On Different Types of Fuzzy Skylines

Allel Hadjali[1], Olivier Pivert[1], and Henri Prade[2]

[1] Irisa – Enssat, University of Rennes 1
Technopole Anticipa 22305 Lannion Cedex France
[2] IRIT, CNRS and University of Toulouse, 31062 Toulouse Cedex 9, France
hadjali@enssat.fr, pivert@enssat.fr, prade@irit.fr

Abstract. This paper deals with database preference queries based on the skyline paradigm, which aim at retrieving the tuples non Pareto-dominated by any other. We propose different ways to fuzzify such queries in order to make them more flexible, to increase their discrimination power, to make them more drastic or more tolerant. In particular, some of these extensions make it possible to reduce the risk of getting many incomparable tuples, even when the number of dimensions is high.

1 Introduction

Numerous approaches have been proposed to make database systems more flexible in supporting user preferences (see [1] for a survey). One of the most well-known approaches is that of skyline queries proposed in [2]. Given a set r of n-dimensional tuples or points, a skyline query returns the set of non-dominated points in r. A tuple t dominates a tuple t' if t is at least as good as t' in all dimensions and strictly better than t' in at least one dimension.

Several research efforts have been made to develop efficient algorithms and to introduce different variants for skyline queries [3,4,5,6,7,8]. In particular, the problem of skyline rigidity is addressed in [9] where a flexible dominance relationship is proposed. It allows the enlarging of the skyline with points that are not much dominated by any other point (even if strictly speaking they are dominated). This issue is also addressed in [10] through an extension of the *winnow* operator initially proposed in [11]. However, many other ways to make skyline queries "fuzzy" can be thought of, and the objective of the present paper is to present and discuss some of them, that we think meaningful.

The paper is structured as follows. Section 2 consists of a reminder about skyline queries. Section 3 describes five different ways in which a skyline may become "fuzzy" when it is refined, relaxed, simplified, extended to uncertain data, or generalized to incompletely stated context-dependent preferences. Section 4 concludes the paper and outlines some perspectives for future research.

2 Reminder About Skyline Queries

The notion of a skyline in a set of tuples is easy to state (since it amounts to exhibit non dominated points in the sense of Pareto ordering). Assume we have:

– a given set of criteria $C = \{c_1, \ldots, c_n\} (n \geq 2)$ associated respectively with a set of attributes A_i, $i = 1, \ldots, n$;
– a complete ordering $\succcurlyeq_i$ given for each criterion i expressing preference between attribute values[3] (the case of non comparable values is left aside).

A tuple $u = (u_1, \cdots, u_n)$ in a database D *dominates* (in the sense of Pareto) another tuple $u' = (u'_1, \cdots, u'_n)$ in D, denoted by $u >_{dom} u'$, iff u is at least as good as u' in all dimensions and strictly better than u' in at least one dimension:

$$u >_{dom} u' \Leftrightarrow \forall i \in \{1, \ldots, n\},\ u_i \succcurlyeq_i u'_i \text{ and } \exists i \in \{1, \ldots, n\} \text{ such that } u_i \succ_i u'_i. \tag{1}$$

A tuple $u = (u_1, \cdots, u_n)$ in a database D belongs to the skyline S, denoted by $u \in S$, if there is no other tuple $u' = (u'_1, \cdots, u'_n)$ in D which dominates it:

$$u \in S \Leftrightarrow \forall u',\ \neg(u' >_{dom} u). \tag{2}$$

Then any tuple u' is either dominated by u, or is non comparable with u. In the following, we denote by $Dm(u)$ those tuples from D that are dominated by u:

$$Dm(u) = \{u' \in D \mid u >_{dom} u'\} \tag{3}$$

and by $Inc(u)$ those tuples which are non comparable with u:

$$Inc(u) = \{u' \in D \mid u' \neq u\ \wedge\ \neg(u >_{dom} u')\ \wedge\ \neg(u' >_{dom} u)\} \tag{4}$$

Table 1. An extension of relation *car*

	make	*category*	*price*	*color*	*mileage*
t_1	Opel	roadster	4500	blue	20,000
t_2	Ford	SUV	4000	red	20,000
t_3	VW	roadster	5000	red	10,000
t_4	Opel	roadster	5000	red	8000
t_5	Fiat	roadster	4500	red	16,000
t_6	Renault	coupe	5500	blue	24,000
t_7	Seat	sedan	4000	green	12,000

Example 1. Let us consider a relation *car* of schema (*make*, *category*, *price*, *color*, *mileage*) whose extension is given in Table 1, and the query:

select * **from** car **preferring**
(*make* = 'VW' **else** *make* = 'Seat' **else** *make* = 'Opel' **else** *make* = 'Ford') **and**
(*category* = 'sedan' **else** *category* = 'roadster' **else** *category* = 'coupe') **and**
(**least** *price*) **and** (**least** *mileage*);

In this query, "$A_i = v_{1,1}$ *else* $A_i = v_{1,2}$" means that value $v_{1,1}$ is strictly preferred to value $v_{1,2}$ for attribute A_i. It is assumed that any domain value which is absent from a preference clause is less preferred than any value explicitly specified in the

[3] $u \succ v$ means u is preferred to v. $u \succcurlyeq v$ means u is at least as good as v, i.e., $u \succcurlyeq v \Leftrightarrow u \succ v \vee u \approx v$, where $\approx$ denotes indifference.

clause (but it is not absolutely rejected). Here, the tuples that are not dominated in the sense of the *preferring* clause are $\{t_3, t_4, t_7\}$. Indeed, t_7 dominates t_1, t_2, and t_5, whereas every tuple dominates t_6 except t_2.

Notice that if we add the preference criterion ($color$ = 'blue' **else** $color$ = 'red' **else** $color$ = 'green') to the query, then the skyline is $\{t_1, t_2, t_3, t_4, t_5, t_7\}$, i.e., allmost all of the tuples are incomparable. ⋄

3 Different Types of Fuzzy Skylines

There may be many different motivations for making skylines fuzzy in a way or another. First, one may want to refine the skyline by introducing some ordering between its points in order to single out the most interesting ones. Second, one may like to make it more flexible by adding points that strictly speaking do not belong to it, but are close to belonging to it. Third, one may try to simplify the skyline either by granulating the scales of the criteria, or by considering that some criteria are less important than others, or even that some criteria compensate each other, which may enable us to cluster points that are somewhat similar. Fourth, the skyline may be "fuzzy" due to the uncertainty or the imprecision present in the data. Lastly, the preference ordering on some criteria may depend on the context, and may be specified only for some particular or typical contexts. We now briefly review each of these ideas.

3.1 Refining the Skyline

The first idea stated above corresponds to refining S by stating that u is in the fuzzy skyline S_{MP} if i) it belongs to S, ii) $\forall u'$ such that $u >_{dom} u'$, $\exists i$ such that u_i is *much preferred* to u'_i, denoted $(u_i, u'_i) \in MP_i$, which can be expressed:

$$u \in S_{MP} \Leftrightarrow u \in S \wedge \forall u' \in Dm(u),\ \exists i \in \{1, \ldots, n\} \text{ s.t. } (u_i, u'_i) \in MP_i \quad (5)$$

(where $\forall i, (u_i, u'_i) \in MP_i \Rightarrow u_i \succ_i u'_i$; we also assume that MP_i agrees with $\succcurlyeq_i$: $u_i \succcurlyeq_i u'_i$ and $(u'_i, u''_i) \in MP_i \Rightarrow (u_i, u''_i) \in MP_i$). (5) can be equivalently written:

$$u \in S_{MP} \Leftrightarrow \forall u' \in D,\ (\neg(u' >_{dom} u) \wedge (u >_{dom} u' \Rightarrow \exists i \in \{1, \ldots, n\} \text{ such that } (u_i, u'_i) \in MP_i)) \quad (6)$$

Note that u is in S_{MP} if it is incomparable with every other tuple or if it is highly preferred on at least one attribute to every tuple it dominates.

When MP_i becomes gradual, we need to use a fuzzy implication such that $1 \to q = q$ and $0 \to q = 1$, which can be expressed as $max(1 - p,\ q)$ (with $p \in \{0, 1\}$). The previous formulas can be translated into fuzzy set terms by:

$$\mu_{S_{MP}}(u) = min_{u' \in D}\ min(1 - \mu_{dom}(u',\ u),\ max(1 - \mu_{dom}(u,\ u'),\ max_i\ \mu_{MP_i}(u_i,\ u'_i))) \quad (7)$$

where $\mu_{dom}(u, u') = 1$ if u dominates u' and is 0 otherwise, and $\mu_{MP_i}(u_i, u'_i)$ is the extent to which u_i is much preferred to u'_i (where $\mu_{MP_i}(u_i,\ u'_i) > 0 \Rightarrow u_i \succ_i u'_i$). Moreover, we assume $u_i \succcurlyeq_i u'_i \Rightarrow \mu_{MP_i}(u_i, u''_i) \geq \mu_{MP_i}(u'_i, u''_i)$. Clearly, we have $S_{MP} = S$ when MP_i reduces to the crisp relation $\succ_i$. S_{MP} may of course be non normalized (no tuple gets degree 1), or even empty.

Example 2. Let us consider again the data and query from Example 1 (without the criterion on *color*). Let us introduce the relations $\mu_{\ll_{price}}(x, y) = 1$ if $y - x \geq 1000$, 0 otherwise, and $\mu_{\ll_{mileage}}(x, y) = 1$ if $y - x \geq 5000$, 0 otherwise. Let us assume that the notion "much preferred" is defined as $\forall(x, y)$, $\mu_{MP}(x, y) = 0$ for attributes *make* and *category*. The result is now the set $\{t_3, t_4\}$. Tuple t_7 does not belong to the skyline anymore since at least one of the tuples that it dominates (here t_5) is not *highly* dominated by it.

Notice that if the *mileage* value in tuple t_7 were changed into 20,000, tuple t_7 would then belong to S_{MP}. This situation occurs when the "weakening" of a tuple makes it incomparable with all the others.$\diamond$

A still more refined fuzzy skyline S^* selects those tuples u from S_{MP}, if any, that are such that $\forall u' \in Inc(u)$, $\nexists j$ such that u'_j is much preferred to u_j:

$$\begin{aligned} u \in S^* \Leftrightarrow u \in S_{MP} \wedge \\ \forall u' \in Inc(u),\ \nexists j \in \{1, \ldots, n\} \text{ such that } (u'_j, u_j) \in MP_i. \end{aligned} \tag{8}$$

Thus, the graded counterpart of Formula (8) is:

$$\mu_{S^*}(u) = min(\mu_{S_{MP}}(u),\ min_{u' \in Inc(u)}\ (1 - max_j\ \mu_{MP_j}(u'_j, u_j)) \tag{9}$$

S^* gathers the most interesting points, since they are much better on at least one attribute than the tuples they dominate, and not so bad on the other attributes (w.r.t. other non comparable points).

Example 3. Using the same data and strengthened preferences as in Example 2, we get $S^* = \emptyset$ since both t_3 and t_4 are much worse than t_2 (and t_7) on *price*. $\diamond$

S^* and S_{MP} do not seem to have been previously considered in the literature.

3.2 Making the Skyline More Flexible

Rather than refining the skyline, a second type of fuzzy skyline (denoted by S_{REL} hereafter) corresponds to the idea of relaxing it, i.e., u still belongs to the skyline to some extent (but to a less extent), if u is only weakly dominated by any other u'. Then, $u \in S_{REL}$ iff it is false that there exists a tuple u' much preferred to u w.r.t. all attributes (this expression was proposed in [9]). Formally, one has:

$$u \in S_{REL} \Leftrightarrow \nexists u' \in D,\ \forall i \in \{1, \ldots, n\},\ (u'_i, u_i) \in MP_i \tag{10}$$

or in fuzzy set terms:

$$\begin{aligned} \mu_{S_{REL}}(u) &= 1 - max_{u' \in D}\ min_i\ \mu_{MP_i}(u'_i, u_i) \\ &= min_{u' \in D}\ max_i\ 1 - \mu_{MP_i}(u'_i, u_i). \end{aligned} \tag{11}$$

Example 4. Let us use the same data and query as in Example 1, and assume that "much preferred" is defined as "preferred" on attributes *make* and *category*, and as in Example 2 for attributes *price* and *mileage*. We get $S_{REL} = \{t_1, t_2, t_3, t_4, t_5, t_7\}$ instead of $S = \{t_3, t_4, t_7\}$ since neither t_1, t_2, nor t_5 is highly dominated by any other tuple on all the attributes.$\diamond$

One has $S \subseteq S_{REL}$, i.e., $\forall u \in D,\ \mu_S(u) \le \mu_{S_{REL}}(u)$, since Formula (2) writes

$$\begin{aligned}\mu_S(u) &= min_{u' \in D}\, 1 - \mu_{dom}(u',\, u)\\ &= min_{u' \in D}\, 1 - min(min_i\, \mu_{\succcurlyeq_i}(u'_i,\, u_i),\, max_i\, \mu_{\succ_i}(u'_i,\, u_i))\\ &= min_{u' \in D}\, max(max_i\, \mu_{\prec_i}(u'_i,\, u_i),\, min_i\, \mu_{\preccurlyeq_i}(u'_i,\, u_i))\end{aligned}$$

in fuzzy set terms and $max(max_i \mu_{\prec_i}(u'_i, u_i), min_i \mu_{\preccurlyeq_i}(u'_i, u_i)) \le max_i \mu_{\preccurlyeq_i}(u'_i, u_i)$ $\le max_i\, 1 - \mu_{MP_i}(u'_i,\, u_i)$ since $1 - \mu_{\preccurlyeq_i}(u'_i,\, u_i) = \mu_{\succ_i}(u'_i,\, u_i) \le \mu_{MP_i}(u'_i,\, u_i)$. So, one finally has:

$$S^* \subseteq S_{MP} \subseteq S \subseteq S_{REL}. \tag{12}$$

Note that Formula (10) may seem very permissive, but in case we would think of replacing $\forall i$ by $\exists i$, we would lose $S \subseteq S_{REL}$, which is in contradiction with the idea of enlarging S.

Another way of relaxing S is to consider that u still belongs to a fuzzily extended skyline S_{FE} if u is close to u' with $u' \in S$. This leads to the following definition

$$u \in S_{FE} \Leftrightarrow \exists u' \in S \text{ such that } \forall i,\ (u_i,\ u'_i) \in E_i \tag{13}$$

where E_i is a reflexive, symmetrical approximate indifference (or equality) relation defined on the domain of A_i, such that $(u_i,\, u''_i) \in E_i$ and $u_i \preccurlyeq_i u'_i \preccurlyeq_i u''_i \Rightarrow (u_i,\, u'_i) \in E_i$. Moreover, it is natural to assume that $(u_i,\, u'_i) \in E_i \Rightarrow (u_i,\, u'_i) \notin MP_i$ and $(u'_i,\, u_i) \notin MP_i$. S_{FE} can be expressed in fuzzy set terms (then assuming $\mu_{E_i}(u_i,\, u'_i) > 0 \Rightarrow \mu_{MP_i}(u_i,\, u'_i) = 0$, i.e. in other words, $support(E_i) = \{(u_i,\, u'_i) | \mu_{E_i}(u_i,\, u'_i) > 0\} \subseteq \{(u_i,\, u'_i) | 1 - \mu_{MP_i}(u_i,\, u'_i) = 1\} = core(\overline{MP_i})$). We also assume $u_i \preccurlyeq_i u'_i \preccurlyeq_i u''_i \Rightarrow \mu_{E_i}(u_i,\, u'_i) \ge \mu_{E_i}(u_i,\, u''_i)$. We have

$$\mu_{S_{FE}}(u) = max_{u' \in D}\, min(\mu_S(u'),\, min_i\, \mu_{E_i}(u_i,\, u'_i)) \tag{14}$$

Then we have the following inclusions.

$$S \subseteq S_{FE} \subseteq S_{REL} \tag{15}$$

Proof. $S \subseteq S_{FE}$. Clearly, $\mu_S \le \mu_{S_{FE}}$, since the approximate equality relations E_i are reflexive (i.e.,$\forall i, \forall u_i, \mu_{E_i}(u_i,\, u_i) = 1$).

$S_{FE} \subseteq S_{REL}$. Let us show it in the non fuzzy case first, by establishing that the assumption $u \notin S_{REL}$ and $u \in S_{FE}$ leads to a contradiction. Since $u \notin S_{REL}$, $\exists \mathring{u} \in D$ s.t. $\forall i, (\mathring{u}_i, u_i) \in MP_i$. Besides, since $u \in S_{FE}$, $\exists u^* \in S, \forall i, (u^*_i, u_i) \in E_i$. Observe that u^* does not dominate $\mathring{u}$ (since $\forall i, u^*_i \succcurlyeq_i \mathring{u}_i$ entails $(u^*_i,\, u_i) \in MP_i$, due to $\forall i, (\mathring{u}_i,\, u_i) \in MP_i$, which contradicts $\forall i, (u^*_i,\, u_i) \in E_i$). $\mathring{u}$ does not dominate u^* either (since $u^* \in S$). Then, $\mathring{u}$ and u^* are incomparable, and $\exists j, \exists k, u^*_j \succcurlyeq_j \mathring{u}_j$ and $u^*_k \prec_k \mathring{u}_k$. But, we know that in particular $(\mathring{u}_j,\, u_j) \in MP_j$ and $(u^*_j,\, u_j) \in E_j$. Let us show by *reductio ad absurdum* that it entails $\mathring{u}_j \succ_j u^*_j$. Indeed, assume $\mathring{u}_j \preccurlyeq_j u^*_j$; $(\mathring{u}_j,\, u_j) \in MP_j$ entails $\mathring{u}_j \succcurlyeq_j u_j$ and $(\mathring{u}_j,\, u_j) \notin E_j$. Hence a contradiction since $u^*_j \succcurlyeq_j \mathring{u}_j \succcurlyeq_j u_j$ and $(u^*_j,\, u_j) \in E_j$ (w.r.t. a property assumed for E_j). But in turn, $\mathring{u}_j \succ_j u^*_j$ contradicts that $\exists j, u^*_j \succcurlyeq_j \mathring{u}_j$, the hypothesis we start with. Thus, assuming $u \notin S_{REL}$ leads to a contradiction. The fuzzy case can be handled similarly by working with the cores and supports of fuzzy relations. ■

3.3 Simplifying the Skyline

On the contrary, it may be desirable to simplify the skyline, for example because it contains too many points. There are many ways to do it. The definitions of S^* and S_{MP} serve this purpose, but they may be empty as already said. We now briefly mention three other meaningful ways to simplify the skyline.

Simplification Through Criteria Weighting First, one may consider that the set of criteria is partitioned into subsets of decreasing importance, denoted by $W_1, \ldots, W_k$ (where W_1 gathers the criteria of maximal importance). Then we may judge that a tuple cannot belong to the skyline only because it strictly dominates all the other tuples on a non fully important criterion. Indeed it may look strange that a tuple belongs to the skyline while it is dominated on all the important criteria, even if its value on a secondary criterion makes the tuple finally incomparable. In this view, less important criteria may be only used to get rid of tuples that are dominated on immediately less important criteria, in case of ties on more important criteria. Let us introduce the following definitions underlying the concept of a hierarchical skyline.

$$\begin{aligned} u >_{dom_{W_i}} u' \Leftrightarrow \forall j \text{ such that } c_j \in W_i, \\ (u_j \succcurlyeq_j u'_j \ \wedge \ \exists p \text{ such that } (c_p \in W_i \ \wedge \ u_p \succ_p u'_p)). \end{aligned} \tag{16}$$

$$\forall i \in \{1, \ldots, k\}, u \in S_{W_i} \Leftrightarrow u \in S_{W_{i-1}} \ \wedge \ \forall u' \in D, \neg(u >_{dom_{W_i}} u') \tag{17}$$

assuming $\forall u \in D,\ u \in S_{W_0}$. The set S_{W_j} gathers the tuples that are not dominated by any other in the sense of the criteria in $W_1 \cup ... \cup W_j$. By construction, one has:

$$S_{W_1} \supseteq S_{W_2} \supseteq \ldots \supseteq S_{W_k}.$$

In the same spirit, in [12], an operator called *cascade* iteratively eliminates the dominated tuples in each level of a preference hierarchy. Prioritized composition of preferences obeying the same concept can also be modeled by the operator *winnow* proposed by Chomicki [11].

Example 5. Let us consider the data from Table 1, and the query:

select * **from** car **preferring**
((category = 'sedan' **else** category = 'roadster' **else** category = 'coupe') **and**
(color = 'blue' **else** color = 'red' **else** color = 'green')) (W_1)
cascade (**least** price) (W_2);

We get the nested results: $S_{W_1} = \{t_1, t_7\}$ and $S_{W_2} = \{t_7\}$. ⋄

An alternative solution — which does not make use of priorities but is rather based on counting — is proposed in [3,4] where the authors introduce a concept called *k-dominant skyline*, which relaxes the idea of dominance to k-dominance. A point p is said to k-dominate another point q if there are k ($\leq d$) dimensions in which p is better than or equal to q and is better in at least one of these k dimensions. A point that is not k-dominated by any other points is in the k-dominant skyline. Still another method for defining an order for two incomparable tuples is proposed in [5], based on the number of other tuples that each of the two tuples dominates (notion of k-representative dominance).

Simplification Through The Use of Coarser Scales A second, completely different idea for simplifying a skyline is to use coarser scales for the evaluation of the attributes (e.g., moving from precise values to rounded values). This may lead to more comparable (or even identical) tuples. Notice that the skyline obtained after simplification does not necessarily contain less points than the initial one (cf. the example hereafter). However, the tuples that *become* member of the skyline after modifying the scale are in fact *equivalent* preferencewise.

Example 6. Let us consider a relation r of schema (A, B) containing the tuples $t_1 = \langle 15.1, 7\rangle$, $t_2 = \langle 15.2, 6\rangle$, and $t_3 = \langle 15.3, 5\rangle$, and the skyline query looking for those tuples which have the smallest value for both attributes A and B. Initially, the skyline consists of all three tuples t_1, t_2, t_3 since none of them is dominated by another. Using rounded values for evaluating A and B one gets $\{t_3\}$ as the new skyline. Let us now consider that relation r contains the tuples $t'_1 = \langle 15.1, 5.1\rangle$, $t'_2 = \langle 15.2, 5.2\rangle$, and $t'_3 = \langle 15.3, 5.4\rangle$. This time, the initial skyline is made of the sole tuple t'_1 whereas the skyline obtained by simplifying the scales is $\{t'_1, t'_2, t'_3\}$. $\diamond$

Simplification Through the Use of k-discrimin Still another way to increase the number of comparable tuples is to use a *2-discrimin* (or more generally an order *k-discrimin*) ordering (see [13] from which most of the following presentation is drawn). A definition of classical *discrimin* relies on the set of criteria not respected in the same way by both tuples u and v, denoted by $D_1(u, v)$ [14]:

$$D_1(u, v) = \{c_i \in C \mid v_i = u_i\} \tag{18}$$

$$u >_{disc} v \Leftrightarrow min_{c_i \notin D_1(u, v)}\, u_i > min_{c_i \notin D(u, v)}\, v_i \tag{19}$$

Discrimin-optimal solutions are also Pareto-optimal but not conversely, in general (see [14]).

Classical discrimin is based on the elimination of identical singletons at the same places in the comparison process of the two sequences. Thus with classical discrimin, comparing $u = (0.2, 0.5, 0.3, 0.4, 0.8)$ and $v = (0.2, 0.3, 0.5, 0.6, 0.8)$ amounts to comparing vectors u' and v' where $u' = (0.5, 0.3, 0.4)$ and $v' = (0.3, 0.5, 0.6)$ since $u_1 = v_1 = 0.2$ and $u_5 = v_5 = 0.8$. Thus, $u =_{min} v$ and we still have $u =_{discrimin} v$. More generally, we can work with 2-element subsets which are identical and pertain to the same pair of criteria. Namely in the above example, we may consider that $(0.5, 0.3)$ and $(0.3, 0.5)$ are "equilibrating" each other. Note that it supposes that the two corresponding criteria have the same importance. Then we delete them, and we are led to compare $u'' = (0.4)$ and

$v'' = (0.6)$. Let us take another example: $u_2 = (0.5,\ 0.4,\ 0.3,\ 0.7,\ 0.9)$ and $v_2 = (0.3,\ 0.9,\ 0.5,\ 0.4, 1)$. Then, we would again delete (0.5, 0.3) with (0.3, 0.5) yielding $u_2' = (0.4,\ 0.7,\ 0.9)$ and $v_2' = (0.9,\ 0.4,\ 1)$. Note that in this example we do not simplify 0.4, 0.9 with 0.9, 0.4 since they do not pertain to the same pair of criteria. Note also that simplifications can take place only one time. Thus, if the vectors are of the form $u = (x,\ y,\ x,\ s)$ and $v = (y,\ x,\ y,\ t)$ (with $min(x,\ y) \leq min(s,\ t)$ in order to have the two vectors min-equivalent), we may either delete components of ranks 1 and 2, or of ranks 2 and 3, leading in both cases to compare $(x,\ s)$ and $(y,\ t)$, and to consider the first vector as smaller in the sense of the order 2-discrimin, as soon as $x < min(y,\ s,\ t)$.

We can now introduce the definition of the (order) 2-discrimin [13]. Let us build a set $D_2(u,\ v)$ as $\{(c_i,\ c_j) \in C \times C$, such that $u_i = v_j$ and $u_j = v_i$ and if there are several such pairs, they have no common components$\}$. Then the 2-discrimin is just the minimum-based ordering once components corresponding to pairs in $D_2(u, v)$ and singletons in $D_1(u, v)$ are deleted. Note that $D_2(u, v)$ is not always unique as shown by the above example. However this does not affect the result of the comparison of the vectors after the deletion of the components as it can be checked from the above formal example, since the minimum aggregation is not sensitive to the place of the terms. Notice that the k-discrimin requires stronger assumptions than Pareto-ordering since it assumes that the values related to different attributes are comparable (which is the case for instance when these values are obtained through scoring functions).

This idea of using *k-discrimin* for simplifying a skyline can be illustrated by the following example, where we compare hotels on the basis of their price, distance to the station, and distance to a conference location (which should all be minimized). Then (80, 1, 3) et (70, 3, 1) are not Pareto comparable, while we may consider that the two distance criteria play similar roles and that there is equivalence between the sub-tuples (1, 3) and (3, 1) leading to compare the tuples on the remaining components.

3.4 Dealing with Uncertain Data

The fourth type of "fuzzy" skyline is quite clear. When attributes values are imprecisely or more generally fuzzily known, we are led to define the tuples that certainly belong to the skyline, and those that only possibly belong to it, using necessity and possibility measures. This idea was suggested in [8].

3.5 Dealing with Incomplete Contextual Preferences

In [15], we concentrate on the last category of "fuzzy" skyline that is induced by an incompletely known context-dependency of the involved preferences. In order to illustrate this, let us use an example taken from [16], which consists of a relation with three attributes *Price*, *Distance* and *Amenity* about a set of hotels (see Table 2). A skyline query may search for those hotels for which there is *no cheaper* and, at the same time, *closer* to the beach alternative. One can easily check that the skyline contains hotels h_4 and h_5. In other terms, hotels h_4 and h_5 represent non-dominated hotels w.r.t. *Price* and *Distance* dimensions.

Table 2. Relation describing hotels

Hotel	Price	Distance	Amenity
h_1	200	10	Pool(P)
h_2	300	10	Spa(S)
h_3	400	15	Internet'I)
h_4	200	5	Gym(G)
h_5	100	20	Internet(I)

Table 3. Contextual Skylines

Context	Preferences	Skyline
C_1: Business, June	$I \succ G, I \succ \{P, S\}$, $G \succ \{P, S\}$	h_3, h_4, h_5
C_2: Vacation	$S \succ \{P, I, G\}$	h_2, h_4, h_5
C_3: Summer	$P \succ \{I, G\}$ $S \succ \{I, G\}$	h_1, h_2, h_4, h_5
C_q: Business, Summer	–	?

Let us now assume that the preferences on attribute *Amenity* depend on the *context*. For instance, let us consider the three contexts C_1, C_2 and C_3 shown in Table 3 (where a given context can be composed at most by two context parameters (*Purpose*, *Period*)). For example, when the user is on a business trip in June (context C_1), hotels h_3, h_4 and h_5 are the results of the skyline query for C_1. See Table 3 for contexts C_2 and C_3 and their corresponding skylines.

Let us now examine situation C_q (fourth row in Table 3), where the user plans a business trip in the summer but states no preferences. Considering amenities *Internet* (I) and *Pool* (P), one can observe that: (i) I may be preferred to P as in C_1; (ii) P may be preferred to I as in C_3, or (iii) I and P may be equally favorable as in C_2. Moreover, the uncertainty propagates to the dominance relationships, i.e., every hotel may dominate another with a certainty degree that depends on the context. In [15], it is shown how a set of plausible preferences suitable for the context at hand may be derived, on the basis of the information known for other contexts (using a CBR-like approach). Uncertain dominance relationships are modeled in the setting of possibility theory. In this framework, the user is provided with the tuples that are not dominated with a high certainty, leading to a notion of possibilistic contextual skyline. It is also suggested how possibilistic logic can be used to handle contexts with conflicting preferences, as well as dependencies between contexts.

4 Conclusion

The paper has provided a structured discussion of different types of "fuzzy" skylines. Five lines of extension have been considered. First, one has refined he skyline by introducing some ordering between its points in order to single out the most interesting ones. Second, one has made it more flexible by adding points that strictly speaking do not belong to it, but are close to belonging to it. Third, one has aimed at simplifying the skyline either by granulating the scales of the criteria, or by considering that some criteria are less important than others, or even that some criteria compensate each other. Fourth, the case where the skyline is fuzzy due to the uncertainty in the data has been dealt with. Lastly, skyline queries has been generalized to incompletely stated context-dependent preferences.

Among perspectives for future research, let us mention: (i) the integration of these constructs into a database language based on SQL, (ii) the study of query optimization aspects. In particular, it would be worth investigating whether some techniques proposed in the context of Skyline queries on classical data (for instance those based on presorting, see, e.g., [17]) could be adapted to (some of) the fuzzy skyline queries discussed here.

References

1. Hadjali, A., Kaci, S., Prade, H.: Database preferences queries: a possibilistic logic approach with symbolic priorities. In: Proc. FoIKS. (2008) 291–310
2. Borzsony, S., Kossmann, D., Stocker, K.: The skyline operator. In: Proc. ICDE. (2001) 421–430
3. Chan, C.Y., Jagadish, H.V., Tan, K.L., Tung, A.K.H., Zhang, Z.: Finding k-dominant skylines in high dimensional space. In: ACM SIGMOD 2006. 503–514
4. Chan, C., Jagadish, H., Tan, K., Tung, A., Zhang, Z.: On high dimensional skylines. In: Proc. of EDBT 2006, LNCS 3896. (2006) 478–495
5. Lin, X., Yuan, Y., Zhang, Q., Zhang, Y.: Selecting stars: The k most representative skyline operator. In: Proc. ICDE. (2007) 86–95
6. Khalefa, M.E., Mokbel, M.F., Levandoski, J.J.: Skyline query processing for incomplete data. Proc. ICDE (2008) 556–565
7. Pei, J., Jiang, B., Lin, X., Yuan, Y.: Probabilistic skylines on uncertain data. In: VLDB 2007. 15–26
8. Hüllermeier, E., Vladimirskiy, I., Prados Suárez, B., Stauch, E.: Supporting case-based retrieval by similarity skylines: Basic concepts and extensions. In: Proc. of ECCBR'08. (2008) 240–254
9. Goncalves, M., Tineo, L.: Fuzzy dominance skyline queries. In: Proc. of DEXA'07. (2007) 469–478
10. Zadrozny, S., Kacprzyk, J.: Bipolar queries and queries with preferences. In: Proc. of FlexDBIST'06. (2006) 415–419
11. Chomicki, J.: Preference formulas in relational queries. ACM Transactions on Database Systems **28**(4) (2003) 427–466
12. Kießling, W., Köstler, G.: Preference SQL — design, implementation, experiences. In: Proc. of the 2002 VLDB Conference. (2002) 990–1001
13. Prade, H.: Refinement of Minimum-Based Ordering in between Discrimin and Leximin. In: Proc. Linz Seminar on Fuzzy Set Theory. (2001) 39–43
14. Dubois, D., Fargier, H., Prade, H.: Fuzzy constraints in job-shop scheduling. Journal of Intelligent Manufacturing **6** (1995) 215–234
15. Hadjali, A., Pivert, O., Prade, H.: Possibilistic contextual skylines with incomplete preferences. In: Proc. of the 2nd IEEE International Conference on Soft Computing and Pattern Recognition (SoCPaR'10), Cergy-Pontoise, France (2010)
16. Sacharidis, D., Arvanitis, A., Sellis, T.: Probabilistic contextual skylines. In: ICDE. (2010) 273–284
17. Bartolini, I., Ciaccia, P., Patella, M.: Efficient sort-based skyline evaluation. ACM Trans. Database Syst. **33**(4) (2008) 1–49