IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693

A Comparative Study on Spatio-Temporal Data Correlation and Pattern Discovery Techniques for Prediction Mining

R. Sarala[1*], V. Saravanan[2]

[1]Hindusthan College of Arts and Science, Coimbatore, Tamil Nadu, India
[2]Department of IT, Hindusthan College of Arts and Science, Coimbatore, Tamil Nadu, India

**Corresponding Author: rsarala3011@gmail.com, Contact No: 9994552286*

Available online at: www.ijcseonline.org

Accepted: 16/May/2018, Published: 31/May/2018

Abstract- A spatiotemporal database handles both the space and time information. A spatiotemporal database includes spatial (i.e., location and geometry of the object) and temporal data (i.e., timestamp or time interval of valid objects) where geometry of object changes over time. Spatio-temporal correlation analysis is used for identifying the spatial and temporal relationships of multiple events. The spatio-temporal objects contain number of features in pattern discovery process. However, the existing spatio-temporal pattern discovery and prediction techniques are failed to predict the future event in accurate manner and time consumption remained unaddressed. Our main objective of the research is the spatio-temporal correlation, spatio-temporal pattern discovery and prediction with higher accuracy. In order to increase the performance of spatio-temporal pattern discovery and prediction, machine learning technique are employed in our work.

***Keywords*:** spatiotemporal database, correlation analysis, features, pattern discovery, prediction, machine learning technique

I. INTRODUCTION

The spatiotemporal object is an object with at least one spatial and one temporal property. The spatial properties are location and geometry of an object. The temporal property is time interval where the object is valid. Spatiotemporal object contains spatial, temporal and non-spatial attributes. Spatiotemporal datasets collects the changing values of spatial and thematic attributes over time period. An event in spatiotemporal dataset explains the spatial and temporal phenomenon at time 't' and location 'x'. Spatiotemporal data mining is a research area for development and for analysis of the large spatiotemporal databases. The patterns with the variations in space and time are termed as the spatiotemporal patterns. Spatiotemporal data mining tasks are employed for determining different types of potentially useful and unknown patterns from the spatiotemporal databases. The patterns and trends are employed for spatiotemporal phenomena and decision making or preprocessing step for analysis and mining.

This paper is organized as follows: Section II studies the review on different spatio-temporal pattern mining, Section III portrays the study and analysis of the existing spatio-temporal pattern prediction techniques, Section IV explains the possible comparison of existing techniques. In Section V, the discussion and limitations of the existing spatio-temporal pattern prediction techniques are studied with future direction and Section VI concludes the paper.

II. LITERATURE SURVEY

A two-level hierarchy with time lag lasso was presented to handle dependency structure learning for mining the dependencies based on time lag [1]. The designed method depends on decomposition of coefficients into products of two-level hierarchical coefficients, namely feature level and time-level. But, the static dependency structure learning were failed to manage the time-varying observations. Socio-spatio-temporal important locations (SSTIL) and SSTIL mining problem were addressed with spatial, temporal and social dimensions of social media datasets [2]. The designed method discovered SSTILs depending on the user and group preferences. But, the spatio-temporal pattern discovery was not carried out with accurate manner.

Cluster-Confidence-Rate-Boosting (CCRBoost) approach was presented to identify the hierarchical structure of spatio-temporal patterns at many resolution levels and construct predictive model depending on identified structure [3]. But, CCRBoost functioned based on the weight value for predictive modeling. In case of different number of training sample patterns, the weight value failed to update in small period of time. Two data mining techniques called support vector machine (SVM) and group method of data handling (GMDH) was employed to recognize the spatiotemporal meteorological correlations [4]. The designed techniques forecast basin scale seasonal droughts. But, SVM classifiers were the weak classifier which failed to select the features. In addition, SVM has lack of transparency problem.

A hierarchical trajectory clustering algorithm based periodic pattern mining addressed problems from traditional approaches, namely hierarchical reference spots and consideration of sequence [5]. But, it was not suitable for multi-level time periods and complex datasets. A model termed Star was introduced for spatio-temporal data in archaeology [6]. Star included vague representation by fuzzy dates and fuzzy relationships between them. A set of rules were described for deriving the temporal knowledge from the topological and stratigraphic relations. But, fuzzy relationships failed to select the spatio-temporal features as it functioned based on if-then rules.

Nonparametric estimation of probability density function for dependent spatio-temporal data was carried out through recursive kernel approach [7]. Adaptive counterparts of classical recursive kernel estimate of spatio-temporal processes were updated when new observations were studied for spatio-temporal processes with weak dependence structures. However, it was very difficult to correlate the spatio-temporal processes by recursive kernel estimators. A new pattern-oriented data aggregation technique was embedded for automated reasoning in real time decision making [8]. The designed methodology included discovering spatio-temporal patterns from time series data by extended rough set based intelligent feature selection technique through multi valued decision systems. But, rough set-based rule induction process failed to increase dependence on complete information system.

An interactive visual analytics system was designed to maintain the exploration of sparsely sampled trajectory data from social media [9]. A heuristic model minimized the uncertainty caused by social media data. Heuristic model was used only for searching process but, the feature selection was not carried out. New Spatio-temporal Congestion algorithms were designed to construct causality trees from congestions and calculated propagation probabilities with temporal and spatial information of congestions [10]. But, the support rate failed to reflect the probability for tree formation given congestion at root segment. Spatial Clustering approach was designed to determine the spatiotemporal patterns of household travel from taxi trajectory dataset with large number of point locations [11]. But, the spatial clustering method failed to improve the performance of spatiotemporal pattern discovery for large number of point locations at different time intervals.

The quality control measures were explained for real-time, spatio-temporal data from aggregator viewpoint [12]. But, spatial and temporal attributes with quality control measures failed to identify data timestamps or the locations in correct manner. An effective data-mining process was introduced to understand the travel patterns of individual metro passengers in Shenzhen [13]. The travel patterns in individual level develop the method to recover them depending on raw smart card transaction data. However, spatio-temporal pattern prediction was not carried out with higher accuracy.

The Functional dependencies theory was developed for minimizing the sizes of databases with the aid of eliminating redundant data [14]. But, spatio-temporal pattern discovery was remained unaddressed. Apriori based algorithm for mining interesting patterns was designed by considering Conjunctive databases in [15]. However, the pattern detection rate was not improved. Data mining repeatedly sorts huge amounts of data for identifying known/unknown patterns and produces valuable new perceptions and makes effective predictions [16]. However, the consumption was high to discover the sequence of patterns. In the proposed Fast Hierarchical Relevance Vector Machine was developed to provide data mining –based maximum detection rate of attacks and minimum false alarm rate [17]. But, the performance of attack detection was not efficiently increased.

III. SPATIOTEMPORAL DATA CORRELATION AND PATTERN DISCOVERY FOR PREDICTION

Spatiotemporal data comprises the states of object, event or position in space over time period. Large amount of spatiotemporal data are collected from application fields like traffic management, environment monitoring and weather forecast. Spatio-temporal data are employed in many fields like environmental sciences, geophysics, oceanography, soil science, econometrics, epidemiology, environmental science, forestry, image processing. The phenomena of interest are continuous in space and time. The data are gathered with time and space. A plethora of procedure like atmospheric pollutant concentrations, precipitation fields and surface winds are classified through the spatial and temporal variability.

A. Discovering Spatio-temporal Dependencies Based on Time-lag in Intelligent Transportation Data

Two-Level Hierarchies with time Lag lasso (TLHL) method is introduced for mining the dependencies based on the time lag. The designed method is depending on coefficients decomposition into two-level hierarchical coefficients. The first level hierarchy denotes the feature level and the second level denotes the time-level. Feature-level component is employed for the space feature learning such as traditional feature learning determination. Time-level component denotes the time feature learning. It correlates with the time lag and limited by the time lag distribution. The merits of two levels are spatiotemporal feature learning to the real value. The decomposition is used from the theory. A regression coefficient is similar to zero when any one of two components is zero.

The feature-level manage the dependency of prediction variables and respond variables. The time-level component denotes the selection of the time lag. TLHL model places the Gaussian and Cauchy distributions for component coefficients as priors to reduce the model complexity. The prior information of time lag is collected in the intelligent transportation data. A probabilistic formulation was employed with probabilistic priors to hierarchical coefficients. An expectation-maximization (EM) algorithm is used to study the model parameters. The synthetic and real-world traffic data in TLHL model is efficient in mining the dependency because of considering the time lag. TLHL model is joins the concept of causality with regression algorithm. The traditional regression coefficient is divided into the products of two-level hierarchical coefficients.

B. Discovering socio-spatio-temporal important locations of social media users

Spatio-temporal important locations (SSTILs) are the locations that are often visited by the social media users in their social media history. Discovering SSTILs is essential application for recommender systems, advertisement applications, urban planning, etc. Discovering SSTILs is demanding task because of spatial, temporal, and social dimensions of datasets, devoid of adequate interest measures and the requirement for computationally-efficient algorithms.

A new spatial, temporal, and social interest measures are employed to compute SSTILs and to determine the locations. The location density, visit lifetime and user prevalence measures are used. The time window incidence, temporal location significance for user and socio-spatio-temporal location significance interest measures are introduced to add the temporal dimension. The locations are classified as user-level socio-spatio-temporal important locations and group-level socio-spatio-temporal important locations. The designed technique employed user-level SSTILs to determine the group-level SSTILs.

A naive algorithm performs the user-level spatial important locations and temporal important locations discovery independently. It joined the spatial and temporal locations to determine the user-level spatio-temporal important locations. SSTILs of user group are determined depending on the user prevalence of locations. The designed algorithm does not have early-pruning operation in naive algorithm. In Temporal-First Socio-Spatio-Temporal Important Location (TF-SSTIL) Miner algorithm, temporal important locations are identified and spatial analysis of temporal important locations is carried out to find out the SSTILs. TF-SSTIL Miner algorithm minimizes the candidate important positions for spatial analysis. The temporal important locations are taken at spatial important locations discovery stage. The pruning of temporal unimportant locations increases the computational complexity.

C. Hierarchical Spatio-Temporal Pattern Discovery and Predictive Modeling

A spatio-temporal (ST) pattern is regarded as repeated sequence or association of ST events or ST features. For recognizing the sequences or associations, ST patterns of crime incidences, suitable distance-based and duration-based measurements are required to limit the size or shape of pattern. The nonstationarity property of ST patterns is identified by Radcliffe in crime patterns and in climate studies. A new approach called Cluster-Confidence-Rate-Boosting (CCRBoost) diminishes the nonstationarity in detecting ST pattern by representing ST pattern as hierarchical structure. The designed approach constructs the predictive model depending on hierarchically learned pattern. The local abstracted patterns are detected from distributed demonstrations and hierarchically study the global ensemble pattern built on detected local patterns. The indicators are collected from original data. The indicators are spatio-temporal features as every indicator denotes primary factor of spatiotemporal context. When pattern of drunk-driving incidents takes place often in locations, one indicator in pattern is number of bars.

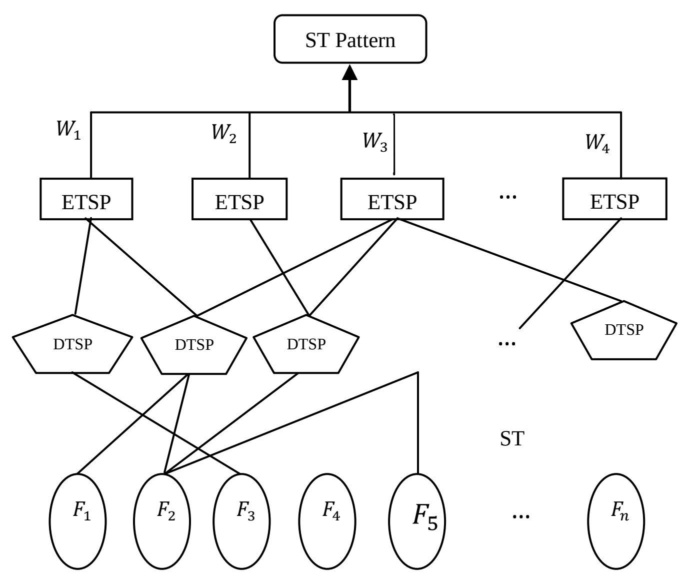

Figure 1 Hierarchical Spatio-Temporal Pattern

Figure 1 explained the hierarchical spatio-temporal pattern. Hierarchical Spatio-Temporal Pattern multi-level clustering method is introduced to detect local distributions at many granularity levels through changing number of clusters. The distributions are not-mutually-exclusive sub-partitions where the features are learned effectively. DSTP discovery is inserted with feature selection process that selects the delegated indicators to denote the ST pattern. It is used for chronologically dissected datasets to detect the DSTPs at various time span. A real-world ST pattern is difficult the phenomenon to collect inside the single ESTP. A boosting approach with greedy search algorithm chooses DSTPs and forms ESTP to construct the predictive model by one layer of ESTP at time. Every layer of ESTP allocated the confidence factor based weight which is correlation in predictive model. Through finding all indicators of one location, the designed model predicts the occurrence of target event at location.

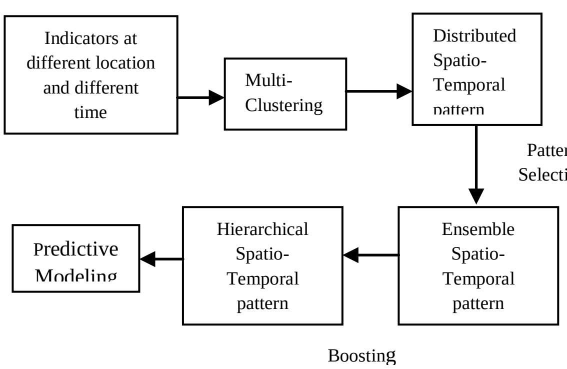

Figure 2 Flow Diagram of CCRBoost approach

As described in figure 2, CCRBoost approach starts with feature construction stage where all indicators of different time periods at every location are generated. With indicators of one location in same time period, a feature vector '$\vec{x}$' is taken. Distributed Spatio-Temporal Pattern (DSTP) is taken out from distribution of locations with same feature vectors. Appropriate DSTPs was used with an Ensemble Spatio-Temporal Pattern (ESTP) through greedy algorithm. By boosting process, confidence-rate 'W' 'is assigned repeatedly to every ESTP based on their weight and join all ESTPs into one model. A strong hypothesis $H(\vec{x})$ depending on hierarchical model considered as global spatio-temporal pattern. $H(\vec{x})$ is employed to predict occurrence of target events.

IV. PERFORMANCE ANALYSIS OF SPATIO-TEMPORAL CORRELATION, PATTERN DISCOVERY AND PREDICTION TECHNIQUES

In order to compare the spatio-temporal correlation, pattern discovery and prediction techniques, no. of spatio-temporal features is taken to perform the experiment. Various parameters are used for increasing the correlation, pattern discovery and prediction performance.

A. Feature Selection Time (FST)

Feature selection time is defined as the amount of time taken to select the spatio-temporal features. It is difference of ending time and starting time of spatio-temporal features. It is measured in terms of milliseconds (ms). FST is formulated by,

FST = Ending time – Starting time of feature selection

When the feature selection time is lesser, the method is said to be more efficient.

Table 1 Tabulation for Feature Selection Time

Number of spatio-temporal features (Number)	Feature Selection Time (ms)		
	TLHL Method	TF-SSTIL Miner algorithm	CCRBoost Approach
10	12	31	25
20	15	33	28
30	17	34	30
40	21	37	31
50	14	32	28
60	11	29	25
70	10	26	22
80	14	28	24
90	18	31	27
100	22	35	31

Table 1 describes the feature selection time with respect to number of spatio-temporal features ranging from 10 to 100. Feature selection time comparison takes place on existing Two-Level Hierarchies with time Lag lasso (TLHL) method, Temporal-First Socio-Spatio-Temporal Important Location (TF-SSTIL) Miner algorithm and Cluster-Confidence-Rate-Boosting (CCRBoost). From table value, it is clear that the feature selection time using TLHL method is lesser when compared to TF-SSTIL Miner algorithm and CCRBoost. The graphical representation of feature selection time is shown in figure 3.

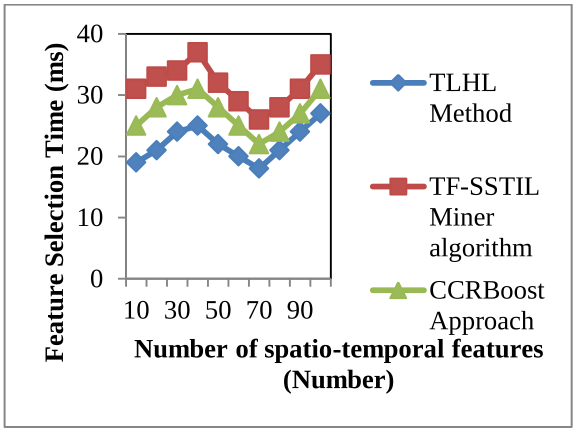

Figure 3 Measurement of Feature Selection Time

From figure 3, feature selection time based on different number of spatio-temporal features is described. It is clear that feature selection time gets increased or decreased due to excess of irrelevant features present in the input number of spatio-temporal features. From the figure, feature selection time of TLHL method is effectively reduced when compared to TF-SSTIL Miner algorithm and CCRBoost. This is because the TLHL method depends on decomposition coefficients into two-level hierarchical coefficients. The first level hierarchy represents the feature level and second level symbolizes the time-level. Feature-level component is used for space feature learning like traditional feature learning determination. Time-level component denotes time feature learning. Therefore, feature selection time of TLHL method is reduced by 30% when compared to TF-SSTIL Miner algorithm and 18% when compared to CCRBoost.

B. Pattern Discovery Rate (PDR)

Pattern discovery rate is defined as the ratio of number of pattern discovered correctly from the number of spatio-temporal features. It is measured in terms of percentage (%). PDR is formulated as,

$$PDR = \frac{number\ of\ pattern\ discovered\ correclty}{number\ of\ spatio - temporal\ features}$$

When the pattern discovery rate is higher, the method is said to be more efficient.

Table 2 Tabulation for Pattern Discovery Rate

Number of spatio-temporal features (Number)	Pattern Discovery Rate (%)		
	TLHL Method	TF-SSTIL Miner algorithm	CCRBoost Approach
10	64	88	72
20	67	90	75
30	62	84	70
40	60	80	68
50	64	86	75
60	68	90	80
70	72	94	82
80	69	88	76
90	75	92	81
100	79	96	85

Table 2 portrays the pattern discovery rate with respect to number of spatio-temporal features ranging from 10 to 100. Feature selection time comparison takes place on existing Two-Level Hierarchies with time Lag lasso (TLHL) method, Temporal-First Socio-Spatio-Temporal Important Location (TF-SSTIL) Miner algorithm and Cluster-Confidence-Rate-Boosting (CCRBoost). From table value, it is clear that the pattern discovery rate using TF-SSTIL Miner algorithm is higher when compared to TLHL method and CCRBoost. The graphical representation of pattern discovery rate is shown in figure 4.

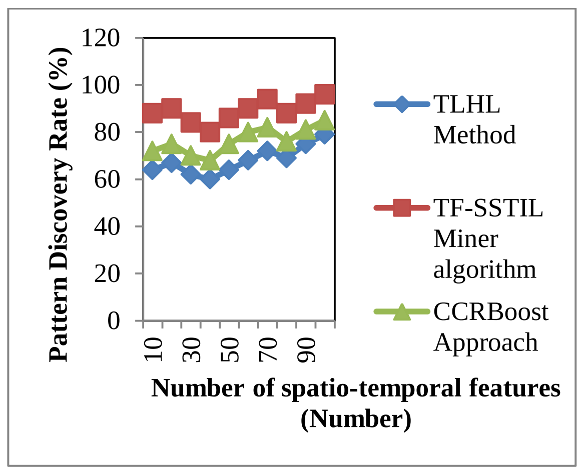

Figure 4 Measurement of Pattern Discovery Rate

From figure 4, pattern discovery rate based on different number of spatio-temporal features is described. From that, pattern discovery rate on TF-SSTIL Miner algorithm is higher than TLHL method and CCRBoost. This is done by detecting temporal important locations and performing spatial analysis of temporal locations to discover SSTILs. TF-SSTIL Miner algorithm reduces the candidate important positions for the spatial analysis. This in turns, important locations only considered at the spatial important location discovery stage. Therefore, pattern discovery rate is improved in TF-SSTIL Miner algorithm by 30% than TLHL method and 16% than CCRBoost.

C. Prediction Accuracy

Predictive accuracy is defined as the rate at which the spatio-temporal patterns get predicted in accurate manner. Ii is measured in terms of percentage (%). It is formulated as,

$$PA = \frac{\text{Number of spatio-temporal patterns predicted correctly}}{\text{Total number of spatio-temporal patterns}}$$

When the prediction accuracy is higher, the method is said to be more efficient.

Table 3 Tabulation for Prediction Accuracy

Number of spatio-temporal data (Number)	Prediction Accuracy (%)		
	TLHL Method	TF-SSTIL Miner algorithm	CCRBoost Approach
10	63	69	81
20	65	72	83
30	66	74	86
40	68	78	88
50	70	80	89
60	65	76	85
70	75	84	92
80	78	86	93
90	80	87	95
100	83	90	98

Table 3 explains the prediction accuracy with respect to number of spatio-temporal data ranging from 10 to 100. Prediction accuracy comparison takes place on existing Two-Level Hierarchies with time Lag lasso (TLHL) method, Temporal-First Socio-Spatio-Temporal Important Location (TF-SSTIL) Miner algorithm and Cluster-Confidence-Rate-Boosting (CCRBoost). From table value, it is clear that the prediction accuracy using CCRBoost is higher when compared to TLHL method and TF-SSTIL Miner algorithm. The graphical representation of prediction accuracy is described in figure 5.

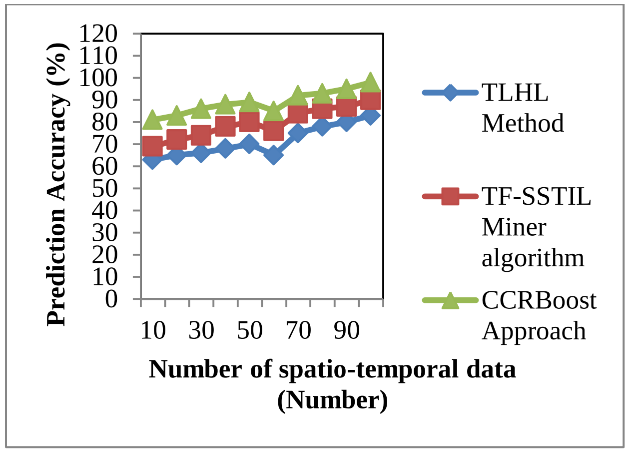

Figure 5 Measurement of Prediction Accuracy

From figure 5, prediction accuracy based on different number of spatio-temporal data is explained. It is clear that prediction accuracy gets increased or decreased due to irrelevant feature selection in earlier stage. From that, prediction accuracy of CCRBoost is higher than TLHL method and TF-SSTIL Miner algorithm. This is because CCRBoost approach starts with the feature construction stage where all indicators from the original data are generated. With indicators of one location, feature vector are taken. Distributed Spatio-Temporal Pattern (DSTP) is extracted from the distribution of locations with similar feature vectors. DSTPs were used with Ensemble Spatio-Temporal Pattern (ESTP) through greedy algorithm. Confidence-rate is assigned to ESTP depending on weight and combines all ESTPs into one technique. The prediction accuracy of CCRBoost is improved up to 25% than TLHL method and 12% than TF-SSTIL Miner algorithm.

V. DISCUSSION ON SPATIO-TEMPORAL PATTERN DISCOVERY AND PREDICTION TECHNIQUES

Two-level hierarchy with time lag lasso method was designed for mining dependencies by time lag. The designed approach was depending on decomposition of coefficients into products of two-level hierarchical coefficients, namely feature level and other denotes the time-level. A probabilistic formulation was carried out through relating the probabilistic priors to hierarchical coefficients and design expectation-maximization (EM) algorithm to detect the model parameter. The static dependency structure learning were failed to manage the time-varying observations. In addition, the evolvement of dependencies and time lag in a network were not monitored.

SSTIL Miner algorithm is introduced to discover the spatial locations and then temporal analysis of spatial important locations is performed. SF-SSTIL Miner algorithm applies spatial pruning at earlier stage and spatial important locations are considered at the temporal important locations discovery stage. The designed algorithms were employed for socio-spatio-temporal important location discovery. The designed interest measures are employed to find out the SSTILs efficiently depending on user and group preferences. However, spatio-temporal pattern discovery was not carried out with accurate manner.

CCRBoost was employed to detect the hierarchical structure of spatio-temporal patterns at many resolution levels and design predictive model depending on identified structure. Distributed spatio-temporal pattern (DSTP) was taken out from the distribution that comprises locations with same indicators at the same time period and by multi-clustering. ESTP symbolizes spatio-temporal pattern of different regularities or non-stationary pattern. But, CCRBoost functioned depending on only weight value for predictive modeling. For different number of training sample patterns, the weight value failed to update in lesser time.

A. Future Direction

The future direction of spatio-temporal pattern discovery can be carried out using correlation and ensemble classification techniques for improving the prediction performance with high accuracy and lesser time consumption.

VI. CONCLUSION

A comparison of different existing spatio-temporal correlation and patter discovery for prediction is studied. From the study, it is observed that the existing techniques failed to improve the prediction performance with better accuracy and minimal time. The survival review shows that the existing two-level hierarchy with time lag lasso method failed to manage the time-varying observations. In addition, spatio-temporal pattern discovery was not carried out with accurate manner. The wide range of experiments on existing methods determines the performance of many predictive scheduling techniques with its limitations. Finally, from the result, the research work can be carried out using machine learning techniques for improving the performance of spatio-temporal pattern discovery and prediction.

REFERENCES

[1] Xiabing Zhou, Haikun Hong, Xingxing Xing, Kaigui Bian, Kunqing Xie, Mingliang Xu, "*Discovering Spatio-temporal Dependencies Based on Time-lag in Intelligent Transportation Data*", Neurocomputing, Elsevier, Vol. 259, pp. 76-84, 2017.

[2] Mete Celik and Ahmet Sakir Dokuz, "*Discovering Socio-Spatio-Temporal Important Locations of Social Media Users*", Journal of Computational Science, Elsevier, Vol. 22, pp. 85-98, 2017.

[3] Chung-Hsien Yu, Wei Ding, Melissa Morabito, and Ping Chen, "*Hierarchical Spatio-Temporal Pattern Discovery and Predictive Modeling*", IEEE Transactions on Knowledge and Data Engineering, Vol. 28, Issue 4, pp. 979 – 993, 2016.

[4] Banafsheh Zahraie, Mohsen Nasseri and Fariborz Nematizadeh, "*Exploring spatiotemporal meteorological correlations for basin scale meteorological drought forecasting using data mining methods*", Arabian Journal of Geosciences, Springer, Vol. 10, Issue 419, pp. 1-15, 2017.

[5] Dongzhi Zhang, Kyungmi Lee and Ickjai Lee "*Hierarchical Trajectory Clustering for Spatio-temporal Periodic Pattern Mining*", Expert Systems with Applications, Elsevier, Vol.92, pp.1-11, 2018.

[6] Alberto Belussi and Sara Migliorini, "*A spatio-temporal framework for managing archeological data*", Annals of Mathematics and Artificial Intelligence, Springer, Vol. 80, Issue. 3-4, pp. 175-218, 2017.

[7] Aboubacar Amiri and Sophie Dabo-Niang, "*Density estimation over spatio-temporal data streams*", Econometrics and Statistics, Elsevier, Vol. 5, pp.148-170, 2017.

[8] Sanchita Mal-Sarkar, Iftikhar U. Sikder and Vijay K. Konangi, "*Spatio-temporal Pattern discovery in sensor data: A multivalued decision systems approach*", Knowledge-Based Systems, Elsevier, Vol. 109, pp.137–146, 2016.

[9] Siming Chen, Xiaoru Yuan, Zhenhuang Wang, Cong Guo, Jie Liang, Zuchao Wang, Xiaolong (Luke) Zhang and Jiawan Zhang, "*Interactive Visual Discovering of Movement Patterns from Sparsely Sampled Geo-tagged Social Media Data*", IEEE Transactions on Visualization and Computer Graphics, Vol. 22, Issue 1, pp. 270 – 279, 2016.

[10] Hoang Nguyen, Wei Liu and Fang Chen, "*Discovering Congestion Propagation Patterns in Spatio-Temporal Traffic Data*", IEEE Transactions on Big Data, Vol. 3, Issue 2, pp.169 – 180, 2017.

[11] Feng Mao, Minhe Ji and Ting Liu, "*Mining spatiotemporal patterns of urban dwellers from taxi trajectory data*", Frontiers of Earth Science, Springer, Vol.10, Issue 2, pp. 205–221, 2016.

[12] Douglas E. Galarus and Rafal A. Angryk, "*Spatio-temporal quality control: implications and applications for data consumers and aggregators"*, Open Geospatial Data, Software and Standards, Springer, Vol.1, Issue 2, pp. 1-23, 2016.

[13] Juanjuan Zhao, Qiang Qu, Fan Zhang, Chengzhong Xu and Siyuan Liu, "*Spatio-Temporal Analysis of Passenger Travel Patterns in Massive Smart Card Data*", IEEE Transactions on Intelligent Transportation Systems, Vol. 18, Issue 11, pp. 3135 – 3146, 2017.

[14] Vijay Kumar Verma and Pradeep Sharma, "*Data Dependencies Mining In Database by Removing Equivalent Attributes*", International Journal of Computer Science International Journal of Computer Science and Engineering, Vol.1, Issue. 1,pp. 13-16, 2013.

[15] Pradeep Chouksey, "*An Apriori based Algorithm for mining interesting patterns Using Conjunctive datasets*", International Journal of Scientific Research in Computer Sciences and Engineering, Vol. 4 , Issue.5, pp.31-36, 2016.

[16] Namrata Ghuse, Pranali Pawar, Amol Potgantwar, "*An Improved Approch For Fraud Detection In Health Insurance Using Data Mining Techniques*", International Journal of Scientific Research in Network Security and Communication, Vol.5 , Issue.3,pp.27-33, 2017.

[17] P. Rutravigneshwaran, "*A Study of Intrusion Detection System using Efficient Data Mining Techniques*", International Journal of Scientific Research in Network Security and Communication, Vol. 5, Issue.6, pp.5-8, 2017.

Authors Profile

Mrs.R.Sarala is a Assistant Professor in Department of Computer Science at KG College of Arts and Science, Coimbatore. She did her MCA degree at the Madras University, Chennai. She started her teaching profession at KG College of Arts and Science, Coimbatore in 2012. Her teaching areas are Data Structures, Computer Networks, Data Mining, Visual Programming, Software Engineering and Software Project Management, C Programming and Java Programming. She did her M.Phil. at Bharathiar University, Coimbatore. She is currently pursuing Ph.D. at Bharathiar University under the guidance of Dr.V.Saravanan. She has published 4 papers in international referred journals. She has presented papers at various conferences and published book.

Dr.V. Saravanan is a Professor and Head in PG Department of Information Technology at Hindusthan College of Arts and Science, Coimbatore. He did his M.Sc in Computer Science at Bharadhidasan University, Trichy. And he did his MCA at the Periyar University, Salem. He started his teaching profession at Thanthai Hans Roever College Perambalur, Trichy in 1999. Later in 2004 he joined in Hindusthan College of Arts and Science, Coimbatore. His teaching areas are Software engineering and mobile computing as it is his main area of interest. He did his M.Phil and Ph.D. at Manonmaniam Sundaranar university, Tirunelveli. His Ph.D. was on Wireless Networking in video streaming. He has over 30 publications in internationals referred journals. He has presented research papers at conferences, published articles and papers in various journals. He is renowned key note address in both national and international conferences.