European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 07, Issue 02, 2020

An Empirical Study of Deep Learning Strategies for Spatial Data Mining

K. Sivakumar[1], A.S. Prakaash[2]

[1]*Professor, Department of Mathematics, Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India.*

[2]*Research Scholar, Department of Mathematics, Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India.*

[1]*siva111k@gmail.com*

[2]*prakaashphd333@gmail.com*

Abstract: The emergence of scalable frameworks for machine learning to efficiently analyse and derive valuable insights from these data has triggered growing volumes of data collected. Huge spatial data frameworks cover a wide variety of priorities, including tracking of infectious diseases, simulation of climate change, etc. Conventional mining techniques, especially statistical frameworks to handling these data, are becoming exhausted due to the rise in the number, volume and quality of spatio-temporal data sets. Various machine learning tasks have recently shown efficiency with the development of deep learning methods. We therefore include a detailed survey in this paper on important impacts in the application of deep learning techniques to the mining of spatial data.

Keyword: Big data, Convolutional Neural Network, Deep Learning, Machine Learning, Spatial Data Mining.

1. INTRODUCTION

Deep learning systems with their unique areas have been widely adopted in modern times in various big data applications, such as health care, information extraction, data processing, among many others. Besides that, in spatial applications, the volumes of data generated and processed have seen an unprecedented explosion. For example, space telescopes starts to 150 GB of remote sensing each week, medical devices produce spatial images (X-rays) at a rate of 50 PB per year, while the NASA satellite earth image dataset contains over 500 TB. Designers and developers throughout have either applied spatial upgrades to existing machine learning for the fast performance of such vast quantities of spatial data. A broad range of options have been inspired by such things as modifications and creative approaches, especially in the fields of genetics, environmental science, climatology Shekhar et al.[2005].

Through this review, we aim to include a detailed overview of deep learning frameworks and techniques that effectively support large spatial data. In data maintenance, vast volumes of data are collected using data collection methods such as barcode technology, remote sensing, satellite telemetry, etc. This huge data analysis requires the need for knowledge discovery to be retrieved, which leads to a fruitful field called information discovery repositories. Database knowledge discovery is the exploration of suitable patterns from huge datasets and is linked to related fields such as data science, database systems, data mining, visualisation, and evolutionary computing. In relational databases, most data analysis instances are tested and are now a significant market in space databases, temporary databases, collaborative databases, object-oriented databases Wang et al.[2019].

Spatial data is linked to objects which, through the relations between them, occupy space. Spatial information is defined by spatial indexing frameworks, and features are provided to access spatial data that pose difficulties in accessing spatial data from information. Spatial data mining is an analysis of indirect knowledge and other trends not specifically discovered in spatial analysis. Exploring knowledge in the fields of machine learning, facility management, and data management is the framework for exploring knowledge in databases. Wide technical databases such as spatial distributed systems, spatial reasoning and so on are also managed in order to lay the foundations for spatial data mining to help Li et al.[2015].

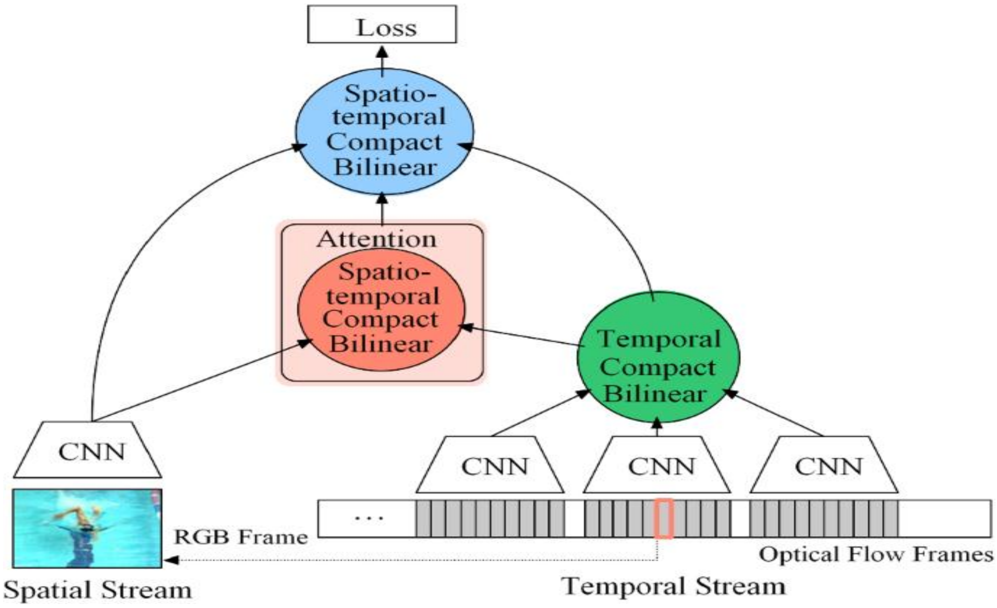

Figure 1 General Architecture of Deep Learning Spatiotemporal Model

Working with spatial knowledge, which undermines the efficiency of user applications with spatial data mining techniques, is extremely complicated and requires spatial data understanding to extract the required patterns from broad spatial databases. In order to customise queries and reorganise information, the especially challenging tasks are to investigate correlations between spatial and sub-spatial data. Typical laws, extraction rules, set of common clusters, spatial data relations, and so on, are different types of spatial knowledge Nagaprasad et al.[2010]. A comprehensive overview of deep learning in spatial data mining can therefore be strengthened in this report. With this detailed introduction on spatial data mining, Section II addresses the relevant analysis, Section III describes the different deep learning techniques in spatial data mining along with possible research directions in Section IV. Section V concludes the review.

2. RELATED STUDY

To analyse spatial data, the spatial statistical approach is used and is now possibly the best-known area of study. Therefore, a number of optimization systems have been introduced. Functional representations of statistical information of spatial phenomena have been well handled with existing optimization techniques. But, since many spatial data are made by the movement of an object, a major drawback has been the statistical consistency of the spatially different data. Regression models may be used to address this limitation, but due to inadequate statistical knowledge, the entire model structure is more difficult and the analysis of spatial data may not be the preferred end-user choice Zeitouni et al.[2000].

Many computational approaches, including fields such as artificial intelligence, databases and statistics, rely on relational databases. Following the statistical cluster analysis method, machine learning methods are commonly used in spatial data mining. Several strategies of generalisation, inference, and labeling are also improved for information systems in the spatial repositories. Deep learning techniques currently play a prominent role in data mining , predictive modelling and visualisation. The main goal of deep learning methods is to first take advantage of the empirical evidence and to be used in situations where the simulation mechanism is obscured or not yet understood. Together with information exploration, spatial data solvent extraction is an effective derivation of tacit, relevant, unknown, very significant, spatial or non-spatial information that may include various laws, precision, trends and restrictions on incomplete, noisy, scattered and unstructured data Suisse et al.[2002].

Deep learning focuses on developing computer programmes that can change when presented with any kind of new information. Deep learning uses that expertise to better understand information dynamics and adjust the behaviours of the system. In previous years, the classification of spatial data has fascinated many researchers, becoming both an important and a challenging task. In current literature, several classification methods are used to classify photos. In previous years, the ANN technique was illustrated as a noteworthy classification method. Any specific neural network's execution utility is described by its topological structural character, along with the active learning used. There is no specific rule for determining the topological structural network behaviour of ANN used in the classification assignment Vatsavai et al.[2012].

If we still need to develop new algorithms or extend existing ways of modelling spatial properties and relations directly, the dilemma now arises. Although it is hard to guess what direction future work will take, both methods seem to be gaining momentum for now. Significant benefits in the development of spatial data mining and frameworks have been explored in recent years, specifically in the context of outer detection, spatial co-location law, classification / prediction, and algorithms for clustering. In the analysis of spatial data, machine learning models presently play a significant role. In recent years, several important techniques for spatial data mining have been established, such as Artificial Neural Network, Support Vector Regression, and Decision Trees Dhaya et al.[2016].Therefore, major research attempts are being taken to assist in efficient analysis and information within such spatial data mining techniques, either by providing spatial extensions to current machine learning solutions or by creating new solutions from scratch. This survey reviews the detailed survey on spatial data mining deep learning techniques and applications.

3. DEEP LEARNING STRATEGIES

Predictive Learning

Based on their historical knowledge, the objective of predictive learning is to correctly estimate spatio-temporal data. All the components will belong to various ST data cases for different applications, resulting in a number of frameworks for predictive learning problems. Predictive problems that depend on the model's ST data instances. Usually, points are merged to produce time series or spatial maps, such as killings, traffic incidents and social events, enabling deep learning models to be applied. By integrating all the events that occurred in the same time city slot and location, the models normalized the raw point data. Then, to train as info, they updated residual convolutionary unit hierarchical structures.

The neural network model called Convolutional Deep Short-Term Memory (ConvLSTM) is used in the prediction issue. First, it also incorporates the point data and is modelled as a 3-D tensor in a spatiotemporal field. For the deep learning model, each tensor input (i;j;t) represents the count of the grid cell (i;j) in the time slot t. Input into Conv LSTM is the historical tensor for prediction. A spatially deep learning system for predicting future events occurring at different locations effectively. Collection Timeline. Traffic flow data can be modelled as a time series on a road or highway in road-level traffic prediction. A variety of studies have worked briefly to forecast road-level traffic. The first use of the stacked auto encoder was to learn road-segment level features that forecast traffic flow.

Traffic flow evidence as historical data on a highway and proposed using Deep Belief Networks (DBNs) to forecast potential traffic flow on the basis of previous assessments of traffic flow. A convolution-layered deep learning model is to extract features from the traditional taxi demand time series, and the technologies are acquired to find trends with other additional information, including weather and social media texts. Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM) are commonly used for prediction of time series spatio temporal results. Cumulative LSTM and sequence design to assess a road section's traffic speed. In order to get data on traffic speeds, other visual characteristics such as geographical structure, highways, public social activities, such as interactive requests for travel information, were also found in their design Ester et al.[2000].

Climate variables such as wind speed are normally interpreted as time series, and for future weather predictions, RNN / LSTM projections are then used. An ensemble model to forecast, for example, probabilistic wind speeds. Usually fMRI time series data are used in the field of fMRI data analysis to diagnose disease. The Long Short Term Memory (LSTM) model for classifying persons with standard controls that are taken from the fMRI resting-state time series. A profoundly deep convolutional auto-encoder model called DCAE to unattendedly learn features from Task-based functional magnetic resonance imaging (tfMRI) time series. Data from time series typically do not provide sensory features, so the spatial similarities between the dataset are not specifically taken into account in deep learning predictions Kanevski et al.[2009].

Representation Learning

In order to enable subsequent data mining or machine learning operations, representation learning aims to understand the conceptual and useful meanings of input data, and explanations are generated by integrating several representations of linear or nonlinear input data. The analysis of trajectory data types and spatial maps has focused on many of the recent studies on ST data representation learning.

- Trajectory

Trajectories are omnipresent in location-based social networks (LBSNs). In order to learn interpretations of the trajectories, neural network models are widely used. For the issue of trajectory similarity computation, a seq2seq-based model is used to obtain trajectory interpretations. The trajectory similarity is resilient to non-linear, low sampling levels and distorted set point depending on the characteristics acquired. Comparably, a trajectory is converted into a series of features to characterise object movements, and then to identify large, fixed-length clustering formulations using a series of auto encoders.

Location-based social network (LBSN) data usually has two essential aspects , i.e. location trajectory data and users ' social network. In order to learn the interpretation of the social network and the representation of trajectories, a neural network model is used to describe the two objects and explore their connexions. To distinguish short- or long-term sequential similarities in mobile trajectories, RNN and GRU models are used. Content-aware hierarchical point of interest embedding model called CAPE is used. In a user's check-in sequence, CAPE, the embedding vectors of POIs are taught to be identical to each other. To learn about location embedding in LBSNs, a geographically convolutional neural tensor network called GeoCNTN is used. To discover how to implement trajectory embedding, RNN and Autoencoder have been used and use the encoding to predict the social circle of the user in LBSNs.

- Spatial Maps

Many studies study how representations of spatial maps can be taught. To learn spatiotemporal characteristics from explicit spatial representations of sensor data, a convolutional framework of the neural network is used. For spatial representation, the urban community scheme is used as a learning feature. A collaborative embedding learning framework was built on human mobility networks to practice urban community structures. A nonlinear representation of brain connectivity patterns from neuroimaging data is used to explain neuropsychiatric disorders. Multi-side-View Guided AutoEncoder (MVAE) is suggested to learn data generated from fMRI and DTI images of the input brain connectome Bengio et al.[2013].

Classification

In the analysis of the fMRI findings, the function of classification is studied extensively. Across the neuroscience arena, brain imaging technology has proven to be a challenging subject. In general, fMRI has been commonly used for different classifiers in neuroscience experiments along with deep learning approaches, such as classification of disease, brain function domain and cognitive function while interpreting words.

ST data from various classification tasks are extracted based on raw fMRI data. Using repetitive long short term neural memory networks (LSTMs) is being used to classify individuals with autism spectrum disorders (ASD) and standard controls specifically from time series fMRI data. The fMRI data was used as spatial maps and classifier feeds. The whole brain functional connectivity matrix is calculated based on the Pearson correlation between each fMRI regression pair, instead of explicitly using each individual fMRI time series data. The matrix of correlation is therefore an input to the DNN model of the ASD classification.

A more general connectome-convolutional neural network (CCNN) can integrate information from a number of functional connectivity metrics, making it simple to respond to a diverse variety of network training connectivity descriptor variations are being used.Many works also specifically use 3D structural MRI brain scanning images as the ST raster data, and then the 3D-CNN model is commonly used to understand the functions of ST raster classification.

Two 3D convolutional communication techniques are used in the brain MRI classification, which are the variants of convolutional neural networks. Their concepts can be extended by intermediate hand-crafted approximation of the item to 3D MRI images. Deep 3D-CNN frame detection and identification of a wide range of subsequent brain networks that are replicated using an insufficient 3-d whole brain fMRI signal model Permal et al.[2019].

Estimation and Inference

Recent research on the estimation and inference of ST data focuses primarily on the forms of spatial maps and trajectories.

- Spatial Maps

But, due to the high cost and limited stations, monitoring stations were constructed to collect pollutant statistics. Suggesting fine-grained details on urban air quality thus becomes an important concern for both government and individuals. The air quality inference query is used for any location centred on air contaminants. A deep neural network model named ADAIN is used to model heterogeneous data and to learn the complex associations between characteristics. Usually, ADAIN combines two types of neural networks: that is, feed forward neural networks to model static data and persistent sequential data modelling neural networks, followed by hidden interaction capture layers.

To approximate the precipitation from remote data, deep neural networks are applied. A stacked denoising auto encoder extracts features, and calculates the amount of precipitation. Predicting the length of a potential journey, given the place of origin, the formative way and the time of departure, is a crucial challenge for intelligent transportation systems. To reduce this issue, to estimate the time of arrival, a deep multi-task model of learning representation is used. This approach yields functional representation that preserves unique travel habits when exploiting the underlying road network and prior information in spatiotemporal terms.

- Trajectories

Trajectories have begun to regulate travel time from mobility direction data. A deep RNN-based model called DEEPTRAVEL that can train to approximate travel time from historical trajectories. A Deep learning method is used for Travel Time Estimation called DeepTTE that calculates the travel time of the entire track explicitly calculates and then summarises track segments or sub-tracks. The issue of inferring the intent of a user visit from trajectory data at a certain spot. For the inference of activity types from GPS trajectory data produced by personal smartphones, a graph convolutional neural network (GCNs) is used.

A user's mobility maps are created on the basis of all his / her public spaces and clearly defined on the basis of trajectory data, and then the spatio-temporal activity graphs are fed into GCNs for the form of estimation activity. The Trajectory-User Linking (TUL) issue, which aims at identifying and linking trajectories to users who create them in the LBSN. A model based on Recurrent Neural Networks (RNN) is used to solve the TUL problem by combining the embedding model for the check-in trajectory and stacked LSTM. An significant research field of travel demand and travel guides is the identification of the implementation of passenger transport modes, such as bikes, trains, walks, etc. To forecast travel modes based only on raw GPS trajectories, a CNN model is used where behaviors are classified as walking , cycling, busing, riding and passenger rail Sahu et al.[2015].

Anomaly Detection

Detection of anomalies attempts to distinguish, by distinguishing from other outcomes, the unusual artefacts or observations that pose questions. Current studies on the detection of ST data phenomena concentrate mostly on event types and spatial maps.

- Events

Events are non - current traffic jams caused by temporary disturbances such as collisions, sporting games, powerful storms, etc. CNN will mention the non-recurring traffic irregularities caused by incidents. To detect traffic accidents from social media data, two deep learning approaches have been used: Deep Belief Network (DBN) and Long Short-Term Memory (LSTM) to recognize tweets related to traffic accidents.

Using traffic flow data, CNN has been used to automatically classify traffic events on urban networks. Broad and diverse data is collected to understand how human mobility affects the probability of traffic collisions, including vehicle trajectory data and fatal crash data. A deep Stack denoise Autoencoder model is used to know about the characteristics of social mobility by hierarchical models, and these characteristics have been used to accurately estimate the impact on traffic accident risk.

- Spatial Maps

As an alternative method to predicting extreme climate events such as hurricanes and heat waves, Deep Learning methods were first applied. With the weather image data as the input, the model was instructed to distinguish tropical cyclones, weather fronts, and atmospheric reservoirs. The architecture focuses on various deep neural network models, the CNNs used to monitor extreme climate events, and Pixel's nonlinear super-resolution system for replicating low-resolution climate data with high-resolution weather data.A multichannel spatiotemporal CNN architecture is used to address the issue of climate events for semi-supervised bounding box prediction. To enhance extreme weather locations, the device can use texture dependencies and unlabeled data Chandola et al.[2009].

4. APPLICATIONS AND RESEARCH DIRECTIONS

Deep Learning Model Selection

Several types of related ST data can often be gathered, and distinct information can be chosen. It is not well understood how to pick the ST data representations correctly, and the corresponding deep learning modes. For example, modelling the local traffic data of each road as a time series in traffic flow forecasts so that RNN, DNN or SAE are used to measure; Some studies are used as spatial maps to build traffic flow data from multiple road connexions so that CNN is used for prediction; and to construct road network traffic flow data as a graph so that GraphCNN is used. There is a lack of in-depth experiments on how to properly pick deep learning models and ST data embodiments to accurately answer the problem of STDM being studied.

Broader Applications to more STDM Tasks

While deep learning interventions are commonly used in diverse STDM tasks mentioned earlier, deep learning models such as standard pattern mining and association mining have not addressed some functions. The key aspect of deep learning is its powerful opportunity to recognise characteristics that are important for some STDM tasks such as predictive learning and classification, which depend heavily on high-quality features. Nonetheless, for other STDM tasks such as standard pattern mining, training high-quality features will not be useful as features are not included in those tasks. There are currently very few, or even no, deep learning models used for the tasks listed above, based on the analysis. How deep learning models can be applied to wider applications or how deep learning models can be combined with conventional models, such as conventional pattern mining, into different STDM functions remains unclear Gangappa et al.[2017].

Research Directions

In spatial data mining, data access methods are distinguished from access methods in relational databases, and complex spatial structures are hard to handle using traditional methods. Spatial data mining algorithms are lacking in order to refine discovery patterns. The error patterns raise the search space of the algorithm, and to remove unnecessary data, an effective knowledge discovery algorithm needs to be developed. An effective method involves the development of a language for database queries. But in the course of knowledge exploration, the domain expert's expertise is not used effectively. The spatial data mining method cannot even be handled by users because the retrieval of data through spatial data mining is limited. At present, the current layout of information has been limited to database areas. It does not involve spatial data mining with an expanded relational database architecture and relational databases.

To formulate a spatial database, inductive reasoning and active databases are established as advanced database systems, such as Object-Oriented (OO). For the recovery of data, powerful R-trees are used to make OO servers. Although the image database mining approach explores the use of OO techniques to make generalisations on complex data structures and complexity under mining. Like the Bayesian methods, an evidence theory can better model uncertainty than traditional representations of probability. Spatial data mining can be applied to fuzzy techniques. Information about each object and suitable object that is the same distance from the medoid may also be held in an alternative clustering process. The use of multiple thematic maps to perform spatial data mining can require some applications. The generalisation-based, spatial and non-spatial generalisations of parallelization are called further expansion. Temporal spatial data generalisation includes the generalisation of map samples obtained at time intervals Fink et al.[2020].

The discovery of parallel knowledge in spatial data extraction would significantly accelerate parallel data mining. In popularising information discovery techniques, the improvement of data mining techniques with advanced computational methods produces innovative new approaches and user interface design. The user interface is expanded to identify objects of interest with the use of 3D tools. The language needs to be sufficiently flexible to fit the inclusion of techniques used by a wide range of data types in spatial databases. Data exploration is not sufficient, but it must also be presented in a comprehensible way. Via data mining, people who enjoy the magnitude of visual data and scenes are exploited too. However, multidimensional visualisation is an area of inexperience. Spatial data mining may utilize computer graphic modelling techniques in this situation.

5. CONCLUSION

Spatial data mining is a rising scientific field, with a wide range of applications in geo-information systems, diagnostic imaging, robot deployment, and so on. While numerous ways to uncover secret information from spatial data have been developed, this paper explored deep learning methods in spatial data mining. The future spatial data mining possibilities for spatial databases discuss undiscovered knowledge discovery topics that make spatial data mining an exciting and demanding area of study.

6. REFERENCES

[1] Bengio, Y. & Courville, Aaron & Vincent, Pascal. (2013). Representation Learning: A Review and New Perspectives. *IEEE transactions on pattern analysis and machine intelligence.* 35. 1798-1828. 10.1109/TPAMI.2013.50.

[2] Chandola, Varun & Banerjee, Arindam & Kumar, Vipin. (2009). Anomaly Detection: A Survey. *ACM Comput. Surv.* 41. 10.1145/1541880.1541882.

[3] R. Dhaya, S.K.B. Sangeetha, T. Vaishnavi, Kavya and Ayeshabanu. (2016). An Energy Efficient Search Scheme on Mobile Cloud. *Middle-East Journal of Scientific Research* 24 (10): 3341-3344, 2016. DOI: 10.5829/idosi.mejsr.2016.3341.3344

[4] Ester, Martin & Frommelt, Alexander & Kriegel, Hans-Peter & Sander, Joerg. (2000). Spatial Data Mining: Database Primitives, Algorithms and Efficient DBMS Support. *Data Min. Knowl. Discover*. 4. 193-216. 10.1023/A:1009843930701

[5] Fink, Olga & Wang, Qin & Svensen, Markus & Dersin, Pierre & Lee, Wan-Jui & Ducoffe, Melanie. (2020). Potential, challenges and future directions for deep learning in prognostics and health management applications. *Engineering Applications of Artificial Intelligence.* 92. 103678. 10.1016/j.engappai.2020.103678

[6] Gangappa, M. & Kiran, C. & Sammulal, P. (2017). Techniques for Machine Learning based Spatial Data Analysis: Research Directions. *International Journal of Computer Applications.* 170. 9-13. 10.5120/ijca2017914643.

[7] Kanevski, Mikhail & Foresti, Loris & Kaiser, Christian & Pozdnoukhov, Alexei & Timonin, Vadim & Tuia, Devis. (2009). Machine learning models for geospatial data.

[8] Li, D. & Wang, Shuliang. (2005). *Concepts, Principles and Applications of Spatial Data Mining and Knowledge Discovery*. 27-29.

[9] Li, D. & Wang, Shuliang & Li, Deyi. (2015). *Spatial Data Mining*. 10.1007/978-3-662-48538-5.

[10] Nagaprasad, Sriramula & Dr.A. Vinaya Babu, & K. Madhukar, & Verghese, D.Marlene & V. Mallaiah, & A. Sreelatha, & A. Pratima, (2010). Spatial Data Mining Using Novel Neural Networks for Soil Image Classification and Processing. *International Journal of Engineering Science and Technology*. 2.

[11] Perumal, Muthu & Raja, S. (2019). Deep Learning Based Image Classification and Abnormalities Analysis of MRI Brain Images. 427-431. 10.1109/IMICPW.2019.8933239.

[12] Sahu, Pradip & Pal, Santi & Das, Ajit. (2015). Estimation and Inferential Statistics. 10.1007/978-81-322-2514-0.

[13] Shekhar, Shashi & Zhang, Pusheng & Huang, Yan. (2005). Spatial Data Mining. 10.1007/0-387-25465-X_39.

[14] Suisse, Martigny & Gilardi, Nicolas & Bengio, Samy. (2002). Local Machine Learning Models for Spatial Data Analysis. *Journal of Geographic Information and Decision Analysis*. 4.

[15] Vatsavai, Ranga & Ganguly, Auroop & Chandola, Varun & Stefanidis, Anthony & Klasky, Scott & Shekhar, Shashi. (2012). Spatiotemporal data mining in the era of big spatial data: Algorithms and applications. *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data, Big Spatial* 2012. 1. 1-10. 10.1145/2447481.2447482.

[16] Wang, Senzhang & Cao, Jiannong & Yu, Philip. (2019). Deep Learning for Spatio-Temporal Data Mining: A Survey.

[17] Zeitouni, Karine. (2000). A survey of spatial data mining methods databases and statistics point of views. 487-491.