

Making the Skyline Larger: A Fuzzy-Neighborhood-Based Approach

Djamal Belkasm^{1,2}, Allel Hadjali^{2(✉)}, and Hamid Azzoune³

¹ DIF-FS/UMBB, Boumerdes, Algeria

² LIAS/ENSMA, Poitiers, France

{[djamal.belkasm](mailto:djamal.belkasm@ensma.fr), [allel.hadjali](mailto:allel.hadjali@ensma.fr)}@ensma.fr

³ LRIA/USTHB, Algiers, Algeria
azzoune@yahoo.fr

Abstract. Skyline queries have gained much attention in the last decade and are proved to be valuable for multi-criteria ranking. They are based on the concept of Pareto dominance. In many real-life applications, the skyline returns only a small number of non-dominated objects which could be insufficient for the user. In this paper, we discuss an approach to enriching the small skyline with particular points that could serve the decision makers' needs. The idea consists in identifying the most interesting non-skyline points belonging to the fuzzy neighborhood of a skyline point and then adding them to the classical skyline. To do so, a particular fuzzy closeness relation is introduced. The relaxed skyline obtained which include the classical skyline, is a discriminated set. Furthermore, an efficient algorithm to compute the relaxed skyline is proposed. Extensive experiments are conducted to demonstrate the effectiveness of our approach and the performance of the proposed algorithm.

Keywords: Fuzzy sets · Databases · Skyline queries · Closeness · Relaxation

1 Introduction

In recent years, preference queries have received a great attention by many database researchers. Skyline queries [1] are specific example of SQL extensions that allow users to express preference in queries. Based on Pareto dominance relationship, skyline queries select all non-dominated objects based on a multi-criteria comparison. This means that, given a set D of d -dimensional points, a skyline query returns, the skyline S , set of points of D that are not dominated by any other point of D . A point p dominates another point q iff p is better than or equal to q in all dimensions and strictly better than q in at least one dimension. One can see that skyline points are incomparable. Several research studies have been conducted to develop efficient algorithms and introduce multiple variants of skyline queries [2–5]. However, querying a d -dimensional data sets using a skyline operator may lead to two possible scenarios: (i) a large number of skyline points

returned, which could be less informative for users, (ii) a small number of skyline points returned, which could be insufficient for users. To solve the two above problems, various approaches have been proposed to refine the skyline, therefore reducing its size [6–13], but only very few works exist to relax the skyline in order to increase the number of skyline results [10, 14–17]. Goncalves and Tıneo [15] propose a flexible dominance relationship using fuzzy comparison operators. This increases the skyline with points that are only weakly dominated by any other point. In [10], Hadjali et al. discuss some ideas of relaxing the skyline. In [14], and taking as starting point the study in [10], we develop an approach, called *MP2R* (*Much Preferred Relation for Relaxation*), for skyline relaxation. This approach relies on a novel fuzzy dominance relationship *Much Preferred (MP)* which makes more demanding the dominance between the points of D .

In this paper, we investigate another way of relaxing the skyline S . The idea is to consider that a non-skyline point p still belongs to a *fuzzily extended skyline* S_{FE} if p is *close* to a skyline point q . We then develop an approach, called *C2R* (*Closeness Relation for Relaxation*), to enlarging the small skyline with points that are closest to skyline points (keep in mind that those points are ruled out from the skyline when applying the classical Pareto dominance). The approach makes use of a particular appropriate fuzzy “*Closeness (C)*” relation. Each element in the relaxed skyline obtained S_{FE} is then associated with a degree ($\in [0, 1]$) expressing the extent to which it belongs to S_{FE} . In summary, the main contributions made are as follows:

- We provide the definition and semantic basis for a relaxed variant of skyline S_{FE} .
- We develop and implement an algorithm to compute S_{FE} efficiently.
- We conduct a set of experiments to study and analyze the relevance and effectiveness of S_{FE} .
- Finally, we present a comparative study between S_{FE} and S_{Relax} (i.e., the relaxed skyline obtained by the *MP2R* approach of [14]).

The paper is structured as follows: Sect. 2 provides some necessary background on skyline queries and on *MP2R*-based approach to skyline relaxation. In Sect. 3, we introduce a new approach for skyline relaxation based on fuzzy closeness relationship. An algorithm to efficiently compute S_{FE} is presented and discussed. Section 4 is devoted to the experimental study. Finally, Sect. 5 concludes the paper and draws some lines for future works.

2 Background

In this section, we recall some notions on skyline queries. Then, we present our *MP2R*-based approach for Skyline relaxation.

2.1 Skyline Queries

The notion of skyline queries was pioneered in [1]. Subsequently, the interest in this area has exploded: [1] has garnered over 1800 citations (Google Scholar,

January 2016). Skyline queries are a specific, yet relevant, example of preference queries. They rely on Pareto dominance principle which can be defined as follows:

Definition 1. Let D be a set of d -dimensional data points and u_i and u_j two points of D . u_i is said to dominate in Pareto sense u_j (denoted $u_i \succ u_j$) iff u_i is better than or equal to u_j in all dimensions and better than u_j in at least one dimension. Formally, we write

$$u_i \succ u_j \Leftrightarrow (\forall k \in \{1, \dots, d\}, u_i[k] \geq u_j[k]) \wedge (\exists l \in \{1, \dots, d\}, u_i[l] > u_j[l]) \quad (1)$$

where each tuple $u_i = (u_i[1], u_i[2], u_i[3], \dots, u_i[d])$ with $u_i[k]$ stands for the value of the tuple u_i for the attribute A_k .

In (1), without loss of generality, we assume that the largest value, the better.

Definition 2. The skyline of D , denoted by S , is the set of points which are not dominated by any other point.

$$u \in S \Leftrightarrow \nexists u' \in D, u' \succ u \quad (2)$$

Skyline queries compute the set of Pareto-optimal tuples in a relation, i.e., those tuples that are not dominated by any other tuple in the same relation.

Example 1. To illustrate the concept of the skyline, let us consider a database containing information on candidates as shown in Table 1. The list of candidates includes the following informations: Code, Age, Management experience (man_exp in years), Technical experience (tec_exp in years) and distance work to Home (dist_wh in Km). Ideally, personnel manager is looking for a candidate with the largest management and technical experience (Max man_exp and Max tec_exp), ignoring other informations. Applying the traditional skyline will return the following candidates: M_5, M_8 . As can be seen, such results are the most interesting candidates (see Fig. 1).

2.2 MP2R-based Approach for Skyline Relaxation

In [14] we have proposed an approach to relax skyline called $MP2R$. Its relies on a new dominance relationship that allows enlarging the skyline with the most interesting points among those ruled out when computing the initial skyline S . This new dominance relationship uses a fuzzy relation, named “*Much Preferred (MP)*” to compare two tuples u and u' . So, u is an element of S_{relax} if there is no tuple $u' \in U$ such that u' is *much preferred* to u (denoted $MP(u', u)$) in all skyline attributes. Formally, we write:

$$u \in S_{relax} \Leftrightarrow \nexists u' \in U, \forall i \in \{1, \dots, d\}, MP_i(u'_i, u_i) \quad (3)$$

where, MP_i is a fuzzy preference relation defined on the domain \mathbb{D}_i of the attribute A_i and $MP_i(u'_i, u_i)$ expresses the extent to which the value u'_i is *much preferred* to the value u_i . Each element u of S_{relax} is associated with a degree ($\in [0, 1]$). The semantics of this relation is represented by the trapezoidal function $(\gamma_{i1}, \gamma_{i2}, \infty, \infty)$, and denoted $MP_i^{(\gamma_{i1}, \gamma_{i2})}$, see Fig. 2. Figure 3 shows the relaxed version, S_{relax} , of the skyline S of the Example 1.

Table 1. List of candidates

Code	Age	man_exp	tec_exp	dist_wh
M1	32	5	10	35
M2	41	7	5	19
M3	37	5	12	45
M4	36	4	11	39
M5	40	8	10	18
M6	30	4	6	27
M7	31	3	4	56
M8	36	6	13	12
M9	33	6	6	95
M10	40	7	9	20

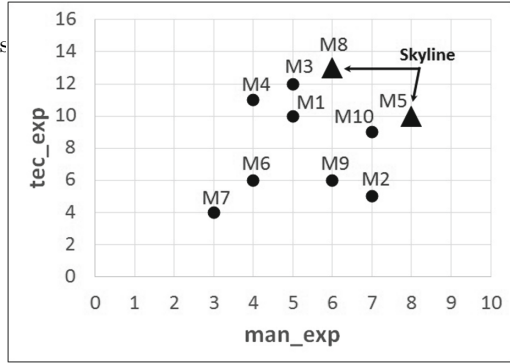


Fig. 1. Skyline of candidates

3 C2R: An Efficient Approach to Enlarging the Skyline

Let $\mathbb{D} = (\mathbb{D}_1, \mathbb{D}_2, \dots, \mathbb{D}_d)$ a d -dimensional space where \mathbb{D}_i is the domain attribute of A_i and $R(A_1, A_2, \dots, A_d)$ a relation defined in \mathbb{D} . We assume the existence of a total order relationship on each domain \mathbb{D}_i . $U = (u_1, u_2, \dots, u_n)$ is a set of n tuples belonging to a relation R . Let S be the skyline of U and S_{FE} the relaxed skyline of U computed by $C2R$ approach.

3.1 Principe of the Approach

Our approach relies on the idea of identifying interesting points that are in the neighborhood of skyline points and adding them to the skyline S . Let u be a tuple of $U - S$, and u' a tuple of S . Then, $u \in S_{FE}$ if u is close to u' . We write:

$$u \in S_{FE} \Leftrightarrow \exists u' \in S, \text{ such that } \forall i \in \{1, \dots, d\}, (u_i, u'_i) \in C_i \tag{4}$$

where, C_i is a reflexive, symmetrical approximate indifference (or equality) relation defined on the domain \mathbb{D}_i of the attribute A_i and $C_i(u_i, u'_i)$ expresses the extent to which the value u_i is close to the value u'_i . Since C_i is of a gradual

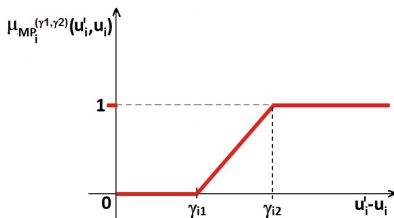


Fig. 2. μ_{MP_i} function

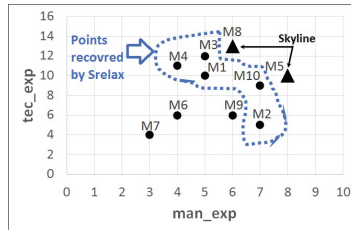


Fig. 3. S_{relax}

nature, each element u of S_{FE} is associated with a degree ($\in [0, 1]$) expressing the extent to which u belongs to S_{FE} . In fuzzy set terms, we write:

$$\mu_{S_{FE}}(u) = \max_{u' \in S} \min_i \mu_{C_i}(u_i, u'_i) \tag{5}$$

As for C_i relation on \mathbb{D}_i , its semantics can be provided by the formulas (6) (see also Fig. 4). In terms of t.m.f., C_i writes $(0, 0, \gamma_{i1}, \gamma_{i2})$, and denoted $C_i^{(\gamma_{i1}, \gamma_{i2})}$. It is easy to check that $C_i^{(0,0)}$ corresponds to the classical equality “=”.

$$\mu_{C_i^{(\gamma_{i1}, \gamma_{i2})}}(u_i, u'_i) = \begin{cases} 0 & \text{if } |u_i - u'_i| \geq \gamma_{i2} \\ 1 & \text{if } |u_i - u'_i| \leq \gamma_{i1} \\ \frac{(\gamma_{i2} - |u_i - u'_i|)}{\gamma_{i2} - \gamma_{i1}} & \text{else} \end{cases} \tag{6}$$

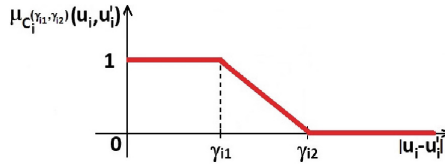


Fig. 4. The membership function $\mu_{C_i^{(\gamma_{i1}, \gamma_{i2})}}$

Let $\gamma = ((\gamma_{11}, \gamma_{12}), \dots, (\gamma_{d1}, \gamma_{d2}))$ be a vector of pairs of parameters where $C_i^{(\gamma_{i1}, \gamma_{i2})}$ denotes the C_i relation defined on the attribute A_i and $S_{FE}^{(\gamma)}$ denotes the extended skyline computed on the basis of the vector γ . One can easily check that the classical Skyline S is equal to $S_{FE}^{(\mathbf{0})}$, where $\mathbf{0} = ((0, 0), \dots, (0, 0))$.

Definition 3. Let γ and γ' be two vectors of parameters. We say that $\gamma \geq \gamma'$ if and only if $\forall i \in \{1, \dots, d\}, (\gamma_{i1}, \gamma_{i2}) \geq (\gamma'_{i1}, \gamma'_{i2})$ (i.e., $\gamma_{i1} \geq \gamma'_{i1} \wedge \gamma_{i2} \geq \gamma'_{i2}$).

Proposition 1. Let γ and γ' be two vectors of parameters. The following property holds: $\gamma \leq \gamma' \Rightarrow S_{FE}^{(\gamma)} \subseteq S_{FE}^{(\gamma')}$.

Proof. Let $\gamma \leq \gamma'$, one can deduce that $\forall i, C_i^\gamma \subseteq C_i^{\gamma'}$. Let $u \in S_{FE}^{(\gamma)}$
 $\Rightarrow \exists u' \in S, \forall i \in \{1, \dots, d\}, (u_i, u'_i) \in C_i^{(\gamma_{i1}, \gamma_{i2})}$
 $\Rightarrow \exists u' \in S, \forall i \in \{1, \dots, d\}, \mu_{C_i^{(\gamma_{i1}, \gamma_{i2})}}(u_i, u'_i) > 0$
 $\Rightarrow \exists u' \in S, \forall i \in \{1, \dots, d\}, \mu_{C_i^{(\gamma'_{i1}, \gamma'_{i2})}}(u_i, u'_i) > \mu_{C_i^{(\gamma_{i1}, \gamma_{i2})}}(u_i, u'_i) > 0$
 $\Rightarrow \exists u' \in S, \forall i \in \{1, \dots, d\}, (u_i, u'_i) \in C_i^{(\gamma'_{i1}, \gamma'_{i2})} \Rightarrow u \in S_{FE}^{(\gamma')}$
 So we have $S_{FE}^{(\gamma)} \subseteq S_{FE}^{(\gamma')}$ □

Lemma 1. Let $\gamma = ((0, \gamma_{12}), \dots, (0, \gamma_{d2}))$ and $\gamma' = ((\gamma'_{11}, \gamma'_{12}), \dots, (\gamma'_{d1}, \gamma'_{d2}))$, the following holds: $S_{FE}^{(\mathbf{0})} \subseteq S_{FE}^{(\gamma)} \subseteq S_{FE}^{(\gamma')}$

Table 2. Degrees of the elements of S_{FE}

Mat	M5	M8	M3	M10	M1	M2	M4	M6	M7	M9
$\mu_{S_{FE}}$	1	1	0.66	0.66	0.28	0	0	0	0	0

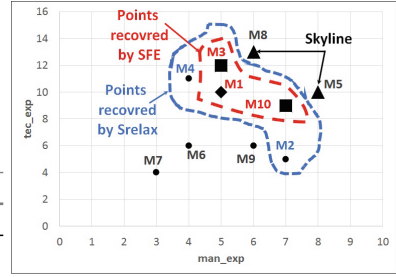


Fig. 5. Points retrieved by S_{FE}

Example 2. Let us come back to the skyline calculated in Example 1. Assume that the fuzzy “Closeness” relations corresponding to the skyline attributes (man_exp and tec_exp) are respectively given by:

$$\mu_{C_{man_exp}^{(1/2,2)}}(u, u') = \begin{cases} 1 & \text{if } |u - u'| \leq 1/2 \\ 0 & \text{if } |u - u'| \geq 2 \\ (-2|u - u'| + 4)/3 & \text{else} \end{cases} \quad (7)$$

$$\mu_{C_{tec_exp}^{(1/2,4)}}(u, u') = \begin{cases} 1 & \text{if } |u - u'| \leq 1/2 \\ 0 & \text{if } |u - u'| \geq 4 \\ (-2|u - u'| + 8)/7 & \text{else} \end{cases} \quad (8)$$

Now, applying our approach to relax the skyline $S = \{M_5, M_8\}$ found in Example 1, leads to the following $S_{FE} = \{(M_5, 1), (M_8, 1), (M_3, 0.66), (M_{10}, 0.66), (M_1, 0.28)\}$, see Table 2. One can note that some candidates that were not in S are now elements of S_{FE} (such M_3, M_{10} and M_1) see Fig. 5. As can be seen, S_{FE} is larger than S and $S_{FE} \subseteq S_{relax}$. Let us now take a glance at the content of S_{FE} , one can observe that (i) the skyline elements of S are still elements of S_{FE} with a degree equal to 1; (ii) Appearance of new elements recovered by our approach whose degrees are less than 1 (such as M_3). Interestingly, the user can select from S_{FE} : (i) the Top- k elements (k is a user-defined parameter), or (ii) the subset of elements, denoted $(S_{FE})_\sigma$, with a degrees higher than a threshold σ provided by the user. In the context of Example 2, it is easy to check that $Top - 5 = \{(M_5, 1), (M_8, 1), (M_3, 0.66), (M_{10}, 0.66), (M_1, 0.28)\}$ and $(S_{FE})_{0.66} = \{(M_5, 1), (M_8, 1), (M_3, 0.66), (M_{10}, 0.66)\}$.

3.2 S_{FE} Computation

To compute S_{FE} , we proceed in two steps (see Fig. 6). Firstly we compute the skyline S using a slightly modified version of **BNL** algorithm [14], then we execute our *FES* algorithm to relax the skyline S (see Algorithm 1).

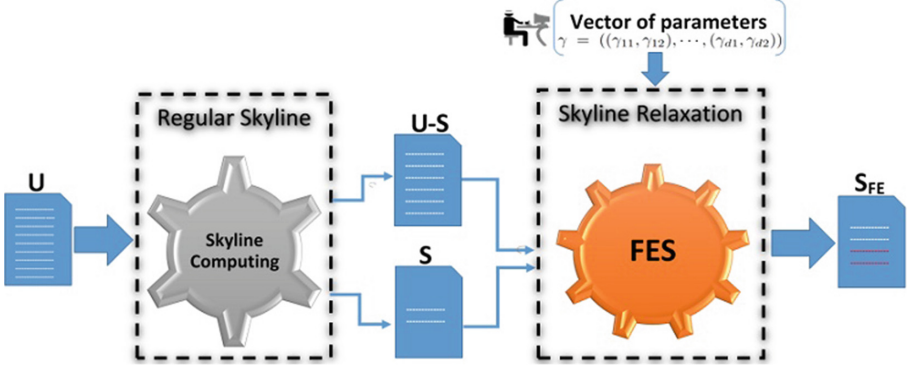


Fig. 6. Enlarging skyline process

Algorithm 1: FES

Input: Set of n tuples $U - S$; a Skyline of m tuples S ; γ a vector of parameters;

Output: The relaxed skyline S_{FE} ;

```

1 begin
2    $S_{FE} = S$ ;
3   for  $i = 1$  to  $n$  do
4      $V_{max} = 0$ ;
5     for  $j = 1$  to  $m$  do
6        $V_{min} = 1$ ;
7       for  $k = 1$  to  $d$  do
8          $V_{min} = MIN(V_{min}, \mu_{C_k}(u_i, u_j))$ ;
9         if  $V_{min} = 0$  then
10          break;
11        $V_{max} = MAX(V_{max}, V_{min})$ ;
12        $\mu_{S_{FE}}(u_i) = V_{max}$ ;
13       if  $V_{max} = 1$  then
14        break;
15     if  $\mu_{S_{FE}}(u_i) > 0$  then
16       $S_{FE} = S_{FE} \cup \{u_i\}$ ;
17   rank  $S_{FE}$  in decreasing order w.r.t.  $\mu_{S_{FE}}(u_i)$ ;
18   - to return top-K;
19   - or to return  $u_i$  satisfying  $\mu_{S_{FE}}(u_i) \geq \sigma$ , ( $\sigma$  is a user-defined threshold).
    
```

4 Experimental Study

The goal of this study is to demonstrate the effectiveness of the approach proposed and its ability to relax small skylines with the most interesting tuples. We also compare the results obtained with those computed by the $\mathcal{MP2R}$ approach.

4.1 Experimental Environment

We use a Linux OS, on a machine with an i7 processor, a RAM of 8 GB and a 250 GB of disk. Algorithms were implemented with Java. Dataset benchmark is generated using method described in [1] following three distribution schema (correlated, anti-correlated and independent). For each dataset, we consider different sizes (5 K to 750 K). Each tuple contains an integer identifier (4 bytes), 12 decimal fields (96 bytes) with values belonging to the interval [0,1], and a string field with length of 10 characters. Therefore, the size of one tuple is 110 bytes.

4.2 Experimental Results

We vary a collection of parameters that could impact the results. This collection includes the dataset size [D] (5 K, 10 K, 50 K, 100 K, 250 K, 500 K, 750 K), dataset distribution schema [DIS] (independent, correlated, anti-correlated), the number of skyline dimensions [d] (2, 4, 6, 8, 10, 12) and the relaxation thresholds $[\gamma = (\gamma_{i1}, \gamma_{i2}), i \in \{1, \dots, d\}]$ where $(\gamma_{i1}, \gamma_{i2} \in [0,1]$ and $\gamma_{i1} \leq \gamma_{i2}$). The default values of these parameters are $D = 5\text{ K}$; $\text{DIS} = \text{“Correlated”}$; $d = 2$; $\gamma = ((0,0.25), (0,0.25))$. In our experiment, we consider that the less the value, the better. Also, we address the issue of comparison between S_{FE} and S_{relax} in terms of Data distribution scheme [DIS], Number of skyline dimension [d], Data size [D] and Variation of the values of $(\gamma_{i1}, \gamma_{i2})$.

$-S_{FE}$ vs S_{relax} w.r.t [DIS]. Figure 7 shows that the particularity of correlated data minimize the seize of S_{FE} and S_{relax} . We observe also that $\mathcal{C}\mathcal{2}\mathcal{R}$ approach retrieves fewer tuples than $\mathcal{M}\mathcal{P}\mathcal{2}\mathcal{R}$ because it is more demanding when relaxation processes. We note that the execution time of $\mathcal{C}\mathcal{2}\mathcal{R}$, for the three distributions, is largely low compared with the time of $\mathcal{M}\mathcal{P}\mathcal{2}\mathcal{R}$ approach.

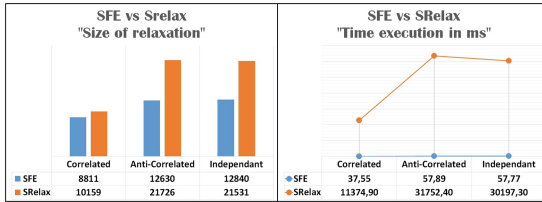


Fig. 7. S_{FE} vs S_{relax} w.r.t [DIS]. (Color figure online)

$-S_{FE}$ vs S_{relax} w.r.t [d]. When dimensionality increases (from 2 to 12) the size of S_{FE} and S_{relax} increases proportionally (see Fig. 8). We also note that S_{FE} outperforms S_{relax} in terms on computing time.

$-S_{FE}$ vs S_{relax} w.r.t [D]. The analysis of Fig. 9 shows that the size of S_{FE} and S_{relax} are proportional to the size of the dataset. While in terms of execution time, the computation of S_{FE} is extremely faster.

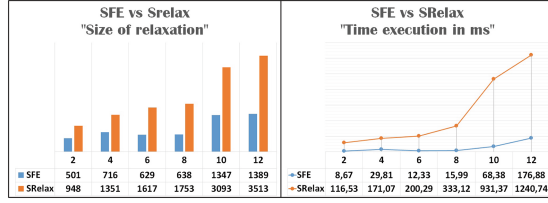


Fig. 8. S_{FE} vs S_{relax} w.r.t [d] (Color figure online)

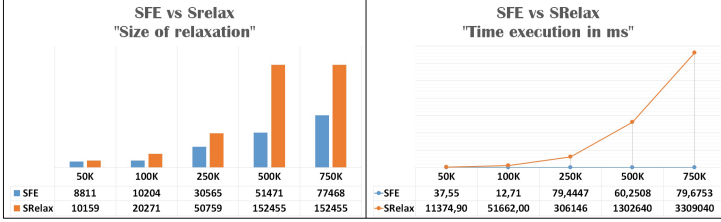


Fig. 9. S_{FE} vs S_{relax} w.r.t [D] (Color figure online)

As can be seen, this first part of the experimental study shows that $\mathcal{C}2\mathcal{R}$ approach is better and more optimal than $\mathcal{MP}2\mathcal{R}$ approach.

Variation of $(\gamma_{i1}, \gamma_{i2})$ values. Now, we show the influence of the variation of $(\gamma_{i1}, \gamma_{i2})$ values on the size and the computation time of S_{FE} and S_{relax} . The idea is to vary both thresholds. For the sake of simplicity, and since the data are normalized, we will apply the same values of (γ_1, γ_2) for all skyline dimensions. Note that the size of the skyline is equal to 1 and we will analyze the variation of the number of tuples whose degree $\mu_{S_{FE}}(u) > 0$. The following scenarios are worth to be discussed:

Scenario 1: In this scenario, we fix γ_{i1} and vary γ_{i2} to increase the relaxation zone. We observe the following cases:

- $\gamma_{i1} = 0$ and $\gamma_{i2} \in \{0; 0.25; 0.5; 0.75; 1\}$ (see Fig. 10)
- $\gamma_{i1} = 0.25$ and $\gamma_{i2} \in \{0.25; 0.5; 0.75; 1\}$ (see Fig. 11)
- $\gamma_{i1} = 0.5$ and $\gamma_{i2} \in \{0.5; 0.75; 1\}$ (see Fig. 12)
- $\gamma_{i1} = 0.75$ and $\gamma_{i2} \in \{0.75; 1\}$ (see Fig. 13)

The analysis of Fig. 10 shows that the size of S_{FE} and S_{relax} increases when the value of γ_{i2} increases. We also note that there are no tuples whose degrees of relaxation is equal to 1 (this is due to the value of $\gamma_{i1} = 0$). In Figs. 11, 12 and 13 we note that the value of γ_{i2} controls the size of relaxation (by S_{FE} or S_{relax}).

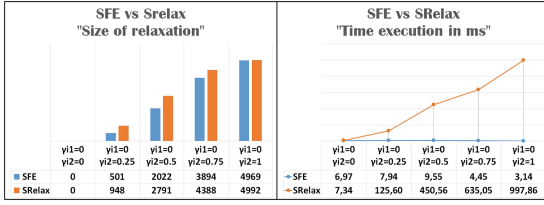


Fig. 10. Scenario 1: Fix γ_{i1} and vary γ_{i2} (case1) (Color figure online)

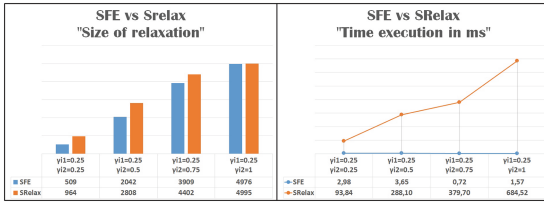


Fig. 11. Scenario 1: Fix γ_{i1} and vary γ_{i2} (case2) (Color figure online)

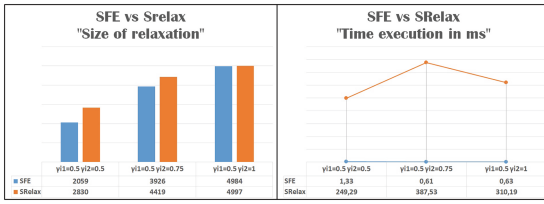


Fig. 12. Scenario 1: Fix γ_{i1} and vary γ_{i2} (case3) (Color figure online)

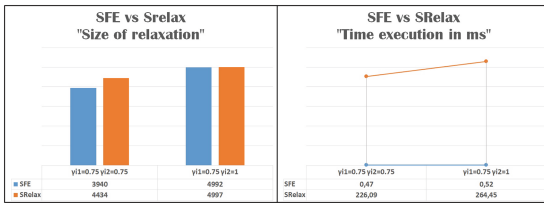


Fig. 13. Scenario 1: Fix γ_{i1} and vary γ_{i2} (case4) (Color figure online)

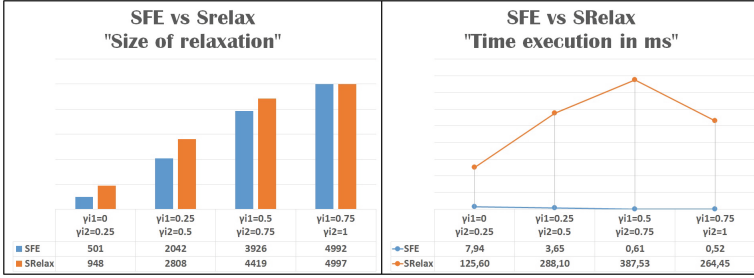


Fig. 14. Varying γ_{i1} and γ_{i2} (Color figure online)

However, we observe the appearance of retrieved tuples with degrees equal 1. It is noted that, whatever the value of γ_{i1} and γ_{i2} , the $\mathcal{C}\mathcal{R}$ is more efficient than $\mathcal{M}\mathcal{P}\mathcal{R}$ in terms of computation time.

Scenario 2: In this scenario, we vary both thresholds. The obtained results are shown in Fig. 14. The analysis of these curves shows that the relaxation process becomes more permissive when thresholds move away from the origin. Nevertheless, S_{FE} is always more selective than S_{relax} on the number of tuples retrieved (i.e., $|S_{FE}| < |S_{relax}|$)¹ and more efficient in terms of computation time. The Fig. 15 illustrates the distribution of tuples recovered by S_{FE} according to their degrees of relaxation.

Scenario 3: In the previous scenarios, the vector $\gamma = (\gamma_{i1}, \gamma_{i2})$ is similar when computing S_{relax} and S_{FE} . Here we will show the impact of using different vectors γ and γ' respectively for S_{FE} and S_{relax} . Table 3 summarizes the results obtained. One can observe that $|S_{FE}| < |S_{relax}|$ if $\gamma \leq \gamma'$, $|S_{FE}| > |S_{relax}|$ otherwise (Fig. 16).

Table 3. Impact of the vector γ and γ' .

	#Tuples		Conclusion
	S_{FE}	S_{relax}	
Cas1: $\gamma < \gamma'$	2043	4998	$ S_{FE} < S_{relax} $
Cas2: $\gamma > \gamma'$	4985	4403	$ S_{FE} > S_{relax} $
Cas3: $\gamma = \gamma'$	2042	2808	$ S_{FE} < S_{relax} $

¹ Even the relation $S_{FE} \subseteq S_{relax}$ holds in this context.

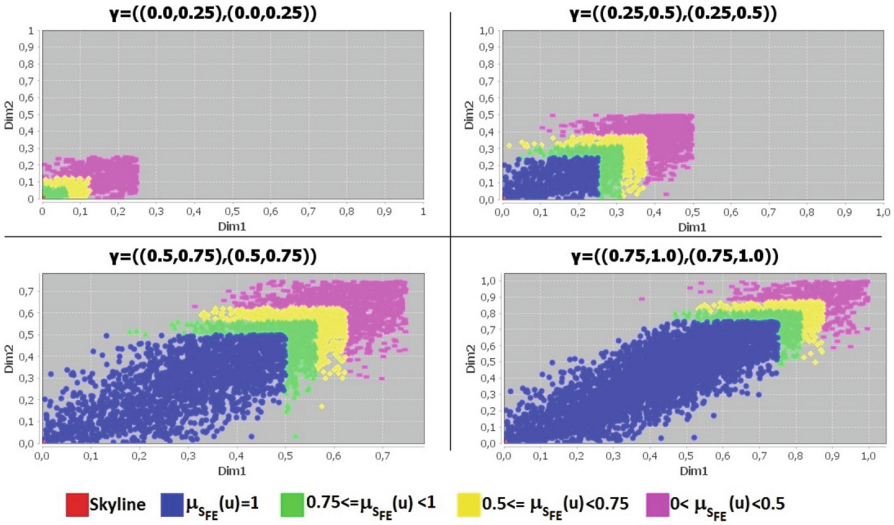


Fig. 15. Distribution of tuples recovered by S_{FE} (Color figure online)

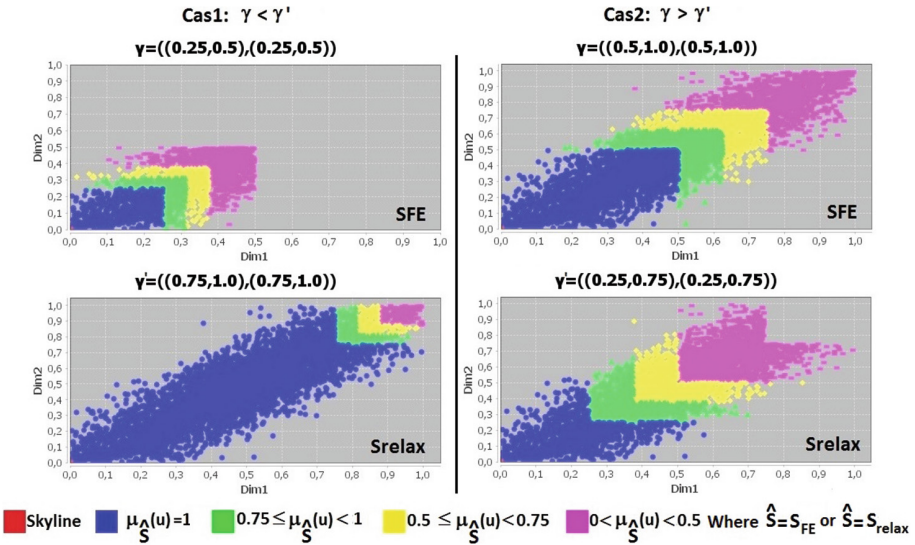


Fig. 16. Distribution of tuples recovered by S_{FE} and S_{relax} (Color figure online)

5 Conclusion

In this paper, we addressed the problem of skyline relaxation, especially less skylines. We propose a new approach for relaxing the skyline, called $\mathcal{C}\mathcal{R}$. This approach is based on a particular fuzzy *Closeness* relation whose semantics is a user-defined. In addition, a new algorithm called **FES** to compute the relaxed skyline is proposed. The experimental study we done has shown that, on the one hand, and in some cases, the $\mathcal{C}\mathcal{R}$ approach is more restrictive than $\mathcal{MP}\mathcal{R}$ approach when relaxing classic skyline and, on the other hand, the computation cost of $\mathcal{C}\mathcal{R}$ is more acceptable. Furthermore, $\mathcal{C}\mathcal{R}$ like $\mathcal{MP}\mathcal{R}$ involves various parameters, which can be used to control the size and the quality of the relaxed skyline. As for future work, we will consider the $\mathcal{C}\mathcal{R}$ approach using a relative fuzzy closeness relation. Then, we will investigate the issue of skyline relaxation in the categorical attributes context.

References

1. Börzsönyi, S., Kossman, D., Stocker., K.: The skyline operator. In: ICDE, pp. 421–430 (2001)
2. Hadjali, A., Pivert, O., Prade, H.: Possibilistic contextual skylines with incomplete preferences. In: SoCPaR, pp. 57–62 (2010)
3. Khalefa, M.E., Mokbel, M.F., Levandoski, J.J.: Skyline query processing for incomplete data. In: IEEE ICDE, pp. 556–565 (2008)
4. Pei, J., Jiang, B., Lin, X., Yuan, Y.: Probabilistic skylines on uncertain data. In: VLDB, pp. 15–26 (2007)
5. Yiu, M.L., Mamoulis, N.: Efficient processing of top-k dominating queries on multi-dimensional data. In: VLDB, pp. 483–494 (2007)
6. Abbaci, K., Hadjali, A., Lietard, L., Rocacher, D.: A linguistic quantifier-based approach for skyline refinement. In: IFSA/NAFIPS, pp. 321–326 (2013)
7. Chan, C.Y., Jagadish, H.V., Tan, K., Tung, A.K.H., Zhang, Z.: Finding k-dominant skylines in high dimensional space. In: ACM SIGMOD, pp. 503–514 (2006)
8. Chan, C.-Y., Jagadish, H.V., Tan, K.-L., Tung, A.K.H., Zhang, Z.: On high dimensional skylines. In: Ioannidis, Y., Scholl, M.H., Schmidt, J.W., Matthes, F., Hatzopoulos, M., Böhm, K., Kemper, A., Grust, T., Böhm, C. (eds.) EDBT 2006. LNCS, vol. 3896, pp. 478–495. Springer, Heidelberg (2006)
9. Endres, M., Kießling, W.: Skyline snippets. In: Christiansen, H., De Tré, G., Yazici, A., Zadrozny, S., Andreasen, T., Larsen, H.L. (eds.) FQAS 2011. LNCS, vol. 7022, pp. 246–257. Springer, Heidelberg (2011)
10. Hadjali, A., Pivert, O., Prade, H.: On different types of fuzzy skylines. In: Kryszkiewicz, M., Rybinski, H., Skowron, A., Raś, Z.W. (eds.) ISMIS 2011. LNCS, vol. 6804, pp. 581–591. Springer, Heidelberg (2011)
11. Hüllermeier, E., Vladimirskiy, I., Prados Suárez, B., Stauch, E.: Supporting case-based retrieval by similarity skylines: basic concepts and extensions. In: Althoff, K.-D., Bergmann, R., Minor, M., Hanft, A. (eds.) ECCBR 2008. LNCS (LNAI), vol. 5239, pp. 240–254. Springer, Heidelberg (2008)
12. Lin, X., Yuan, Y., Zhang, Q., Zhang, Y.: Selecting stars: the k most representative skyline operator. In: ICDE, pp. 86–95 (2007)

13. Papadias, D., Tao, Y., Fu, G., Seeger, B.: An optimal and progressive algorithm for skyline queries. In: ACM SIGMOD, pp. 467–478 (2003)
14. Belkasmi, D., Hadjali, A.: $\mathcal{MP}\in\mathcal{R}$: a human-centric skyline relaxation approach. In: Proceedings of FQAS, Cracow, Poland, 26–28 October, pp. 227–241 (2015)
15. Goncalves, M., Tineo, L.J.: Fuzzy dominance skyline queries. In: Wagner, R., Revell, N., Pernul, G. (eds.) DEXA 2007. LNCS, vol. 4653, pp. 469–478. Springer, Heidelberg (2007)
16. Balke, W., Güntzer, U., Siberski, W.: Restricting skyline sizes using weak pareto dominance. Inform. Forsch. Entwickl. **21**(3–4), 165–178 (2007)
17. Jin, W., Han, J., Ester, M.: Mining thick skylines over large databases. In: Boulicaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D. (eds.) PKDD 2004. LNCS (LNAI), vol. 3202, pp. 255–266. Springer, Heidelberg (2004)