CLUSTER ANALYSIS FOR OLAP IN ONLINE DECISION SUPPORT SYSTEMS

Kiruthika S1, Umamaheswari E2, Karmel A3, Kanchana Devi V4

1Department of Computer Science and Engineering, Sona College of Technology, Salem.
kiruthika.cse@sonatech.ac.in
2Associate Professor Grade - II, Center Faculty - Cyber Physical Systems, Vellore Institute of Technology, Chennai, 600 127, Tamilnadu, India umamaheswari.e@vit.ac.in
3Associate Professor Grade - I, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, 600 127, Tamilnadu, India Karmel.a@vit.ac.in
4Associate Professor Grade - I, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, 600 127, Tamilnadu, India kanchanadevi@vit.ac.in

ABSTRACT

Online Decision Support Systems serve people of different segments of the society viz., traders, business enthusiasts, entrepreneurs, etc., Decision Support Systems are used in almost every business segment, spanning from micro-level to High level and even Heavy industries. This phenomenon is one of the results of Globalization. Proper decisions have to be made in every business industry to cope up with the prevailing market conditions. These Decision support systems are fed with large volumes of data. These data volumes are generally huge and heterogeneous. Cluster Analysis is a technique used to find out the group of data which is similar to each other and dissimilar to other data. This technique is very important for the efficient functioning of any Online Decision Support System. This paper focuses on using the Cluster Analysis Techniques to make the Decision Support Systems to work efficiently.

Keywords: Online Analytical Processing, Decision Support Systems, Cluster Analysis.

I. INTRODUCTION

The role of Decision Support Systems has become vital in almost all industries world-wide. Taking appropriate decisions in every walk of business is the first right deed, done by any business concern. Expecting a success without taking the right decision is really not viable. Deploying an efficient Decision Support System is the significant task in any business. Data used in these Decision Support Systems are generally in the form of extremely large cubes called Hypercubes [9]. The data in these hypercubes need proper analysis and categorization. Online Analytical Processing (OLAP) has to be done on these huge data cubes. Cluster analysis is being used as a pre-processing technique for the categorization of these huge data. The OLAP system used here is the Multidimensional OLAP system.

II. CLUSTER ANALYSIS

Cluster analysis is the technique of finding groups of objects – Objects are similar to one another in the same group and are dissimilar in the objects in the other groups. This technique is generally used in Unsupervised Learning strategy in Data mining [6]. The objects are generally unlabeled, as the Fig.1 depicts.
Cluster analysis is finding similarities between data according to the characteristics found in the data and grouping similar data objects into clusters. A lot of web data has to be analyzed and clustered. Techniques depicted in [7] can be employed for caching web data and then can be applied clustering techniques.

**Need of Cluster Analysis in DSS**

This powerful technique has a handful of outcomes, when used in any unsupervised learning systems. Starting from Hypothesis Generation and testing to finding the K-nearest neighbour, this technique has its own applications. We focus on Data reduction and finding the K-nearest neighbour, so as to analyze the huge data cubes. Data reduction varies from drill down/up technique which was discussed in previous proposed systems [9].

**Applying Cluster Analysis in Data Cubes**

OLAP systems use very large sized unlabeled data in the form of data cubes [5], in which the analysis has to be done. Decisions are made based on these analyses and are made available to the users. The most important aspect of this analysis is the control over the size of the data cube. Reducing the size is done by cluster analysis technique, whereas the relevance can also be obtained.

**Data Reduction**

Connectivity clustering technique is used in this system. This technique initially considers each data point as its own cluster [4]. It is an iterative technique, where points closer to each other are more related. Thus analysis is done upon more accurate data, to get more precise decisions. This connectivity clustering technique [7] is best suited for OLAP systems, as the hierarchical decomposition is done to reduce the number of irrelevant connections are eliminated as shown in Fig.3. As compared with techniques such as Centroid, Density and
Distribution clustering techniques, Connectivity clustering serves good or OLAP based Decision Support Systems.

![Connectivity Cluster Analysis technique](image)

**Finding the K-nearest neighbour**

The search identifies the top k nearest neighbours to the query. It classifies a point based on the consensus of its neighbours. This is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure. The business data, which is considered as the background data, has a lot of heterogeneity. Honoring the heterogeneity, we may have to find out the real connectivity on these data to utilize the data completely, as shown in Fig.5.

**III. PROPOSED SYSTEM**

The proposed system uses the ETL processed background business data, which are in the form of huge data cubes.

![Online Business Decision Support System](image)

The Extract, Transform and Load operations are carried out upon the underlying data. The data cubes obtained using these operations are very huge and the relevancy is also considered to be limited. The OLAP tools are used to carry out preliminary operations, so as to obtain certain degree of relevancy. Following the Online Analytical Processing operations [8], Clustering Analysis is done. The queries are obtained from the users and the requested business decisions are given out in the same User Interface.

The user interface is designed such that the user places the query to the Business Decision Support System. The OLAP operations are carried out to obtain the huge data cubes, upon which the Clustering analysis is done – where the data size is reduced and the KNN operation is also done to obtain the relevancy of data to be analyzed.
The user interface is designed such that the system also reads the expertise level of the user. Keeping all these criteria and supporting data in hand, the analysis is carried out. The more relevant and appropriate decision is given out to the user through the same interface. E-mailing the decisions for the future reference by the users can also be done, if requested.

REFERENCES


Dr. S. Kiruthika is working as an Assistant Professor in the Department of Computer Science and Engineering at Sona College of Technology. She has completed her PhD in Information and Communication Engineering at Anna University in 2016. She has completed her ME from Anna University in the year 2006 and her BE from Bharathiyar University in the year 2003. Her Research interests include Natural Language Processing, Data Science and Business Analytics.
Dr. E. Umamaheswari is an Associate professor Grade II, Centre for Cyber Physical Systems, VIT University, Chennai. She completed her PhD in Anna University, 2015. Her area of specialization includes, Software Engineering (Testing, Metrics), Cloud Security, ERP, Machine Learning, Network Security, IoT. She has published various papers in SCI, Elsevier, Springer indexed journals. She is guiding more number of research scholars in the said specialization. She is an active ACM member and IEEE member. She is a gold medalist and a star performer in her master of science. She is an editor for many journals in IGI global, Wiley publications etc., She is a reviewer in many journals to name a few Mobile Network and Applications. She was a PhD and Poster Track Chair in ICBCCC from 2016 onwards, ICRTAC from 2019 onwards. She is a reviewer in Journal of software, Future generation of computer system, MONET, wireless network, IGI global. She is also an editor of International Journal of web portal.

Dr. Karmel Arockiasamy is an Associate professor, School of Computing Science and Engineering, VIT University, Chennai. She completed her PhD in Anna University, 2016. Her area of specialization includes, Mobile Adhoc Networks, Network Security, IoT and Cyber Security. She has published various papers in Scopus indexed journals and guided more number of projects. She is an active ACM member. She is a star cyber secure user-R11; She received best faculty award in the year 2011-2012. She is a reviewer in many journals to name a few Wireless Network, Mobile Network and Applications. She was a PhD and Poster Track Chair in ICBCCC’18.

Dr. Kanchana devi V, is a Distinguished Professor in Vellore Institute of Technology, Chennai Campus, Tamilnadu, India. In March 2006, she received her Bachelor Degree in Computer Science and Engineering (College Topper), Mount Zion College of Engineering and Technology, Tamilnadu, India. In May 2009, She received her Master Degree in Computer and Communication (Gold Medal - Anna University) in Sri Sairam College of Engineering, Tamilnadu, India. In October 2019, she received her Ph. D in Computer Science and Engineering, Vellore Institute of Technology, Chennai Campus, Tamilnadu, India. Her research interests include Wireless Sensor Networks, Internet of Things, Cyber Physical Systems, and Distributed Computing.