Using G-skyline to improve Decision-Making

Sana Nadouri
LIRE/LIAS laboratories.
Constantine-Futuroscope, 25000-86360
sana.nadouri@univ-constantine2.dz/ensma.fr

Zaidi Sahnoun
LIRE Laboratory UC2.
Constantine, 25000
zaidi.sahnoun@univ-constantine2.dz

Allel Hadjali
LIAS Laboratory ENSMA.
Futuroscope, 86360
Allel.hadjali@ensma.fr

Abstract

Skyline is an operator that can help users making decisions using a multidimensional data and conflicting criteria. Skyline is based on Pareto dominance relationship, it returns objects that are not dominated by any other object in the database. Recently, the skyline definition was expanded to group decision making to meet complex real life needs encountered in many modern domain applications. We used the groups skyline in our architecture to reinforce the Multi-agent distributed decision support system by integrating the process to the comparison agent. In this paper, we introduce the Skyline operator, Groups skyline and we propose to integrate groups skyline to our internal distributed decision support system architecture.

Keywords - Skyline, Groups skyline, Decision support system, Distributed decision support system

1. Introduction

The Skyline operator (Maxima or Pareto dominance relationship) is a multi-criteria analysis operator that manages query complexity. Skyline extracts tuples from database using user preferences and returns the best response. The skyline is very successful in the database filed since its introduction by Borzsony in 2001 [Borzsony01], many algorithms were proposed to retrieve objects that present the optimal combination of the dataset characteristics in a local and distributed environment.

Recently, the skyline definition and the local algorithms become inadequate to answer various queries requiring the analysis of not only individual tuples but also their combinations, for that reason, Groups skyline was introduced.

In this paper, we propose a method to reinforce our Multi-agent distributed decision support system using groups skyline [Nadouri18].

The remainder of the paper is organized as follows. Section 2, introduces the Skyline operator, Groups Skyline, Decision support systems and Distributed decision support systems. Section 3, provides an initial description of the approach using a sequence diagram. Finally, Section 4 concludes the paper and draws some future research directions.

2. Background and state of art

2.1. The Skyline operator

The skyline operator returns records in dataset that provide optimal trade-offs of multiple dimensions, since its introduction to the database community in [Borzsony01], the skyline operator had a real interest [Hose16],[Tiakas15],[Paolo18], which allows it to stand out of many other types of query preferences. Skyline is based on Paredo dominance concept that can be defined as follows:

Definition (Dominance or Pareto), noted:$\prec$, When having two tuples: p and q, if p is as good as q in all dimensions and better than q in at least one, then p dominate q ($p \prec q$), if $p \prec q$ and simultaneously $q \prec p$, then they are incomparable.

Formally (assuming that the smallest value is better):

$$p \prec q \Rightarrow \forall i \in [1,d] : pi \leq i\ qi \text{ and } \exists j \in [1,d] : pj <j\ qj$$

A set of algorithms were proposed, the most used are indexed or not indexed algorithms:

- **Indexed**: Index - proposed in 2001 [Berrouigat15], [Tan01], Bitmap - proposed in 2001 [Berrouigat15], [Tan01] NN - proposed in 2002 [Kossmann02], [Nguyen18], BBS - proposed in 2003 [Papadias03], [Papadias05], [Nguyen18].
- **Not indexed**: BNL - proposed in 2001 [Borzsony01], [Nguyen18], D&C – proposed in 2001 [Borzsony01], SFS – proposed in 2005 [Chomicki03], LESS – proposed in 2005 [Godfrey05].

In: Proceedings of the 3rd Edition of the International Conference on Advanced Aspects of Software Engineering (ICAASE18), Constantine, Algeria, 1,2-December-2018, published at http://ceur-ws.org

2.2. Group Skyline

Real-world applications require choosing groups of objects rather than individual objects, sport is one of them, selecting the best team based on a list of athletes requires comparing teams players by performance. Another example is choosing a group of experts to review and evaluate papers based on the experts collective strength on multiple desired skills. Group Skyline is also important in other domains e.g. group recommendation, investments selection, detection of most dangerous places when fire or a crime is made, etc. The combination points or Groups Skyline is defined as follows [Liu15]:

Definition 1: Based on Skyline (called: G-Skyline). Given a dataset P of n points. p and p' are two different points in P, p dominates p' ($p \prec p'$), if for all i, $p[i] \leq p'[i]$, and for at least one i, $p[i] < p'[i]$ in $1 \leq i \leq d$.

Definition 2: Based on Group dominance (called G-Dominance). When having 2 groups: G=p1,p2,...,pk and G'=p'1,p'2,...,p'k, we say group G g-dominates group G', denoted by $G \prec gG'$, if we can find two permutations of the k points for G and G', G=pu1,pu2,...,puk and G'=p'v1,p'v2,...,p'vk, such that $pui \leq p'vi$ for all i ($1 \leq i \leq k$) and $pui \prec p'vi$ for at least one i.

We divide Groups Skyline algorithms intro two classes:

1. G-Skyline using static data
   - **Top-k Skyline Groups Queries:** [Zhu17] returns k skyline groups, it combines skyline groups and top-k queries using Bit vector to store the dominated number of each point.
   - **Finding Pareto Optimal Groups:** [Liu15] two algorithms were proposed: the point-wise and the unit group-wise algorithm. The authors present a structure that represents the points in a directed skyline graph and captures all the dominance relationship among the points based on the notion of skyline layers.
   - **On skyline groups:** [Zhang14] they identified two anti-monotonic properties with varying degrees of applicability: order-specific as well as weak candidate-generation property. The authors propose three techniques, namely output compression, input pruning, and search space pruning.
   - **Group skyline computation:** [im12] they proposed GDynamic an equivalent to a dynamic algorithm that fills a table of skyline groups. It generates candidate groups in a progressive manner and updates the resultant groups skyline dynamically.
2. G-Skyline using stream data
   - **Finding Group-Based Skyline over a Data Stream in the Sensor Network:** [Dong18] they invoke the problem of Computing G-Skyline when a new point p arrives. First, they check which layer the point p belongs to, and then update the graph to construct the new relationships between all the points, finally, they compute the G-Skyline based on the sharing strategy.
   - **Efficient Processing of Skyline Group Queries over a Data Stream:** [guo16] authors store dominance information that could be reused. For each active object p, they maintain (1) the number of dominators, denoted by p:num and (2) objects that could be dominated by p, denoted by p:dominatee. When an object is added or removed, they update p:num and p:dominatee of each object influenced by p. Objects having fewer than k dominators are reported as candidates.

2.3. Distributed decision support system (DDSS)

DSS is defined in different ways, it is a system that assist decision makers when making their decisions, in order to confirm or correct the decision [Poleto15]. In the same way that definitions vary by authors, there is no standard architecture to define these systems, DSS contains several parts and sub-parts that are listed below. We have 5 basic components [Simon60], [Otero18], [Chandiok16]: *The database management system, The model management system, The knowledge engine, The user interface and the user*. The DSS process has 4 essential phases: **Intelligence, Design, Choice, Implementation.**

The system complexity and the distribution of environments and systems require the distribution of the decision. Currently, there is no definition to illustrate DDSS structure since existing architectures depend on the problem to be solved. We define a DDSS as a set of Decision support systems that communicate in a distributed environment and share a common goal in different sites, in other words, it is an extended version of a DSS.

3. The proposition

We proposed a new approach called ADS2, which is based on the definition of the distributed decision support system modeled in the article [Nadouri18].

The internal architecture is modeled using Multi-Agent approach and contains mainly 6 components, one of these components is the comparison Agent, the internal behavior of the comparison Agent is based on the Groups Skyline process because each external DSS will give a partial decision or a decision based on its internal knowledge.

Using the same process of Groups Skyline we propose to integrate it to obtain a better decisions in time.

As shown in Figure.1. the Agent compares the different Decisions received from other DSSs (in this example, we have 2 DSSs) using a G-Skyline algorithm, the system returns the best combination of the proposed decisions, the decisions are then sent to the DSSs to confirm or reinforce the process, the algorithm returns the final decision if and only if the different DSSs agree and the predefined decision time is not achieved.

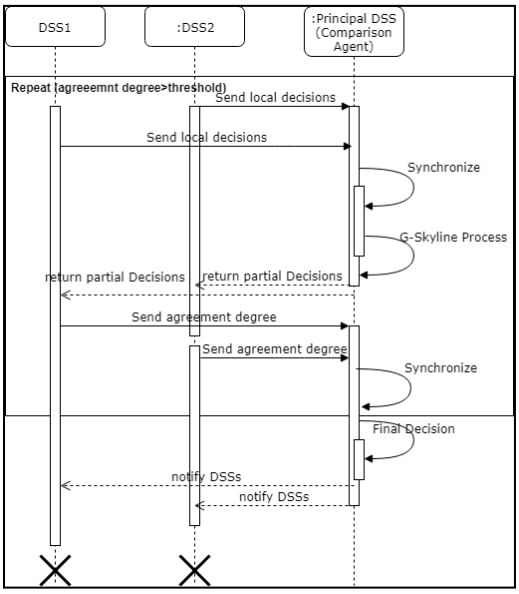

Figure 1: The approach behavior

4. Conclusion and future work

In this paper, we introduced the skyline and groups skyline concepts. We also proposed to integrate the groups skyline concept to our Multi-agent Distributed decision support systems.

We are still developing the method. For validation, we will implement the GSM method, we will also improve the groups skyline algorithm, many challenges need to be considered, we can cite the issue of different group size, groups are not extracted progressively, the large number of possible points combination and the large number of output groups. Some of these issues can be solved using relaxation methods and progressive algorithms used in individual skyline algorithms. Finally, we think about revisiting the dominance relationship for groups skyline definition.

References

[Berrouigat15] Khadidja Berrouigat, Fatima Bensaha. *Recherche des skyline à base de l'algorithme bitmap - Skyline search based on the bitmap algorithm.* Algeria: Abou Bekr Belkaid University, 2015.

[Borzsony01] S. Borzsony, D. Kossmann, and K. Stocker. «The skyline operator.» *17th International Conference on Data Engineering.* 2001.

[Chandiok16] A. Chandiok and D. K. Chaturvedi. «Cognitive Decision Support System for medical diagnosis.» *International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT).* 2016. 337-342.

[Chomicki03] J. Chomicki and P. Godfrey and J. Gryz and D.Liang. «Skyline with presorting.» *19th International Conference on Data Engineering.* 2003. 717-719.

[Dong18] Dong, Leigang and Liu, Guohua and Cui, Xiaowei and Li, Tianyu,. «Finding Group-Based Skyline over a Data Stream in the Sensor Network.» *Information*, 2018: 33.

[Godfrey05] Godfrey, Parke and Shipley, Ryan and Gryz, Jarek,. «Maximal Vector Computation in Large Data Sets.» *the 31st International Conference on Very Large Data Bases.* Trondheim, Norway: VLDB Endowment, 2005. 229-240.

[guo16] guo. «Efficient processing of skyline group queries over a data stream.» *Tsinghua Science and Technology*, 2016: 29-39.

[Hose16] Hose, Katja. «Skyline Queries.» *Datenbank-Spektrum*, 2016: 247-251.

[im12] im. «Group skyline computation.» *Information Sciences*, 2012: 151-169.

[Kossmann02] Kossmann, Donald and Ramsak, Frank and Rost, Steffen. «Shooting Stars in the Sky: An Online Algorithm for Skyline Queries.» *28th International Conference on Very Large Data Bases.* Hong Kong, China: VLDB Endowment, 2002. 275-286.

[Liu15] Liu, Jinfei and Xiong, Li and Pei, Jian and Luo, Jun and Zhang, Haoyu,. «Finding pareto optimal groups: Group-based skyline.» *Proceedings of the VLDB Endowment*, 2015: 2086-2097.

[Nadouri18] Sana Nadouri, Yassine Ouhammou, Zaidi Sahnoun and Allel Hadjali. «Towards a multi-agent approach for distributed decision support systems.» *27th IEEE International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises.* Paris: IEEE, 2018. 6.

[Nguyen18] Nguyen Ngoc Vinh. *Spatial Skyline Query Algorithms.* 13 04 2018. viblo.asia/p/spatial-skyline-query-algorithms.

[Otero18] Abraham Otero. *https://pdfs.semanticscholar.org.* 13 04 2018. https://pdfs.semanticscholar.org.

[Paolo18] Paolo Ciaccia. *http://www-db.deis.unibo.it/courses/SI-M/slides/02.1.Skyline.pdf.* 13 04 2018. http://www-db.deis.unibo.it/courses/SI-M/slides/02.1.Skyline.pdf.

[Papadias03] Papadias, Dimitris and Tao, Yufei and Fu, Greg and Seeger, Bernhard. «An Optimal and Progressive Algorithm for Skyline Queries.» *ACM SIGMOD International Conference on Management of Data.* San Diego, California: ACM, 2003. 467-478.

[Papadias05] Papadias, Dimitris and Tao, Yufei and Fu, Greg and Seeger, Bernhard. «Progressive Skyline Computation in Database Systems.» *ACM Trans. Database Syst.*, 2005: 41-82.

[Poleto15] Poleto, Thiagoand de Carvalho, Victor Diogho Heuerand Costa, Ana Paula Cabral Seixas,. «The Roles of Big Data in the Decision-Support Process: An Empirical Investigation.» *Decision Support Systems V - Big Data Analytics for Decision Making*, 2015: 10-21.

[Simon60] Simon, H.A,. *The New Science of Management Decision.* Harper & Row, 1960.

[Tan01] Tan, Kian-Lee and Eng, Pin-Kwang and Ooi, Beng Chin. «Efficient Progressive Skyline Computation.» *27th International Conference on Very Large Data Bases.* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001. 301-310.

[Tiakas15] Tiakas, Eleftherios and Papadopoulos, Apostolos N and Manolopoulos, Yannis. «Skyline queries: An introduction.» *6th International Conference on Information, Intelligence, Systems and Applications (IISA).* IEEE, 2015. 1-6.

[Zhang14] Zhang, Nan and Li, Chengkai and Hassan, Naeemul and Rajasekaran, Sundaresan and Das, Gautam,. «On Skyline Groups.» *IEEE Transactions on Knowledge and Data Engineering*, 2014: 942-956.

[Zhu17] Zhu, Haoyang and Zhu, Peidong and Li, Xiaoyong and Liu, Qiang,. «Top-k Skyline Groups Queries.» *EDBT*, 2017: 442-445.