International Journal for Modern Trends in Science and Technology, 8(12): 52-57

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0812009
Available online at: http://www.ijmtst.com/vol8issue12.html

Recommendation of Spatial Data Queries using Clustering Techniques

Sudhakar J[1] | Venkata Ramana Reddy Busireddy[2] | Subhash Chandra N[3]

[1]Research Scholar, Computer Science and Engineering, JNTUH, Hyderabad, Telangana, India.
[2]Professor, Department of Computer Science and Engineering, KSRM College of Engineering, Kadapa, Andhra Pradesh, India.
[3]Professor, Department of Computer Science and Engineering, CVR College of Engineering, Pocharam, Telangana, India.

To Cite this Article
Sudhakar J, Venkata Ramana Reddy Busireddy and Subhash Chandra N. Recommendation of Spatial Data Queries using Clustering Techniques. International Journal for Modern Trends in Science and Technology 2022, 8(10), pp. 52-57. https://doi.org/10.46501/IJMTST0812009

Article Info
Received: 15 November 2022; Accepted: 06 December 2022; Published: 11 December 2022.

ABSTRACT

Spatial data warehouses store large volumes of integrated and historized multidimensional spatial data in order to be explored and analyzed by various users. The data exploration is a process of searching relevant information in a data set. The data set to explore is a spatial data cube taken out from the spatial data warehouse that users interrogate by launching sequences of SOLAP (Spatial On-Line Analytical Processing) queries. However, this volume of information can be very large and diversified; it is thus necessary to help the user to face this problem by guiding them in their spatial data cube exploration in order to find relevant information.

Keywords: Spatial Query Recommendation, Spatial Queries, Query Recommendation, Clustering, Spatio-Semantic Similarity, Spatial Datamining.

1. INTRODUCTION

The BI system is realized by applying two different steps. The first step is the Extract, Transform and Load data. The ETL tools are responsible for extracting data from different heterogeneous sources, providing the integration and data cleansing according to a target schema or data structure, loading and storing data in a data warehouse Franklin.C *et al* [17]. The second step is to analyze data by using an analysis server such as Spatial OLAP server. It is a rapid and flexible way for analysts to navigate, explore and analyze the large amount of data stored in the data warehouse. Indeed, the user can make analysis reports by using some reporting tools, dashboards, navigation and statistical tools. These tools offer capabilities to explore data and support the analysis process.

To analyze data, users interactively navigate a data cube by launching sequences of SOLAP queries over spatial data warehouse. The problem appeared when the user may have no idea of what the forthcoming query should be. As a solution and to help the user in his navigation, we need a recommendation system Song et al [11]. The remainder of this paper is organized as follows: Section 2 presents the related work. Section 3 presents an overview of proposed approach for query

recommendation based on the exploration of data. Section 4 presents the evaluation of results. Section 5 concludes the paper.

2. RELATED WORK

In this section we present previous work that has influenced our design and implementation of recommendation system. It includes techniques for the design and implementation of spatial query recommendation system. For this reason, we present the various methods that have been proposed to explore data. A recommendation system [1-5] is categorized into

1) Content-based method: The user is recommended elements similar to the ones the user preferred in the past.

2) Collaborative method: The current user is recommended elements similar to the preferences of the previous users and the preferences of the current user. Aligon *et al* [8].

3) Hybrid method: This method combines both the content-based and the collaborative method.

S. Aissi, *et al.* [1] proposed a SOLAP recommendation methodology that aims to help users better exploit spatial data warehouses and retrieve relevant information by recommending personalized spatial–MDX queries. The proposed methodology detects implicitly the preferences and needs of SOLAP users using a spatio-semantic similarity measure. Finally, the proposed methodology was described and validated by employing large set of data.J. Aligon, *et al.* [8] presented a recommendation methodology stemming from collaborative filtering. In this literature, author claim that the entire sequence of queries belonging to an OLAP session, because it gives the user a synergic view of information. Similar to other collaborative methodologies, proposed features have three stages, (i) Examine the log for sessions that currently issued by the user (ii) Extract the most appropriate sub-sessions and (iii) Acclimate the top-ranked sub-session to the current user's session. After describing the proposed approach, the experimental outcome confirmed that the proposed methodology worked effectively on large set of data.

S. Bimonte, *et al*. [3,12] investigated the combination of Volunteered Geographic Information (VGI) in Spatial-OLAP (SOLAP) system. In this paper, author addressed some similarities and differences among these two types of systems such as conceptual quality-oriented framework for warehousing and OLAPing VGI data. In specific, to address precision and credibility difficulties associated to VGI data, author proposed two new Extract-Transform-Load (ETL) operators such as aggregation based on the VGI credibility and a filter based on the historical precision along with a new spatio-multidimensional model that provides decision makers with a global description of the quality of the aggregated data. Experimental outcome demonstrated that the proposed framework can achieve effective and favourable performance for data warehouse. Finally, we summarise that the methods proposed by [1-8] used the collaborative filtering that uses log of queries, the methods proposed by [10, 11, 13] used content-based method that uses the user profile. However, the guiding step was applied in the methods proposed by [4, 5].

3. PROPOSED SYSTEM

The main objective of the research is to improve the effectiveness of spatial data warehouse exploitation and also to achieve higher efficiency of Query Recommendation System. To help the user to go forward in the exploration of the spatial data, we propose an approach for recommending SOLAP queries. The input would be SOLAP query, the SOLAP queries stored in the query log and number of desired clusters.

The proposed method makes use of semantic and spatial similarity measure to compare the similarity of the queries. After finding out the similarity measure K Means clustering is applied on the data-set to group the queries into clusters. K-means is a centroid based clustering algorithm, where we calculate the distance between each point and a centroid to assign it to a cluster. The goal is to identify the K number of groups in the dataset. It is an iterative process of assigning each data point to the groups and slowly data points get clustered based on similar features. The objective is to minimize the sum of distances between the data points and the cluster centroid, to identify the correct group each data point should belong to.

Here, we divide a data space into K clusters and assign a mean value to each. The data points are placed in the clusters closest to the mean value of that cluster.

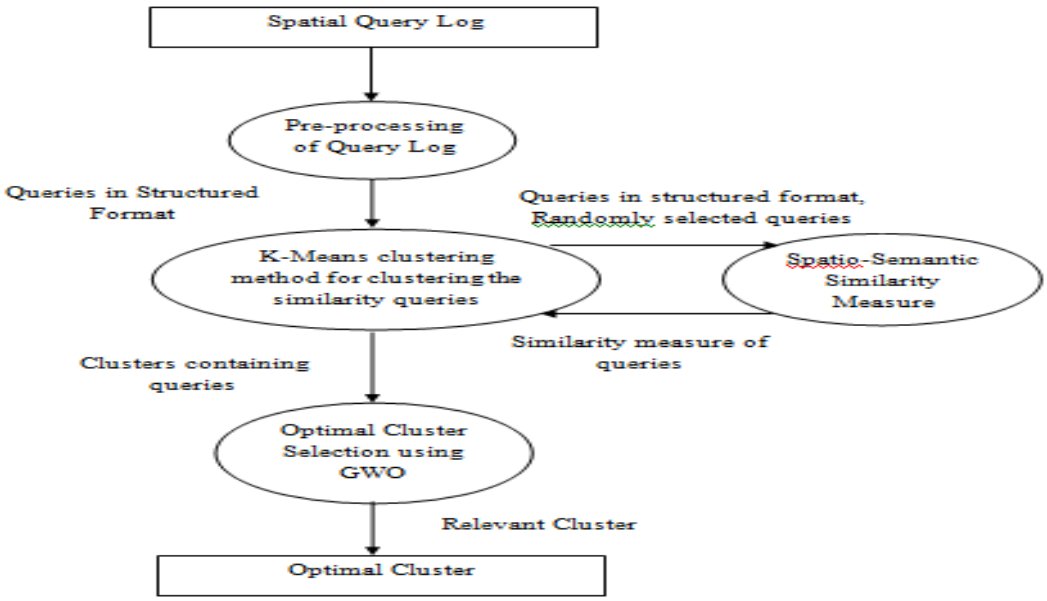

Figure 1: Data flow diagram for Recommendation Using K-Mean Clustering

3.1 Spatial Similarity Measures

In spatial queries, we can find the use of the three main categories of spatial distance, which are defined in the literature as follows: topological distance, direction distance and metric distance [19, 20]. So, to compute the distance between two spatial queries, we need to measure the similarity between the topological distance, direction distance and metric distance relation.

Topological Distance

We used the Topological distance proposed by [19] given in the equation (1)

$$Dist_{top}(TR(a, b), TR'(a', b')) = Dist_{top}(TR, TR') + \sum_{i=1}^{n} Dist_{top}(r1, r2) \quad (1)$$

Direction Distance

We used the direction distance proposed by [20] given in the equation (2)

$$Dist_{Dir}(q, q') = \sum_{i=1}^{2} Cost(TOR(a, b), TOR'(a', b')) \quad (2)$$

Metric Distance

We used the Metric distance proposed by [20] given in the equation (3)

$$Dist_{MetD}(q, q') = Dist_{MetD}(TD(a, b), TD'(a', b')) = \sum_{i=1}^{2} \sum_{j=1}^{2} a_{ij} \quad (3)$$

Spatial Distance

The spatial distance given by equation (4) is the sum of Topological Distance, Direction Distance and Metric Distance

$$Dist_{SpatialR}(q, q') = Dist_{top}(q, q') + Dist_{Dir}(q, q') + Dist_{MetD}(q, q') \quad (4)$$

3.2 Spatial Query Recommendation System using K-Means clustering Algorithm

Input: L: Pre-processed spatial queries from the query log.

K: The number of desired clusters.

Output: K clusters consisting of spatial queries.

Procedure

Step-1: Select the number K to decide the number of clusters.

Step-2: Select random K points or centroids.

Step-3: Assign each data point to their closest centroid, which will form the predefined K clusters.

Step-4: Calculate the variance and place a new centroid of each cluster.

Step-5: Repeat the third steps, reassign each data point to the new closest centroid of each cluster.

Step-6: If any reassignment occurs, then go to step-4 else go to step-7.

Step-7: End.

4. EXPERIMENTAL EVALUATION

In this section, we present the results of the experiment that we have conducted to assess the capabilities of our approach. This system gives the possibility to recommend an ordered set of spatial queries. First, to navigate in the spatial data cube the current user launched a sequence of spatial queries by using the SOLAP server over a spatial data warehouse. All the previous sessions of spatial queries are stored in the log. The proposed system recommends relevant set of spatial queries to the current user. Our experiment evaluates the efficiency of our approach to recommend spatial queries.

4.1 Execution Time

The proposed method illustrates a mathematical model for recommending queries equation (5). The proposed technique computes Execution time based on time taken to recommend the queries

Execution Time (ET) is calculated as:

$$E_{RT} = T_{CD} * T_{AR} \quad (5)$$

Where, T_{CD} is a total number of spatial record and T_{AR} is average retrieval time for a query processing of spatial user.

The performance is presented in Figure 2 according to various log sizes.

4.2 Precision

The purpose of this test is to evaluate the precision according to the number of candidate queries generated. The five queries are used for this process and precision is measured. The time required for this method to recommend the queries to the user is evaluated for the filtered log with the different file size. The log size is the number of queries presented in the log after pre-processing. The time has been evaluated for the recommending the 5 queries.

The precision is used to derive the quality and performance of the recommendation in many areas. It is calculated by Equation (6).This is used to find the value relevant recommendation for the user preference.

$$precision = \frac{|\{Relevant\ Recommendation\} \cap \{Proposed\ Recommendation\}|}{\{Proposed\ Recommendation\}} \quad (6)$$

4.3 Recall

Recall expressed as the total number of relevant information which is extracted based on search queries and divided by the total number of available relevant information. Recall is the ratio of the total amount of relevant searched information to the total number available relevant records in centralized database. It is described in Equation (7).

$$Recall = \frac{T_P}{T_P + F_N} \quad (7)$$

Where, T_P is truly positive and F_N is a false negative.

Table 1 illustrates the precision (P), recall(R) and Execution Time (ET) in milliseconds for PostGIS queries. Here, proposed system is compared with previous methodologies such as SOLAP Queries Recommendation System (Spatial-OLAP-RS) [2] and Candidate Queries Generation using the threshold Similarity (CQGS) [1] the results are presented in Table 1.

Recommendation System	Queries in PostGIS		
	P	R	ET
Spatial-OLAP-RS	0.45	0.48	190
CQGS	0.73	0.71	120
Proposed	0.75	0.8	110

Table 1: Precision (P), Recall (R), and Execution Time (ET) in milliseconds for Queries in PostGIS

Table shows the precision, recall and execution time for PostGIS query dataset. According to table 1 outcomes, it noticed that the proposed system performed well on PostGIS queries. Finally, the article claims that the proposed system is the better approach than the previous existing methodologies.

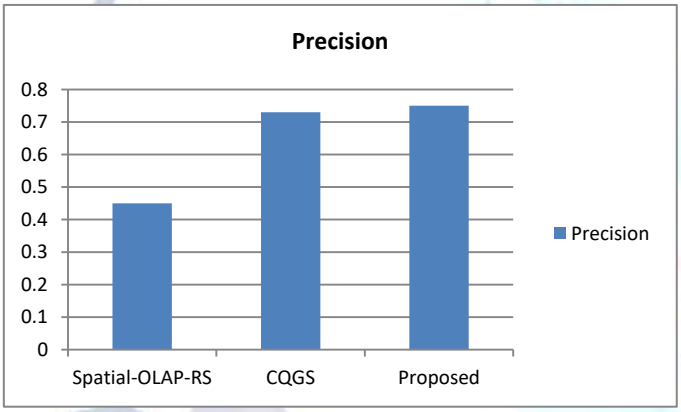

Figure 2: Precision for Queries in PostGIS

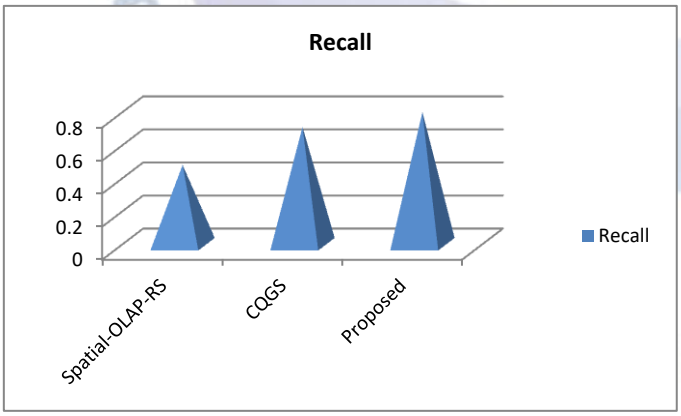

Figure 3: Recall for Queries in PostGIS

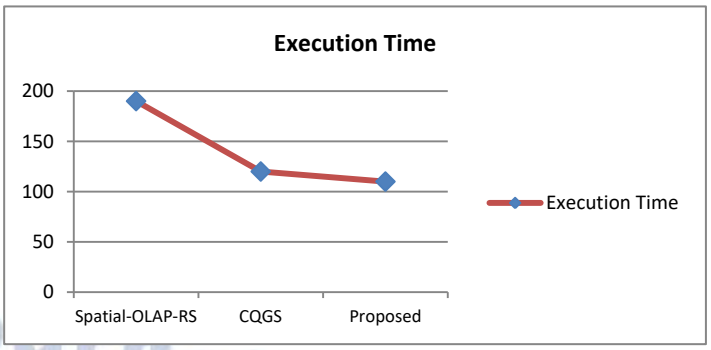

Figure 4: Execution Time for Queries in PostGIS

From the observations made from Figure 2 to 4 for the Queries in PostGIS dataset, the proposed approach computes precision, recall and execution time for identifying the effectiveness of technique. The proposed system is compared with CQGS and Spatial-OLAP-RS existing approach with respect to execution time precision and recall. CQGS is the nearest competitor on precision and recall constraints. Finally, the paper announces that the proposed approach performs better on every parameter & respective input constraints.

5. CONCLUSION

Satellite and remote sensing methods have produced a huge amount of geographical data. This amount of data creates a need for the development of an efficient spatial data exploration and construction of spatial query recommendation systems. Spatial data has been analyzed by statisticians and geographers for a long time. Most of the methods traditionally used in spatial data analysis do not take into account challenges posed by large amounts of data and do not make much use of the spatial query processing methods.

In this paper, we proposed a recommendation system to help users in their exploration of a spatial data. For that purpose, we suggested an approach for generating recommendation of SOLAP queries using K-Means clustering in the context of the collaborative exploration of spatial data.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Aissi, S., Gouider, M.S., Sboui, T. and Said, L.B. A spatial data warehouse recommendation approach: conceptual framework and experimental evaluation. Human-centric Computing and Information Sciences a Springer Open Journal, vol.5, no.1, p.30, 2015.

[2] Olfa Layouni, Jalel Akaichi. Spatio-Temporal OLAP Queries Similarity Measure and Algorithm. International Journal of Data Warehousing and Mining, Volume 15, Issue 2, 2019

[3] S. Bimonte and E.Negre(2014), "Evaluation of user satisfaction with OLAP recommender systems: an application to RecoOLAP on a agricultural energetic consumption datawarehouse", International Journal of Business Information Systems. Volume 21 Issue 1, Pages 117-136, December 2016.

[4] Olfa Layouni, Jalel Akaichi. Query Recommendation Systems Based on the Exploration of OLAP and SOLAP Data Cubes. Springer International Publishing AG, p.333, 2018,

[5] Olfa Layouni, Fahad Alahmari and Jalel Akaichi: "Recommending Multidimensional Spatial OLAP Queries". Springer International Publishing Switzerland 2016

[6] J. Sudhakar, N. Subhash Chandra, B. V. Ramana Reddy. Optimal Query Recommendation in Spatial Data Warehouse Using Democratic Grey Wolf Optimization. International Journal of Advanced Research in Engineering and Technology, Volume 12, Issue 3, pp. 626-636, March 2021.

[7] J. Sudhakar, B. V. Ramana Reddy, N. Subhash Chandra. An Approach on SOLAP Query Recommendation System Review. Universal Review, Volume VIII, Issue VI, pp 528 – 533, June 2019.

[8] Aligon, J., Gallinucci, E., Golfarelli, M., Marcel, P. and Rizzi, S. A collaborative filtering approach for recommending OLAP sessions. Decision Support Systems, 69, 2015, pp. 20-30.

[9] Wang, J., Huang, J.Z., Wu, D., Guo, J. and Lan, Y. An incremental model on search engine query recommendation. Neurocomputing, 218, 2016, pp. 423-431.

[10] Wang, J., Huang, J.Z., Guo, J. and Lan, Y. Recommending high-utility search engine queries via a query-recommending model. Neurocomputing, 167, 2015, pp. 195-208.

[11] Song, W., Liang, J.Z., Cao, X.L. and Park, S.C. An effective query recommendation approach using semantic strategies for intelligent information retrieval. Expert Systems with Applications, 41(2), 2014, pp. 366-372.

[12] Bimonte, S., Boucelma, O., Machabert, O. and Sellami, S. A new Spatial OLAP approach for the analysis of Volunteered Geographic Information. Computers, Environment and Urban Systems, 48, 2014, pp. 111-123.

[13] Zhang, Z. and Nasraoui, O. Mining search engine query logs for social filtering-based query recommendation. Applied Soft Computing, 8(4), 2008, pp. 1326-1334.

[14] Giacometti.A, Marcel. P, Negre. E, Soulet. A. Query recommendations for OLAP discovery-driven analysis. IJDWM 7(2), 1–25 (2011)

[15] Diansheng Guo, Jeremy Mennis, Spatial data mining and geographic knowledge discovery - An introduction, Computers, Environment and Urban Systems 33, 403–408, 2009.

[16] Min-Ju Kyung, Jae-Hong Yom, Seung-Yong Kim, Spatial Data Warehouse Design and Spatial OLAP Implementation for Decision Making of Geospatial Data Update, KSCE Journal of Civil Engineering, Surveying and Geo-Spatial Information Engineering, springer, P.1023, 2012

[17] Franklin.C. An introduction to geographic information systems: linking maps to databases. Database 15(2), 12–21 (1992)