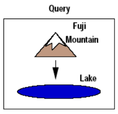
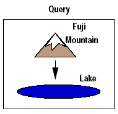
A multimedia query



select image P, *object* object1, *object* object2
where P *contains* object1
and P *contains* object2
and object1.semantical_property *s_like* "mountain"
and object1.image_property *image_match* "Fuji_mountain.gif"
and object2.semantical_property *is* "lake"
and object2.image_property *image_match* "lake_image_sample.gif"
and object1.position *is_above* object2.position

Maria Luisa Sapino (BDM 2018)

A multimedia query



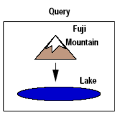
Crisp

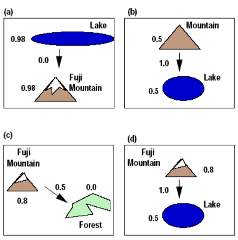
select image P, *object* object1, *object* object2
where P *contains* object1
and P *contains* object2
and object1.semantical_property *s_like* "mountain"
and object1.image_property *image_match* "Fuji_mountain.gif"
and object2.semantical_property *is* "lake"
and object2.image_property *image_match* "lake_image_sample.gif"
and object1.position *is_above* object2.position

Fuzzy
(imperfect)

Maria Luisa Sapino (BDM 2018)

Query…and results…





Maria Luisa Sapino (BDM 2018)

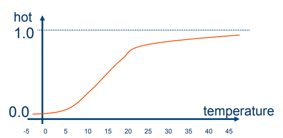
Reasons for imperfection

- Similarity between features (yellow/orange)
- Imperfections in the feature extraction algorithms
- Imperfections in the query formulation methods
- Partial match requirements
- Imperfections in the index structures and clustering algorithms

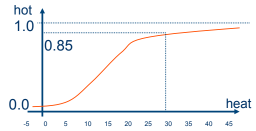
Fuzzy set..

- Fuzzy set *F* with domain *D* is defined using a membership function $$\mu_F : D \rightarrow [0, 1].$$
- A crisp (conventional) set *C* with domain *D* is defined using a membership function $$\mu_C : D \rightarrow \{0, 1\}.$$
- A fuzzy set corresponds to a fuzzy predicate

Example



Example



Hot(29°) = 0.85

Empty fuzzy set

$$\forall x \in X : f_{\phi}(x) = 0$$

Universal fuzzy set

$$\forall x \in X : f_{u}(x) = 1$$

α-Universal fuzzy set

$$\forall x \in X : f_{X^{[\alpha]}}(x) = \alpha$$

Maria Luisa Sapino (BDM 2018)

Support of a fuzzy set, F

$$\text{supp}(F) = def\{x \in X \mid f_F(x) > 0\}$$

Maria Luisa Sapino (BDM 2018)

Height of a fuzzy set, F

$$\text{height}(F) = \max_{x \in X}\{f_F(x)\}$$

Maria Luisa Sapino (BDM 2018)

Height of a fuzzy set, F

$$\text{height}(F) = \max_{x \in X}\{f_F(x)\}$$

height(F)=1; normal

Maria Luisa Sapino (BDM 2018)

Height of a fuzzy set, F

$$\text{height}(F) = \max_{x \in X}\{f_F(x)\}$$

height(F)<1; subnormal

height(F)=1; normal

Maria Luisa Sapino (BDM 2018)

Normalized fuzzy set, F*

$$F^* = F/\text{height}(F)$$

Maria Luisa Sapino (BDM 2018)

Normalized fuzzy set, F*

$$F^* = F/\text{height}(F)$$

$$\text{supp}(F^*) = \text{supp}(F)$$

Maria Luisa Sapino (BDM 2018)

Cardinality of F

$$card(F) = \sum_{x \in X} f_F(x)$$

$$card(F) = \int_{x \in X} f_F(x)dx$$

Maria Luisa Sapino (BDM 2018)

Cardinality of F

$$card(F) = \sum_{x \in X} f_F(x)$$

$$card(F) = \int_{x \in X} f_F(x)dx$$

Probabilistic databases:
cardinality is 1

Maria Luisa Sapino (BDM 2018)

α-cut a fuzzy set, F

$$F^{\geq\alpha} = \{x \in X \mid f_F(x) \geq \alpha\}$$

Maria Luisa Sapino (BDM 2018)

Strong α-cut a fuzzy set, F

$$F^{>\alpha} = \{x \in X \mid f_F(x) > \alpha\}$$

Maria Luisa Sapino (BDM 2018)

Kernel of F

$$\text{kernel}(F) = F^{\geq 1}$$

Maria Luisa Sapino (BDM 2018)

Set operations

$$A \subseteq B \leftrightarrow \forall x \quad f_A(x) \leq f_B(x)$$

Maria Luisa Sapino (BDM 2018)

Set operations

$$A \subseteq B \leftrightarrow \forall x \quad f_A(x) \leq f_B(x)$$

A understimates B

Maria Luisa Sapino (BDM 2018)

Set operations

$$A \subseteq B \leftrightarrow A^{>\alpha} \subseteq B^{>\alpha}$$

$$A \subseteq B \leftrightarrow A^{\geq\alpha} \subseteq B^{\geq\alpha}$$

Maria Luisa Sapino (BDM 2018)

Set operations

$$A \subseteq B \leftrightarrow \text{supp}(A) \subseteq \text{supp}(B)$$

$$A \subseteq B \leftrightarrow height(A) \leq height(B)$$

Example query

Fuzzy
(imperfect)

$$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{”}).$$

Example query

Fuzzy
(imperfect)

$$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{”}).$$

0.84 0.68

Example query

Fuzzy (imperfect)

$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{"}a.gif\text{"}).$

0. 76 0.84 0.68

Fuzzy logical operator

Example query

Fuzzy (imperfect)

$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{"}a.gif\text{"}).$

0. 76 0.84 0.68

Fuzzy logical operator

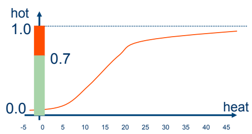
$$Q(Y_1, \ldots, Y_n) \leftarrow \Theta(p_1(Y_1, \ldots, Y_n), \ldots, p_m(Y_1, \ldots, Y_n)),$$

- Fuzzy and crisp predicates
- Fuzzy logical expression and a merge function
- Results is a ranked list (with the associated fuzzy values!)

How to process a fuzzy query?

$$Q(Y_1, \ldots, Y_n) \leftarrow \Theta(p_1(Y_1, \ldots, Y_n), \ldots, p_m(Y_1, \ldots, Y_n)),$$

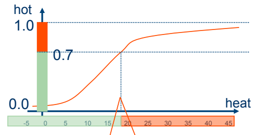
- First approach…make predicates crisp!!!
  - Use thresholds….



How to process a fuzzy query?

$$Q(Y_1, \ldots, Y_n) \leftarrow \Theta(p_1(Y_1, \ldots, Y_n), \ldots, p_m(Y_1, \ldots, Y_n)),$$

- First approach…make predicates crisp!!!
  - Use thresholds….



Good for processing in traditional databases

Not good for multimedia applications

How to process a fuzzy query?

$$Q(Y_1, \ldots, Y_n) \leftarrow \Theta(p_1(Y_1, \ldots, Y_n), \ldots, p_m(Y_1, \ldots, Y_n)),$$

- Second approach…use suitable fuzzy logic!!!
  - Merge (or scoring) functions….

$$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{”}).$$

0. 76 0.84 0.68

0.76 = $\mu_{\wedge}$(0.84,0.68)

Example merge functions…

Min semantics
$\mu_{P_i \wedge P_j}(x) = min\{\mu_i(x), \mu_j(x)\}$
$\mu_{P_i \vee P_j}(x) = max\{\mu_i(x), \mu_j(x)\}$
$\mu_{\neg P_i}(x) = 1 - \mu_i(x)$

Example merge functions…

Min semantics	*Product* semantics
$\mu_{P_i \wedge P_j}(x) = min\{\mu_i(x), \mu_j(x)\}$	$\mu_{P_i \wedge P_j}(x) = \frac{\mu_i(x) \times \mu_j(x)}{max\{\mu_i(x), \mu_j(x), \alpha\}} \quad \alpha \in [0,1]$
$\mu_{P_i \vee P_j}(x) = max\{\mu_i(x), \mu_j(x)\}$	$\mu_{P_i \vee P_j}(x) = \frac{\mu_i(x)+\mu_j(x)-\mu_i(x)\times\mu_j(x)-min\{\mu_i(x),\mu_j(x),1-\alpha\}}{max\{1-\mu_i(x),1-\mu_j(x),\alpha\}}$
$\mu_{\neg P_i}(x) = 1 - \mu_i(x)$	$\mu_{\neg P_i}(x) = 1 - \mu_i(x)$

Example merge functions…

Min semantics	*Product* semantics
$\mu_{P_i \wedge P_j}(x) = min\{\mu_i(x), \mu_j(x)\}$	$\mu_{P_i \wedge P_j}(x) = \frac{\mu_i(x) \times \mu_j(x)}{max\{\mu_i(x), \mu_j(x), \alpha\}} \quad \alpha \in [0,1]$
$\mu_{P_i \vee P_j}(x) = max\{\mu_i(x), \mu_j(x)\}$	$\mu_{P_i \vee P_j}(x) = \frac{\mu_i(x)+\mu_j(x)-\mu_i(x)\times\mu_j(x)-min\{\mu_i(x),\mu_j(x),1-\alpha\}}{max\{1-\mu_i(x),1-\mu_j(x),\alpha\}}$
$\mu_{\neg P_i}(x) = 1 - \mu_i(x)$	$\mu_{\neg P_i}(x) = 1 - \mu_i(x)$

Arithmetic average (N-ary)

$\mu_{P_i \wedge \ldots \wedge P_j}(x)$	$\mu_{\neg P_i}(x)$	$\mu_{P_i \vee \ldots \vee P_j}(x)$
$\frac{\mu_i(x)+\ldots+\mu_j(x)}{\lvert\{P_i,\ldots,P_j\}\rvert}$	$1 - \mu_i(x)$	$\frac{\lvert\{P_i,\ldots,P_j\}\rvert-\mu_i(x)+\ldots+\mu_j(x)}{\lvert\{P_i,\ldots,P_j\}\rvert}$

(mostly used in information retrieval)

Triangular norms (and co-norms)

- How to emulate the properties of a crisp predicate

	T-norm binary function N (for $\wedge$)	T-conorm binary function C (for $\vee$)
Boundary conditions	$N(0,0) = 0$, $N(x,1) = N(1,x) = x$	$C(1,1) = 1$, $C(x,0) = C(0,x) = x$

Triangular norms (and co-norms)

- How to emulate the properties of a crisp predicate

	T-norm binary function N (for $\wedge$)	T-conorm binary function C (for $\vee$)
Boundary conditions	$N(0,0)=0,\ N(x,1)=N(1,x)=x$	$C(1,1)=1,\ C(x,0)=C(0,x)=x$
Commutativity	$N(x,y)=N(y,x)$	$C(x,y)=C(y,x)$

Triangular norms (and co-norms)

- How to emulate the properties of a crisp predicate

	T-norm binary function N (for $\wedge$)	T-conorm binary function C (for $\vee$)
Boundary conditions	$N(0,0)=0,\ N(x,1)=N(1,x)=x$	$C(1,1)=1,\ C(x,0)=C(0,x)=x$
Commutativity	$N(x,y)=N(y,x)$	$C(x,y)=C(y,x)$
Monotonicity	$x \leq x', y \leq y' \rightarrow N(x,y) \leq N(x',y')$	$x \leq x', y \leq y' \rightarrow C(x,y) \leq C(x',y')$

Triangular norms (and co-norms)

- How to emulate the properties of a crisp predicate

	T-norm binary function N (for $\wedge$)	T-conorm binary function C (for $\vee$)
Boundary conditions	$N(0,0)=0,\ N(x,1)=N(1,x)=x$	$C(1,1)=1,\ C(x,0)=C(0,x)=x$
Commutativity	$N(x,y)=N(y,x)$	$C(x,y)=C(y,x)$
Monotonicity	$x \leq x', y \leq y' \rightarrow N(x,y) \leq N(x',y')$	$x \leq x', y \leq y' \rightarrow C(x,y) \leq C(x',y')$
Associativity	$N(x,N(y,z)) \leq N(N(x,y),z)$	$C(x,C(y,z)) \leq C(C(x,y),z)$

Triangular norms (and co-norms)

- How to emulate the properties of a crisp predicate

	T-norm binary function N (for $\wedge$)	T-conorm binary function C (for $\vee$)
Boundary conditions	$N(0,0)=0,\ N(x,1)=N(1,x)=x$	$C(1,1)=1,\ C(x,0)=C(0,x)=x$
Commutativity	$N(x,y)=N(y,x)$	$C(x,y)=C(y,x)$
Monotonicity	$x \leq x', y \leq y' \rightarrow N(x,y) \leq N(x',y')$	$x \leq x', y \leq y' \rightarrow C(x,y) \leq C(x',y')$
Associativity	$N(x,N(y,z)) \leq N(N(x,y),z)$	$C(x,C(y,z)) \leq C(C(x,y),z)$

- Bellman and Giertz: "The unique aggregation functions for evaluating AND and OR that preserve logical equivalence of queries involving only conjunction and disjunction and that are monotonic in their arguments are **min** and **max**."

Triangular norms (and co-norms)

- Emulating the properties of a crisp predicate may not be good for multimedia applications!!!
- Boundary conditions prevent partial matches

$$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{''}).$$

0. 00 0.99 0.00

$0.00 = \mu_{\wedge min}(0.99,0.00)$

Triangular norms (and co-norms)

- Emulating the properties of a crisp predicate may not be good for multimedia applications!!!
- Monotone condition is weak!!

$$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{''}).$$

0. 70 0.71 0.70

0. 70 0.99 0.70

$0.70 = \mu_{\wedge min}(0.71,0.70)$

$0.70 = \mu_{\wedge min}(0.99,0.70)$

Triangular norms (and co-norms)

- Emulating the properties of a crisp predicate may not be good for multimedia applications!!!
- N-ary semantics may be enough !!!
  - consider all relevant features at the same time, instead of in pairs!

Arithmetic average (N-ary)

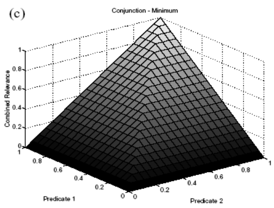
$\mu_{P_i \wedge \ldots \wedge P_j}(x)$	$\mu_{\neg P_i}(x)$	$\mu_{P_i \vee \ldots \vee P_j}(x)$
$\frac{\mu_i(x)+\ldots+\mu_j(x)}{\lvert\{P_i,\ldots,P_j\}\rvert}$	$1-\mu_i(x)$	$\frac{\lvert\{P_i,\ldots,P_j\}\rvert-\mu_i(x)+\ldots+\mu_j(x)}{\lvert\{P_i,\ldots,P_j\}\rvert}$

Geometric average (N-ary)

$\mu_{P_{i1} \wedge \ldots \wedge P_{in}}(x)$	$\mu_{\neg P_i}(x)$	$\mu_{P_{i1} \vee \ldots \vee P_{in}}(x)$
$(\mu_{i1}(x) \times \ldots \times \mu_{in}(x))^{\frac{1}{n}}$	$1-\mu_{i1}(x)$	$1-((1-\mu_{i1}(x)) \times \ldots \times (1-\mu_{in}(x)))^{\frac{1}{n}}$

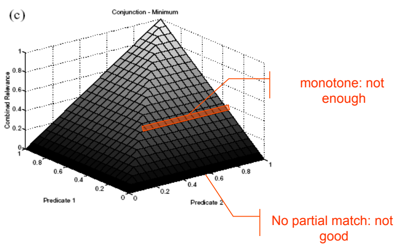
M 2018)

Visualisation of minimum



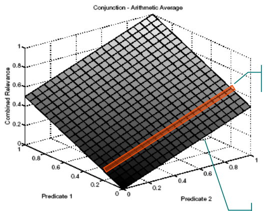
Maria Luisa Sapino (BDM 2018)

Visualisation of minimum



Maria Luisa Sapino (BDM 2018)

Visualisation of arithmetic average

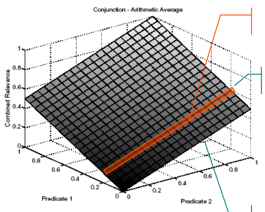


strictly monotone:better!!

partial matches: better!!

Maria Luisa Sapino (BDM 2018)

Visualisation of arithmetic average



constant slope: not enough

strictly monotone:better!!

partial matches: better!!

Maria Luisa Sapino (BDM 2018)

Interestingness…..

- An increase in the score of a subquery
  - with a lower score is more interesting than

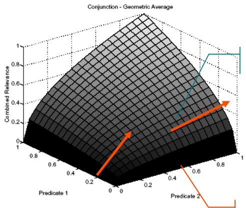
a similar increase in the score of a subquery
  - with a large score

$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{"}a.gif\text{"}).$

0. 40	0.00	0.80
0. ??	0.10	0.80
0. ??	0.00	0.90

Maria Luisa Sapino (BDM 2018)

Visualisation of geometric average



Maria Luisa Sapino (BDM 2018)

….parametric geometric average

$$\mu_{(P_{i1}\wedge\ldots\wedge P_{in})}(t, r_{true}, \beta) = \frac{((\prod_{\mu_k(t)\geq r_{true}} \mu_k(t)) \times (\prod_{\mu_k(t)<r_{true}} \beta))^{1/n} - \beta}{1-\beta}$$

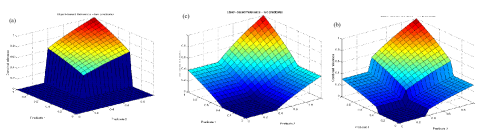
Truth cutoff

Falsehood value

Maria Luisa Sapino (BDM 2018)

….parametric geometric average

$$\mu_{(P_{i1}\wedge\ldots\wedge P_{in})}(t, r_{true}, \beta) = \frac{((\prod_{\mu_k(t)\geq r_{true}} \mu_k(t)) \times (\prod_{\mu_k(t)<r_{true}} \beta))^{1/n} - \beta}{1-\beta}$$



r_{true}=0.4; β=0.4 r_{true}=0.4; β=0.2 r_{true}=0.4; β=0.0

Maria Luisa Sapino (BDM 2018)

How to put weigths?????

- How do I state that image properties are more important than semantic properties??
- What do we mean?:
  - A change in the value of image property should have a larger impact than a similar change in the value of the semantic property.

$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{''}).$

$Q(X)$	s_like	$image\_match$
0. 60	0.60	0.60
0. 65	0.70	0.60
0. 68	0.60	0.70

Maria Luisa Sapino (BDM 2018)

How to put weigths?????

$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{''}).$

- How do I state that image properties are more important than semantic properties??

Maria Luisa Sapino (BDM 2018)

Fagin’s proposal

- Desiderata
  - If all weigths are equal the result should be equal to no weight case

Maria Luisa Sapino (BDM 2018)

Fagin's proposal

- Desiderata
  - If all weigths are equal the result should be equal to no weight case
  - If one of the weigths is zero, the subquery can be dropped without effecting the rest

Fagin's proposal

- Desirata
  - If all weigths are equal the result should be equal to no weight case
  - If one of the weigths is zero, the subquery can be dropped without effecting the rest
  - ..the result should be a continuous function of the weigths

Fagin's proposal

- Let $\theta_1 + \theta_2 + ... + \theta_m = 1$

  $\theta_1, \theta_2, ..., \theta_m \geq 0$

  $\theta_1 \geq \theta_2 \geq ... \geq \theta_m$

Fagin's proposal

- Let $\theta_1 + \theta_2 + \ldots + \theta_m = 1$
  $\theta_1, \theta_2, \ldots, \theta_m \geq 0$
  $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_m$
- then

$$
\begin{aligned}
f_{(\theta_1,\theta_2,\ldots,\theta_m)}(x_1,x_2,\ldots,x_m) = {} & (\theta_1-\theta_2) f(x_1) + \\
& 2(\theta_2-\theta_3) f(x_1,x_2) + \\
& 3(\theta_3-\theta_4) f(x_1,x_2,x_3) + \\
& \ldots\ldots\ldots \\
& (m-1)(\theta_{(m-1)}-\theta_m) f(x_1,x_2,x_3,\ldots,x_{(m-1)}) + \\
& m\theta_m f(x_1,x_2,x_3,\ldots,x_m)
\end{aligned}
$$

Fagin's proposal

- Let $\theta_1 + \theta_2 + \ldots + \theta_m = 1$
  $\theta_1, \theta_2, \ldots, \theta_m \geq 0$
  $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_m$
- then

$$
\begin{aligned}
f_{(\theta_1,\theta_2,\ldots,\theta_m)}(x_1,x_2,\ldots,x_m) = {} & (\theta_1-\theta_2) f(x_1) + \\
& 2(\theta_2-\theta_3) f(x_1,x_2) + \\
& 3(\theta_3-\theta_4) f(x_1,x_2,x_3) + \\
& \ldots\ldots\ldots \\
& (m-1)(\theta_{(m-1)}-\theta_m) f(x_1,x_2,x_3,\ldots,x_{(m-1)}) + \\
& m\theta_m f(x_1,x_2,x_3,\ldots,x_m)
\end{aligned}
$$

If *f* is continuous, then the weighted function is also continuous

Fagin's proposal

- Let $\theta_1 + \theta_2 + \ldots + \theta_m = 1$
  $\theta_1, \theta_2, \ldots, \theta_m \geq 0$
  $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_m$

- then

$$
\begin{aligned}
f_{(\theta_1,\theta_2,\ldots,\theta_m)}(x_1,x_2,\ldots,x_m) = {} & (\theta_1-\theta_2) f(x_1) + \\
& 2(\theta_2-\theta_3) f(x_1,x_2) + \\
& 3(\theta_3-\theta_4) f(x_1,x_2,x_3) + \\
& \ldots\ldots\ldots \\
& (m-1)\theta_{(m-1)} f(x_1,x_2,x_3,\ldots,x_{(m-1)})
\end{aligned}
$$

If lowest weigth is 0, then the corresponding sub query can be omitted

Fagin's proposal

- Let $\theta_1 + \theta_2 + \ldots + \theta_m = 1$
  $\theta_1, \theta_2, \ldots, \theta_m \geq 0$
  $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_m$
- then $f_{\left(\frac{1}{m},\frac{1}{m},\ldots,\frac{1}{m}\right)}(x_1, x_2, \ldots, x_m) = \left(\frac{1}{m} - \frac{1}{m}\right) f(x_1) + 2\left(\frac{1}{m} - \frac{1}{m}\right) f(x_1, x_2) +$

If all weigths are equal…

$$3\left(\frac{1}{m} - \frac{1}{m}\right) f(x_1, x_2, x_3) + \quad \ldots\ldots\ldots$$

$$(m-1)\left(\frac{1}{m} - \frac{1}{m}\right) f(x_1, x_2, x_3, \ldots, x_{(m-1)}) +$$

$$m\frac{1}{m} f(x_1, x_2, x_3, \ldots, x_m)$$

Fagin's proposal

- Let $\theta_1 + \theta_2 + \ldots + \theta_m = 1$
  $\theta_1, \theta_2, \ldots, \theta_m \geq 0$
  $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_m$
- then $f_{\left(\frac{1}{m},\frac{1}{m},\ldots,\frac{1}{m}\right)}(x_1, x_2, \ldots, x_m) = f(x_1, x_2, x_3, \ldots, x_m)$

If all weigths are equal then the result is equal to the no-weighted case

Example (arithmetic average)

$$score(a \wedge b) = \frac{score(a) + score(b)}{2}$$

Example (arithmetic average)

$$score(a \wedge b) = \frac{score(a) + score(b)}{2}$$

$$score(a \wedge b) = (\theta_a - \theta_b)score(a) + 2\theta_b \frac{score(a) + score(b)}{2}$$

Maria Luisa Sapino (BDM 2018)

Example (arithmetic average)

$$score(a \wedge b) = \frac{score(a) + score(b)}{2}$$

$$score(a \wedge b) = (\theta_a - \theta_b)score(a) + 2\theta_b \frac{score(a) + score(b)}{2}$$

$$score(a \wedge b) = \theta_a score(a) + \theta_b score(b)$$

Maria Luisa Sapino (BDM 2018)

Example (product)

$$score(a \wedge b) = score(a) \times score(b)$$

$$score(a \wedge b) = (\theta_a - \theta_b)score(a) + 2\theta_b score(a) \times score(b)$$

Maria Luisa Sapino (BDM 2018)

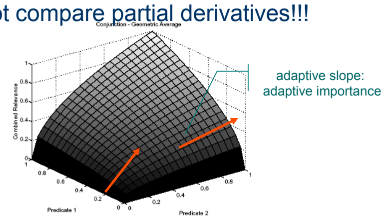
Are Fagin’s desiderata enough?

- It does not compare partial derivatives!!!

Maria Luisa Sapino (BDM 2018)

Are Fagin’s desiderata enough?

- It does not compare partial derivatives!!!



Maria Luisa Sapino (BDM 2018)

Are Fagin’s desiderata enough?

- It does not compare partial derivatives!!!
- Importance: Given a function *f(x,y), x* has a higher contribution than *y* iff

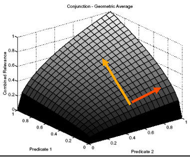
$$\forall a,b \quad \left.\frac{\partial f}{\partial x}\right|_{\langle a,b\rangle} > \left.\frac{\partial f}{\partial y}\right|_{\langle a,b\rangle}$$

Maria Luisa Sapino (BDM 2018)

Are Fagin's desiderata enough?

- It does not compare partial derivatives!!!
- Importance: Given a function *f(x,y), x* has a higher contribution than *y* iff

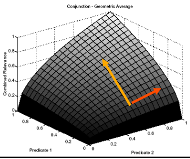
$$\forall a,b \ \left.\frac{\partial f}{\partial x}\right|_{\langle a,b\rangle} > \left.\frac{\partial f}{\partial y}\right|_{\langle a,b\rangle}$$



Are Fagin's desiderata enough?

- It does not compare partial derivatives!!!
- Importance: Given a function *f(x,y), x* has a higher contribution than *y* iff

$$\forall a,b \ \left.\frac{\partial f}{\partial x}\right|_{\langle a,b\rangle} > \left.\frac{\partial f}{\partial y}\right|_{\langle a,b\rangle}$$



$$\text{relimp}(x,y)\Big|_{\langle a,b\rangle} = \frac{\left.\frac{\partial f}{\partial x}\right|_{\langle a,b\rangle}}{\left.\frac{\partial f}{\partial y}\right|_{\langle a,b\rangle}}$$

Are Fagin's desiderata enough?

- Importance: Given a function *f(x,y), x* has a higher contribution than *y* iff $\forall a,b \ \left.\frac{\partial f}{\partial x}\right|_{\langle a,b\rangle} > \left.\frac{\partial f}{\partial y}\right|_{\langle a,b\rangle}$
- Example:

$$score(x \wedge y) = \theta_x score(x) + \theta_y score(y)$$

$$\left.\frac{\partial score(x \wedge y)}{\partial score(x)}\right|_{\langle a,b\rangle} = \theta_x\Big|_{\langle a,b\rangle} = \theta_x$$

OK!

$$\left.\frac{\partial score(x \wedge y)}{\partial score(y)}\right|_{\langle a,b\rangle} = \theta_y\Big|_{\langle a,b\rangle} = \theta_y$$

Are Fagin's desiderata enough?

- Importance: Given a function *f(x,y), x* has a higher contribution than *y* iff $\forall a,b \; \left.\frac{\partial f}{\partial x}\right|_{\langle a,b\rangle} > \left.\frac{\partial f}{\partial y}\right|_{\langle a,b\rangle}$
- Example:

$$score(x \wedge y) = (\theta_x - \theta_y)score(x) + 2\theta_y score(x) \times score(y)$$

$$\left.\frac{\partial score(x \wedge y)}{\partial score(x)}\right|_{\langle a,b\rangle} = (\theta_x - \theta_y) + 2\theta_y b$$

NOT OK!

$$\left.\frac{\partial score(x \wedge y)}{\partial score(y)}\right|_{\langle a,b\rangle} = 2\theta_y a$$

Maria Luisa Sapino (BDM 2018)

$$\left.relimp(x,y)\right|_{\langle a,b\rangle} = \frac{(\theta_x - \theta_y) + 2\theta_y b}{2\theta_y a}$$

Are Fagin's desiderata enough?

- Importance: Given a function *f(x,y), x* has a higher contribution than *y* iff $\forall a,b \; \left.\frac{\partial f}{\partial x}\right|_{\langle a,b\rangle} > \left.\frac{\partial f}{\partial y}\right|_{\langle a,b\rangle}$
- Example:

$$score(x \wedge y) = (\theta_x - \theta_y)score(x) + 2\theta_y score(x) \times score(y)$$

$$\left.\frac{\partial score(x \wedge y)}{\partial score(x)}\right|_{\langle a,b\rangle} = (\theta_x - \theta_y) + 2\theta_y b$$

NOT OK!

$$\left.\frac{\partial score(x \wedge y)}{\partial score(y)}\right|_{\langle a,b\rangle} = 2\theta_y a$$

Maria Luisa Sapino (BDM 2018)

$$\left.relimp(x,y)\right|_{\langle a,a\rangle} = \frac{(\theta_x - \theta_y) + 2\theta_y a}{2\theta_y a} = 1 + \frac{(\theta_x - \theta_y)}{2\theta_y a} > 1$$

Are Fagin's desiderata enough?

- Importance: Given a function *f(x,y), x* has a higher contribution than *y* iff $\forall a,b \; \left.\frac{\partial f}{\partial x}\right|_{\langle a,b\rangle} > \left.\frac{\partial f}{\partial y}\right|_{\langle a,b\rangle}$
- Example:

$$score(x \wedge y) = (\theta_x - \theta_y)score(x) + 2\theta_y score(x) \times score(y)$$

$$\left.\frac{\partial score(x \wedge y)}{\partial score(x)}\right|_{\langle a,b\rangle} = (\theta_x - \theta_y) + 2\theta_y b$$

NOT OK!

$$\left.\frac{\partial score(x \wedge y)}{\partial score(y)}\right|_{\langle a,b\rangle} = 2\theta_y a$$

Maria Luisa Sapino (BDM 2018)

$$\text{relimp}(x,y)\Big|_{\langle a,b\rangle} = \frac{\theta_x b}{\theta_y a}$$

How about?

- Importance: Given a function *f(x,y), x* has a higher contribution than *y* iff $\forall a,b \quad \frac{\partial f}{\partial x}\Big|_{\langle a,b\rangle} > \frac{\partial f}{\partial y}\Big|_{\langle a,b\rangle}$
- Example:

$$score(x \wedge y) = score(x)^{\theta_x} \times score(y)^{\theta_y}$$

$$\frac{\partial score(x \wedge y)}{\partial score(x)}\Big|_{\langle a,b\rangle} = \theta_x a^{\theta_x - 1} b^{\theta_y}$$

NOT OK!

$$\frac{\partial score(x \wedge y)}{\partial score(y)}\Big|_{\langle a,b\rangle} = \theta_y a^{\theta_x} b^{\theta_y - 1}$$

$$\text{relimp}(x,y)\Big|_{\langle a,a\rangle} = \frac{\theta_x a}{\theta_y a} = \frac{\theta_x}{\theta_y} > 1$$

How about?

- Importance: Given a function *f(x,y), x* has a higher contribution than *y* iff $\forall a,b \quad \frac{\partial f}{\partial x}\Big|_{\langle a,b\rangle} > \frac{\partial f}{\partial y}\Big|_{\langle a,b\rangle}$
- Example:

$$score(x \wedge y) = score(x)^{\theta_x} \times score(y)^{\theta_y}$$

$$\frac{\partial score(x \wedge y)}{\partial score(x)}\Big|_{\langle a,b\rangle} = \theta_x a^{\theta_x - 1} b^{\theta_y}$$

NOT OK!

$$\frac{\partial score(x \wedge y)}{\partial score(y)}\Big|_{\langle a,b\rangle} = \theta_y a^{\theta_x} b^{\theta_y - 1}$$

Ranking

$$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{''}).$$

Ranking

$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{''}).$

0.90	X2	0.85	X3
0.80	X5	0.80	X5
0.70	X6	0.75	X2
0.60	X4	0.74	X6
0.50	X1	0.74	X1
0.40	X3	0.70	X4

Ranking

$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{''}).$

??

0.90	X2	0.85	X3
0.80	X5	0.80	X5
0.70	X6	0.75	X2
0.60	X4	0.74	X6
0.50	X1	0.74	X1
0.40	X3	0.70	X4

Ranking (and first-k retrieval)

$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{''}).$

??

0.90	X2	0.85	X3
0.80	X5	0.80	X5
0.70	X6	0.75	X2
0.60	X4	0.74	X6
0.50	X1	0.74	X1
0.40	X3	0.70	X4

First solution…join based on X

$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{''}).$

??

0.90	X2	0.85	X3
0.80	X5	0.80	X5
0.70	X6	0.75	X2
0.60	X4	0.74	X6
0.50	X1	0.74	X1
0.40	X3	0.70	X4

⋈ X=X

- Join the two information sources based on X
- Sort all results based on the merged score
- Select the first k

First solution…join based on X

$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{''}).$

??

0.90	X2	0.85	X3
0.80	X5	0.80	X5
0.70	X6	0.75	X2
0.60	X4	0.74	X6
0.50	X1	0.74	X1
0.40	X3	0.70	X4

⋈ X=X

Need to access the entire database at least once!!!

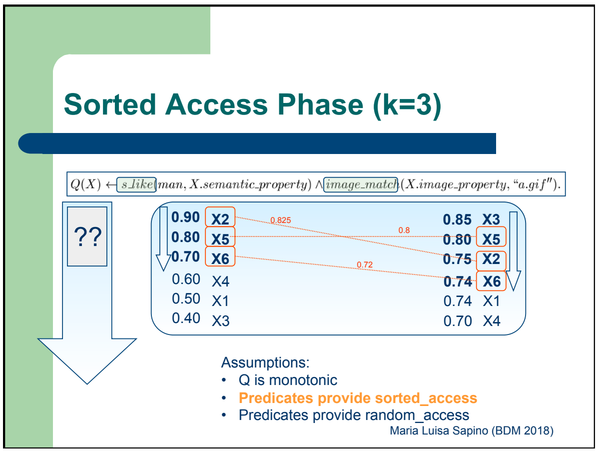
Ranked join for top-k retrieval (Fagin)

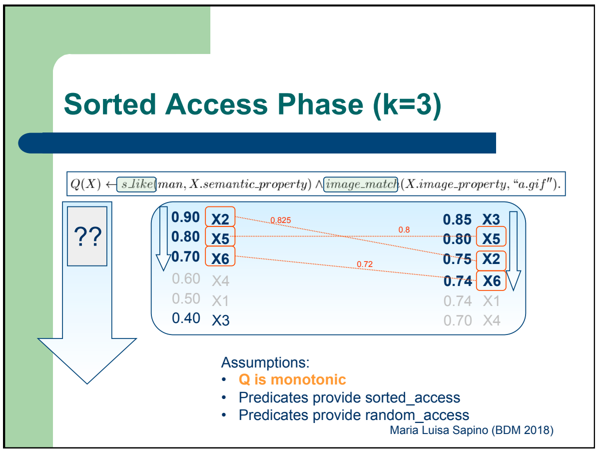
$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{''}).$

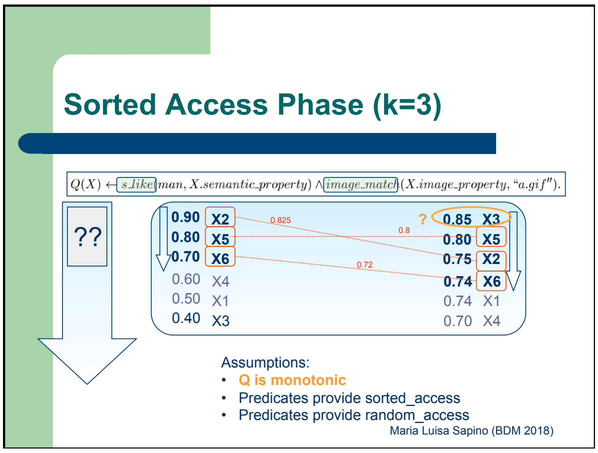
??

0.90	X2	0.85	X3
0.80	X5	0.80	X5
0.70	X6	0.75	X2
0.60	X4	0.74	X6
0.50	X1	0.74	X1
0.40	X3	0.70	X4

Assumptions:
- Q is monotonic
- Predicates provide sorted_access
- Predicates provide random_access

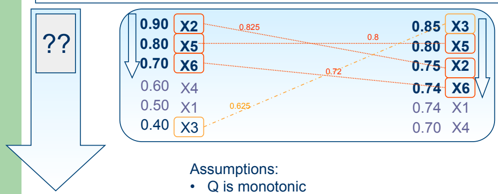

Sorted Access Phase (k=3)
$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{”})$.
??
0.90 X2
0.80 X5
0.70 X6
0.60 X4
0.50 X1
0.40 X3
0.825
0.8
0.72
0.85 X3
0.80 X5
0.75 X2
0.74 X6
0.74 X1
0.70 X4
Assumptions:
• Q is monotonic
• Predicates provide sorted_access
• Predicates provide random_access
Maria Luisa Sapino (BDM 2018)


Sorted Access Phase (k=3)
$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{”})$.
??
0.90 X2
0.80 X5
0.70 X6
0.60 X4
0.50 X1
0.40 X3
0.825
0.8
0.72
0.85 X3
0.80 X5
0.75 X2
0.74 X6
0.74 X1
0.70 X4
Assumptions:
• Q is monotonic
• Predicates provide sorted_access
• Predicates provide random_access
Maria Luisa Sapino (BDM 2018)


Sorted Access Phase (k=3)
$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{“}a.gif\text{”})$.
??
0.90 X2
0.80 X5
0.70 X6
0.60 X4
0.50 X1
0.40 X3
0.825
0.8
0.72
?
0.85 X3
0.80 X5
0.75 X2
0.74 X6
0.74 X1
0.70 X4
Assumptions:
• Q is monotonic
• Predicates provide sorted_access
• Predicates provide random_access
Maria Luisa Sapino (BDM 2018)

Random Access Phase (k=3)

$Q(X) \leftarrow \mathit{s\_like}(man, X.semantic\_property) \wedge \mathit{image\_match}(X.image\_property, \text{"a.gif"}).$

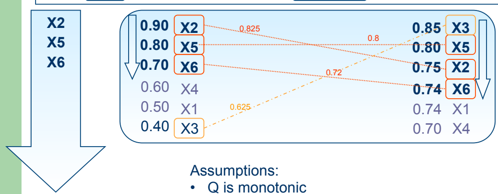


Assumptions:
- Q is monotonic
- Predicates provide sorted_access
- **Predicates provide random_access**

Maria Luisa Sapino (BDM 2018)

Result… (k=3)

$Q(X) \leftarrow \mathit{s\_like}(man, X.semantic\_property) \wedge \mathit{image\_match}(X.image\_property, \text{"a.gif"}).$

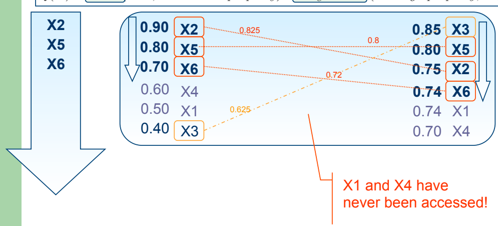


Assumptions:
- Q is monotonic
- Predicates provide sorted_access
- **Predicates provide random_access**

Maria Luisa Sapino (BDM 2018)

Advantage!!

$Q(X) \leftarrow \mathit{s\_like}(man, X.semantic\_property) \wedge \mathit{image\_match}(X.image\_property, \text{"a.gif"}).$



Maria Luisa Sapino (BDM 2018)

If the merge function is min… Use only one pred. for sorted access (k=3)

$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{"a.gif"})$.

??

0.90	**X2**	0.85	X3
0.80	X5	0.80	X5
0.70	X6	0.75	X2
0.60	X4	0.74	X6
0.50	X1	0.74	X1
0.40	X3	0.70	X4

Maria Luisa Sapino (BDM 2018)

Sorted+Random Access (k=3)

$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{"a.gif"})$.

??

0.90	**X2**	0.85	X3
0.80	**X5**	**0.80**	**X5**
0.70	**X6**	**0.75**	**X2**
0.60	X4	**0.74**	**X6**
0.50	X1	0.74	X1
0.40	X3	0.70	X4

0.75 0.8 0.7

- Stop when the next value is smaller than the third candidate

Maria Luisa Sapino (BDM 2018)

Sorted+Random Access (k=3)

$Q(X) \leftarrow s\_like(man, X.semantic\_property) \wedge image\_match(X.image\_property, \text{"a.gif"})$.

X2
X5
X6

0.90	**X2**	0.85	X3
0.80	**X5**	**0.80**	**X5**
0.70	**X6**	**0.75**	**X2**
0.60	X4	**0.74**	**X6**
0.50	X1	0.74	X1
0.40	X3	0.70	X4

0.75 0.8 0.7

X1, X3, and X4 have never been accessed!

Maria Luisa Sapino (BDM 2018)