LARODEC-ISG, Tunis (Tunisia) / LIAS-ENSMA, Poitiers (France)

Advanced Models for Graph Data Exploitation and Analysis

PhD. Student: Amna ABIDI

LIAS, IDD Seminar, 18th May., 2017

Under the guidance of

LARODEC, ISG, Tunis

- Mohamed Anis BACH TOBJI, Univ. Manouba, ESEN. Tunisia.
- Boutheina BEN YAGHLANE, University of Carthage, IHEC, Tunisia.

LIAS, ENSMA, Poitiers

- Allel HADJALI, University of Poitiers, ENSMA, France.

Outline

Due to the openness of the web and variety of sources in internet, the reliability of collected data is questioned.

Introduction

- Several researchers enriched the basic RDF data model with trust information (Hartig, 2009; Tamaszuk et al., 2012; Fionda and Greco, 2015).
- To reason in presence of trust information, we need new methods to query RDF data.
- Skyline operator is the most used preference queries when data are perceived with uncertainty.

Introduction

- Several researchers enriched the basic RDF data model with trust information (Hartig, 2009; Tamaszuk et al., 2012; Fionda and Greco, 2015).
- To reason in presence of trust information, we need new methods to query RDF data.
- Skyline operator is the most used preference queries when data are perceived with uncertainty.

Introduction

- Several researchers enriched the basic RDF data model with trust information (Hartig, 2009; Tamaszuk et al., 2012; Fionda and Greco, 2015).
- To reason in presence of trust information, we need new methods to query RDF data.
- Skyline operator is the most used preference queries when data are perceived with uncertainty.

- RDF is a W3C framework to represent information in the Web in a meaningful (semantic) way.
- An RDF statement is a triple $< subject, predicate, object >$ or $< s, p, o >$.

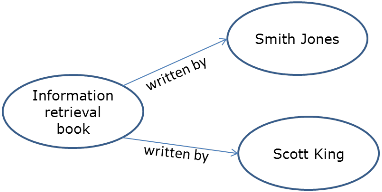

Figure: RDF Graph example.

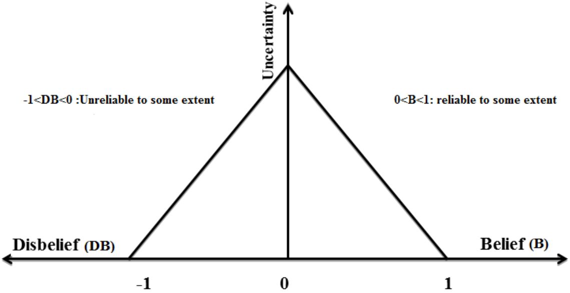

Figure: Meaning of trust values (inspired from Hartig. O, 2009).

RDF SPOT

An RDF SPOT X is a quadruple $< s, p, o, t >$, where o is a value of a predicate p related to a subject s, with a trust t. The triple $< s, p, o >$ is denoted by X^*.

The Skyline operator

Pareto Dominance: A key notion

Let P and Q be two points in a set of points denoted O with n attributes. A point Q dominates a point P denoted by $Q \succ P$, if $\forall i \in$ [1, n] $q_i \leq p_i \wedge \exists j, q_j < p_j$.

$$Q \succ P = \bigwedge (\bigwedge_{1 \leq i \leq n} q_i \leq p_i, \bigvee_{1 \leq i \leq n} q_i < p_i)$$

The Skyline preference relation

Skyline operator

The skyline is the set of points that are dominated by no other points (Börzsönyi, 2001).

Skyline operator

Let O be a set of points having n attributes. The skyline of O denoted by S is defined as:

$$S = \{P \in O / \nexists Q \in O, Q \succ P\}$$

The Skyline preference relation

Skyline operator

The skyline is the set of points that are dominated by no other points (Börzsönyi, 2001).

Skyline operator

Let O be a set of points having n attributes. The skyline of O denoted by S is defined as:

$$S = \{P \in O / \nexists Q \in O, Q \succ P\}$$

Table: Example: hotel properties

	Price	Distance
h_1	20	100
h_2	30	110
h_2	20	100
h_2	10	110
h_2	40	120

- $h_1 \succ h_2 = 1 \Rightarrow h_2$ is pruned.
- $h_1 \succ h_3 = 0$
- Skyline: $\mathbf{S} = \{h_1, h_3, h_4\}$

Trust Dominance

Trust dominance degree

Let P and Q be two subjects having n properties p_i and q_i, respectively with $1 \leq i \leq n$. The degree of dominance between P and Q, denoted by $d(Q \succ P)$.

$$d(Q \succ P) = min(\min_{1 \leq i \leq n} Trust(q_i \leq p_i), \max_{1 \leq i \leq n} Trust(q_i < p_i))$$

Table: Example of hotels properties.

Hotels	case 1		case 2		case 3		case 4	
	price	distance	price	distance	price	distance	price	distance
h_1	20(0.2)	100(0.4)	20(0.6)	80(0.7)	20(0.3)	100(0.5)	20(0.3)	70(0.5)
h_2	30(0.3)	110(0.5)	25(0.3)	70(0.1)	20(0.4)	100(0.6)	25(0.4)	70(0.5)

We proceed on computing the Trust-Skyline over those cases:

- case 1: $d(h_1 \succ h_2)$= min(min(0.2,0.4), max(0.2,0.4))=0.2
- case 2: $d(h_1 \succ h_2)$= min(min(0.3,-1), max(0.3,-1))=-1
- case 3: $d(h_1 \succ h_2)$= min(min(0.3,0.5), max(-1,-1))=-1
- case 4: $d(h_1 \succ h_2)$= min(min(0.3,0.5), max(0.3,-1))=0.3

Table: Example of hotels properties.

Hotels	case 1		case 2		case 3		case 4	
	price	distance	price	distance	price	distance	price	distance
h_1	20(0.2)	100(0.4)	20(0.6)	80(0.7)	20(0.3)	100(0.5)	20(0.3)	70(0.5)
h_2	30(0.3)	110(0.5)	25(0.3)	70(0.1)	20(0.4)	100(0.6)	25(0.4)	70(0.5)

We proceed on computing the Trust-Skyline over those cases:

- case 1: $d(h_1 \succ h_2)$= min(min(0.2,0.4), max(0.2,0.4))=0.2
- case 2: $d(h_1 \succ h_2)$= min(min(0.3,-1), max(0.3,-1))=-1
- case 3: $d(h_1 \succ h_2)$= min(min(0.3,0.5), max(-1,-1))=-1
- case 4: $d(h_1 \succ h_2)$= min(min(0.3,0.5), max(0.3,-1))=0.3

Trust Dominance

Trust of a point

Given an RDF point P with n properties p_i such that $1 \leq i \leq n$. Each property is associated with a trust value t_i. The trust of a point, denoted by $P.t^-$ is the minimum trust degree among all its properties.

$$P.t^- = \min_{1 \leq i \leq n} (p_i.t)$$

Dominance degree

Given two points P and Q having the trusts $Q.t^-$ and $P.t^-$.

$$d(Q \succ P) = \begin{cases} min(Q.t^-, P.t^-) & if\ Q* \succ P* \\ -1 & else \end{cases}$$

Dominance degree

Dominance degree properties

The trust dominance is transitive. Given two RDF triples P and Q, and a threshold $\alpha \in [-1, 1]$

$$\text{if } d(R \succ Q) > \alpha \text{ and } d(Q \succ P) > \alpha; \longrightarrow d(R \succ P) > \alpha$$

Dominance degree properties

The trust dominance is asymmetric. Given two RDF triples P and Q, and a threshold $\alpha \in [-1, 1]$
$d(Q \succ P) > \alpha$ Then $d(P \succ Q) = -1$

Dominance degree

Dominance degree properties

The trust dominance is transitive. Given two RDF triples P and Q, and a threshold $\alpha \in [-1, 1]$

$$\text{if } d(R \succ Q) > \alpha \text{ and } d(Q \succ P) > \alpha; \longrightarrow d(R \succ P) > \alpha$$

Dominance degree properties

The trust dominance is asymmetric. Given two RDF triples P and Q, and a threshold $\alpha \in [-1, 1]$
$d(Q \succ P) > \alpha$ Then $d(P \succ Q) = -1$

Trust-Skyline

Definition

The T-Skyline of a data set D, denoted by $T - Sky^{\alpha}$, contains each point P in D such there is no point Q that dominates P with a trust degree greater than a user defined threshold $\alpha \in [-1, 1]$.

$$T - sky^{\alpha} = \{P \in D / \nexists Q \in D, d(Q \succ P) \geq \alpha\}$$

Example

Table: Example of hotels candidate list of T-Sky, α=0.1

Hotel	Price	Distance
h_1	23 (0.5)	5 (0.3)
h_2	50 (0.2)	4 (0.6)
h_3	50 (0.7)	3 (0.5)
h_4	40 (0.1)	1 (0.3)
h_5	50 (0.6)	2 (0.4)

- $d(h_1 \succ h_2) = 0.2 \geq \alpha$, h_2 is pruned.
- $d(h_1 \succ h_3) = 0.3$, thus h_3 is also pruned.
- $d(h_1 \succ h_4) = -1$ and $d(h_4 \succ h_1) = -1$, no pruning.
- $d(h_1 \succ h_5) = 0.3$. h_5 is pruned $\Rightarrow$ $T - Sky^{\alpha}$=$\{h_1, h_4\}$.

TRDF-Skyline Algorithm

- SQL query
- Naive T-Skyline algorithm: optimization using α measure.
- TRDF-Skyline algorithm: a second optimization based on the transitivity property.

Conference Article

Skyline Modeling and Computing over Trust RDF Data, Proc. of the 19th International Conference on Enterprise Information Systems (ICEIS'2017), 26-29 April, 2017, Porto, Portugal. Best Paper Award, Area: Software Agents and internet Computing.

Possibility theory

- A possibility distribution is a function π: $X \rightarrow [0, 1]$ and $\pi(a)$ expresses the degree to which a is a possible value for the considered variable.
- The normalization condition imposes that at least one of the values of the domain (a_0) is completely possible, i.e., $\pi(a_0) = 1$.

- $\Pi(E_1 \cup E_2) = \max(\Pi(E_1), \Pi(E_2))$
- $\Pi(E_1 \cap E_2) = \min(\Pi(E_1), \Pi(E_2))$ if E_1 and E_2 are logically independent.

Possibility theory

- The sum of the degrees from a possibilistic distribution is different than 1 makes dealing with incompletely known distributions possible.
- Possibility theory constitutes an alternative to capture different kind of uncertainty of a qualitative nature.

Possibilistic RDF database

- A possibilistic RDF database D is a set of possibilistic triples. Each triple t is associated with a possibility value $P(t)$ indicating its ability to occur.
- In the possibilistic distribution we extend the RDF triple $< S, P, O >$ to a quadruple $< S, P, O, P_i >$ where O is a value of a predicate P related to a subject S, with a possibility measure P_i.

Possibilistic RDF Graph Data

A possibilistic RDF graph data $\tilde{G}^P$=(V, E, P) is a graph represented by the triple $(V(G), E(G), P(G))$, where:

- $V(G)$ represents a finite set of vertices,
- $E(G)$ is a finite set of edges e_{ij},
- $P(G)$ is the possibility associated to each triple of G.

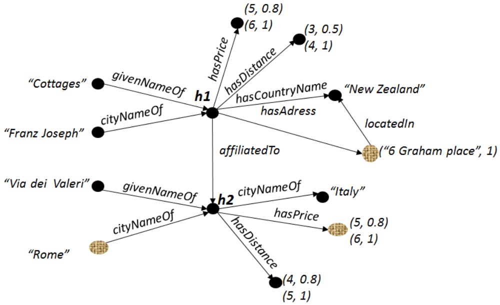

Figure: Graph representation of uncertain RDF data.

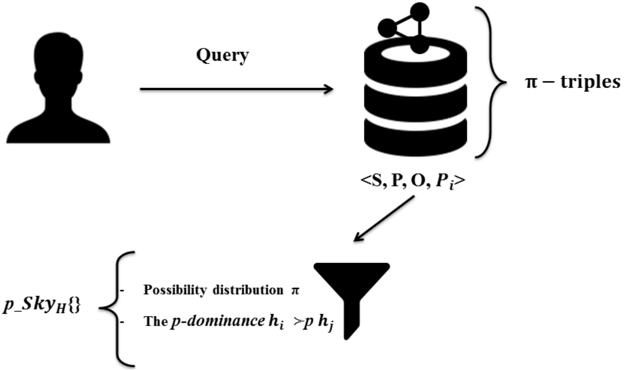

Figure: Skyline over Possibilistic RDF data.

Conclusion and perspectives

- We have extended Skyline operator over trust weighted RDF data: Trust-skyline model.
- We investigated the extension of possibility theory over uncertain RDF data (redefinition of dominance relationship).

Conclusion and perspectives

- The RDF query language SPARQL is of declarative nature, we need to extend SPARQL into a possibility-aware query language.
- The subgraph solution are those having the most possibility to occur or to be part of the result.

Published articles/ Journal paper

[C1] Amna Abidi, Mohamed Anis Bach Tobji, Allel Hadjali , Boutheina Ben Yaghlane, *Skyline Modeling and Computing over Trust RDF Data*, Proc. of the 19th International Conference on Enterprise Information Systems (ICEIS'2017), 26-29 April, 2017, Porto, Portugal.

[C2] Amna Abidi, Nassim Barhri, Mohamed Anis Bach Tobji, Allel Hadjali , Boutheina Ben Yaghlane, *First steps towards an electronic meta-journal platform based on crowdsourcing*, Proc. of the 2nd International Conference on Digital Economy, (ICDEc'2017), Springer-LNBIP, 04-06 May, 2017, Sidi Bou Said, Tunisia.

[J1] Amna Abidi, Sayda Elmi, Mohamed Anis Bach Tobji, Allel Hadjali, Boutheina Ben Yaghlane, *"Possibilistic Skyline queries over RDF data"*, International Journal of Approximate Reasoning (IJAR), under evaluation.

Thank you for your attention