Decathlon, Conflicting Objectives and User Preference Querying

Peter Vojtáš

Faculty of Mathematics and Physics, Charles University,
Malostranské náměstí 25, Prague, Czech Republic
Peter.Vojtas@mff.cuni.cz

Extended Abstract

First motivation of our approach is Decathlon as an athletic discipline and its development in the last century (`http://www.decathlon.ee`). During this period, motivated by a desire of a fair competition, a development in scoring tables occurred. Today we can say it is stabilized. It shows single disciplines ordered (and scored) in direction of better (harder) achievement. Comparison with single discipline world records shows that current decathlon world record holder (Roman Sebrle, CZ) was able to achieve about 65-92% of physical measurement and about 59-88% of point achievements of single discipline world records. Of course the point of decathlon is that all disciplines should be done by a single athlete in two consecutive days. Total achievement in decathlon is evaluated by the sum of point in single disciplines.

Another motivation comes from the paper IHAB F. ILYAS, GEORGE BESKALES and MOHAMED A. SOLIMAN. A Survey of Top-k Query Processing Techniques in Relational Database Systems, To appear in ACM Computing Surveys. A person looking for a house evaluates market offers by an aggregation of house price and tuition price in a school near to location. Nevertheless the price for house and tuition price cannot be simply added, the aggregation is, as in following SQL query

```
SELECT h.id, s.id
FROM House h; School s
WHERE h.location=s.location
ORDER BY h.price + 10 * s.tuition
LIMIT 5
```

Last motivation comes from multicriterial decision (Source [RTC] R.T.Clemen. Making hard decisions. Brooks/Cole Publ. Comp. 1996). Here, several examples are used, especially for conflicting objectives (like price and durability of a car). The solution is, that single objectives are represented by a objective function $Ui(xi)$ and aggregation (either linear or nonlinear) is here called utility function, gives a ordering of products (decisions) by total score equal to $U(x_1, \ldots, x_m) = k_1U_1(x_1) + \cdots + k_mU_m(x_m)$.

V. Snášel, K. Richta, J. Pokorný (Eds.): Dateso 2008, pp. 76–78, ISBN 978-80-248-1746-0.

Main goal of this talk is to point to similarities in all of these applications: IAAF – "combined events", Decision analysis and rank aware Querying.

Similarities can be characterized as follows: Incomparable disciplines (attributes) are mapped to points, score, ... - hence comparable values. Best, top k – preserves ordering if better in all axes (disciplines, attributes). There is some monotone aggregation, combination (e.g. (weighted) sum)

In what follows a class of aggregation functions is discussed - Sum and/or average, Weighted average, ... general. Common characteristics are: Monotone in attribute score, ideal point given by application, user dependent, implicit learning (in athletics took about 100 years), user similarity - collaborative. Our contribution reflects explicit learning, adaptation during the query cycles. A possibility was discussed: to fix aggregation and tune attribute scoring or attribute score fixed - tuning aggregation.

We have presented a unifying approach: (local) attribute preference represented by a scoring (fuzzy) function $f : DA \rightarrow [0, 1]$, Combination @ : $[0, 1]^n \rightarrow [0, 1]$, (global) score $score(o) = @(f_1(o.A_1), \ldots, f_m(o.A_m))$ We have a model-theoretic semantics based on – fuzzy logic, fixpoint semantics – fuzzy Data(/Pro)log Proof-theoretic semantics – best, top-k, heuristics.

We discussed also a form of data (in one table, several tables, locally or distributed, on the Web, frequently changing versus rather stable data, in relational, XML, HTML, text, ...). Further issues are preprocessing, indexes, Query optimization, Top-k versus table scan (experiments) Fagin instance optimal TA/NRA algorithm.

Different models were considered for User (One, many, different, ... User profile, Group decision, Changing intention during querying, Query formulation – clicking conjunctive query or sample evaluation).

We have a procedure how to learn users "Decathlon principle aggregation (combination). One for learning $score(o) = @(f_1(o.A_1), \ldots, f_m(o.A_m))$ either with Fixed @, tuning f_i (like IAAF @=+), or fixed f_i, learning @ - fuzzy ILP. Learning both -local preferences (user can have different order of preference, e.g. close-far) global preferences. For this new inductive task – ordinal classification we can either compare orderings or generalize precision, recall.

Special focus in different objectives for different user, e.g. (as a variation of above example)

```
SELECT h.id, s.id
FROM House h; School s
WHERE h.location=s.location
ORDER BY MAX h.price + 10 * s.tuition
LIMIT 5

SELECT h.id, s.id
FROM House h; School s
WHERE h.location=s.location
ORDER BY MIN h.price + 10 * s.tuition
LIMIT 5
```

```
SELECT h.id, s.id
FROM House h; School s
WHERE h.location=s.location
ORDER BY h.price + 60 * s.tuition
LIMIT 5

SELECT h.id, s.id
FROM House h; School s
WHERE h.location=s.location
ORDER BY @(f(h.price), g(s.tuition))
LIMIT 5
```

We have concluded with open problems and future work. Unified framework does not cover nominal data, multidimensional data, hierarchical data, ...