



(RESEARCH ARTICLE)

Applicability of skyline query algorithm for career identification of k-12 graduates

Custodio Eunice Bondoc ^{1,2,*} and Tumibay Gilbert Malawit ²

¹ *Bulacan State University, City of Malolos, Bulacan, Philippines.*

² *Angeles University Foundation, Graduate School, Angeles City, Philippines.*

Publication history: Received on 18 February 2020; revised on 24 February 2020; accepted on 25 February 2020

Article DOI: <https://doi.org/10.30574/gjeta.2020.2.3.0012>

Abstract

In making the student highly proficient in a particular field, choosing the right program track in the Senior High School (SHS) program is critical. The additional 2 years in high school allow them to choose a specialization that is related to the field or profession that they wish to pursue in College. Computing user preference using skyline query algorithm can be a tool in determining the right career intended for a student since it can be used for various applications that include decision making, personalized services, and search pruning.

This paper focused on reviewing significant literature concerning the applicability of this algorithm in identifying the right course based on the student's academic qualifications and K to 12 program tracks, strand and specialization as entry requirements in higher education. Different skyline query algorithm and processing were presented. Other significant variables like academic performance, examination taken and other significant variables will be considered in evaluating the capability of the student before entering to a college degree program. Identification of the data sets of the senior high school program tracks will be categorized into Academic tracks, Technical-Vocational-Livelihood, and Arts, Design and Sports which can be used to classify all the courses available related to the students' chosen career.

Keywords: Skyline Query Algorithm; Senior High School Program Tracks; User Preference; Entry Requirement in Higher Education

1. Introduction

The Department of Education (DepEd) in 2011, administered a shift to a fresh learning scheme which is the K to 12 basic education program which aims at enhancing the learners' basic skills, producing more competent citizens, and preparing graduates for lifelong learning and employment. The students of the new system will graduate at the age of 18 and will be ready for employment, entrepreneurship, middle-level skills development, and higher education upon graduation. The new curriculum is designed to enable graduates to join the labor force right after high school, and suitably prepare those who want to go on to higher education.

Ideally, the K to 12 graduates is already equipped to able to join the workforce right away. This is through the help of the electives which are usually offered during grades 11 to 12. The electives, or areas of specialization, include the following: Academics for those who wish to pursue higher studies, Technical-vocational for those who want to acquire employable skills after high school, and Sports and Arts for those who are inclined in the said fields.

In the new program of DepEd, students are given the chance to choose among three tracks: Academic, Technical-Vocational-Livelihood, Sports and Arts where they will undergo immersion, which provides relevant exposure and actual experience in their chosen track. Choosing the right track in the Senior High School program is critical in making the student highly proficient in a particular field. The additional 2 years allow them to choose a particular track that is related to the field or profession that they wish to pursue in the future.

* Corresponding author: Custodio Eunice Bondoc

However, there could be a mismatch in the program track and the student's interests, personality and passion which may lead to a poorly knowledgeable graduate who will find it hard to compete in the job market or keep up with the other students when they go to college. The big challenge, therefore, is making the K to 12 work correctly and efficiently, in order to deliver what it meant to. Definitely, the academe needs to make sure that the graduates, year after year, will be able to find the jobs they prepared for.

This study is an attempt to determine whether skyline queries algorithm can be used for the computation of user-preference like students' choice of career or course to be taken in College. The researcher is considering using skyline operation, which can filter out a set of interesting points from a potentially large set of data points in determining different factors that somehow affects or influence the choice of career of the students.

In this study, the researcher looked into the different literature to see if the use of skyline queries algorithm is applicable in analyzing user preferences considering the number of courses available and universities in the Philippines which can be chosen by the student for higher education.

2. Related Work: The applicability of skyline query algorithm

Selecting the best possible course by a student is a key decision to make, and considered challenging. This requires an individual to develop certain degree of self-awareness which entails decision making. This process usually consists of considering several factors like interests, personality type, work-related values, skills and even possible compensation in the future. The researcher believes that using skyline query algorithm, student preferences in selecting the course to be taken in College can be computed.

Skyline query helps in identifying the best objects in a multi-attribute dataset. To understand the concept of the skyline query, consider an example of a person going on holiday to *Boracay, Philippines* and looking for a hotel that is less cost and close to the beach. This means that the person is interested in all such hotels that are not worse than any other hotel on both dimensions. This set of interesting hotels is the "Skyline". From the skyline, one can make the decision after weighing the personal or the imposed preferences.

In the present information age, everyone has an access to a large amount of information, data and options to choose from [1]. The amount of information available and the rate of change may hide the optimal and truly desired solution. This reveals the need of a mechanism that will highlight the best options to choose among every possible scenario [1].

Recently, there has been much interest in processing skyline queries for various applications that include decision making, search pruning and even for personalized services. Skyline queries aim to prune a search space of large numbers of multi-dimensional data items to a small set of interesting items by eliminating items that are dominated by others. Existing skyline algorithms assume that all dimensions are available for all data items. A point is considered as interesting, if there are no any other points better than the interesting one in all the evaluation criteria. The popularity of the skyline operator is mainly due the paradigm's simplicity and its applicability on multi-criterion decision support with respect to user preferences.

To be specific, consider a typical skyline query example for a house purchase. Suppose that a house might be interesting for somebody if no other house is both cheaper and closer to a metro-station. It is considered that as the distance of a house from a point of highly interest is decreased, in this case a metro-station, the objective value which is the price of the house is increased. So, the user will try to find the best money-to-value ratio that can satisfies one's own preferences.

Another example is given by Borzsony et al., [2], where some indicative applications areas for which skyline queries are useful are the customer information services, decision support and decision-making systems. For instance, a skyline query can be used by travel agencies to find a reasonable priced hotel near the sea or to find good salespersons which have low salary.

To Dellis [3], reverse skyline queries can assist in market research applications in order to find if a specific product is appealing to consumers or to identify the best location for a new branch where can support microeconomic data mining or even in continuous data stream environments such as stock exchange systems [4]. This algorithm can also be used on location-based systems (LBS) in identifying the shortest route to a destination or the closest point of interest among many [5]. Another application is the distributed query optimization. This can be particularly useful in cloud architectures where data are scattered among servers or in the case where Quality of web services is the primary goal. Skyline queries can also be used to focus on a subspace of attributes in order to identify the skyline on a small subset of the dimensions of the dataset that are defined. Skyline queries have also applications in computer security and

especially on problems concerning privacy and authentication [6]. Skyline computation in metric space can assist the DNA searching problem in bioinformatics. Finally, skyline queries are applicable in a wide variety of data types such as partial ordered, and even for incomplete or uncertain data [7].

2.1. Skyline query processing

The increasing size of multidimensional data and the rapid growth of decision support systems lead to seek for new efficient methods of data processing in order to retrieve useful insights. Skyline queries assume that every user has a series of preferences over the attributes of data. Those preferences indicate what user's likes and dislikes (e.g. "I like the sea more than the mountains" or "I prefer to go vacations on an island rather than on a mountain). All the preferences are considered equivalent and will help to discard the items of the dataset that will not be preferred by anyone. This results in a small subset that contains the most interesting and preferred items based on all the references of all users. This set will be the skyline set or the pareto optimal set.

In recent years, skyline query processing has become an important issue in database research for extracting interesting objects from multi-dimensional datasets. The skyline query processing is applicable in many applications that require multi-criteria decision making without using cumulative functions in order to define the best results but based on user's preferences. The skyline operator filters out a set of interesting points based on a set of evaluation criteria from a potentially large dataset of points. A point is considered as interesting, if there is not any other point better than that in all the evaluation criteria. The popularity of the skyline operator is mainly due the paradigm's simplicity and its applicability on multi-criterion decision support with respect to user preferences.

In this study, the researcher will consider set of interesting courses. The student may consider the compensation that can be received after finishing the degree course. Skyline queries are focused on filtering out a set of points from a large set of data points. The points which are not dominated by other points are interesting. Skyline operator is firstly proposed by Borzsonyi et al. [2]. They presented two algorithms, say Block-Nested-Loops (BNL) and Divide and-Conquer (D&C). They concluded that the Skyline operation is useful for a number of database applications, including decision support and visualization. Their experimental results indicated that a database system should implement a block-nested-loops algorithm for good cases and a divide-and-conquer algorithm for tough cases. More specifically, they propose to implement a block-nested loops algorithm with a window that is organized as a self-organizing list and a divide-and-conquer algorithm that carries out m-way partitioning and "Early Skyline" computation. Furthermore, they also discussed Btrees and R-trees on evaluating skyline queries.

Chomicki et al. [8] proposed a skyline algorithm, Sort-Filter-Skyline (SFS), based on presorting BNL. Tan et al. [9] presented two progressive algorithms, Bitmap and Index, to improve the performance of skyline computing. A widely used effective algorithm nowadays, Brach and Bound Skyline (BBS), which is based on nearest neighbor search which was developed by Papadias et al.[10].

Whereas, Borzsonyi et al.,[2] proposed to extend database systems by a Skyline operation. According to them, this operation filters out a set of interesting points from a potentially large set of data points where a point is interesting if it is not dominated by any other point. Looking back at their example, a hotel which might be interesting for somebody traveling to Nassau if no other hotel is both cheaper and closer to the beach. Even if there are hotels available which provides online customer reservation and information system like the one developed by Castro and Custodio [11], still, user preferences should be the priority. They showed how SQL can be extended to pose Skyline queries, present and evaluate alternative algorithms to implement the Skyline operation, and show how this operation can be combined with other database operations. One of the procedures that they did is translating a Skyline Query into a Nested SQL Query. They showed how Skyline queries can be implemented on top of a relational database system by translating the Skyline query into a nested SQL query. They also used two-dimensional Skyline operator through the block-nested-loops algorithm and they compared every tuple with every other tuple. This is essentially happening if a Skyline query is implemented on top of a database system.

For Chester and Assent [12], skyline queries are well-studied problem for multidimensional data, wherein points are returned to the user and no other point is preferable across all attributes. This leaves only the points most likely to appeal to an arbitrary user. While Yin, Gu, Wei, Zhou and Liu [13] analyzed the customer's information for marketing insights. They analyzed the problem based on reverse skyline queries, and propose a new algorithm that significantly reduces the query cost by pruning unqualified customers without any false positive and identifies reverse skyline points as early as possible based on decision region and effect region. They further target on arbitrary k prospective customers and formulate the problem as top-k reverse skyline (top-k RD) queries. By extending the notion of reverse skyline to reverse skyline, it then support arbitrary k and group-based promotions. They evaluated the framework that they

formulated with extensive experiments, and the results demonstrate that the framework has promising results for reverse skyline queries which can efficiently support top-k RS queries. Yin et al., proposed OneTraversal to find the most prospective customers based on the notion of reverse skyline queries. The OneTraversal significantly reduces query cost by pruning unqualified points without any false positive and enables progressive results outputting. We then formulated the problem of finding k prospective customers as the top-k reverse skyline (top-k RS) queries and extended the notion of reverse skyline to reverse skyline order to support arbitrary k and group-based promotion.

Han, Jung, Eon and Yeom [14] presented a skyline-based matchmaking framework. According to them, the current method of carrying out the matchmaking procedure identifies items based on users' specifications. They further rethink that matchmaking procedures in such a way that they can find candidates among the identified items. This endows a user with the right of choice on deciding the best-possible items. They used the approach from the perspective of skyline computation and present an efficient skyline algorithm that gathers interesting item candidates efficiently. To devise an efficient sequential skyline algorithm where they adopt the lattice-based indexing using a lattice composition technique and an optimized dominance-check algorithm. Moreover, they parallelize the algorithm using breadth-first-search (BFS). Their extensive experimental results show that their algorithm outperforms current state-of-the-art algorithms, and the speedup factor of the parallelized algorithm is near-linear.

Li, Annisa, Zaman, Asif and Morimoto [15] conceptualized the MapReduce-Based Computation of Area Skyline Query for Selecting Good Locations in a Map. According to them, selection of good locations in a map is an indispensable function in many applications. To select specific locations, there is a need to specify detailed selection criteria. However, it is not easy especially for users of mobile devices. Therefore, an idea of skyline queries can be used, which are known to be easy and effective to retrieve interesting data from a database like selecting good locations in a map. However, the query is not fast enough for handling big data. In relation to this, Li et al. simplified and revised algorithm of the query using MapReduce framework so that it can be used for big data. Experiments' results demonstrate that the performance and scalability are superior to previous area skyline algorithm and can handle big data.

Lastly, Dwivedi and Rajput [16] investigated the evaluation of skylines over disparate sources via joins in efficient manner. The basic idea of the approach is that without computing skyline in the entire joined table, joined skyline only can still be processed based on the property of being skyline to quickly identify the skyline object for the joined tuple. They proposed an algorithm to build on top of the traditional relational Block Nested-Loop join algorithms, which fuses the computation of the join and the skyline in order to outputs the correct skyline without computing the full join. The experimental results demonstrate the applicability of interweaving join and skyline together.

The presented related literature on skyline algorithm became an avenue for the analysis of a possible algorithm computation of user preferences for career identification of the K to 12 graduates.

3. Methodology

This paper searched for literature in some useful databases journals and articles which provides an overview of Skyline query algorithm and be able to identify the methodology used by the authors. Different skyline algorithm techniques were presented that can be used as basis for the analysis of user preferences in identifying particular career to be taken by the senior high school graduates in College.

Identifying the data sets of the senior high school program tracks into Academic tracks, Technical-Vocational-Livelihood, and Arts, Design and Sports will be done to classify all the courses available related to the students' chosen program track. These data sets will be analyzed together with the other significant variables to be inputted by the graduates using the appropriate skyline algorithm technique. This could be a mechanism to highlight the best options for the student to choose among every possible scenario.

This study is an exploratory research which according to Saunders & Lewis, [17] intends to explore the research statement. Singh [18] said that exploratory research design does not aim to provide the final and conclusive answers to the research questions, but merely explores the research topic with varying levels of depth. In this study, the researcher explores the use of skyline algorithm where its processing maybe applicable in many applications that require multi-criteria decision making without using cumulative functions in order to define the best results but based on user preferences.

4. Results and discussion

With the above discussion of the background information about the K to 12 senior high school program tracks and the significance of the skyline queries algorithm, the researcher will attempt to establish an algorithm that can compute students' preference in selecting a particular career through their chosen program track and discovering relational queries as well as extracting interesting points from a set of points of the program tracks. These data sets will be first analyzed using block-nested-loops algorithm of Borzsonyi et al. [2] as the skyline algorithm technique.

Another consideration is the algorithm of Dwivedi and Rajput [17] for skyline evaluation using block nested loop join. This algorithm can be used in analyzing and computing user preferences applicable in career identification of the senior high school graduates.

5. Conclusion

Different authors discussed several skyline queries algorithm and query processing. Algorithm were analyzed just to highlight the best options to choose and for database searching in order to extract interesting objects from multi-dimensional datasets. The proposed methods of skyline combined with basic nested loop join is already part of all database system and its adaptation and performance is obvious. In identifying the course to be taken by the K to 12 graduates, together with the other significant variables, these data sets will then be analyzed using the block-nested-loops skyline algorithm.

Compliance with ethical standards

Acknowledgments

The authors would like to express their gratitude and thanks to Angeles University Foundation and to Bulacan State University Administration for all the kind support and help. Thanks and appreciations to their family, colleagues, friends and relatives who supported them for the completion of this research and to the Almighty God who is always there to guide and help them.

Disclosure of conflict of interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Kalyvas C and Tzouramanis T. (2017). A Survey of Skyline Query Processing. Department of Information and Communication Systems Engineering, University of the Aegean, Karlovassi, Samos, 83200.
- [2] Borzsonyi S, Kossmann D and Stocker K. (2001). The skyline operator. In Proceedings of the 17th International Conference on Data Engineering, 421–430.
- [3] Dellis E and Seeger B. (2007). Efficient computation of reverse skyline queries. In VLDB, 291–302.
- [4] Lin Y Yuan, W Wang and H Lu. (2005). Stabbing the sky: Efficient skyline computation over sliding windows. ICDE.
- [5] Kriegel HP, Renz M and Schubert M. (2010). Route skyline queries: A multi-preference path planning approach. In ICDE, 261–272.
- [6] L Chen, K Hwang and Wu J. (2012). MapReduce skyline query processing with new angular partitioning approach. In IPDPS Workshops, 2262-2270.
- [7] J Pei, B Jiang, X Lin and Y Yuan. (2007). Probabilistic skylines on uncertain data. In VLDB, 15–26.
- [8] Chomicki J, Godfrey P, Gryz J and Liang D. (2003). Skyline with presorting. Proceedings of the 19th international conference on data engineering.
- [9] Hoffner Y, Facciorusso C, Field S and Schade A. (2000). Distribution issues in the design and implementation of a virtual market place.

- [10] Papadias D, Tao Y, Fu G and Seeger B. (2005). Progressive skyline computation in database systems. *ACM TransDatabaseSyst*.
- [11] Castro MD B and Custodio EB. (2016). Managing customer reservations of BulSU hostel through the development of Online Information and Reservation System. *Journal of Advances in Information Technology*, 7 (3).
- [12] Chester S Sidlauskas, D Assent I and Bogh KS. (2015). Scalable parallelization of skyline computation for multi-core processors.
- [13] Yin B, Gu K, Wei X, Zhou S and Liu Y. (2018). A cost-effective framework for finding prospective customers based on reverse skyline queries. *Knowledge-Based Systems, Elsevier*, 117-135.
- [14] Han H, Jung H, Eom H and Yeom HY. (2011). An efficient skyline framework for matchmaking applications. *Journal of Network and Computer Applications, Elsevier*, 102-115.
- [15] Li Annisa, Zaman Asif and Morimoto. (2018). MapReduce Algorithm for Location Recommendation by Using Area Skyline Query.
- [16] Dwivedi S and Rajput A. (2014). Skyline Evaluation within Join Operation, Block Nested Loop Join Implementation. *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(3).
- [17] Saunders M, Lewis P and Thornhill A. (2012). *Research Methods for Business Students*, 6th edition, Pearson Education Limited.
- [18] Singh K. (2007). *Quantitative Social Research Methods*. SAGE Publications, 64.

How to cite this article

Custodio EB and Tumibay GM. (2020). Applicability of skyline query algorithm for career identification of k-12 graduates. *Global Journal of Engineering and Technology Advances*, 2(3), 10-15.
