Skyline Operator over *Tripadvisor* Reviews within the Belief Functions Framework

Fatma Ezzahra Bousnina[1], Sayda Elmi[14] , Mouna Chebbah[12], Mohamed Anis Bach Tobji[13], Allel HadjAli[4], Boutheina Ben Yaghlane[15]

[1] Université de Tunis, ISG, LARODEC fatmaezzahra.bousnina@gmail.com
[2] Université de Jendouba, FSEGJ mouna.chebbah@fsjegj.rnu.tn
[3] Univ. Manouba, ESEN anis.bach@isg.rnu.tn
[4] Université de Poitiers, ENSMA, LIAS allel.hadjali@ensma.fr; saida.elmi@ensma.fr
[5] Université de Carthage, IHEC boutheina.yaghlane@ihec.rnu.tn

Abstract. The crowdsourcing *Tripadvisor* platform do not offer a multi-criteria filtering functionality for their users. Thus, these users are obliged to choose only one criteria to filter a query's results. In this paper, we introduce a new skyline operator, in the context of belief functions theory, to meet the multi-criteria filtering objective. The queried data, modeled with the theory of belief functions, takes into account all reviews and also reviewers' reliabilities. Experiments show interesting results of the proposed skyline operator in terms of size and performance.

Keywords: Evidence Theory ; Evidential Databases; Reliability Estimation; Combination; Discounting; Crowdsourcing; Skyline; Tripadvisor

1 Introduction

Crowdsourcing is a practice that asks the crowd or consumers via a questionnaire to propose or create a marketing policy. It provides a powerful system for creating data from real life participants. Partakers contribute with their feedbacks/reviews about a defined task. Crowdsourcing can be very challenging when it comes to gather and process information. *Tripadvisor* platform is one of the most well known crowdsourcing sites where travelers express their opinions about hotels they visited through an evaluation form. The collected reviews are used later to answer users' queries about the best hotels regarding some criteria like distance, price, etc. However, *Tripadvisor* can not answer to a multi-criteria query (for example, the cheapest and closest hotel to the beach). Skyline queries [2] are defined as preference queries that offer the possibility of multi-criteria filtering. Nevertheless, this kind of queries are not adapted to the crowdsourcing platforms. They query databases where each tuple corresponds to a different hotel. In fact, they do not combine reviews about the same question. Thus, the use of the theory of belief functions to assess reviewers' reliabilities, to combine reviews and also to consider reviewers' reliabilities.

In this paper, we model the reviewers' feedbacks of *Tripadvisor* with the theory of belief functions [5]. First, we combine them to produce a data set of rates per hotel. Reviewers' scores are considered in the combination operation as the sources' reliabilities. Then, we introduce a new evidential skyline operator that deals with the particular type of obtained data. Finally, we implement the new operator and we lead experiments to compare its performance with the classic technique. Throughout this paper, example of table 1 will be used. In this table, travelers give their evaluations about hotels $\{h_1; h_2; h_3; h_4\}$ over a scale of 6 notes.

In the sequel of this paper, some basic concepts of the theory of belief functions, evidential databases and skyline operator are presented in section 2. Aggregation of travelers' reviews considering their reliabilities are presented in section 3. In section 4, the new skyline operator applied over the obtained *Tripadvisor* data is proposed. Experimental results and comparison with the classic evidential skyline [7] are also presented in the same section. Conclusion and future works are held in section 5.

Table 1. Reviews about Hotels

Reviewers	*Hotels*	*Price*	*Place*	*Service*	*Score*
R_1	h_1	3	4	3	1510
R_2	h_1	-1	4	2	22800
R_3	h_2	4	-1	5	400
R_4	h_2	3	5	-1	8140

2 Background Material

In this section, some basic concepts relative to the belief functions theory, evidential databases and skyline operator are presented.

2.1 Theory of Belief Functions

The theory of belief functions was introduced by Dempster and Shafer [5,6,11], it is also called *theory of evidence* or *the Dempster-Shafer theory*. In one hand, evidence theory provides an explicit representation of uncertainty and imprecision. In the other hand, it models other types of imperfection such the partial and the total ignorance. Let Θ be a finite set of exhaustive and mutually exclusive hypotheses called *frame of discernment*. The *power set* $2^{\Theta} = \{\{\varnothing\}, \{\theta_1\}, \{\theta_2\}, ..., \{\theta_n\}, \{\theta_1, \theta_2\}, .., \{\theta_1, \theta_2, ..., \theta_n\}\}$ includes all subsets of Θ. A *basic belief assignment* (*bba*), also called a *mass function*, is a mapping $m : 2^{\Theta} \longrightarrow [0, 1]$ such that

$$m(\emptyset) = 0 \quad \text{and} \quad \sum_{A \subseteq \Theta} m(A) = 1 \tag{1}$$

When $m(A) > 0$, A is called a *focal element*. The mass $m(A)$ is the belief committed exactly to A and to none of its subsets.

The *belief function*, denoted *bel*, represents the minimal degree of faith committed exactly to an hypothesis A, such that:

$$bel(A) = \sum_{B,A \subseteq \Theta : B \subseteq A} m^{\Theta}(B) \tag{2}$$

The *plausibility function*, denoted *pl*, is the maximal degree of faith committed to an hypothesis A, such that:

$$pl(A) = \sum_{B,A \subseteq \Theta : A \cap B \neq \varnothing} m^{\Theta}(B) \tag{3}$$

The belief function, *bel*, quantifies the degree of faith on a proposition A justified by degrees of supports (masses) of its subsets. It quantifies also the degree of faith on a comparison. Thus, comparing two independent probability distributions is easy in the framework of probability theory. However, standard *bel* and *pl* functions are not able to manage comparisons. Indeed, their definitions were extended [1,8,9] to meet the aim of comparing two independent (*bbas*).

Let X and Y be two independent variables $m_X, m_Y : 2^{\Theta} \rightarrow [0,1]$ their respective evidential values. The *bel* of inequalities are defined as follows:

Definition 1. *(Inequality bel($X \leq Y$))*

$$bel(X \leq Y) = \sum_{A \subseteq \Theta} (m_X(A) \sum_{B \subseteq \Theta, A \leq^{\exists} B} m_Y(B)). \tag{4}$$

Definition 2. *(Inequality bel($X < Y$))*

$$bel(X < Y) = \sum_{A \subseteq \Theta} (m_X(A) \sum_{B \subseteq \Theta, A <^{\forall} B} m_Y(B)) \tag{5}$$

In the theory of belief functions, a large set of combination rules [10] merge *bba*s in the aim of improving decisions. The first one is the Dempster's rule of combination [5] that generalizes the Bayes rule. It is normalized and it combines mass functions produced from different and independent sources. The joint mass is obtained from merging two *bba*s using the orthogonal sum. This rule of combination is commutative, associative but not idempotent [6].

Definition 3. *Let m_1 and m_2 be two independent mass functions, the* joint mass $m_{1\oplus 2}$ *is computed such that:*

$$m_{1\oplus 2}(A) = \begin{cases} \dfrac{\sum_{B \cap C = A} m_1(B).m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B).m_2(C)} & \forall A \neq \emptyset \\ 0 & \forall A = \emptyset \end{cases} \tag{6}$$

[6] $m \circledcap m \neq m$

A particular combination is the discounting which considers sources' reliabilities into their mass functions. It is a specific mechanism to the belief functions theory that discounts masses proportionally to their sources' reliabilities. However, sources' reliabilities need to be learned before the discounting.

The reliability factor α in [0, 1] characterizes the credibility of a source. Note that (i) $\alpha = 1$ represents a fully reliable source, (ii) $\alpha = 0$ represents an unreliable one and (iii) $1 - \alpha$ is the discounting. The discounted mass m^α is computed as follows:

$$\begin{cases} m^\alpha(A) = \alpha.m(A) & \forall A \subset \Theta \\ m^\alpha(\Theta) = \alpha.m(\Theta) + (1-\alpha) \end{cases} \tag{7}$$

The theory of belief functions, is used to model imperfect data in many domains like medicine and weather forecasting. Such data need to be stored in order to be later queried. Thereby, specific database models that can handle data modeled with belief functions theory were introduced.

2.2 Evidential Databases

An *Evidential database* denoted (*EDB*), also named *Dempster-Shafer database*. The evidential database model was firstly introduced by Lee [8,9]. Later on, other models were proposed [1,3,4]. An *EDB* stores perfect and imperfect information, modeled using the evidence theory. It has N tuples and D attributes. An *evidential value*, denoted V_{ta} is the value of an attribute a for a tuple t that represents a *bba*, m_{ta}, such that:

$$V_{ta} : 2^{\Theta_a} \rightarrow [0, 1] \tag{8}$$

$$\text{with } m_{ta}(\varnothing) = 0 \;\text{ and } \sum_{A \subseteq \Theta_a} m_t(A) = 1 \tag{9}$$

The set of focal elements of a *bba* V_{ta} is noted F_{ta} such that:

$$F_{ta} = \{x \subseteq \Theta / m_{ta}(x) > 0\}$$

2.3 Skyline Operator

Skyline operator over an *EDB* introduced by [2] is based on the formal model of Pareto dominance also called Pareto preference.

Let $\mathcal{H}$ be a collection of objects defined on a set of attributes $A = \{a_1, a_2, \ldots, a_d\}$ such that:

Definition 4. *(Pareto Dominance) Given two objects $h_t, h_l \in \mathcal{H}$, h_t dominates h_l (in the sense of Pareto), denoted by $h_t \succ h_l$, if and only if h_t is as good or better*[7] *than h_l in all attributes and strictly better in at least one attribute, i.e.,*

[7] To make simple and without loss of generality, we assume through all the paper that the smaller the value the better it is.

$\forall a_r \in A : h_t.a_r \leq h_h.a_r \wedge \exists a_\ell \in A : h_t.a_\ell < h_l.a_\ell$ where $h_t.a_r$ and $h_l.a_r$ stand for the r^{th} attribute of h_t and h_l, respectively.

Definition 5. *(Skyline) The skyline of $\mathcal{H}$, denoted by $Sky_{\mathcal{H}}$, includes objects of $\mathcal{H}$ that are not dominated by any other object, i.e., $Sky_{\mathcal{H}} = \{h_t \in \mathcal{H} \mid \nexists\ h_l \in \mathcal{H}, h_l \succ h_t\}$.*

In this paper, we propose a new optimized evidential skyline operator that we apply over the *Tripadvisor* travelers' reviews; however, earlier these responses are treated with the belief functions' tools. In the next section, we present the modeling of given responses as *bba*s and then the discounting with sources' reliabilities. Finally, these discounted *bba*s are combined per attribute for the different hotels.

3 Elicitation of Reviewers' Feedbacks as Belief Functions

Tripadvisor provides a reviewing form for travelers in order to evaluate hotels according to several criteria. A response about one criteria for a specific hotel can be in {-1;1;2;3;4;5}. A response in {1;2;3;4;5} is precise and certain. It induces a precise and certain belief function. The response -1 reflects the total ignorance. All responses to the same review (same hotel and same criteria) are combined in order to provide one *bba* that summarizes all the reviewers' evaluations. Note that, responses need to be discounted to take into account reviewers' reliabilities. A reviewer response is translated into a *bba*, in the context of belief functions theory.

3.1 Construction of Mass Functions

Belief functions theory allows the construction of basic belief assignments (*bbas*) from the set of hypotheses. The mass of an hypothesis A as modeled in equation (1) and denoted, $m(A)$, is interpreted as the degree of support given by an expert and that reflects his belief on that response A. This mass can not be divided on subsets of A. In *Tripadvisor* platform, each traveler chooses one rate from 1 to 5. If he does not provide a rate, his response is interpreted as -1. From the theory of belief functions' point of view, the frame of discernment is $\Theta = \{1, 2, 3, 4, 5\}$. We recall that -1 is interpreted as total ignorance, $m(\Theta) = 1$. Each non empty response is interpreted as certain and precise belief functions over Θ.

Example 1. The first reviewer R_1 gives a rate 3 for the service of hotel h_1. This response is interpreted as $m(3) = 1$.

Table 2 is an interpretation of table 1, in the context of belief functions theory for criteria: Price, Place and Service.

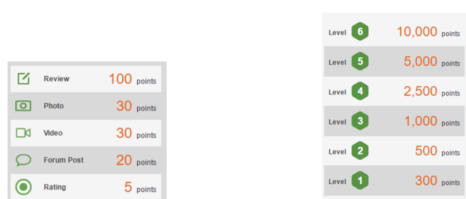
These mass functions are combined in order to have only one *bba* for each hotel. Before combining these reviews, they have to be discounted to take into account the travelers' reliabilities. Therefore, reviewers' reliabilities are firstly estimated.

Table 2. Construction of mass functions

Reviewers	*Hotels*	*Price*	*Place*	*Service*	*Score*
R_1	h_1	m(3) = 1	m(4) = 1	m(3) = 1	0.136
R_2	h_1	m(Θ) = 1	m(4) = 1	m(2) = 1	0.99
R_3	h_2	m(4) = 1	m(Θ) = 1	m(5) = 1	0.036
R_4	h_2	m(3) = 1	m(5) = 1	m(Θ) = 1	0.733

3.2 Reliability Estimation and Discounting

One of the most interesting challenges in crowdsourcing is quantifying the reliability of reviewers. The conflict between two experts' opinions reflects the unreliability of at least one of them. The estimated reliability of each reviewer is used later to weaken their given opinions modeled through the basic belief assignments(*bbas*). The *Tripadvisor* platform attributes to each reviewer a number of points depending to its contributions. These points are accumulated when the traveler (reviewer) gives an opinion about a hotel that he visited. Figure 1(a) shows how the *Tripadvisor* rewards reviewers that add photos, videos, helpful reviews, etc. Added to that, *Tripadvisor* divides its reviewers into 6 levels, shown in figure 1(b): the first level is assigned to travelers having 300 to 2499 points and the final and the sixth level is affected to travelers with points starting from 10.000. Method of rewarding travelers as illustrated in figure 1 is fixed by the *Tripadvisor* platform.



(a) Point (b) Levels

Fig. 1. Computation of points in *Tripadvisor* and their corresponding levels

We propose to estimate the reliability of each reviewer based on points and levels given by the *Tripadvisor* platform. Thus, we propose two methods: the first is to calculate a reliability for each reviewer having points from 300 to 10.000 relatively to the sixth level, as shown in equation (10), and the second is to compute the reliability score for reviewers having more than 10.000 point (i.e, travelers that acquire the last level and accumulating more points), as shown in equation (11).

The maximal score is fixed to 0.9 for the 10.000 points. Based on that, a reliability is computed for reviewers having points under 10.000, such that:

$$Score = (points * 0.9)/10.000 \tag{10}$$

When the number of points accumulated by a reviewer are greater that 10000, the reliability is computed such that:

$$Score = 1 - (1/points) \tag{11}$$

Figure 3.2 shows the reviewers' reliabilities according to accumulated points.

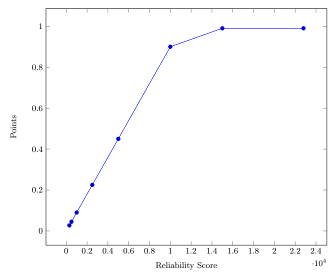


Fig. 2. Estimated Reviewers' Reliabilities

Example 2. the first reviewer R_1 in table 1 has accumulated 1510 points and since his number of points is lower than 10000 then his reliability score is computed using method (i): $Score_{R_1}$ = 1510 * 0.9 / 10000 = 0.136. The second reviewer R_2 has accumulated more points than 10000 then his reliability score is computed using method (ii): $Score_{R_2}$ = 1 - (1 /22800) = 0.99. Estimated reliabilities for all reviewers are shown in table 2.

The reliability estimated for each reviewer is used to discount the basic belief assignments that reflect their reviews about hotels using equation (7).

Example 3. The reviewer R_1, the reliability degree is $\alpha = 0.136$. Thus: $m^{\alpha}_{Price}(3) = 0.136 * 1 = 0.136$
$m^{\alpha}_{Price}(\Theta) = 0.136 * 0 + (1 - 0.136) = 0.864$

Results of discounted mass functions are shown in table 3.

Once the reviews, modeled as *bba*s, are discounted, they may be combined.

3.3 Combination of Reviews

In theory of belief functions, combination rules aggregate data from different and independent sources to get one mass function that reflects all sources' opinions.

Table 3. Discounting of mass functions

Reviewers	*Hotels*	*Price*	*Place*	*Service*
R_1	h_1	m(3) = 0.136 m(Θ) = 0.864	m(4) = 0.136 m(Θ) = 0.864	m(3) = 0.136 m(Θ) = 0.864
R_2	h_1	m(Θ) = 1	m(4) = 0.99 m(Θ) = 0.01	m(2) = 0.99 m(Θ) = 0.01
R_3	h_2	m(4) = 0.036 m(Θ) = 0.964	m(Θ) = 1	m(5) = 0.036 m(Θ) = 0.964
R_4	h_2	m(3) = 0.733 m(Θ) = 0.267	m(5) = 0.733 m(Θ) = 0.267	m(Θ) = 1

Table 4. Combination of *bba*s about the Price of h_2

Price	$m^{h_2}_{R_3}(\Theta) = 0.964$	$m^{h_2}_{R_3}(4) = 0.036$
$m^{h_2}_{R_4}(\Theta) = 0.267$	$m_{34}(\Theta) = 0.26$	$m_{34}(4) = 0.01$
$m^{h_2}_{R_4}(3) = 0.733$	$m_{34}(\Theta) = 0.7$	$m_{34}(\emptyset) = 0.03$

Example 4. Reviews about hotel h_2 for attribute *Price* are combined as shown in table 4.

The joint mass of reviewers R_3 and R_4, $m_{3\oplus 4}$ about the price of hotel h_2 is: (i) $m_{3\oplus 4}(3) = 1/(1 - 0.03) * 0.7 = 0.72$; (ii) $m_{3\oplus 4}(4) = 1(1 - 0.03) * 0.03 = 0.012$; (iii) $m_{3\oplus 4}(\Theta) = 1/(1 - 0.03) * 0.26 = 0.268$.

Similarly, we combine all *bba*s for each attribute for the different hotels. The obtained evidential database EDB is in table 5.

Table 5. Evidential Database

Hotels	*Price*	*Place*	*Service*
h_1	m(3) = 0.136 m(Θ) = 0.864	m(4) = 0.9814 m(Θ) = 0.0086	m(2) = 0.98 m(3) = 0.01 m(Θ) = 0.01
h_2	m(3) = 0.72 m(4) = 0.012 m(Θ) = 0.268	m(4) = 0.992 m(Θ) = 0.008	m(5) = 0.036 m(Θ) = 0.964

The obtained database is evidential with either precise *bba*s, or partial ignorance *bba*s. This EDB is then queried with preference conditions using the skyline operator. Preference conditions may deal either with one attribute like Price, Place or Service or with a combination of these attributes leading to the multi criteria filtering.

4 Skyline operator in Tripadvisor and Experimental results

Applying the evidential skyline operator [7], we can apply an existing method according to the *Tripadvisor* data. The dominance relationship extended to evidential data can be defined as follows:

Definition 6. *(The b-dominance) Given two objects $h_i, h_j \in \mathcal{H}$ and a belief threshold b, h_i b-dominates h_j denoted by $h_i \succ_b h_j$ if and only if h_i is believably as good or better than h_j in all attributes a_r in A $(1 \leq r \leq d)$ and strictly better in at least one attribute a_{r_0} $(1 \leq r_0 \leq d)$ according to a belief threshold b, i.e., $\forall a_r \in A : bel(h_i.a_r \geq h_j.a_r) \geq b$ and $\exists a_l \in A : bel(h_i.a_l > h_j.a_l) \geq b$.*

Given an object h_i, we denote by $h_i.a_r^-$ and by $h_i.a_r^+$ respectively the minimum value and the maximum value of the *bba* defined on the attribute a_r denoted by $h_i.a_r$.

Property 1. Let b be a belief threshold, if the mass function affected to the partial ignorance of a *bba* $h_i.a_r$ is greater than (1-b), i.e., $m_{h_i.a_r}(\Theta) > (1 - b)$ then $bel(h_i.a_r \geq h_j.a_r) < b$. In this case, the object h_i can not b-dominate h_j.

Example 5. Suppose we have $b = 0.6$. Let $h_i.a_r$ and $h_j.a_r$ be two *bbas* defined on objects h_i and h_j, respectively, and defined on the attribute a_r such that $h_i.a_r = \{3\}(0.4), \{\Theta\}(0.6)$ and $h_j.a_r = \{2\}(0.3), \{\Theta\}(0.7)$. $bel(h_i.a_r \geq h_j.a_r) = 0.12 < 0.6$ since $m_{h_i.a_r}(\Theta) = 0.6 > (1 - b)$.

Property 2. Let b be a belief threshold, if $m(h_i.a_r^+)$ is greater than b, i.e., $m(h_i.a_r^+) \geq b$ then $bel(h_i.a_r \geq h_j.a_r) \geq b$.

Property 3. Let b be a belief threshold, if $m(h_i.a_r^-)$ is greater than b, i.e., $m(h_i.a_r^-) \geq b$ then $bel(h_i.a_r \geq h_j.a_r) < b$.

Intuitively, an object is in the believable skyline if it is not believably dominated by any other object.

Based on the b-dominance relationship, the notion of b-$sky_{\mathcal{H}}$ is defined as follows.

Definition 7. *(The b-skyline) The b-skyline of $\mathcal{H}$ denoted by b-$sky_{\mathcal{H}}$, comprises those objects in $\mathcal{H}$ that are not b-dominated by any other object, i.e.,*

$$b\text{-}sky_{\mathcal{H}} = \{h_i \in \mathcal{H} \mid \nexists\, h_j \in \mathcal{H}, h_j \succ_b h_i\}.$$

Property 4. Given two belief thresholds b and b', if $b < b'$ then the b-$sky_{\mathcal{H}}$ is a subset of the b'-$sky_{\mathcal{H}}$, i.e., $b < b' \Rightarrow b\text{-}sky_{\mathcal{H}} \subseteq b'\text{-}sky_{\mathcal{H}}$.

Proof. Assume that there exists an object h_i such that $h_i \in b\text{-}sky_{\mathcal{H}}$ and $h_i \notin b'$-$sky_{\mathcal{H}}$. Since $h_i \notin b'\text{-}sky_{\mathcal{H}}$, there must exists another object, say h_j, that b'-dominates h_i. Thus, $\forall a_r \in A : bel(h_j.a_r \geq h_i.a_r) > b'$. But, $b < b'$. Therefore, $\forall a_r \in A : bel(h_j.a_r \geq h_i.a_r) > b$. Hence, $h_j \succ_b h_i$, which leads to a contradiction as $h_i \in b\text{-}sky_{\mathcal{H}}$.

4.1 Experiments

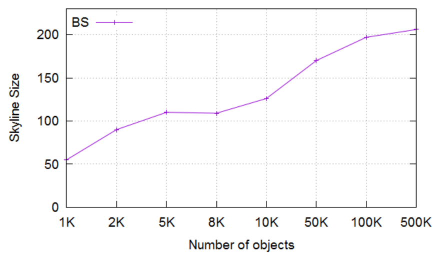
We present now an extensive experimental evaluation of our approach. More specifically, we focus on two issues: (i) the size of the evidential skyline in the context of *Tripadvisor* data; and (ii) the scalability of our proposed properties for the Belief Skyline algorithm denoted by BS. We also implemented, for comparison purposes, a basic algorithm denoted by BBS (baseline belief Skyline). This later is the basic version of the BS algorithm, i.e., it does not use the properties presented in section 4. The generation of the sets of evidential data is controlled by the parameters in Table 6, which lists parameters under investigation, their examined and default values. In each experimental setup, we investigate the effect of one parameter, while we set the remaining ones to their default values. The data generator and the algorithms, i.e., BS and BBS were implemented in Java, and all experiments were conducted on a 2.3 GHz Intel Core i7 processor, with 8GB of RAM.

Table 6. Parameters and Examined Values

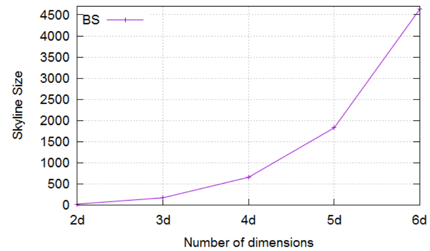
Parameter	Symbol	Values	Default
Number of objects	n	1K, 2K, 5K, 8K, 10K, 50K, 100K, 500K	10K
Number of attributes	d	2, 3, 4, 5, 6	4
belief threshold	b	0.01, 0.1, 0.3, 0.5, 0.7, 0.9	0.5
Number of focal elements	f	2, 3, 5, 7, 8, 9, 10	4

Figure 3 shows the size (i.e., the number of objects returned) of the b-skyline w.r.t. n, d, b and f. Fig. 3(a) shows that the size on the evidential skyline increases with higher n since when n increases more objects have chances not to be dominated. As shown in figure 3(b) the cardinality of the evidential skyline increases significantly with the increase of d. In fact, with the increase of d an object has better opportunity to be not dominated in all attributes. Figure 3(c) shows that the size of the evidential skyline increases with the increase of the b since the b-skyline contains the b'-dominant skyline if $b > b'$; see Property 4.

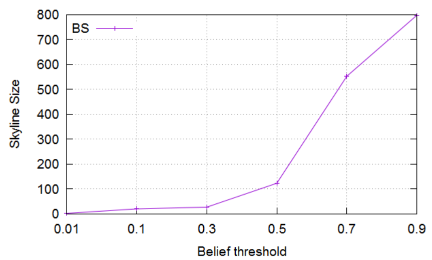
Figure 4 depicts the execution time of the implemented algorithms with regard to n, d, b and f. Overall, BS outperforms BBS. More specifically, BS is faster than BBS thanks to the properties used to improve our algorithm. As expected, figure 4(a) shows that the performance of the algorithms deteriorates with the increase of n. Observe that BS is one order of magnitude faster BBS since it can quickly identify if an object is dominated or not. As shown in figure 4(b) BBS does not scale with d. In fact, when d increases the size of the evidential skyline becomes larger. Hence, BBS performs a large number of dominance checks with a basic function. As shown in figure 4(c), BS is also affected by b. Figure 4(d) shows that BS is more than one order of magnitude faster than BBS.



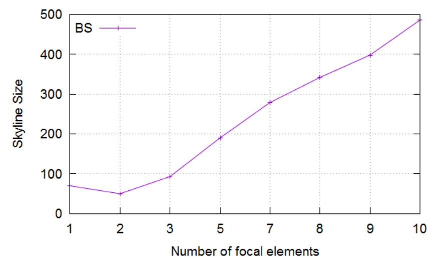
(a) Effect of n



(b) Effect of d

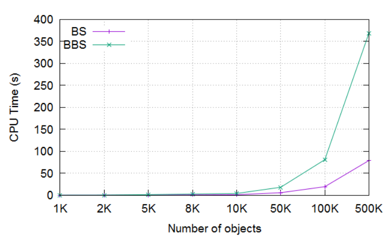


(c) Effect of b

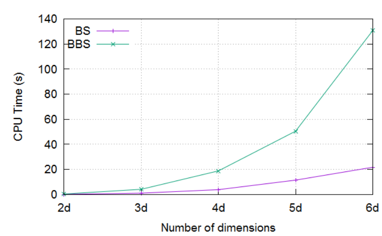


(d) Effect of f

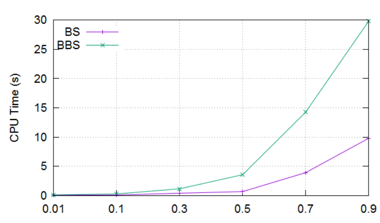
Fig. 3. Skyline Size



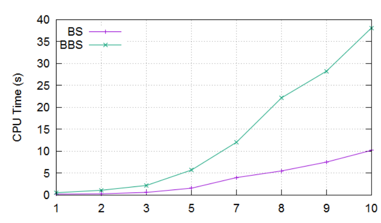
(a) Effect of n



(b) Effect of d



(c) Effect of b



(d) Effect of f

Fig. 4. Skyline Performance

5 Conclusion and Future Works

In this paper, the *Tripadvisor* reviewers' feedbacks about hotels are treated. First, reviews are modeled as basic belief assignments. Then, the belief functions tools are used to discount and combine reviews considering the travelers' reliabilities. Since the *Tripadvisor* do not offer the multi-criteria filtering, we proposed a new Evidential Skyline operator that deals with particular type of data. The proposed Skyline operator represent an optimization of the classic skyline [7]. Finally, our skyline method is evaluated on synthetic data whose properties are similar to *Tripadvisor* ones. Experimental results are very interesting in comparison with the classic skyline method. It showed a clear optimization in terms of performance and skyline size.

Combining feedbacks when a traveler gives more than one review about a specific hotel is a promising perspective, especially that the theory of belief functions offers several combination rules for different use cases.

References

1. D. A. Bell, J. W. Guan, and S. K. Lee. Generalized union and project operations for pooling uncertain and imprecise information. *Data & Knowledge Engineering, DKE*, 18:89–117, 1996.
2. S. Borzsony, D. Kossmann, and K. Stocker. The skyline operator. In *17th International Conference on Data Engineering, ICDE*, pages 421–430, 2001.
3. F. E. Bousnina, M. A. Bach Tobji, M. Chebbah, L. Liétard, and B. Ben Yaghlane. A new formalism for evidential databases. In *22nd International Symposium on Foundations of Intelligent Systems, ISMIS, Lyon, France, October 21-23*, pages 31–40, 2015.
4. F. E. Bousnina, S. Elmi, M. A. Bach Tobji, M. Chebbah, A. HadjAli, and B. Ben Yaghlane. Object-relational implementation of evidential databases. In *International Conference on Digital Economy, ICDEc, Carthage, Tunisia, April 28-30*, pages 80–87, 2016.
5. A. P. Dempster. Upper and lower probabilities induced by a multiple valued mapping. *The Annals of Mathematical Statistics*, 38(2):325–339, 1967.
6. A. P. Dempster. A generalization of bayesian inference. *Journal of the Royal Statistical Society, Series B*, 30:205–247, 1968.
7. S. Elmi, K. Benouaret, A. HadjAli, M. A. Bach Tobji, and B. Ben Yaghlane. Computing skyline from evidential data. In *8th International Conference on Scalable Uncertainty Management, SUM, Oxford, UK, September 15-17*, pages 148–161, 2014.
8. S. K. Lee. An extended relational database model for uncertain and imprecise information. In *8th Conference on Very Large Data Bases, VLDB, Canada*, pages 211–220, 1992.
9. S. K. Lee. Imprecise and uncertain information in databases : an evidential approach. In *8th International Conference on Data Engineering, ICDE*, pages 614–621, 1992.
10. E. Lefevre, O. Colot, and P. Vannoorenberghe. Belief function combination and conflict management. *Information Fusion*, 3(2):149–162, 2002.
11. G. Shafer. A mathematical theory of evidence. *Princeton University Press*, 1976.