LaBRI

université de BORDEAUX

Using Multidimensional Skylines for Regret Minimization

Alami Karim
Maabout Sofian

June 23rd 2020

Context

- Investigate Top-K frequent and Top-K priority skylines as candidate sets for regret minimization queries[1]
- Experiment the speedup provided by NSC(Negative Skycube)[2] to regret minimization queries

[1] D. Nanongkai, A. D. Sarma, A. Lall, R. J. Lipton, and J. Xu. Regret-minimizing representative databases. Proceedings of the VLDB Endowment, 3(1-2):1114–1124, 2010.
[2]Alami, K., Hanusse, N., Kamnang-Wanko, P., & Maabout, S. (2020). The negative skycube. Information Systems, 88, 101443.

Preference queries

Given a set of tuples T, it returns a subset of tuples that suits the user preference

- Top-K query: based on scoring function
- Skyline query: based on dominance

Hotel	Price	Distance
h_1	75	100
h_2	45	150
h_3	50	300
h_4	65	450
h_5	25	500
h_6	50	400
h_7	100	150
h_8	30	300

E.g

- Skyline query result:

h_1, h_2, h_5 and h_8

- Top-2 query result:

Scoring function $f(t) = P + D$

h_1, h_2

Regret minimization query

Limitations

- Top-K query: requires to define a scoring function
- Skyline query: the size of the output is not controlled

Regret minimization query[1]: bounds the output without requiring a scoring function

We consider L the family of linear scoring function

Let $f \in L$, $f_1(T)$ the highest score

- Given $S \subset T$ the maximum regret ratio, $\mathrm{mrr}(\mathrm{S}, \mathrm{L}) = max_{f \in L} \frac{f_1(T) - f_1(S)}{f_1(T)}$

Problem RMS: Given a dataset T, the family of linear function L, an integer r, compute a set $S \subset T$ of size r that minimizes the maximum regret ratio $\mathrm{mrr}(\mathrm{S}, \mathrm{L})$

The regret represents how much users are statisfied with S.

[1] D. Nanongkai, A. D. Sarma, A. Lall, R. J. Lipton, and J. Xu. Regret-minimizing representative databases. *Proceedings of the VLDB Endowment*, 3(1-2):1114–1124, 2010.

Regret minimization query

RMS is NP Hard[1]

sphere[2] is the state of the art heuristic algorithm with guarantees

Skyline points as candidates

Let Sky be the skyline set of T. Let S^* be the optimal solution of an RMS instance such that $r \leq |Sky|$, then $S^* \subseteq Sky$

[1] S.Chester,A.Thomo,S.Venkatesh,andS.Whitesides.Computingk-regret minimizing sets. *Proc. VLDB Endow.*, 7(5):389–400, 2014
[2] M. Xie et al.. Efficient k-regret query algorithm with restriction-free bound for any dimensionality. *Proceedings of SIGMOD Conference 2018*

Mulitidemensional Skylines

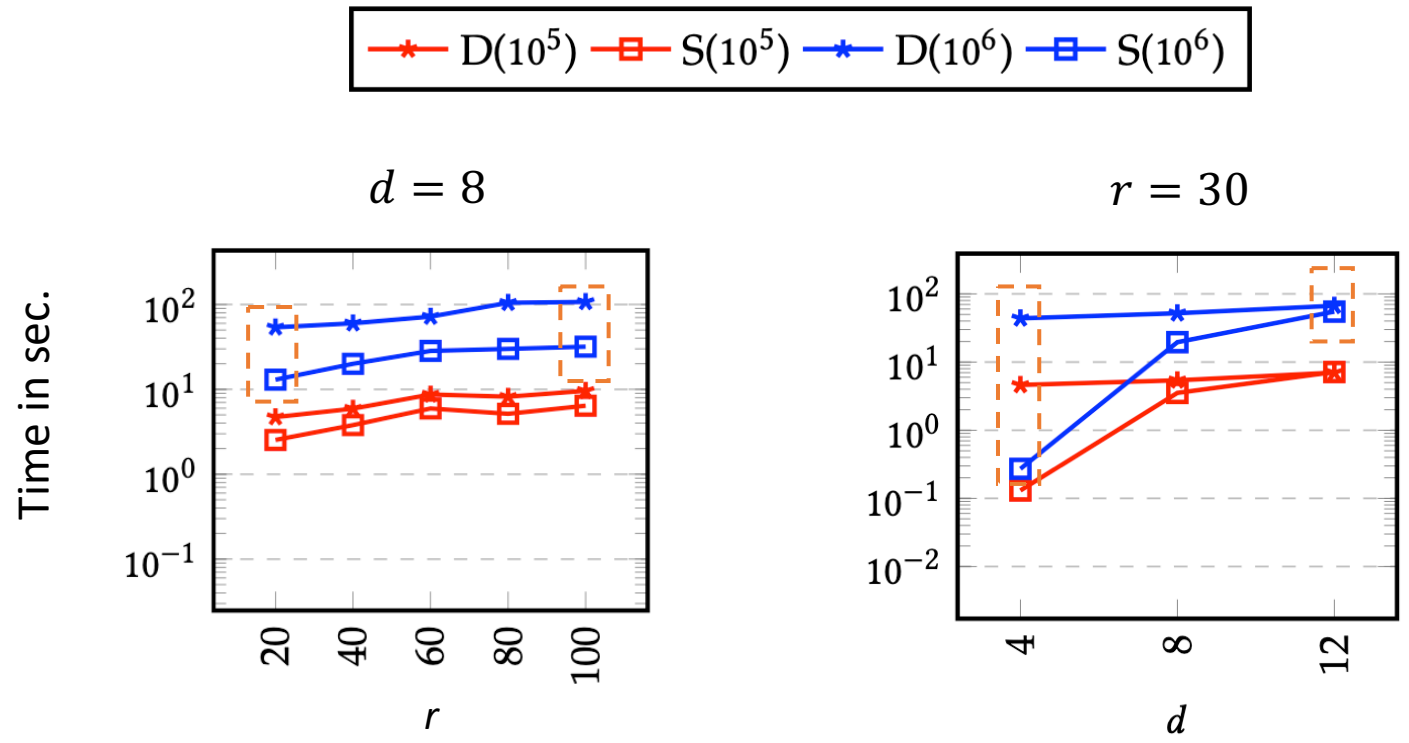
Skyline set may be of the size of the whole dataset

We investigate Top-K Frequent skylines and Top-K priority skyline as candidate sets[1]

Top-K Frequent (Top-KF)	Top-K Priority (Top-KP)
Frequency: the number of subspaces[2] where a tuple is in the skyline	Priority: the cardinality of the smallest subspace where a tuple is in the skyline

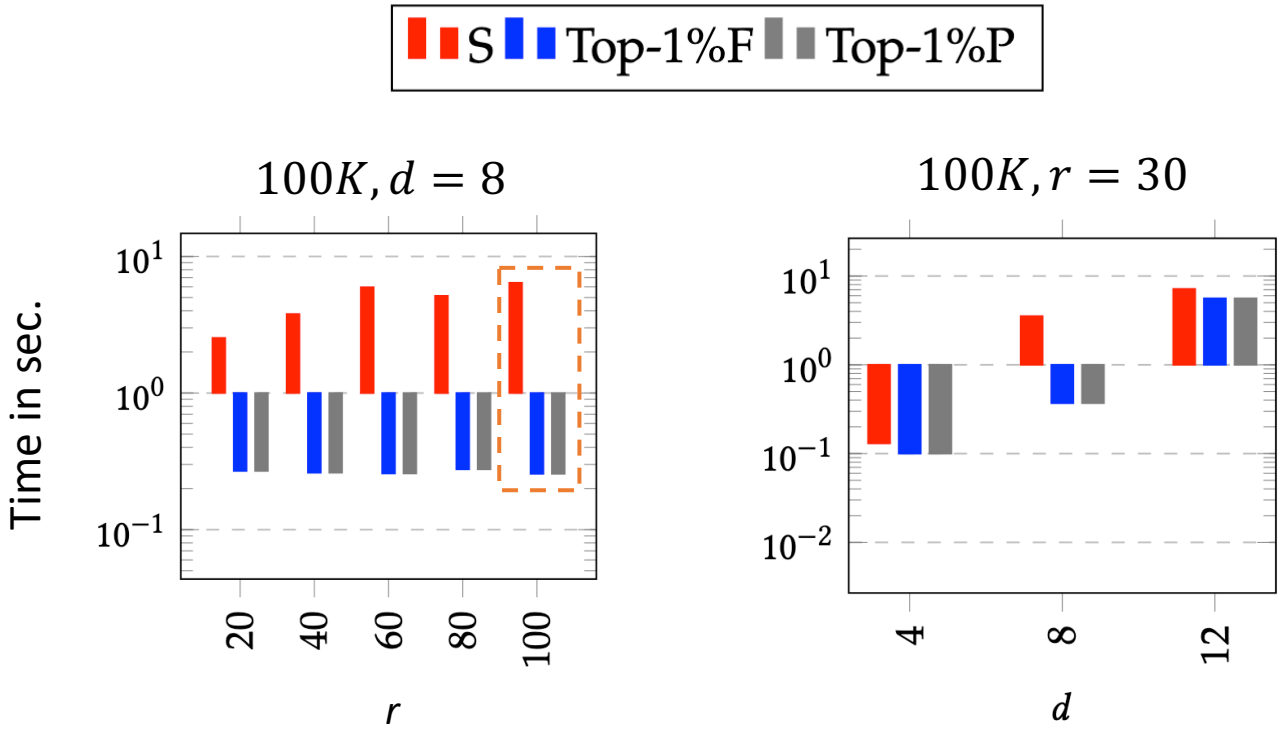
[1] We compute Top-KF and Top-KP with index structure NSC
[2] Subspace: subset of dimensions

Speedup with skyline candidate set



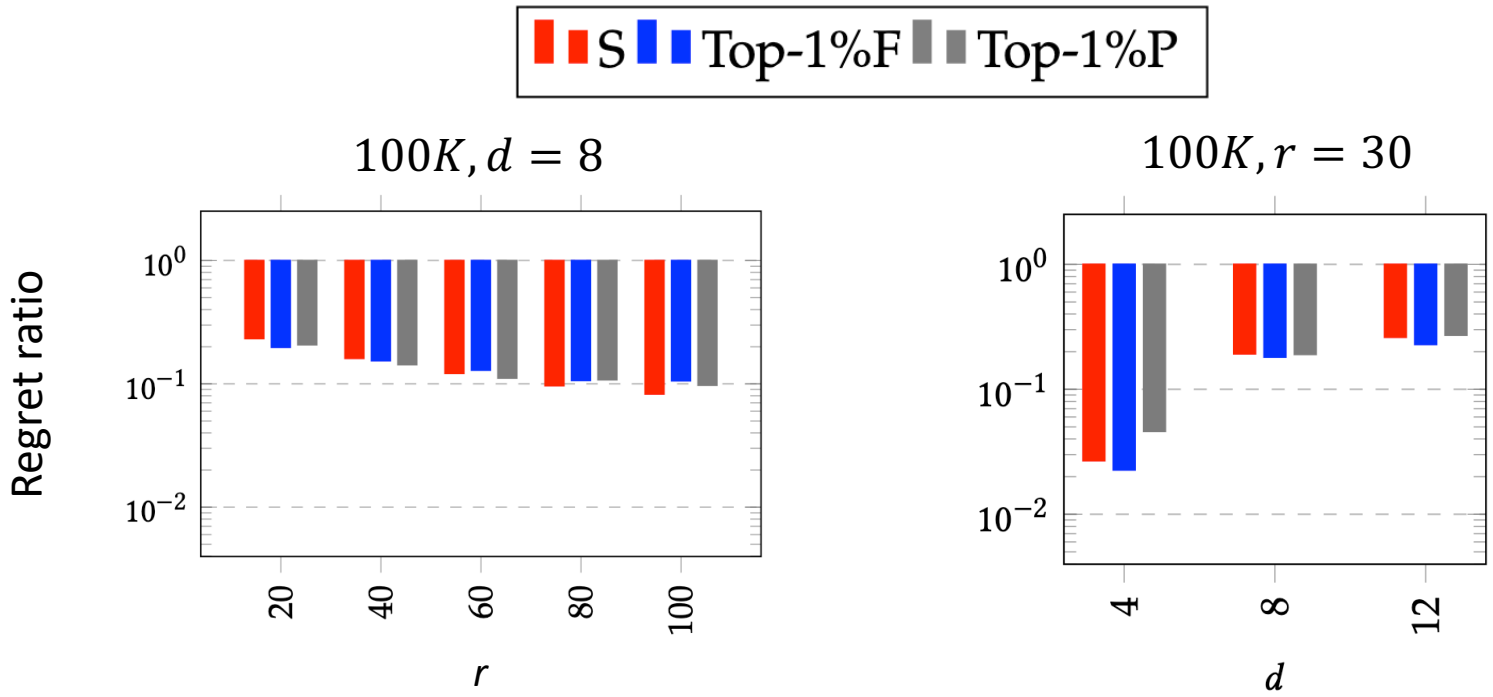
r: size of the output by sphere
d: number of dimensions
D: dataset
S: Skyline set

Speedup with Top-KF and Top-KP



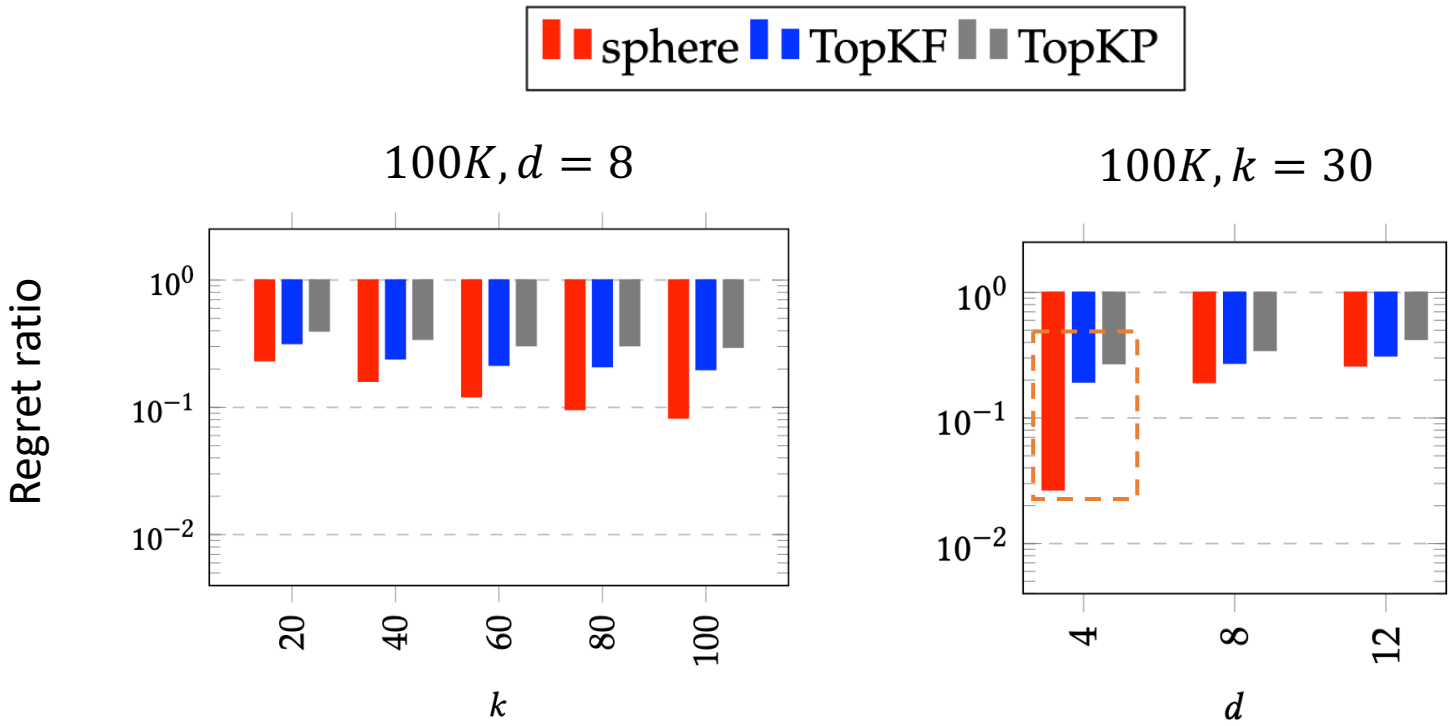
r: size of the output by sphere
d: number of dimensions

Regret ratio with Top-KF and Top-KP



r: size of the output by sphere
d: number of dimensions

Regret ratio of *sphere* vs Top-KF vs Top-KP



k: size of the output by sphere, TopKF ,and TopKP
d: number of dimensions

Conclusion and Perspectives

Conclusion:

- Top-KF computes a good candidate set for *sphere*
- NSC speedup *sphere* by optimizing the computation of candidate sets

Perspective:

- A theoretical guarantee on the regret ratio by using the candidate sets Top-KF and Top-KP

Questions