Chapter 6

A Tour of Lattice–Based Skyline Algorithms

Markus Endres
University of Augsburg, Germany

Lena Rudenko
University of Augsburg, Germany

ABSTRACT

A skyline query retrieves all objects in a dataset that are not dominated by other objects according to some given criteria. There exist many skyline algorithms which can be classified into generic, index-based, and lattice-based algorithms. This chapter takes a tour through lattice-based skyline algorithms. It summarizes the basic concepts and properties, presents high-performance parallel approaches, shows how one overcomes the low-cardinality restriction of lattice structures, and finally presents an application on data streams for real-time skyline computation. Experimental results on synthetic and real datasets show that lattice-based algorithms outperform state-of-the-art skyline techniques, and additionally have a linear runtime complexity.

INTRODUCTION

The Skyline operator (Börzsönyi, Kossmann, & Stocker, 2001) has emerged as an important and very popular summarization technique for multi-dimensional datasets. A Skyline query, also known as Pareto preference query, selects those objects from a dataset $D$ that are not dominated by any others. An object $p$ having $d$ attributes (dimensions) dominates an object $q$, if $p$ is better than $q$ in at least one dimension and not worse than $q$ in all other dimensions, for a defined comparison function. This dominance criterion defines a partial order and therefore transitivity holds. The Skyline is the set of points which are not dominated by any other point of $D$. Without loss of generality, the Skyline with the $\text{min}$ function for all attributes is used in this chapter.

The most cited example on Skyline queries is the search for a hotel that is cheap and close to the beach (Börzsönyi, Kossmann, & Stocker, 2001). Unfortunately, these two goals are complementary, as the hotels near the beach tend to be more expensive. 

DOI: 10.4018/978-1-5225-5396-0.ch006
the two-dimensional space of price and distance to the beach. Interesting are all hotels that are not worse than any other hotel in both dimensions.

The hotels \( p_5, p_6, p_8 \) are dominated by hotel \( p_3 \). The hotel \( p_7 \) is dominated by \( p_6 \), while the hotels \( p_1, p_7, p_8, p_5 \) are not dominated by any other hotels and build the Skyline. From the Skyline, one can now make the final decision, thereby weighing the personal preferences for price and distance to the beach.

Algorithms of the block-nested-loop class (BNL) (Börzsönyi, Kossmann, & Stocker, 2001) are the most prominent algorithms for computing Skylines. In fact, the basic operation of collecting maxima during a single scan of the input data can be found at the core of several Skyline algorithms, cp. (Godfrey, Shipley, & Gryz, 2007; Chomicki, Ciaccia, & Meneghetti, 2013). There are also algorithms utilizing an index structure to compute the Skyline, e.g., (Tan, Eng, & Ooi, 2001; Papadias, Tao, Fu, & Seeger, 2003; Lee, Zheng, Li, & Lee, 2007; Endres & Weichmann, 2017). However, index-based algorithms are not capable of processing arbitrary data without any preparations.

This chapter takes a tour of lattice-based Skyline algorithms, which goes from sequential to parallel approaches to an application in data streams. Lattice Skyline algorithms do not depend on tuple comparisons as BNL does, but on the lattice structure constructed by a Skyline query over low-cardinality domains. Following (Morse, Patel, & Jagadish, 2007; Chomicki, Ciaccia, & Meneghetti, 2013) many Skyline applications involve domains with small cardinalities – these cardinalities are either inherently small (such as star ratings for hotels), or can naturally be mapped to low-cardinality domains (such as price ranges on hotels). However, since there are still many applications where high-cardinality domains are involved, the Scalagon algorithm which overcomes the annoying low-cardinality restriction is also presented.

The remainder of this chapter is organized as follows: Section 2 revisits the concept of Skyline computation. Afterwards the basic sequential lattice-based algorithms Hexagon and LS-B are discussed in Section 3. Based on this background the high-performance parallel versions of these algorithms, namely ARL-S and HPL-S, are presented in Section 4. Section 5 describes Scalagon, a lattice-based

Figure 1. Skyline example
Related Content

A General Knowledge Representation Model for the Acquisition of Skills and Concepts
[www.igi-global.com/article/general-knowledge-representation-model-acquisition/46143?camid=4v1a](www.igi-global.com/article/general-knowledge-representation-model-acquisition/46143?camid=4v1a)

A Machine Learning Based Meta-Scheduler for Multi-Core Processors
[www.igi-global.com/chapter/machine-learning-based-meta-scheduler/56161?camid=4v1a](www.igi-global.com/chapter/machine-learning-based-meta-scheduler/56161?camid=4v1a)

An Action Guided Constraint Satisfaction Technique for Planning Problem
[www.igi-global.com/article/an-action-guided-constraint-satisfaction-technique-for-planning-problem/172126?camid=4v1a](www.igi-global.com/article/an-action-guided-constraint-satisfaction-technique-for-planning-problem/172126?camid=4v1a)

Multi-Fractal Analysis for Feature Extraction from DNA Sequences
[www.igi-global.com/article/multi-fractal-analysis-feature-extraction/43895?camid=4v1a](www.igi-global.com/article/multi-fractal-analysis-feature-extraction/43895?camid=4v1a)