ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 12, April 2018

A SURVEY ON INSIGHT OF SPATIAL DATA AND ANALYTICS

K.Bommi [1], D.J. Evanjaline [2]

[1]Research Scholar, Department of Computer Science, Rajah Serfoji Govt. College(Autonomous), Tanjore, Bharathidasan University, Trichy, India.

[2]Assistant Professor, Department of Computer Science, Rajah Serfoji Govt. College (Autonomous), Tanjore, Bharathidasan University, Trichy, India.

bhommisankar@gmail.com
evanjalinedj@gmail.com

Abstract--- **This paper documents the close study of geospatial data analytics review to bring out traditional geospatial data handling theory and methods giving more concentration in the process such as storing, managing, processing, analyzing, visualizing, and verifying the quality of data. Existing data handling theories and methods are not enough to handle emerging geospatial big data. It is an extremely large data set that give the attention of academia, industry, natural disaster, agriculture and other organization.**

Keywords—Big data, Spatial data, GIS, Data handling, Spatial Data Analytics.

I. INTRODUCTION

Spatial thinking described as “spatial literacy”, and addresses the central issues and problems associated with spatial data that need to be considered in any analytical exercise. In practice, real-world applications are likely to be governed by the organizational *practices* and *procedures* that prevail with respect to particular *places*. Different data are collected, assembled and disseminated like general purpose censuses versus statistical modelling of social surveys, property registers and tax payments. There are also differences in the ways in which different data holdings can legally be merged and the purposes for which data may be used particularly with regard to health and law enforcement data. Some organizations, such as the US Geological Survey, are bound by statute to limit charges for data to several costs such as media used for delivering data while others, such as most national mapping organizations in Europe, are required to extract much heavier charges in order to recoup much or all of the cost of data creation. In spatial data, because of its volume and variety, it is difficult to extract knowledge and perform decision making in an efficient time. According to the International Data Corporation, the digital data could increase by 40 times from 2012 to 2020. So, we should develop tools to handle huge amount of big spatial data. GIS is the framework for gathering, managing, and analysing data. This is the root of the Geography Science. It analyses spatial location and visualize information using map and 3D scenes.

II. BIG DATA

publications and in many conferences. According to Manyika et al. 2011, Big data suggested as a predominant source of innovation, competition and productivity. The

“data stream” become a “flood of data” from different kinds of sources like social networks, sensor, satellite images, messaging system and online resources.

III. WHY SPATIAL DATA ANALYZES TODAY?

Spatial data provides a lot of benefits during climate change, disease surveillance, criminology, intelligence utility and defense etc., They are governed by organizational practices and procedures with respect to particular places. Data are legally merged and used particularly with regarding to health and law of enforcement. GIS applications represents unique location on earth surface at the same time cost effective. The time has come for us to know the importance of Geospatial Data analysis and its benefits to the betterment of the society where we live in.

IV. SPATIAL DATA AS BIG DATA

Laney (2001) [1] first proposed three dimension that characterize the challenges and opportunities of increasingly large data volumes.3 VS

1. Volume
2. Velocity
3. Variety

In additions to the above, another V, which represents veracity has been added to describe data integrity and quality. Further V’s are variability validity, volatility, visibility, value or visualization. These deal with important concepts related to entire pipeline of big data collection, processing & presentation.

In Suthaharan (2014) view, he argued that 3VS cannot support early detection of bigdata characteristics for its classification and proposed 3 CS:

1. Cardinality.
2. Continuity.
3. Complexity.

Let us analyze spatial data systematically,

Volume, Variety and Velocity.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 12, April 2018

The term "big data" first appeared in the mid1990, attained popularity around 2008 and got recognized in 2010. Now it is a buzzword on internet, trade, scientific Volume: Satellites observes rich information. These images with increased resolution show increase in size of the image.

2. Variety: It consists of 3 basic modules.
   a) Raster - image(satellite).
   b) Vector - encompassing points, lines &polygons.
   c) Graph - spatial network.

Other than the above mentioned, there are different types of formats available to store spatial representation. They are,

- Geo Tiff - Geotagged Image File Format(.tif).
- IMG - (JPG, JPEG)
- HDF - Hierarchical Data Format (.hdf5)
- NETCDF - Network Common Data Form (.nc)
- BIL - uncompressed file with the actual pixel of image.
- AAIGRID - ASCII Interchange format for Arc/Info Grid (.prj)
- VECTOR - (EPS, SVG, PDF, AI, DXF, JPEG, JPG, PNG, BMP etc)
- TWITTER - Geo located tweets.
- TEXT - Geo located text.
- GIS - Images.

3. Velocity: The real - time monitoring of earth or any other objects means a continuous flow of data.

4. Veracity: Level of accuracy varies depending on data sources, thus varying in accuracy causes issues in quality assessment of source data and has to be improved statistically.

Geospatial data are collected using ground surveying, photogrammetry and remote sensing. Now it is done through raw scanning, mobile mapping geo-located sensors, geo-tagged web contents. Volunteer geographic information(VGI) global navigation satellite system (GNSS) tracking & so on. According to spatial data first four V's are very important and more fundamental. Further V's are described in the following

5. Volume: Continuous increase of data of VGI, location based social media data, ever increasing real time sensor observation raise not only data storage issues but also a massive analytical issue.

6. Variety: As the name suggest variety, it includes variation of data, such as map data, geotagged text data, structured and unstructured data, raster & vector data calls for more efficient models, indexes and technologies.

For example, use of NOSQL

7. Velocity: Imagery data with frequent revisits at high resolution, continuous streaming of sensor observations, internet of things (IOT) real-time GINSS trajectory & social media data all require matching the speed of data generation and the speed of data processing to meet demand (Dasgupta 2013).

8. Visualization: Here the analysis can identify patterns [illegible] and relationships that emerge from big data analysis.

9. Visibility: Geospatial big data is efficiently accessed and processed by cloud computing and cloud storage. Big data and cloud go hand in hand in reinforcing technologies.

V. COLLECTION OF GEOSPATIAL BIG DATA

Collection occurs though data acquired by the public, so called volunteered Geographic Information (VGI) and data from geo-sensor network resulting in increased availability of spatial information. Now, authoritative datasets were dominating in topographic domain, these new data types extend and enrich geographic data types extend and enrich geographic data in terms of thematic variation and by the fact that it is more user-centric. The latter is especially true for VGI collected by social data (Sester et al. 2014).

Geospatial data collection is a commodity implemented in everyday devices, which are capable of acquiring information at an unprecedented level with respect to geometric and temporal resolution and thematic granularity. Its benefit is it easy to handle, small and able to acquire data even unconsciously.

Geospatial data collection is shifting form a data sparse to a data rich paradigm. This is similar to the situation in topographic data collection for digital terrain models by capturing significant topographic points with morphological characteristics on the one hand ("qualified" points ie,. points with semantics) - as opposed to the collection of point clouds using LiDAR sensors or stereo matching, leading to masses of "unqualified" points (Ackermann 1994). It has two approaches.

First Approach: It require manual selection and measurement and guarantees that the topographic reality can be interpolated by from the sparse measurements.

Second Approach: It assumes that the topographic reality can be captured by the dense measurements and constructed from them. Thus, object formation and identification are shifted to analysis process.

Let us distinguish the following sensor configurations:

1. Objects equipped with sensors moving through space and capturing their own trajectories and local environment. For example, Humans & moving devices like car heavy vehicle etc,
2. Static sensors constantly observe the changing environment. For example, Climate

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

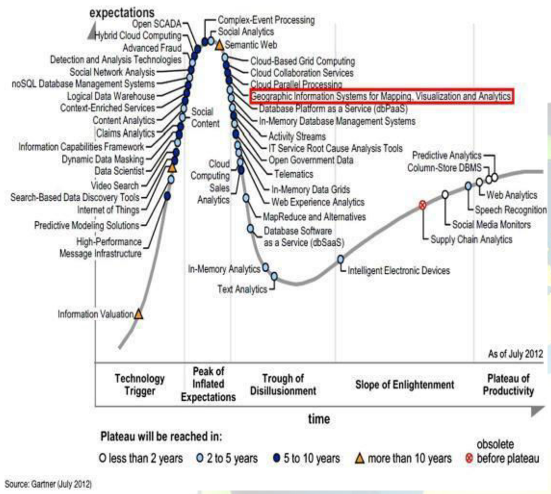
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 12, April 2018

change, Coastal plain area change.

Data acquired by the above mentioned new sensors and new stake holders can be characterized as,

- data streams
- arbitrary high density
- close sensing
- different degrees of positional accuracy.

VI. SIGNIFICANCE OF GEOSPATIAL BIG DATA



Due to modernization and advanced development in computing data analyses, its significance is abundant. For example, instrumentation of cities provide vast amount of real - time data through the likes of smart card ticketing systems, vehicle tracking devices, CCTV, toll systems, inductions loops and other sensors. Though twitter (Social media) we can geotag and use it during disaster management and emergency relief. From environmental perspective, Landsat repositories from NASA provide petabytes of geospatial data (Riebeek 2015). Geospatial big data assist in public health centers. The data are routinely collected and stored. For example patient with doctors or health centers. "Sensors" help to capture traffic or mobility related VGI style information which include participatory data and opportunistic data.

1. Participatory data: This is conducted in a conscious process by a user, who selects objects & their features and give information.

e.g.: Open Street Map

2. Opportunistic data: This occurs unconsciously, with no specific purpose or even a completely different purpose.

e.g.: Exploitation of mobile phone to determine traffic information.

VII. TYPES OF SPATIAL DATA

Spatial data can be discrete or continuity data of space on which the variables are measured. A topology of spatial data based on the conception of space is provided by Manfred et al. [20]. The four types of spatial data are:

- **Point pattern data**: A data set consisting of point locations.
- **Field data / Geostatistical data**: The data set consisting of variable which are conceptually continuous fixed set of points.
- **Area data**: The data set is a fixed set of area objects that may form a lattice.
- **Spatial interaction data**: The data set is a pair of point locations or a pair of areas.

VIII. TAXONOMY FOR ANALYSIS

The large volume of geospatial data raises the question of efficient processing architectures or data processing methodologies for acquiring useful knowledge. Hence, in this study, the available literature is studied based on:

- Spatial data models/infrastructures.
- Spatial data analytics platforms / data processing frameworks/systems.
- Algorithms for spatial data.
- Types of Spatial Data Analysis
- Applications of spatial data.

A. Spatial Data Models/Infrastructure

The available literature for spatial data model is more concentrated towards improving the spatial query performance and throughput. Wang et. al., [2] propose a new geospatial data model X3DORGDM (X3D-based Oriented Relation Geospatial Data Model). It aims to meet the requirements of geo visualization. As a data model, X3DORGDM consists of three components:

1) a collection of geographic data types.

2) a collection of operating algorithms.

3) a collection of integration and consistency rules to define the consistent geo-database or change of state or both.

X3D-based ORGDM has been implemented based on several open source packages like POSTGRESQL, OpenGIS Simple Features Data Model, Computational Geometry Algorithm Library(CGAL), Geographic Data Abstraction Library(GDAL) and PROJ.

Lacasta et al. [3] propose a process to construct a Linked Data model of geospatial resources that allows semantic searching and browsing. There are some initiatives by the standardization bodies Open Geospatial Consortium (OGC) and the International Organization for Standardization

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 12, April 2018

(ISO) to standardize the way geospatial information is created, provided and transformed according to the user needs. However, there are issues like: geo-service creators have to manually describe and provide annotations for their services. The publicly available geospatial catalogue has to be manually annotated, for a search of spatial data to yield better results. This work proposes a methodology that combines and adapts a set of information retrieval and natural language processing techniques to the geospatial web service context. It also shows how to use these techniques to create an automatic system that can identify, classify and interrelate and facilitate the access to geospatial web services.

Chi-Ren Shyu et al. [4] propose a coherent system: GeoIRIS, that allows image analysts to rapidly identify relevant imagery. GeoIRIS ability to answers analyst's questions such as given a query image, show database satellite images that have similar objects and spatial relationship that are within a certain radius of landmark. Their architecture consists of modules namely: Feature Extraction(FE), Indexing Structures(IS), Semantic Framework(SF), GeoName Server(GS), Fusion and Ranking(FR), and Retrieval Visualization(RV). They use Tile based feature extraction and Object based feature extraction for feature extraction. Indexing of continuous valued features based on Entropy Balance Statistical (EBS) k-Dimensional (k-D) tree and indexing the binary-valued features is performed with the Entropy Balanced Bitmap (EBB) tree. It also proposes novel approaches for information ranking, semantic modelling and advanced queries.

S.Roy et al. [6] discuss the metadata issue related to Spatial Data Infrastructure; and they attempt to propose a three-tier infrastructure towards the enhancement of metadata catalogue services in regards of three aspects.

1) Incorporating of various geographic information metadata elements and provision of necessary support for spatial data infrastructures;

2) Achieving interoperability between different metadata standards and those essential for spatial data infrastructures;

3) Enhancement of information retrieval techniques for spatial data infrastructure using disambiguated vocabularies.

B) Spatial Analytics Platform

A geospatial analytics platform uses Hadoop, Hive, other NoSql technologies along with Relational Database to process geo spatial data. The architecture is a proposal and is not available as an implementation. The proposal also describes the required functionality of different components of the architecture. The architecture as shown in is below:

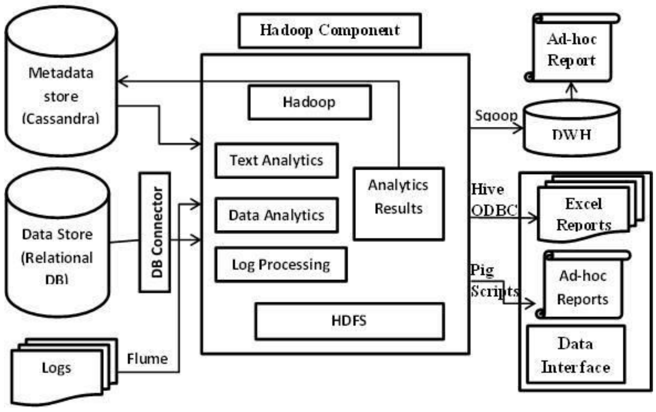
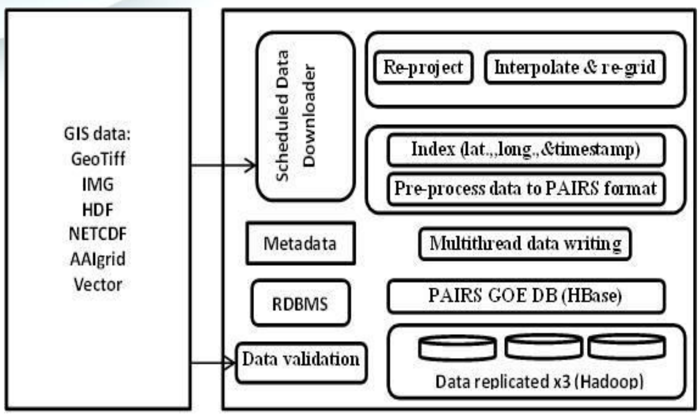


Fig. 1. Logical Architecture of Big Data Platform

C) Data Processing Frameworks

Klein et al. [7] proposed Physical Analytics Integrated Repository and Services (PAIRS) a scalable geo-spatial data analytics platform. It enables rapid data discovery by automatically updating, joining and homogenizing data layers in space and time. It helps in automatic data download, data curation and scalable storage and a computational platform for running physical and statistical models on the curated datasets.

It claims that the key differentiator is its capability in multilayer query, to search multiple data layers and filter based on multiple search criteria. It uses HBase to store data and index built is based on latitude, longitude and timestamp. It uses open source tools to convert data layer projections into WGS84 co-ordinate system. It helps to manage data from multiple sources in a scalable fashion on distributes compute resources. The main component of PAIRS is its Data Integration Engine. It can download, re-project and index data. If a data format is not raster, then it rasterized and handled as a large matrix.



ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 12, April 2018

Fig. 2. Data Integration Layer of PAIRS

XinChen et al. [8] propose a high performance integrated spatial big data analytics framework based on MapReduce paradigm and present a few use cases. Their data integration mainly focuses on spatial datasets that relies on these following query types:

1) Point-in-polygon queries;

2) Cross-matching queries;

3) Nearest neighbour queries.

These kinds of queries are both data and compute intensive. Hence their spatial data integration is based on extending Hadoop- GIS with scalable spatial clustering and spatial regression capabilities.

Data quality that can be characterised through,

- Accuracy
- Completeness
- Timeliness and
- Consistency of data.

In the data integration of layers of PARIS, Spatial Big Data have to take an account to provide appropriate data quality. PARIS have the feature of capability for fast cross-layer data discovery.

COMPARISON OF HADOOP AND SPARK BASED SYSTEMS [9] [10][11][12][13]

	Hadoop-GIS	SpatialHadoop	GeoSpark	SpatialSpark	STARK
Query Language / DSL	Yes	Yes	No	No	Yes
Spatio-Temporal Data	No	No	No	No	Yes
Spatial Partitioning	Yes	Yes	Yes	Yes	Yes
Indexing	Yes	Yes	Yes	Yes	Yes
Persistent Indexes	Yes	Yes	No	Yes	Yes
Filter				no partitioning	
Contains	Yes	Yes	Yes	Yes - w/o Index	Yes
ContainedBy	Yes	Yes	No	Yes - w/o Index	Yes
Intersects	Yes	Yes	Yes	Yes - w/ Index	Yes
WithinDistance	Yes	Yes	No	Yes - w/o Index	Yes
Join	Yes	Yes	Yes - pred. limit.	Yes - returns IDs	Yes
k Nearest Neighbors	Yes	Yes	Yes	No	Yes
Clustering	No	No	No	No	Yes

D) Systems

David Haynes et al. [15] propose Terra Populus a

6. system that provides three web applications that allows to access, analyze and tabulate different datasets under a common platform.

1. Paragon is a prototype parallel spatial database that extends the functionality of PostgreSQL and PostGIS onto multimode systems.
2. Terra Populus tabulator application builds dynamic queries for analyzing large population survey data.
3. Terra Explorer is an exploratory analysis tool for visualizing the spatial datasets within the repository.

Barik et al. [16] propose a Fog Computing based framework called FogGIS for mining from geospatial data. It is built as a prototype using Intel Edison, an embedded microprocessor. This work claims the following contributions:

- FogGIS framework proposes improved throughput and reduced latency for analysis and
transmission of geospatial data.

- Different compression techniques for reducing data size, thereby reducing transmission power.

Bosch et al. [17] propose a geo-spatial document analysis system for the VAST 2011. Their system equips the user to interact with the data in a visual, direct and scalable fashion, which offers diverse views and data management components. Wang et al. [25] have proposed TerraFly Geo-Spatial Cloud platform. This system provides comprehensive spatial analysis methods and visualization.

E) Algorithms for Spatial Data

Depending on the type of conception of space and the measurement level, different algorithms are applied. The widely used data mining algorithms for spatial data are:

1) Spatial Auto-regressive model
2) Markov Random Field model
3) Geographically weighted regression
4) Fractal models
6) Map-Reduce algorithm for polygon retrieval
The General Spatial Interaction Model

F) Types of Spatial Analysis

Spatial analysis is an analytics technique to determine the distribution of the spatial variable, the relationship between the spatial variable, and association of the variables of an area. It refers to the analysis of phenomena distributed in space and having physical dimensions. It is the process of creating new information about a set of geographic features to perform routine examination, assessment, evaluation, modelling, analysis.

It has five basic types. They are,

1. Spatial overlay
2. Contiguity analysis

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 12, April 2018

3. Surface analysis
4. Linear analysis
5. Raster analysis

a. Spatial Overlay

Spatial overlay is a set of methods for transferring data between object in drawings based on their spatial relationships to each other. For example, suppose we have a drawing that shows areas and we have also a set of points showing the locations of cities. Suppose that each city record has a field called population, that gives the population in that city, but that we have no values for the population for each state. We could use spatial overlays to automatically add up the values of the population for all cities in each state and place, that combined into the population field for each field. It transfers data from field in a source object set to a target object set using some method.

b. Contiguity analysis

It is used to study spatial structures contained in seismic- images. It is more efficient for multivariate description and spatial filtering of this kind of images. Contiguity measures evaluate characteristics of spatial units that are connected. These units share one or more characteristics with adjacent units and form a group the term UNBROKEN is the key concept. Different adjacent features may have more than one attribute but they must all have a COMMON attribute to be considered as reflecting contiguity. Contiguity is used to measure shortest and longest straight-line distances across and area and to identify areas of terrain with specified shape characteristics. For example, an area of continuous pasture land with an area of on more than 10 acres with no part it wider than the sound of the ache pig call can be heard.

c. Surface analysis

Surface analysis is the use of microscopic chemical and physical probes that give information about the surface region of a sample. The probed region may be the extreme top layer of atoms, or it may extend up to several microns beneath the sample surface, depending on the technique used. The analysis is providing information about such characteristics as the chemical composition, the level of trace impurities, or the physical structure or appearance of the sampled region

.

d. Linear analysis

The finite element method is the ideal tool for solving static and dynamic problems in engineering and the sciences. Linear analysis assumes linear elastic behaviour and infinitesimally small displacements and strains. To establish a model, it is necessary to become familiar with the finite element methods.

1. Assume that we have a beam with load 'W' acting on it causing a moment 'M', now if we increased the load to '2×W', Then the moment will be '2×M'.
2. Assume that we have a beam with certain dead load & live load, which are causing a deflection in a beam by 10mm & 5mm respectively. If we combine the deflection caused by two loads will be 15mm.

Now what we do in in above Examples is we related the terms directly i.e. load vs. moment in first example &whereas load vs deflection in second example. This called as linear approach. If we combine the deflection caused by two loads will be 15mm.

e. Raster analysis

A **raster data** structure is based on a (usually rectangular, square-based) tessellation of the 2D plane into cells. In the example the cells of tessellation A are overlaid on the point pattern B resulting in an array C of quadrant counts representing the number of points in each cell. For purposes of visualization a lookup table has been used to color each of the cells in an image D. Here are the numbers as a simple vector in row/column order:

1 3 0 0 1 12 8 0 1 4 3 3 0 2 0 2 1 7 4 1 5 4 2 2 0 3 1 2 2 2 2 3
0 5 1 9 3 3 3 4 5 0 8 0 2 4 3 2 8 4 3 2 2 7 2 3 2 10 1 5 2 1 3 7

Finally, here is a run-length encoded representation of the raster, which has 55 positions:

values: 1 3 0 1 12 8 0 1 4 3 ...
lengths: 1 1 2 1 1 1 1 1 1 2 ...

This process clearly results in a loss of information, from the real-valued coordinates of the points, through the integer cell counts, to the ordinal colors, but there are also gains:

The data structure is usually more compact,

The raster is easy to visualize, and It can be related to other raster provided the locations and resolutions are properly conflated.

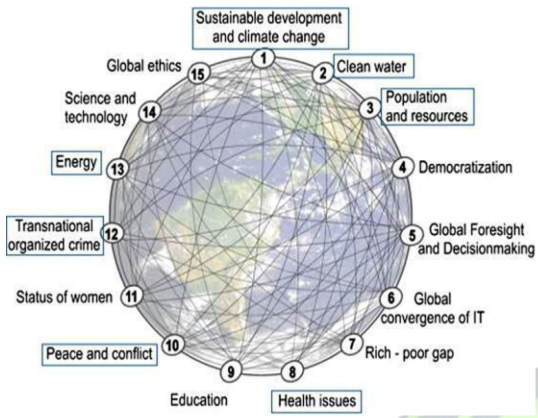
Analysis Environment Some GIS packages require that all input raster layers have the same *resolution* (cell size) and *extents*. Some of the more recent GIS software allow for layers with a variety of cell size and extents to be analyzed. In this case it is important to be sure to set these parameters for the output layer prior to doing the analysis.

G) Applications of Spatial Data

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 12, April 2018

Global challenges facing humanity



1. GIS in Mapping.
2. Telecom and Network Services.
3. Hot spot Analysis.
4. Urban Planning.
5. Transportation Planning.
6. Environmental Impact Analysis.
7. Weather Analysis for Disaster management.
8. Determine land use/land cover change.
9. Navigation.
10. Crime Analysis.
11. Disease Analysis.
12. Natural Resource Management.
13. GIS in Geology.
14. Tourism Information System.
15. Deforestation. etc.,

IX.SPATIAL ANALYSIS SOFTWARE

Spatial Analysis Techniques in Popular Software

Technique	Software Tool
Co-location	Mining Oracle10g
Spatial clustering	ArcGIS 9.3 Spatial Statistics tool, Oracle 10g, CrimeStat, Terra Seer
Spatial hotspots	ArcGIS 9.3 Spatial Statistics tool, CrimeStat, GeoDa
Spatial outliers	ArcGIS 9.3 Spatial Statistics tool, GeoDA
Spatial network hotspots	CrimeStat, SANET
Kriging Spatial autoregression	ArGIS 9.3 Geostatistical Analyst, S+ Spatial Stats, fields package and geoR in R
onditional autoregression	S+ Spatial stats, GeoDa CrimeStat
Geographically weighted regression	ArcGIS 9.3

a) Co-location Mining

Mining the presence of two or more spatial objects are in the same location or they are in close distance to each other. There are two types

1. Co-location of items.
2. Co-location with thematic layer.

b) Spatial Clustering

Spatial clustering which groups similar spatial objects into classes. It is an important concept in spatial data mining. (Han and Hamber 2000). Ex. Similar weather pattern identification.

c) Spatial Hotspots

Hotspots analysis tool assesses whether high or low values cluster spatially. For ex., Accident severity, the number of crimes.

d) Spatial Outliers

A Spatial Outlier is spatially referenced object whose non-spatial attribute values are significantly different from those of other spatially referenced objects in its spatial neighbourhood. For ex., Anomaly.

X. USES SPATIAL DATA

1. To refine sales and marketing.
2. To upgrade asset management.
3. To augment situational awareness and intelligence.
4. To hone risk analysis.
5. To enhance transportation and logistic planning.
6. To sharpen strategic location determination.
7. To improve fraud detection and prevention.

XI. OPEN PROBLEMS AND CHALLENGES

This study reviewed a diverse of theory and methods for geospatial data analysis. Given the unique characteristics of spatial data the following were identified as some of the existing open challenges:

1) Algorithms to handle spatial streaming data
2) New spatial data indexing schemes
3) Volunteered Geographic information systems
4) Mobile mapping and location based services
5) Object based data models for continuous data

The GPS (Global Positioning System) traces of

Public Transit Systems which can be considered as spatial stream data are being stored in the huge volumes in their databases. Newer indexing techniques are required to store and retrieve the data back efficiently. These kinds of datasets may include a time series of attributes such as vehicle location, vehicle speed, fuel levels, emissions of greenhouse gases etc.

XII. FURTHER DEVELOPMENT

This study reviews a variety of geospatial theory and methods that have been used for traditional data but that can be extended to handle geospatial big data.

- The development of new spatial indexing and algorithms to handle real-time, streaming data and to support topology for real-time analytics.
- The development of conceptual and methodological approaches to move big data from descriptive and correlation research and applications to ones that explore casual and explanatory relationships.
- The development of efficient methods to display data integrated in the three dimensions of geographic and one dimension of continuous time. There is a strong need in understanding human capacity to deal with visual information and identifying which visualization type is a good fit for the task at hand, and the target user group. Furthermore, interdisciplinary studies and communication is of critical importance. The advances in scientific visualization and information visualization are both beneficial to geographic visualization; but geographic visualization has also a lot to offer to other domains. Novel visualization paradigms, especially developed for big data tend to be information-rich (thus complex); therefore, we find that highlighting and summarizing approaches should be further investigated. Additionally, and in relation to managing complex visualization displays, technology research in terms of level of detail management remains important.

The development of novel approaches for error propagation so as to effectively assess data quality requires. The challenge is not only the handling the many different types of data for real-time analytics, but rather the ad-hoc combination of data streams in real-time, which may include the capture of the "whole" picture (or "complete population") instead of sampling a small portion of the whole population. In this case quick assessments are preferable that may come out of varying the input data and simulating variability.

XII. CONCLUSION

Geospatial big data presents both challenges and opportunities. This paper outlines some of these challenges from a technical and conceptual perspective, and also provides priority areas that need to be addressed in the future. Once Geospatial big data research matures, the opportunities for overall societal management and decision-making become enormous.

REFERENCES

[1] Laney, Douglas. *3D Data Management:* Controlling Data Volume, Velocity and Variety. Gartner. Retrieved 6 February 2001.

[2] Tao Wang and Xi Chen and An-ming Bao and Wei-sheng Wang, *A new geospatial data model to facilitate geographic data mining and knowledge discovery,* IEEE International Conference on Systems, Man and Cybernetics,3642-3646,2008

[3] J. Lacasta and F. J. Lopez-Pellicer and W. Renteria-Agualimpia and J. Nogueras-Iso, *Improving the visibility of geospatial data on the Web*, IEEE/ACM Joint Conference on Digital Libraries, 155-164, 2014

[4] Chi Ren Shyu and Matthew N. Klaric and Grant J. Scott and Adrian, S. Barb and Curt H. Davis and Kannappan Palaniappan, *GeoIRIS: Geospatial Information Retrieval and Indexing System Content Mining, Semantics Modeling, and Complex Queries,IEEE Trans. Geoscience and Remote Sensing,* 45, 4, 839–852, 2007

[5] A. Subramanian, *IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Empowering geo spatial analysis with big data platform*: Natural resource management, 1-6, 2016

[6] S. Roy and S. Das, 2015 4th International Symposium on *Emerging Trends and Technologies in Libraries and Information Services, Spatial data infrastructures: Its metadata and analysis*, 43-51, 2015

[7] Klein, Levente J. and Marianno, Fernando J. and Albrecht, Conrad M. and Freitag, Marcus and Lu, Siyuan and Hinds, Nigel and Shao, Xiaoyan and Rodriguez, Sergio Bermudez and Hamann, Hendrik F., Big Data, IEEE, *PAIRS: A scalable geo-spatial data analytics platform,* 1290-1298, 2015

[8] Xin Chen and Hoang Vo and Fusheng Wang, BigSpatial'14 3rd ACM SIGSPATIAL International Workshop on Analytics for *BigGeospatial Data, High Performance integrated spatial big data analytics,* 11-14, 2014

[9] Stefan Hagedorn and Philipp Gotze and Kai-Uwe Sattler, International Conference on Extending Database Technology, *Big Spatial Data Processing Frameworks: Feature and Performance Evaluation,* 2017

[10] A. Aji, F. Wang et al., *Hadoop-GIS: A High Performance Spatial Data Warehousing System over Mapreduce, VLDB,* pp. 1009 -1020, 2013.

[11] A. Eldawy and M. F. Mokbel, *A Demonstration of SpatialHadoop:An Efficient MapReduce Framework for Spatial Data, VLDB*, 2013.

[12] S. You, J. Zhang, and L. Gruenwald, *Large-Scale Spatial Join Query Processing in Cloud, ICDE,* 2015.

[13] J. Yu, J. Wu, and M. Sarwat, *GeoSpark: A Cluster Computing Framework for Processing Large-Scale Spatial Data. SIGSPATIAL*, p. 70, 2015

[14] Jae-GilLee, Minseo Kang, *Big Data Research, Geospatial Big Data: Challenges and Opportunities,* 2, 74-81, 2015

[15] David Haynes and Suprio Ray and Steven M. Manson and Ankit Soni, *High performance analysis of big spatial data*, IEEE International Conference on Big Data, Big Data 2015, 1953-1957, 2015

[16] Rabindra K. Barik and Harishchandra Dubey and Arun B. Samaddar and Rajan D. Gupta and Prakash K. Ray, *FogGIS: Fog Computing for Geospatial Big Data Analytics,* CoRR, abs/1701.02601, 2017

[17] Harald Bosch and Dennis Thom and Michael Worner and Steffen Koch and Edwin Puttmann and Dominik Jackle and Thomas Ertl, ScatterBlogs: *Geo-spatial document analysis,* IEEE Conference on Visual Analytics Science and Technology, 309-310, 2011

[18] Qiulei Guo and Balaji Palanisamy and Hassan A. Karimi, *A MapReduce Algorithm for Polygon Retrieval in Geospatial Analysis*, 8th IEEE International Conference on Cloud Computing, 901-908, 2015

[19] Zhen LIU and Huadong GUO and Changlin WANG, *Considerations on Geospatial Big Data,* IOP Conference Series: Earth and Environmental Science, 46, 1, 12-58, 2016

[20] Manfred M. Fischer and Jinfeng Wang, *Spatial Data Analysis: Models, Methods and Techniques* (Springerbriefs in Regional

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 12, April 2018

Science) (1st ed.), Springer Publishing Company, Incorporated, 2011

[21] A.B. Mendes and I.H. Themido, *Multi-outlet retail site location assessment, In International Transactions in Operational Research*, vol. II, no. 1, pp. 118, 2004.

[22] MR Luaces, NR Brisaboa, JR Parama, and JR Viqueira *"A Generic Architecture for Geographic Information Systems."* Universidade da Corua (Tese de Doutoramento) (2004).

[23] Ranga Raju Vatsavai, Auroop Ganguly, Varun Chandola, Anthony Stefanidis, Scott Klasky, and Shashi Shekhar, *Spatiotemporal data mining in the era of big spatial data: algorithms and applications*. In Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data (BigSpatial '12) ACM, New York, NY, USA, 1-10, 2012

[24] Kaiser, Mark S, Soumendra N Lahiri, and Daniel J Nordman, *Goodness of Fit Tests for a Class of Markov Random Field Models.* The Annals of Statistics, 40(1), Institute of Mathematical Statistics, 10430, 2012.

[25] H. Wang, Y. Lu, Y. Guang, E. Edrosa, M. Zhang, R. Camarca, Y. Yesha, T. Lucic, N. Rishe, *Epidemiological Data Analysis in TerraFly Geo-Spatial Cloud, International Conference on Machine Learning and Applications,* 485-490, 2013