decision criterion (1) since these effects do not always seem appropriate. With regard to accumulation, for example, undesirable effects can be caused by a case that is stored several times in a database, e.g., because the same act is chosen routinely for a frequently occurring problem.

As a special realization of (2) the valuation

$$V^{\downarrow}_{p_0,\mathcal{M}}(a) \stackrel{\text{df}}{=} \min_{(p,a,r)\in\mathcal{M}} \max\{n(h(\sigma_{\mathcal{P}}(p, p_0))), u(r)\}$$

is proposed, where $h(\cdot)$ is an order-preserving function which ensures the linear scales of similarity and preference to be commensurable and $n(\cdot)$ is the order-reversing function of the similarity scale. By taking $n(\cdot)$ as $x \mapsto 1 - x$ in $[0,1]$ and (without loss of generality) $h(\cdot)$ as the identity, we obtain

$$V^{\downarrow}_{p_0,\mathcal{M}}(a) \stackrel{\text{df}}{=} \min_{(p,a,r)\in\mathcal{M}} \max\{1 - \sigma_{\mathcal{P}}(p, p_0), u(r)\} \qquad (3)$$

(with $\min \emptyset = 1$ by definition). This criterion can obviously be seen as a qualitative counterpart to (1). Besides, the criterion

$$V^{\uparrow}_{p_0,\mathcal{M}}(a) \stackrel{\text{df}}{=} \max_{(p,a,r)\in\mathcal{M}} \min\{\sigma_{\mathcal{P}}(p, p_0), u(r)\} \qquad (4)$$

(with $\max \emptyset = 0$ by definition) has been introduced in Dubois et al. (1997) as an *optimistic* counterpart to (3). It can be seen as a formalization of the idea to find an act a for which there is at least one problem which is similar to p_0 and for which a has led to a good result. Let us mention that expressions (3) and (4) are closely related to decision criteria that have recently been derived in Dubois and Prade (1995) in connection with an axiomatic approach to qualitative decision making under uncertainty.

In order to make the basic principles underlying the above criteria especially obvious, consider the special case where the qualitative utility scale is given by

[3] More precisely, we consider finite subsets of $[0,1]$ that are invariant under the negation operation $x \mapsto 1 - x$.

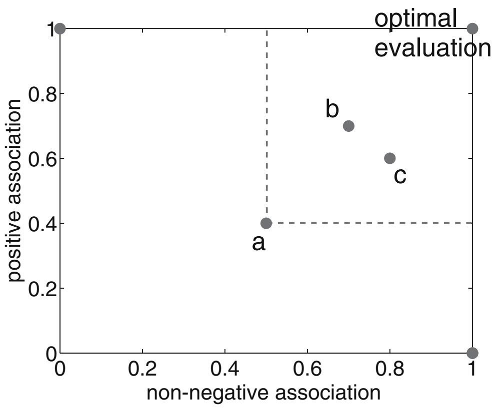

Fig. 1 A "bipolar" evaluation of acts according to a non-negative and a positive association. In the example, act *a* is dominated by *b* and *c*

$\mathcal{U} = \{0, 1\}$. That is, only a crude distinction between "bad" and "good" outcomes is made. Under this assumption, (3) and (4) simplify to

$$V^{\downarrow}_{p_0,\mathcal{M}}(a) = 1 - \max_{(p,a,r)\in\mathcal{M}:u(r)=0} \sigma_{\mathcal{P}}(p, p_0), \qquad (5)$$

$$V^{\uparrow}_{p_0,\mathcal{M}}(a) = \max_{(p,a,r)\in\mathcal{M}:u(r)=1} \sigma_{\mathcal{P}}(p, p_0) \qquad (6)$$

According to (5), the decision maker looks only to cases (p, a, r) with bad outcomes. An act a is discounted whenever it has led to a bad outcome for a problem p which is similar to p_0. Thus, the decision maker is cautious and looks for an act that did not cause a bad experience. According to (6), he only considers the cases with good outcomes. An act a appears promising as soon as it has successfully been applied to a problem p similar to p_0. In other words, the decision maker is more adventurous and looks for an act that it associates with a good experience.

More generally, $V^{\downarrow}_{p_0,\mathcal{M}}(a)$ and $V^{\uparrow}_{p_0,\mathcal{M}}(a)$ reflect a kind of *non-negative* and *positive* association, respectively. This "bipolar" (Dubois, Kaci, & Prade, 2004) evaluation of actions $a \in \mathcal{A}$ gives rise to a partial preference ordering $\preceq$ on $\mathcal{A}$:

$$a \preceq a' \;\Leftrightarrow\; V^{\downarrow}_{p_0,\mathcal{M}}(a) \le V^{\downarrow}_{p_0,\mathcal{M}}(a') \wedge V^{\uparrow}_{p_0,\mathcal{M}}(a) \le V^{\uparrow}_{p_0,\mathcal{M}}(a').$$

Consequently, rational choices correspond to acts that are *non-dominated* in the sense of $\preceq$. In Figure 1, for example, act a is dominated by b, whereas b and c are not directly comparable. Of course, the best evaluation possible is given by $(V^{\downarrow}_{p_0,\mathcal{M}}(a), V^{\uparrow}_{p_0,\mathcal{M}}(a)) = (1, 1)$.

On the one hand, a complete ordering on the set of actions would of course be desirable. On the other hand, incomparability of actions seems to be unavoidable in a *case-based* setting. To illustrate, consider an act a about which nothing is known and that is hence evaluated by $(1, 0)$. Moreover, suppose a second act b, evaluated by $(0, 1)$, to be associated with both a completely negative and a completely positive experience. Is a to be preferred to b or vice versa? In the classical CBDT framework, the positive and negative experience will simply compensate each other.[4] In fact, both acts receive the same score according to (1) (if the positive and negative experience are quantified, respectively, by a utility degree of $-c$ and c, where $c > 0$ is any real number). However, such a compensation is clearly questionable. It is even impossible in the case of ordinal similarity and utility scales, as assumed in our framework.

Even though we are not going to detail the problem in this paper any further, let us note that there are of course alternative ways to refine the partial ordering $\preceq$ to a complete ordering. The simplest option is perhaps to grant priority to one of the two measures, i.e., to concentrate on (3) alone, or to employ a lexicographic ordering. In that case, however, one has to be careful since (3) alone is difficult to interpret if the memory does not contain similar problems to which the action under consideration has already been applied. For example, an action about which nothing is known will receive the highest evaluation, a property that is hardly reasonable. In order to cope with this difficulty, a modification of (3) has been proposed in (Dubois & Prade, 1997a).

4.3. Interpretation of the decision criteria

We already pointed out that the criteria (3) and (4) do not focus on some kind of average performance, since this does hardly make sense within an ordinal setting. Rather, they should be considered from the same point of view as qualitative decision rules such as MAXIMIN (Brafmann & Tennenholtz, 1996; Dubois & Prade, 1995). In fact, we shall now propose an interpretation of (3) as an approximation resp. a similarity-based generalization of a MAXIMIN evaluation.[5]

Observe that we can write (3) as

$$V^{\downarrow}_{p_0,\mathcal{M}}(a) = \min_{0 \le k \le m} \max\{1 - \sigma_k, v_k\}, \tag{7}$$

where the values $0 = \sigma_0 < \sigma_1 < \ldots < \sigma_m = 1$ constitute the (finite) set $\{\sigma_{\mathcal{P}}(p,p') \mid p,p' \in \mathcal{P}\}$ of possible similarity degrees of problems and

$$v_k = \min V_k = \min\{\, u(r) \,|\, (p,a,r) \in \mathcal{M}, \sigma_{\mathcal{P}}(p,p0) = \sigma_k \,\}$$

is the lowest utility obtained in connection with act a for problems which are σ_k-similar to p_0 ($v_k = 1$ by definition if $V_k = \emptyset$).

According to (7), the valuation (3) of an act is completely determined by the lower bounds $v_k (0 \le k \le m)$ which are derived from the memory $\mathcal{M}$ (and discounted according to respective degrees of similarity). This reveals that (3) can indeed be seen as some kind of "experience-based" approximation of the MAXIMIN principle. The case in which all problems are completely similar makes this especially apparent. Then, (3) evaluates an act a simply according to the worst consequence observed so far. More generally, the value v_k can be seen as an estimation of the lower utility bound

$$w_k = \min\{\, u(r(p,a)) \,|\, p_0 \neq p \in \mathcal{P},\ \sigma_{\mathcal{P}}(p,p_0) = \sigma_k \,\},$$

[4] This is to some extent comparable to the standard framework of expected utility theory, where each action is evaluated by its expected utility.

[5] This interpretation can be transferred to (4) in a straightforward way.

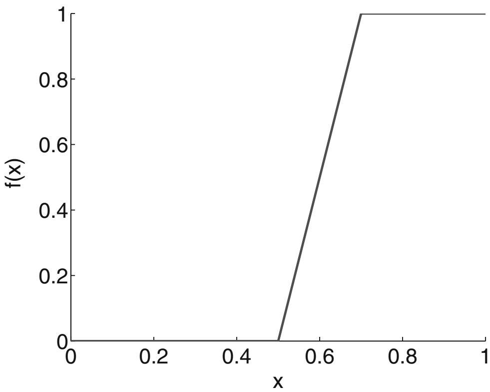

Fig. 2 Example of a function $f(\cdot)$ specifying the linguistic FOR MOST quantifier via $\mu(k) = f(k/n)$

i.e., the smallest degree of utility which can be obtained in connection with act a for (not necessarily encountered) problems from $\mathcal{P}$ which are σ_k-similar to p_0. Then, $V^{\downarrow}_{p_0,\mathcal{M}}(a)$ can be interpreted as an approximation of

$$W^{\downarrow}_{p_0}(a) \; = \min_{0 \le k \le m} \max\{\, 1 - \sigma_k, w_k \,\}.$$

The latter defines a similarity-based generalization of a MAXIMIN-evaluation in the sense that minimum is not taken over the lower utility bounds w_k directly but rather over the bounds that are discounted according to the similarity degrees σ_k: the larger σ_k, the higher the influence (importance) of w_k.

4.4. Fuzzy quantification in act evaluation

In some situations, the extremely pessimistic and optimistic nature of the criteria (3) and (4), respectively, might appear at least as questionable as the accumulation in (1). Here we shall propose a generalization of the decision rule (3) which is a weakening of the demand that an act a has *always* produced good results for similar problems. In fact, one might already be satisfied if a turned out to be a good choice for *most* similar problems, thus allowing for a few exceptions (Dubois & Prade, 1997a). In other words, the idea is to relax the universal "for all" quantifier. A similar weakening of (4) can be obtained by generalizing the existential quantifier.

Consider a finite set A of cardinality $m = |A|$. In connection with propositions of the form "most elements of A do have property B" the fuzzy quantifier MOST can be formalized by means of a fuzzy set (Dubois, Prade, & Testemale, 1988; Yager, 1985),[6] the membership function $\mu : \{0, 1 \ldots m\} \to [0, 1]$ of which satisfies $\mu(m) = 1$ and $\mu(k) \le \mu(k+1)$ for all $1 \le k \le m-1$. The special case "for all" then corresponds to $\mu(k) = 0$ for $0 \le k \le m-1$ and $\mu(m) = 1$. In order to guarantee the comparability of MOST-quantifiers for different values m, the membership function $\mu(\cdot)$ can reasonably be specified in the form $\mu(k) = f(k/m)$, where $f(\cdot)$ is a

[6] Other possibilities of expressing a fuzzy quantifier exist as well, including the use of order-statistics (Prade & Yager, 1994) and an ordered weighted minimum or maximum (Dubois & Prade, 1996).

$[0,1] \to [0,1]$ mapping the input of which is a *percentage* (rather than an absolute number) of elements. Figure 2 shows an example of such a function.

Now, suppose that B is a fuzzy property, i.e., an element x can have the property under consideration to some degree $B(x) \in [0,1]$. Moreover, let $A = \{x_1 \ldots x_m\}$ and suppose (without loss of generality) that $B(x_1) \geq B(x_2) \geq \ldots \geq B(x_m)$. The truth degree of the proposition "most elements of A do have property B" can then be evaluated as

$$E(A) \stackrel{\text{df}}{=} \min_{1 \leq k \leq m} \max\{B(x_k), 1 - \mu(k)\}, \tag{8}$$

where $\overline{\mu}(\cdot)$ is defined by $\overline{\mu}(k) = 1 - \mu(k-1)$ for $1 \leq k \leq m$ (Yager, 1985; Dubois et al., 1988); a membership degree $\overline{\mu}(k)$ can be interpreted as quantifying the importance (and hence $1 - \overline{\mu}(k)$ as the unimportance) that the property X is satisfied for k (out of the m) elements. In the case where A is a fuzzy set as well, (8) can further be generalized to

$$\begin{aligned} E(A) &\stackrel{\text{df}}{=} \min_{0 < \alpha \leq 1} \max\{1 - \alpha, E(A_\alpha)\} \\ &= \min_{0 < \alpha \leq 1} \max\left\{1 - \alpha, \quad \min_{1 \leq k \leq |A_\alpha|} \max\left\{B(x_k^\alpha), f\left(\frac{k-1}{|A_\alpha|}\right)\right\}\right\} \end{aligned} \tag{9}$$

where $A_\alpha = \{x \mid A(x) \geq \alpha\} = \{x_1^\alpha, x_2^\alpha \ldots x_{|A_\alpha|}^\alpha\}$ is the α-cut of A (Bosc, Lietard, & Prade, 1998).

Coming back to our context of case-based decision making, consider again a memory $\mathcal{M}$ of cases and a new problem p_0. We are interested in quantifying the degree to which an act a has induced good outcomes for most of the similar problems. To this end, we can employ (9), with A being the fuzzy set of similar problems and B the fuzzy set of good outcomes. More precisely, for a case $x = (p, a', r)$ we have $B(x) = u(r)$ and $A(x) = \sigma_{\mathcal{P}}(p, p_0)$ if $a = a'$ and $A(x) = 0$ if $a \neq a'$.

4.5. An application: cost-sensitive classification

An interesting problem to which our approach of case-based decision making can be applied is cost-sensitive classification. Predicting the class (label) λ_{x_0} of a query instance x_0 on the basis of a given set of examples of the form $\langle x_\imath, \lambda_{x_\imath} \rangle \in \mathcal{X} \times \mathcal{L}$, $1 \leq \imath \leq n$, where $\mathcal{X}$ is a set of instances and $\mathcal{L}$ a (finite) set of classes, is a standard problem in supervised learning. In this connection, one is usually interested in minimizing the prediction error, i.e., the probability of a misclassification. In some cases, however, it is necessary to distinguish not only between a correct or an incorrect prediction but rather between different types of misclassification. To this end, a *misclassification cost* or *loss* can be associated with each prediction: If λ_{x_0} is the true class, a prediction $\lambda_{x_0}^{est}$ yields a loss of $\ell(\lambda_{x_0}, \lambda_{x_0}^{est})$. A standard example is medical diagnosis, where a false positive (a healthy patient classified as diseased) is usually less bad than a false negative (a sick patient classified as healthy).

Within our framework, an instance $x_\imath$ corresponds to a problem ($\mathcal{X} = \mathcal{P}$) and a predicted class $\lambda_{x_0}^{est}$ to a decision ($\mathcal{L} = \mathcal{A}$). Moreover, it is of course possible to quantify the benefit of a classification in terms of utility degrees $u(\lambda_{x_0}, \lambda_{x_0}^{est})$ instead of loss values. Now, consider an example $\langle x_\imath, \lambda_{x_\imath} \rangle$, i.e., the true class of instance $x_\imath$ is given by $\lambda_{x_\imath}$. For each $\lambda \in \mathcal{L}$, this example (experience) can be transformed into a *case*$\langle x_\imath, \lambda, u(\lambda_{x_\imath}, \lambda)\rangle$. Note that these cases are "hypothetical" in the sense that not all classifications $\lambda \in \mathcal{L}$ have really been tried. Given the utility function and the true class $\lambda_\imath$, we simply know that assigning class λ to $x_\imath$ *would* yield a utility of $u(\lambda_{x_\imath}, \lambda)$.

Considering the classification problem from a case-based decision point of view, each potential class $\lambda \in \mathcal{L}$ can be evaluated by the two complementary measures $V^{\downarrow}_{x_0,\mathcal{M}}(\lambda)$ and $V^{\uparrow}_{x_0,\mathcal{M}}(\lambda)$, respectively, and the final estimation can be made on the basis of these evaluations. This can be seen as an extension of the standard NEAREST NEIGHBOR classifier that will be discussed in more detail in Section 5.1 below. This classifier assigns to a query x_0 the class of the nearest (in terms of a similarity measure $\sigma X(\cdot)$) example $x_\imath$. In fact, it is easy to verify that exactly the same decision is made when assuming a 0/1-utility function (i.e., $u(\lambda, \lambda^{est}) = 1$ if $\lambda = \lambda^{est}$ and $= 0$ otherwise) and choosing the class λ that maximizes $V^{\uparrow}_{x_0,\mathcal{M}}(\lambda)$. This changes, however, when more general utility functions are allowed or when the complementary measure $V^{\downarrow}_{x_0,\mathcal{M}}(\lambda)$ is also taken into account.

The *qualitative* nature of our case-based decision criteria has already been emphasized. Therefore, let us again mention that classifying an instance x_0 on the basis of $V^{\downarrow}_{x_0,\mathcal{M}}(\lambda)$ and $V^{\uparrow}_{x_0,\mathcal{M}}(\lambda)$ is definitely *not* in line with a statistical or probabilistic point of view, i.e.,, a *Bayes decision* that maximizes the (posterior) *expected utility*. A probabilistic approach is often useful and of course a viable alternative. However, in many applications, the underlying statistical assumptions (e.g., independent and identically distributed data) are not satisfied or the number of observations is not large enough to guarantee reasonable probability estimates. Apart from that, a probabilistic (expected utility) approach can again be criticized for its compensatory nature, especially in delicate domains such as e.g., evidence-based medicine.

To illustrate this point, we consider a small toy example where the problem is to predict the disease of a patient. Suppose the following four examples to be given:

patient	1	2	3	4
true class (disease)	A	B	C	A
similarity to new patient	0.8	0.7	0.5	0.3

Moreover, let the utility degrees of decisions be specified as follows:

		prediction		
		A	B	C
true class	A	1	0.7	0
	B	0	1	0.6
	C	0.7	0.7	1

As can be seen, predicting A if the true disease is B (or C if the true disease is A) is really bad (it gives zero utility). Now, the simple nearest neighbor rule would suggest A, which is the disease of the most similar patient. The same decision would be made by every probabilistic variant of the nearest neighbor approach, e.g., by estimating the probability of classes by the (similarity-weighted) relative frequency among the k nearest neighbors. The following table shows the evaluation of the potential predictions by our case-based decision criteria:[7]

	A	B	C
$V^{\downarrow}_{x_0,\mathcal{M}}(y)$	0.3	0.7	0.2
$V^{\uparrow}_{x_0,\mathcal{M}}(y)$	0.8	0.7	0.6

Taking both criteria into consideration, it becomes obvious that decision A is rather risky. Indeed, there is not only a similar A-patient, but also an almost as similar B-patient. In other words, there is not only positive support in favor of decision A, as indicated by a high evaluation $V^{\uparrow}_{x_0,\mathcal{M}}(A) = 0.8$, but also negative evidence, as suggested by a low evaluation $V^{\downarrow}_{x_0,\mathcal{M}}(A) = 0.3$. In this regard, decision B appears to be much safer. A cautious decision maker granting priority to $V^{\downarrow}_{x_0,\mathcal{M}}(y)$ would therefore prefer this classification.

5. Case-based elicitation

In this section, we shall introduce *case-based elicitation* as a generalization of the well-known NEAREST NEIGHBOR (NN) classification principle (Dasarathy, 1991). Therefore, we start with a brief review of this principle, which will then be reinterpreted from a constraint satisfaction point of view. Roughly speaking, case-based elicitation will then be introduced as a fuzzy version of constraint-based NN estimation.

The following setting– common to most instance-based learning methods– will be used throughout this section: $\mathcal{X}$ denotes an instance space, where an instance corresponds to the description x of an object (in attribute-value form). $\mathcal{X}$ is endowed with a similarity measure $\sigma X(\cdot)$. $\mathcal{L}$ is a set of labels, and $\langle x, \lambda_x \rangle \in \mathcal{X} \times \mathcal{L}$ is called a labeled instance (or a case). In classification tasks, $\mathcal{L}$ is a finite (usually small) set comprised of m classes $\{\lambda_1 \ldots \lambda_m\}$. S denotes a sample (case base) that consists of n labeled instances $\langle x_\imath, \lambda_{x_\imath} \rangle$, $1 \leq \imath \leq n$. Finally, a new instance $x_0 \in \mathcal{X}$ (a query) is given, whose label λ_{x_0} is to be estimated.

Let us make a brief comment on the similarity measure $\sigma X(\cdot)$. Since $\mathcal{X}$ is a potentially high-dimensional space, the specification of this measure will usually not be trivial. In this connection, it is common practice to begin with specifying individual similarity measures for each attribute (dimension). A global measure $\sigma X(\cdot)$ is then defined as a linear combination (weighted average) of the individual measures, with the weight of each attribute reflecting its importance.

5.1. Nearest neighbor estimation

The NEAREST NEIGHBOR (NN) principle prescribes to estimate the label of the yet unclassified point x_0 by the label of the nearest sample point, i.e., the one which maximizes the similarity to x_0. The k-NEAREST NEIGHBOR (kNN) approach is a slight generalization which takes the $k \geq 1$ nearest neighbors of a new sample point x_0 into

[7] Recall that these criteria require only a qualitative (ordinal) specification of similarity and utility degrees.

account. That is, an estimation $\lambda_{x_0}^{est}$ of λ_{x_0} is derived from the set $\mathcal{N}_k(x_0)$ of the k nearest neighbors of x_0, most often by means of the *majority vote* decision rule:

$$\lambda_{x_0}^{es} = \arg\max_{\lambda \in \mathcal{L}} \operatorname{card}\{x \in \mathcal{N}_k(x_0) \mid \lambda_x = \lambda\}. \tag{10}$$

Several conceptual modifications and extensions of the basic NN principle have been devised, such as distance weighting of neighbored instances:

$$\lambda_{x_0}^{est} = \arg\max_{\lambda \in \mathcal{L}} \sum_{x \in \mathcal{N}_k(x_0) : \lambda_x = \lambda} \omega_x,$$

where ω_x is the weight of the instance x. The latter is usually an increasing function of $\sigma_{\mathcal{X}}(x, x_0)$. In order to take the uncertainty related to an estimation into account, the simple majority vote (10) is often replaced by the following probability distribution:[8]

$$p_{x_0}(\lambda_\imath) = p^{est}(\lambda_\imath | x_0) = \frac{k_\imath}{k} \tag{11}$$

where $k_\imath = \operatorname{card}\{x \in \mathcal{N}_k(x_0) \mid \lambda_x = \lambda_\imath\}$.

5.2. NN estimation as constraint satisfaction

A label $\lambda_\imath$ receives a high rating according to (11) if it coincides with many among the labels of the k nearest neighbors. In fact, the number $p_{x_0}(\lambda_\imath)$ might also be considered as the degree to which $\lambda_\imath$ satisfies the following constraint:

"Many among the k neighbors have the label $\lambda_\imath$." (12)

According to this point of view, the NN principle is associated with a constraint, $C(\cdot)$, and the evaluation of a label $\lambda_\imath$ is given by the degree to which $\lambda_\imath$ satisfies this constraint:

$$v_{x_0}(\lambda) \stackrel{\mathrm{df}}{=} C(\lambda). \tag{13}$$

The constraint (12) is obviously "fuzzy" in the sense that is can be satisfied to a certain extent. Clearly, one might also think of expressing the NN principle in terms of an ordinary (non-fuzzy) constraint which is either satisfied or not. For example, one might define

$$v_{x_0}(\lambda_\imath) \stackrel{\mathrm{df}}{=} \begin{cases} 1 & \text{if} \quad \operatorname{card}(X_\imath) > k/2 \\ 0 & \text{if} \quad \operatorname{card}(X_\imath) \le k/2 \end{cases}, \tag{14}$$

where $X_\imath$ denotes the subset of neighbors with label $\lambda_\imath$. This kind of constraint has indeed been advocated in connection with NN classification. In fact, (14) is a special kind of *reject option* (Chow, 1970; Hellman, 1970): The maximally supported label is accepted only if the (absolute) support is indeed large enough. Otherwise, a classification is simply refused ($v_{x_0}(\lambda) = 0$ for all $\lambda \in \mathcal{L}$).

As already mentioned above, fuzzy or flexible constraints are much more expressive than ordinary constraints and appear quite reasonable in many appli-

[8] It should be noted that this type of estimation requires further (statistical) assumptions on the classification problem.

cations. As will be seen below, the use of fuzzy constraints is also natural in the context of NN classification. Particularly, the concept of similarity is "fuzzy" by itself: A (normalized) similarity measure can formally be interpreted as a fuzzy relation, namely as a graded (fuzzy) equality relation.[9]

5.3. Flexible constraints

Once having accepted the idea of evaluating a label by the degree to which it satisfies a flexible constraint, one can easily think of alternative specifications of the NN principle. Fuzzy set-based modeling techniques offer a large repertoire for translating such characterizations, given in linguistic form, into logical expressions. This repertoire includes generalized logical connectives such as t-norms and t-conorms, fuzzy relations such as `SIMILAR`, linguistic modifiers such as `VERY`, and fuzzy quantifiers such as `A FEW`.

A t-norm is a binary operator $\otimes : [0,1]^2 \rightarrow [0,1]$ which is commutative, associative, monotone increasing in both arguments and which satisfies the boundary conditions $x \otimes 0 = 0$ and $x \otimes 1 = x$ (Klement, Mesiar, & Pap, 2002). An associated t-conorm $\oplus$ is defined by the mapping $(\alpha, \beta) \mapsto 1 - (1-\alpha) \otimes (1-\beta)$. An important t-norm is the min-operator with associated the t-conorm max. Other important operators include the product $\otimes_P : (\alpha, \beta) \mapsto \alpha\beta$ with related t-conorm $\oplus_P : (\alpha, \beta) \mapsto \alpha + \beta - \alpha\beta$ and the Lukasiewicz t-norm $\otimes_L : (\alpha, \beta) \mapsto \max\{0, \alpha + \beta - 1\}$ with related t-conorm $\oplus_L : (\alpha, \beta) \mapsto \min\{1, \alpha + \beta\}$. In the context of multiple-valued logic, a t-norm and a t-conorm serve, respectively, as a generalized logical conjunction and disjunction. The most common negation operator is the mapping $\neg : x \mapsto 1 - x$, which underlies the De Morgan duality between t-norms and t-conorms.

A linguistic modifier (Zadeh, 1972; Lakoff, 1973) is a function $m : [0,1] \rightarrow [0,1]$ which depicts the effect of linguistic hedges. For example, if $m : x \mapsto x^2$ models the effect of `VERY` and the fuzzy set A represents the concept `BIG`, then $m \circ A$ represents the concept `VERYBIG`. That is, $m(A(x))$ is the degree to which x is very big (the degree to which the proposition that x is big is very true). Other mappings $m(\cdot)$, such as translations, can also be advocated (MacVicar-Whelan, 1978).

Formally, a language of flexible constraints can be specified on the basis of an alphabet consisting of

- a set $\mathcal{V}$ of variable symbols,
- generalized logical connectives $\otimes, \oplus, \rightsquigarrow, \neg$ (conjunction, disjunction, implication, negation),
- a set $\mathfrak{P}$ of predicate symbols, where a predicate of rank k is defined as a fuzzy relation $\mathcal{P} : (\mathcal{X} \times \mathcal{L})^k \rightarrow [0,1]$,
- a set $\mathcal{M}$ of modifier functions,
- a set $\mathcal{Q}$ of fuzzy quantifiers.[10]

The design of a concrete language that can easily be employed by a user is of course highly application-dependent and therefore beyond the scope of this paper. Anyway, by using the above modeling concepts, one can specify sophisticated classification principles that go beyond the simple NN rule. For example, given a new query x_0, a "good" label λ might be specified as one satisfying the following

[9] Recall the similarity-based semantics of fuzzy sets discussed in Section 2.

[10] See Section 4 for the definition of such quantifiers.

Table 2 Database of apartments: `SIZE` is the size in square meters, `DIST` is the distance from the query apartment in kilometers, `PRICE` is the price in dollars

Nr.	SIZE	DIST	PRICE
1	57	5	940
2	128	19	919
3	98	7	880
4	55	17	870
5	123	18	950
6	121	3	965
7	74	5	904
8	101	16	983
9	55	17	925
10	81	2	936

constraint: "Most of the instances close to x_0 have the label λ, and none of the (at least) moderately close instances has a very different label." This constraint could be expressed in logic terms as follows:

$$(\text{FOR MOST } x : \neg\text{CLOSE}(x, x_0) \oplus (\lambda_x = \lambda))$$
$$\otimes(\text{FOR ALL } x : \neg\text{MODERATELY CLOSE}(x) \oplus \neg\text{VERY DIFFERENT}(\lambda_x, \lambda)).$$

More generally, we define the evaluation of a label λ as

$$v_{x_0}(\lambda) \stackrel{\text{df}}{=} C[S, x_0](\lambda), \tag{15}$$

where C is the flexible constraint associated with the specification (Zadeh, 1996). The notation $C[S, x_0](\cdot)$ indicates that the constraint is always evaluated against the background of a given sample S of instances and a new query x_0. (Likewise, a quantifier $Q \in \mathcal{Q}$ always refers to S).

It is worth mentioning that (15) can also mimic the original kNN rule: Consider again the constraint (12) requiring that "λ is supported by many among the k nearest neighbors of x_0." By modeling the fuzzy quantifier `MANY` (out of k) in an appropriate way (namely by the relative frequency), one obtains $v_{x_0}(\lambda) = \imath/k$ iff $\imath$ among the k

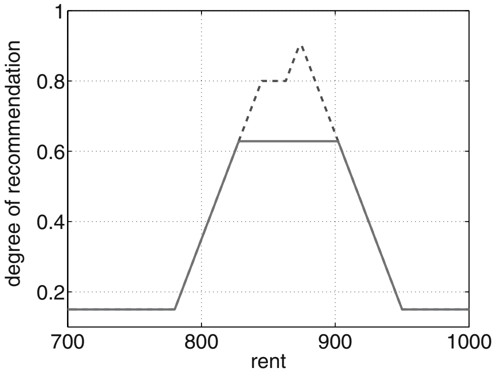

Fig. 3 The function $v_{x_0}(\cdot)$ that specifies the degree of recommendation for the rent λ of an apartment (solid, red line). The dashed (blue) line corresponds to the related `FOR MOST` constraint

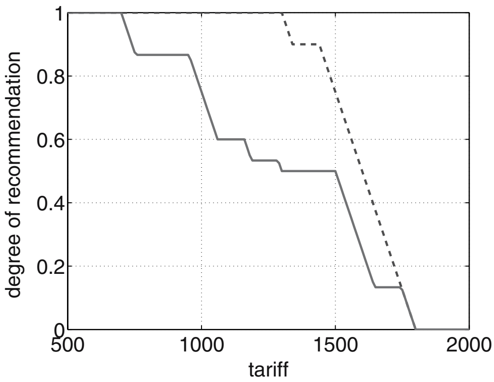

Fig. 4 The function $v_{x_0}(\cdot)$ that specifies the degree of recommendation for the tariff in June (in French francs). The dashed (blue) line corresponds to the related `FOR MOST` constraint

nearest neighbors do have the label λ. In this case, $v_{x_0}(\lambda)$ formally coincides with a degree of probability.

5.4. Case-based elicitation

We look at case-based elicitation as a generalization of instance-based classification: The task shall be to recommend a label, e.g., a remaining attribute value, which is *adequate* in a certain sense, rather than to estimate a *correct* label. As an aside, note that there is no need to distinguish one specific attribute (of an object in attribute-value representation) as the class label. Actually, each of the attributes characterizing an object (or even several of them) can play the role of a label to be estimated or recommended.

Case-based recommendation can simply be put into practice on the basis of the valuation principle (15): The requirements concerning the label are specified in terms of a flexible constraint. Then, the adequacy of each label $\lambda \in \mathcal{L}$ is derived by means of (15). These evaluations can be taken as a point of departure for deciding on the further line of action. Usually, one among the maximally supported labels will be selected, at least if a definite decision must be made. (As mentioned before, the rejection of a decision might be a viable alternative in the case where none of the labels is sufficiently supported).

To illustrate, consider the following toy example: Suppose that we want to let an apartment, and that we are looking for a *reasonable* rent (which plays the role of the label). As we want to avoid any risk of being left with the apartment, we might then quantify the adequacy of a rent λ (in dollars) by the degree to which "λ is slightly lower than the prices for all of the closely located apartments of similar size." Again, it should be emphasized that the aim is not to estimate a "correct" or, say, "fair" price. Rather, we want to find the rent that meets our special demands in an optimal way (and this rent is probably lower than the one which is fair). As a concrete example, suppose the (small) database of apartments as shown in Table 2 to be given. Moreover, let the above constraint be formalized as follows:

$$\texttt{FOR ALL}\ x : (\texttt{CLOSE}(x, x_0) \otimes \texttt{SIMILAR SIZE}(x, x_0)) \rightsquigarrow \texttt{SLIGHTLY LOWER}(\lambda, \lambda_x).$$

Modeling the t-norm by the minimum operator, the implication by the Kleene-Dienes implication $\alpha \rightsquigarrow \beta = \max\{1-\alpha, \beta\}$, and the FOR ALL quantifier also by the minimum, one obtains

$$v_{x_0}(\lambda) = \min_x \max \{1 - \min\{\mathtt{CLOSE}(x, x_0), \mathtt{SIMILAR\,SIZE}(x, x_0)\}, \mathtt{SLIGHTLY\,LOWER}(\lambda, \lambda_x)\} \tag{16}$$

Finally, suppose that the query apartment has a size of $x_0 = 120\,m^2$ and let the fuzzy relations CLOSE, SIMILAR SIZE and SLIGHTLY LOWER be specified, respectively, by $A_{0,20}(\cdot)$, $A_{0,35}(\cdot)$, and $A_{100,100}(\cdot)$, where

$$A_{c,d} : (x, y) \mapsto \max\left\{0, 1 - \frac{|c + x - y|}{d}\right\}. \tag{17}$$

Figure 3 shows the function $v_{x_0}(\cdot)$ that specifies a degree of recommendation for different rents λ. As can be seen, prices around 850 dollars satisfy the constraint quite well, whereas a rent below 750 or above 950 dollars is out of the question.

As the example shows, it may happen that none of the alternatives completely satisfies the constraint, which indicates that the latter is apparently too strong. Of course, one might still choose one among the maximally supported labels, e.g., a rent of 850 dollars.

Another possibility is to weaken the constraint, e.g., by replacing the FOR ALL quantifier by a FOR MOST quantifier. For example, the function $v_{x_0}(\cdot)$ shown by the dashed line in Figure 3 is obtained for the quantifier depicted in Figure 2. As can be seen, there is now a unique optimal recommendation of $\lambda = 870$ for which the constraint is almost completely satisfied.

As a second example, we applied the case-based elicitation method to the ("real-world") database underlying the PRETI system, an experimental platform in information processing maintained at the Insitut de Recherche en Informatique de Toulouse (http://www.irit.fr/PRETI). This database currently maintains information about 650 houses to be let for vacation in the south of France, where each house is described in terms of about 25 attributes (including binary attributes like availability of washing machine, integer attributes like number of rooms and number of beds, and numerical attributes like distance to the beach, low season weekly tariff, ...).

Suppose that we (as a hotel manager) are seeking "a tariff for the month of June which is not much higher than the tariff for most other hotels with similar distances to the closest beach and to the closest swimming pool." We formalized this query in the same way as above, specifying the two Similar Distance relationships by $A_{0,15}(\cdot)$ and $A_{0,5}(\cdot)$, respectively, and the NOT MUCH HIGHER predicate by $(x, y) \mapsto \max\{1 - (x - y)/400, 0\}$ for $x > y$ and 1 for $x \leq y$ (the tariffs are still given in French francs). Moreover, we used again the FOR MOST quantifier shown in Figure 2. The result of the query is depicted by the dashed line in Figure 4. (For comparison, the same figure also shows the result for the more restrictive universal (FOR ALL) quantifier.) As can be seen, tariffs up to 1,300 francs are completely acceptable, while tariffs above 1,800 are out of the question.

6. Summary and conclusion

Extending previous works on fuzzy case-based reasoning, we have proposed two types of case-based decision support, both of which might be useful in the context of information and recommender systems. The first approach is closely related to classical (statistical) decision theory. Here, a case is a triple consisting of a decision problem, the decision that has been made, and the resulting outcome (utility). Having to solve a new problem, a user evaluates potential decisions on the basis of the experience provided by the stored cases.

The second approach, referred to as *case-based elicitation*, is somehow in-between decision making and prediction, and has also much in common with flexible querying of databases. However, as opposed to classical decision making and prediction, the user is neither looking for a decision which is *good* in a normative sense, nor for a prediction which is *correct* in the sense of corresponding to a true outcome. Rather, he specifies his individual preferences and desires in terms of a flexible constraint. By evaluating this constraint on the basis of the cases stored in a database, the system recommends a choice that meets these requirements in an optimal way.

As already mentioned previously, the framework of case-based elicitation as outlined in Section 5 needs to be operationalized in terms of an information system that allows a user to specify elicitation problems in a convenient way. Among other things, this includes a concrete language of flexible constraints, comparable to a fuzzy version of SQL for databases querying (Bosc & Pivert, 1995), as well as (graphical) interfaces for specifying fuzzy concepts such as e.g., similarity measures and fuzzy quantifiers. As future work, it is planned to extend the aforementioned PRETI system in this direction [complementary modules such as flexible database querying and case-based prediction are already available for PRETI (de Calmès et al. 2003)].

The paper clearly reveals the benefits of fuzzy sets and related inference techniques in case-based reasoning. Especially two points are worth mentioning in this connection. The first point concerns the potential of fuzzy sets to model the concepts of similarity, uncertainty and preference within a common framework. Similarity and uncertainty (plausibility) are key concepts in case-based reasoning or, more precisely, in generalizing observed cases. In this paper, we have particularly emphasized the concept of preference, which is not less important in "personalized" case-based reasoning but which has received less attention so far. The second point concerns the possibility of modeling the inference principle underlying CBR in an explicit way and, hence, of contributing to the formal foundations of CBR. In fact, it turned out that case-based inference can adequately be formalized in the context of fuzzy set-based approximate reasoning (Dubois et al., 2002). By drawing on techniques and principles from decision theory, this paper contributes to a formal basis of case-based decision making and recommendation, both of which can be seen as special types of CBR which takes the preferences of the user into account.

References

Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning*, *6*(1), 37–66.

Bellmann, R. E., & Zadeh, L. A. (1970). Decision-making in a fuzzy environment. *Management Science*, *17*, 141–164.

Bosc, P., & Pivert, O. (1992). Some approaches for relational databases flexible querying. *Journal of Intelligent Information Systems*, *1*, 323–354.

Bosc, P., & Pivert, O. (1995). SQLf: A relational database language for fuzzy querying. *IEEE Transactions on Fuzzy Systems*, *3*(1), 1–17.

Bosc, P., Lietard, L., & Prade, H. (1998). An ordinal approach to the processing of fuzzy queries with flexible quantifiers. In A. Hunter & S. Parsons (Eds.), *Applications of uncertainty formalisms*, volume 1455 of *Lecture Notes in Computer Science* (pp. 58–75). Springer-Verlag. Q2

Brafmann, R., & Tennenholtz, M. (1996). On the foundations of qualitative decision theory. In *Proceedings AAAI-96, 13th National Conference on Artificial Intelligence* (pp. 1291–1296). AAAI-Press. Q2

Breese, J. S., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive algorithms for collarborative filtering. In *Proceedings UAI–98*. Madison, WI.

Chow, C. K. (1970). On optimum recognition error and reject tradeoff. *IEEE Transactions on Information Theory*, *IT-16*, 41–46.

Cross, V., & Sudkamp, T. (2002). Similarity and computability in fuzzy set theory: Assessments and applications, *93*. Q3

Dasarathy, B. V. (Ed.) (1991). *Nearest Neighbor (NN) norms: NN pattern classification techniques*. Los Alamitos, California: IEEE Computer Society Press.

de Calmès, M., Dubois, D., Hüllermeier, E., Prade, H., & Sèdes, F. (2003). Flexibility and case-based evaluation in querying: An illustration in an experimental setting. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *11*(1), 43–66.

Dubois, D., & Prade, H. (1995). Possibility theory as a basis for qualitative decision theory. In *Proceedings IJCAI-95, 14th International Joint Conference on Artificial Intelligence* (pp. 1924–1930). Montreal.

Dubois, D., & Prade, H. (1996). Semantics of quotient operators in fuzzy relational databases. *Fuzzy Sets and Systems*, *78*, 89–93.

Dubois, D., & Prade, H. (1997a). A fuzzy set approach to case-based decision. In R. Felix (Ed.), *EFDAN-97, 2nd European Workshop on Fuzzy Decision Analysis and Neural Networks for Management, Planning and Optimization* (pp. 1–9). Dortmund, Germany.

Dubois, D., & Prade, H. (1997b). The three semantics of fuzzy sets. *Fuzzy Sets and Systems*, *90*(2), 141–150.

Dubois, D., Prade, H., & Testemale, C. (1988). Weighted fuzzy pattern matching. *Fuzzy Sets and Systems*, *28*, 313–331.

Dubois, D., Fargier, H., & Prade, H. (1994). Propagation and satisfaction of flexible constraints. In R. R. Yager & L. A. Zadeh (Eds.), *Fuzzy sets, neural networks and soft computing* (pp. 166–187). Q4

Dubois, D., Fargier, H., & Prade, H. (1996a). Possibility theory in constraint satisfaction problems: Handling priority, preference and uncertainty. *Applied Intelligence*, *6*, 287–309.

Dubois, D., Fargier, H., & Prade, H. (1996b). Refinements of the maximin approach to decisionmaking in fuzzy environment. *Fuzzy Sets and Systems*, *81*, 103–122.

Dubois, D., Esteva, F., Garcia, P., Godo, L., de Mantaras, R. L., & Prade, H. (1997). Fuzzy modelling of case-based reasoning and decision. In D. B. Leake & E. Plaza (Eds.), *Case-based reasoning research and development, Proceedings ICCBR-97* (pp. 599–610). Springer-Verlag. Q2

Dubois, D., Esteva, F., Garcia, P., Godo, L., Lopez de Mantaras, R., & Prade, H. (1998). Fuzzy set modelling in case-based reasoning. *International Journal of Intelligent Systems*, *13*, 345–373.

Dubois, D., Prade, H., & Sèdes, F. (2001). Fuzzy logic techniques in multimedia database querying: A preliminary investigation of potentials. *IEEE Transactions on Knowledge and Data Engineering*, *13*(3), 383–392.

Dubois, D., Hüllermeier, E., & Prade, H. (2002). Fuzzy set-based methods in instance-based reasoning. *IEEE Transactions on Fuzzy Systems*, *10*(3), 322–332.

Dubois, D., Kaci, S., & Prade, H. (2004). Bipolarity in reasoning and decision: An introduction. The case of the possibility framework. In *IPMU–04, 10th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Perugia, Italy.

Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave and information tapestry. *Communications of the ACM*, *35*(12), 61–70.

Gilboa, I., & Schmeidler, D. (1995). Case-based decision theory. *Quarterly Journal of Economics*, *110*(4), 605–639.

Hellman, M. E. (1970). The nearest neighbor classification rule with a reject option. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-6*, 179–185.

Kautz, H. (1998). *Recommender systems*. Menlo Park, CA: AAAI Press.

Klement, E. P., Mesiar, R., & Pap, E. (2002). *Triangular norms*. Kluwer Academic Publishers. Q2

Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, *2*, 458–508.

Larsen, H., Kacprzyk, J., Zadrozny, S., Andreasen, T., & Christiansen, H. (Eds.) (2001). *Flexible query answering systems, recent advances*. Physica Verlag. Q2

Lin, W., Alvarez, S. A., & Ruiz, C. (2002). Efficient adaptive-support association rule mining for recommender systems. *Data Mining and Knowledge Discovery*, *6*, 83–105.

MacVicar-Whelan, P. J. (1978). Fuzzy sets, the concept of height, and the hedge very. *IEEE Transactions on Systems, Man, and Cybernetics*, *8*, 507–511.

Prade, H., & Yager, R. R. (1994). Estimations of expectedness and potential surprize in possibility theory. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *2*, 417–428.

Q3 Resnik, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, *40*(3).

Ruspini, E. H. (1991). On the semantics of fuzzy logic. *International Journal of Approximate Reasoning*, *5*, 45–88.

Yager, R. R. (1985). Aggregating evidence using quantified statements. *Information Sciences*, *36*, 179–206.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*, 338–353.

Zadeh, L. A. (1972). A fuzzy-set theoretic interpretation of linguistic hedges. *Journal of Cybernetics*, *2*(3), 4–32.

Q3 Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, *1*(1).

Zadeh, L. A. (1996). Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems*, *2*, 103–111.