

Discovering Outliers from spatiotemporal data in large databases

In modern applications, Spatiotemporal databases are growing rapidly both in size and number. Users are necessary to access the data which have Spatial as well as Temporal information. Spatiotemporal data gives information about current location of an object on the Earth's surface in given time. Users need to identify the objects which behaves dissimilar to other objects. Strangeness behavior of an object is considered as an Outlier. Tracking system of Moving Objects.

Ex: National Train Enquiry System (NTES)
Air Cargo Tracking System (ACTS)

Today's rapid advancements in computing hardware, tracking devices such as GPS receivers and sensors have become pervasive and generating a large amount of spatiotemporal data. Measurements of temperature, pressure, air quality, traffic data etc., GPS data from mobile phones and data from radars that capture location information about people and other moving objects such as cars, trains produce a large amount of data. spatiotemporal data includes in the areas like Geological disaster monitoring, Geophysical exploration, predicting earthquakes and hurricanes, determining global warming trends, and public health (disease spread) .

Many indexing methods are proposed to index spatial, temporal and spatio-temporal data.

A spatiotemporal outlier is an observation whose attribute values are significantly different from those of other spatially and temporally referenced objects in a spatio-temporal neighbourhood.

For efficient retrieval of data, many Indexing techniques are used. The indexing techniques reduce the frequent disk access. COVID-19, COVID effected people information in different cities in a particular given time which gives large amount of data, history about recovered patients' information and updating data requires a proper storage structure.

Problem Statement :

Detecting outliers in spatiotemporal data by preprocessing this data using different encoding schemes, converting three-dimensional data into one dimensional data for indexing.

Discovering Outliers from spatiotemporal data in large databases

Objectives :

Collecting spatiotemporal data that need to be preprocessed and stored in database.

Generating one dimensional data from 3D space filling curves for indexing.

Proposed a hybrid Indexing method to reduce disk access time.

Evaluate query performance by point, range, time-slice and KNN Queries.

Finding outliers in spatiotemporal data.

Methodology:

Data need to be collected and store it in database after pre-processing technique (Data cleaning).

One dimensional data is going to be indexed from 3D data by using space filling curves.

A Hybrid Indexing technique is proposed to reduce frequent disk access.

Query processing is going to perform by query type and need to be identified about its similarity feature.

Detecting an object which behaves differently from other objects.

Dataset :

Taxi Service Trajectory - Prediction Challenge, ECML PKDD 2015 Data Set (UCI Repository).

Taxi Service Trajectory data set [4,5,39] An accurate dataset describing trajectories performed by all the 442 taxis running in the city of Porto, in Portugal. It identified different trajectory points of moving taxi. Partial trajectories are identified with latitude and longitude values. If any vehicle is moving out of the specified trip destination, it is identified as Outlier.