

Efficient Evaluation of Numerical Preferences: Top k Queries, Skylines and Multi-objective Retrieval

Wolf-Tilo Balke
EECS, University of California,
Berkeley, USA
balke@eecs.berkeley.edu

Abstract

Query processing in databases and information systems has developed beyond mere SQL-style exact matching of attribute values. Scoring database objects according to numerical user preferences and retrieving only the top k matches or Pareto-optimal result sets (skyline queries) are already common for a variety of applications.

Recently a lot of database literature has focussed on how to efficiently evaluate queries based on numerical preferences. Specialized algorithms using either top k retrieval (assuming a single compensation function defined over all query predicates, i.e. a global utility function) or computing skylines (assuming all query predicates as pairwise incomparable) have been shown to be capable of avoiding naïve linear database scans by pruning large numbers of database objects and thus vastly improve scalability. However, both paradigms are only two extreme cases of exploring viable compromises for each user's objectives, which may or may not be comparable.

To find the correct result set for arbitrary cases of multi-objective query processing in databases a novel algorithm for computing sets of objects that are non-dominated with respect to a set of monotonic objective functions representing a user's notion of utility, has recently been presented. Naturally containing top k and skyline retrieval paradigms as special cases, this algorithm maintains scalability also for all cases in between. To be more precise, in both special cases the multi-objective retrieval algorithm will behave exactly like the most efficient known evaluation algorithms for top k and skyline queries respectively. This algorithm has also been proved to be correct and instance-optimal in terms of necessary object accesses. Moreover, it improves the psychological response behaviour by progressively producing result objects as quickly as possible, while the algorithm is still running, so user can deal with result objects at the earliest point in time.

Our tutorial will discuss all state of the art algorithms for top k retrieval, skyline queries and multi-objective retrieval and point to open problems, future extensions of the paradigm and research in numerical preferences.

(Dagstuhl Seminar Series Nr 04271. 27.06.-02.07.04

*G. Bosi, R. Brafman, J. Chomicki, W. Kießling:
Preferences: Specification, Inference, Applications*