Skyline with Presorting
Jan Chomicki, Parke Godfrey, Jarek Gryz and Dongming Liang

Abhishek Srivastava

Student ID: 861307778

The problem:

The paper discusses the importance of Block-nested Loops approach to find the skyline and the improvement can be applied to it so that its performance can be increased.

The contribution:

The authors propose *Sort-Filter-Skyline* Algorithm, an improved version of block-nested loop algorithm to find skyline values with better efficiency. They describe the problems with originally proposed techniques that its performance was very dependent on method used to discard the compared tuples either from window or query page table. It was also dependent on the size of the buffer because with larger buffer this approach becomes CPU bound instead of I/O bound. Authors also suggest that results are not guaranteed after single pass.

The method:

The Authors approach to algorithm presented was to limit the existing problems from the BNL technique. Author suggested that presorting the table records in topological manner such that it takes consideration of skyline criteria as well. It improves the existing discarding method. Compared tuple can be added or discarded and since the table is sorted decisions can be taken to further compare or not.

Authors presented following advantages of SFS over BNL:

- SFS uses good optimization over BNL.

- SFS behaves better than BNL in relational engine environment because it works in optimal number of pass and is not CPU-bound as well where BNL lacks both.

- Query optimization can also be done with the ordering provided by the SFS.

- Output pipelining can be applied with SFS which BNL does not support.

Since SFS presorts the data, its effectiveness is controlled by mechanism used for sorting, In this paper authors presented an entropy function, where higher entropy tuple discards more tuples hence increases discard efficiency. Also because of the sorting property SFS is input order independent whereas BNL performance depends on the order of its input.

Comments:

The paper presents a new improved algorithm Sort-Filter Skyline. It is compared with existing BNL methods with randomly generated, uniformly distributed values in terms of number of pages used with different dimensional inputs. SFS outperforms BNL in every test.

However, Some short comings of the algorithms were:

- Additional cost of Sorting a very large database with very high dimensionality.

- Comparison with Divide and Conquer method when dimensions are very high.