A Spatio-Temporal Approach to Identify Variable Size/Shape Collocation

M.Nagabhushana Rao +

+professor of Computer, C.R..Engineering .College , Tirupati, India.

Dr. P. Govindarajulu ++

++Professor of Computer , S. V. University, Tirupati, India.

ABSTRACT: *The goal of data mining is to discover nuggets. Spatial data mining discovers collocation rules. Especially in spatial data mining, when spatial data is relatively represented with time series, a spatio-temporal significance is inferred. In this context the collocation rule that is a quintessence for the spatial data, obtains changes to its size and shape with temporal influence. Thus, the changes that occur to the spatial knowledge are the spatio-temporal transactions. Mining the spatio-temporal transactions and determining the various behavioral aspects of collocation is one of the substantial activities of GIS.*

KEYWORDS: *Spatial Data Mining, Temporal Mining, Spatial Knowledge, Collocation.*

1. INTRODUCTION

Spatial data mining is an important extension that discovers the interesting and previously unknown, but potentially useful patterns from large spatial datasets. Extracting interesting and useful patterns from spatial data sets, which covers various technical overheads like, spatial data infrastructures, spatial relationships, spatial autocorrelation and some others related to spatio-geometry [8].

Temporal data mining is an important extension as it has the capability of mining activity rather than just states and, thus, inferring relationships of contextual and temporal proximity, some of which may also indicate a cause-effect association [7].

Combining the semantic support from spatial data mining and temporal data mining, with a slender correlation, the concept of spatio-temporal mining has ushered to give, the temporal meaning to the evolving and ever changing collocation rules.

The problem of discovering association rules is related to the need of finding the patterns in the data sets. Patterns are identified in the spatial domain, with respect to the physical features and the non-physical features. Physical features are purely spatial features of the spatial objects that are closely related to the spatial geometry. Non-physical features are non-spatial features that describe the properties of the spatial objects. In erstwhile publications and works, boolean spatial features and spatial features with fuzzy set are predominant. Granularity exemplification of the nature of the collocation is clear with fuzzy sets representation of features. Fuzzy sets describe the high-degree of lucidity in representing the spatial features. However, the feature fuzzy or boolean, the problem of collocation is whether a static or a dynamic is predicament to many GIS application users. If static, the heights of utility would go with simple claim of mining adequate static algorithms. If dynamic, there is the need of the understanding of dynamicity of the collocation. The basic aspect that conveys the dynamic nature of the collocation is the change of its nature with respect to its size and shape. The collocation is dynamic in its size and shape, due to the nature of the features. It would be more conventional if the features are considered as fuzzy for the dynamic nature of the collocation.

In most of the cases, using age as an obsolescence factor for rules helps reduce the number of rules to be presented to the user, but in contrast of the realm of understanding the interesting aspects of the domain. One way to solve this crisis is by incorporating time in the model of discovery of the rules. However, the rules are characterized as outdated, obsolete and which does not satisfy the decisive factors of GIS apprehension, such are eliminated. And with the historical rules that describe the incidence, which are mandatory, in addition to which are used to identify the state transformations in terms of size and shape of the elementary parts of the rules are of very important concern, and when related to spatial, many algorithms for mining the spatio-temporal inference encompass much significance. Most typically the potentiality of transformations occurred to the collocation, is to understand the time components related to the transformations and mining the temporal sequences that identify the series of transformations to the collocation.

2. RELATED WORK

The problem of discovering the collocation rules from spatial data is introduced by Shashi Shekhar *et. al.* in [1][2][3]. It was followed by successive refinement and improvement and given a discrete data model representation in the [4][5][6]. In [4], the conceptual notation for the collocation that is designed for non-spatial features of the spatial objects has been discussed, supporting to that, a transformation has be deduced in [5]. To support further experimental work the semantic representation of the data structure to store the collocation has designed in [6].

According to [9], the spatio-temporal mining will be the subjective principle that will establish the correlation between the time components and the spatial aspects. But the experiment carried over in this work, is related to the time components which are define at even with a broader intervals and the spatial knowledge than the spatial objects. It is a direct compact relative discussion about the collocation pattern that changes over a period of longer time.

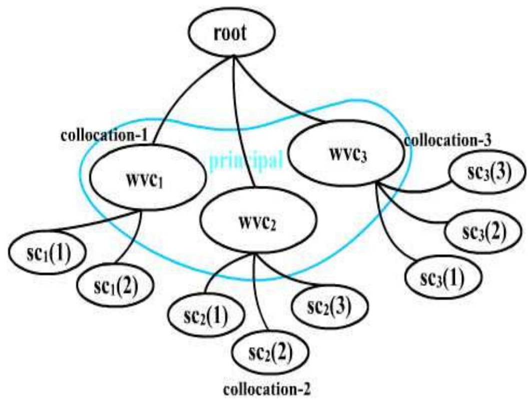
The remainder of this paper is organized as follows: Related work on representing spatial knowledge, identifying the array of collocations as spatial knowledge nuggets, associating the temporal model to the collocations, to obtain mining sequences and progress identification of trends. Spatio-Temporal Associations are also discussed to compare and contrast the intensity of temporal mining of transformations on collocations.

The total work is three fold, first to design the spatio-temporal datasets for spatio-temporal mining , second step is dedicated in fabrication of spatio-temporal association rules, third is to structure the inference that collocation with transformations. [As matter of fact the temporal mining traditionally is implemented on the direct data sets that have temporal significance, in this work, it is proposed that knowledge transformations change very frequently by changing the data sets. For example if a pattern is observed as a collocation among the spatial objects considering all its non-spatial features with fuzzy sets, the same pattern may evolve again when the intensity of the spatial features change, thus occurring the shape or change transformation in the collocation.

The collocation rules mined in this way can help the GIS users thus observing the sequence of events that have been occurred in a specific spatial domain]. The synthetic data set is given in the table, which is generated by `Random Semantic Map Data Generator Tool` which uses a random distribution. If we consider the same data set generated for a given period of time, and the collocation mining is undertaken, then several versions of the same collection of collocation instance would evolve.

3. SPATIAL KNOWLEDGE REPRESENTATION

The collocation is a basic pattern and the collocation rule is the spatial knowledge [1][2][3]. As the collocation of our previous experiments has dealt with fuzzy set combination of values for the features, the intensity of feature and its participation grows enormously. So the collocation cannot be represented as simple antecedent and consequent, rather a set of features participate as the antecedent set and consequent set. The antecedent set of the features are considered to have more weight and the consequent set are having moderate weight to rely in the structure of the collocation. Thus, the representation has been concluded in our previous works as a *principal-ranking structure* [6]. Where the weighted features are coagulated as *principal* and others are connected to the principal as groups of sets called as *rank* of the principal. The feature grouped as in the rank describes the inevitability and the strong relationship of the features in the principal.



`Figure 1.`

Illustrating the A principal-ranking structure for spatial knowledge representation .

Such structures are developed for a collection of collocation instances mined about a spatial layer of a spatial domain. Considering the same spatial layer, If mining is collocation rule is performed instantaneously with regular intervals several structure would evolve, that contain the same collocation instance collections, with few or more changes. More changes can be any way concluded as more surprising or interesting, but the collections of collocation with few changes have to be studied carefully, how they grow or how they descend.

Following **Table 1.** is the synthetic data set generated for the previous work of designing the principal-ranking structure, to fabricate a set of collocation instances which is attached to one instance of time.

sno	sid	sx	sy	sdesc	scd	M	C	P	D	Cr	T	Md	R	O	A	Tr
1	20	60	200	Water-Resource	WA	0	40	20	0	0	0	0	0	0	0	0
2	21	60	203	Water-Resource	WA	90	0	80	0	0	0	0	0	0	0	0
3	19	62	193	Water-Resource	WA	40	20	40	0	0	0	0	0	0	0	0
4	15	66	428	Hospital	HO	0	0	0	0	40	10	0	0	0	0	0
5	18	67	189	Water-Resource	WA	0	40	40	0	0	0	0	0	0	0	0
6	17	71	119	Water-Resource	WA	0	50	70	0	0	0	0	0	0	0	0
7	14	86	338	Hospital	HO	0	0	0	0	40	60	70	0	0	0	0
8	1	115	73	Fuel-Station	FS	50	50	40	70	0	0	0	0	0	0	0
9	11	130	139	Hospital	HO	0	0	0	0	90	60	80	0	0	0	0
10	10	153	313	Fuel-Station	FS	90	0	70	50	0	0	0	0	0	0	0
11	7	157	212	Fuel-Station	FS	40	10	20	20	0	0	0	0	0	0	0
12	36	167	330	Water-Resource	WA	40	90	50	0	0	0	0	0	0	0	0
13	37	167	330	Water-Resource	WA	40	80	80	0	0	0	0	0	0	0	0
14	13	170	204	Hospital	HO	0	0	0	0	30	40	70	0	0	0	0
15	9	172	283	Fuel-Station	FS	0	70	90	60	0	0	0	0	0	0	0
16	39	173	172	School	SC	0	0	0	0	20	30	90	80	90	70	0
17	8	177	249	Fuel-Station	FS	40	90	70	30	0	0	0	0	0	0	0
18	46	185	366	School	SC	0	0	0	0	60	20	30	40	90	70	0
19	2	194	83	Fuel-Station	FS	10	40	80	40	0	0	0	0	0	0	0
20	45	199	341	School	SC	0	0	0	0	50	30	90	70	40	70	0

Table 1.

Output generated of Random Semantic Map Data Generator Tool. (Describes the spatial objects and their values of the fuzzy features in rows).

4. TEMPORAL MODEL

In the domain of temporal data mining, the types of temporal data being analyzed is of prime importance for the temporal knowledge discovery process.

From the large corpus of literature, it is nevertheless important to note that a fully temporal database is not essential for temporal knowledge discovery and that temporal rules can also be derived from sequences of static data sets. Temporal reasoning can be applied to multiple static snapshots of the collocations in order to store several rule sets that are later compared to draw conclusions regarding the change in data over time [10].

However, data are not stored in temporal structures; rules describing the change in the data over time can only be derived indirectly from changes in the stored rule set. The existence of some temporal knowledge can be used to make mining easier. The existence of calendars is used to segment a pattern, thus making the problem more tractable. Temporality can be broadly classified within data as static, sequences, time-stamped, fully-temporal.

Many memory optimistic trees and connection optimistic trees have been available which give more economic importance to store the data sets. However in order to save disk space, common paths and links between the features are maintained only once, since they are shared among the structures: The collection of structures can be viewed as an acyclic graph rather than a collection of independent tree structures [10].

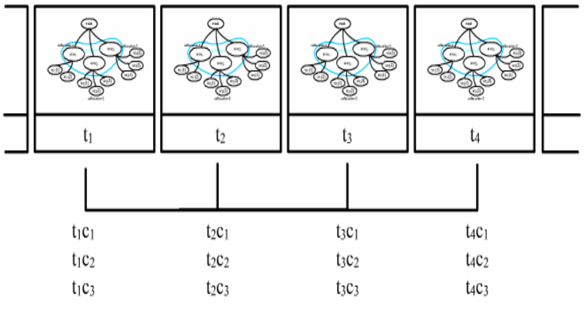


Figure 2.

Temporality of collection of Collocation instances.

The temporality of the datasets is observed with various important specifications according to the previous works. The specifications address with respect to data type and datasets as: (*i*) support of the data types - consider the collocation storage structures, (*ii*) support of the time dimension(s) – valid, transaction and bi-temporal time components, (*iii*) mobility of the datasets - with respect to the changing of collocation and the cardinality of the dataset through time, namely *evolving*, *growing* and *full-dynamic*.

5. SPATIOTEMPORAL APPROACH

One of the important task in spatiotemporal is building an efficient method for spatiotemporal access methods that should at least include modules for generating synthetic datasets, storing datasets, collecting and running access structures and finally visualizing the experimental results.

Extensions of existing spatial access methods for non-spatial features of spatial objects with temporal significance, or new 'from-the-scratch' methods could be reasonable candidates. However, while there are a large family of temporal /multi-version data access methods was already prevailed in various paradigms.

A fundamental issue on generating synthetic spatiotemporal datasets is the definition of a complete set of parameters that control the evolution of spatial objects.

Towards this idea, the two important operations have to be addressed. First is the duration of a collocation instance and second is the resizing / reshaping of collocation instance. The ultimate goal is vested to calculate the consecutive instances of a collocation rule starting from initial instance.

The following definitions give clear preliminary understanding of the work carried over:

Definition (5.1): A spatiotemporal object, is a collocation, identified by *o*, is a time-evolving spatial entity, i.e., its evolution or history is represented by a set of instances (*o*, *c*, *t*) where *c* is the change of the collocation *o* at instant *t*. (where *c* and *t* are called *change-stamp*, *time-stamp* respectively).

Definition (5.2): A spatiotemporal event for a collocation, is one instance of the transformation occurred to the collocation, is represented by $\alpha = (o, c, t)$.

For a collocation, the number of states into which it renounces forms a collocation sequence; which is a simple spatial event.

Definition (5.3): A spatiotemporal sequence is a collection of collocation sequences, where a collocation sequence is formed by set of features, and each item in the sequence is the instance of changes made temporally.

Definition (5.4): A spatiotemporal *sequence* for a collocation, α is a sequence, where α is denoted as $(\alpha_1, \alpha_2, \ldots \alpha_n)$, where α_i is called as *event.*

Definition (5.5): A spatiotemporal cardinality of a collocation (*evolving*, *growing* and *full-dynamic*) is the nature of the change of shape or size transformation occurred to the collocation.

Definition (5.6): A spatiotemporal k-*sequence* for a collocation, if the sum of cardinalities of the α_i is k.

Generally, the sequence mining task is to discover a set of attributes, shared across time among a large number of objects in a given database. For *m* attributes there are $O(m^k)$ potentially frequent sequences of length *k*. A sequence is an ordered list of events, A sequence α is denoted as $(\alpha_1 \rightarrow \alpha_2 \rightarrow \ldots \rightarrow \alpha_q)$ where α_i is an event.

In our problem, for a given a set of collocations and features for each collocation forming a collocation sequence, the support of a sequence s(S) is the percentage of total collocations whose collocation sequence is S. The confidence for a sequence association rule S=> T is the ratio of the number of collocations that contain both sequences S and T to the number that contain S.

TID	*Coll.Seq.*	*Feature*
1	C1	A=10, B=20, C=30
2	C2	A=15, B=15
3	C1	A=15, B=20, C=20
4	C3	A=10, C=45
5	C1	A=20, B=20, C=10
6	C1	A=25, B=20, C=5
7	C3	A=30, B=20, C=65
8	C1	A=30, B=25, C=5
9	C1	A=35, B=25, C=5
10	C3	A=25, B=15, C=60
11	C1	A=40, B=25, C=5
12	C2	A=10, B=20
13	C2	A=20, B=10
14	C2	A=25, B=5
15	C3	A=35, B=25, C=70
16	C2	A=30, B=0
17	C2	A=35
18	C3	A=15, B=5, C=50
19	C3	A=20, B=10, C=55
20	C2	A=40

Assuming from the above mentioned input data, the features A, B, C are forming collocations sequences C1, C2 and C3.

- C1 → {A→B→C}
- C2 → {A, A→B}
- C3 → {A→C, A→B→C}

We supply several weight thresholds and find the weaker and stronger collocations and build sequences. Following is the database D prepared for the collocation sequence C1 containing all the events of the sequence arranged chronologically.

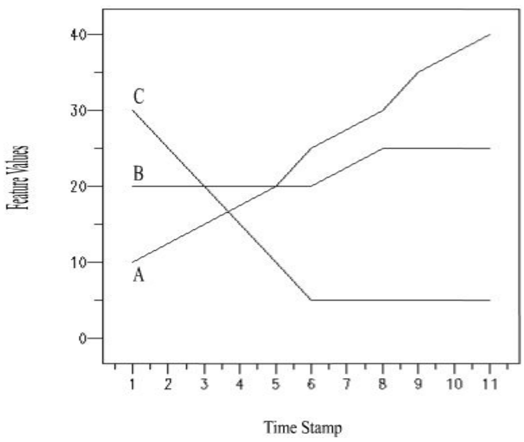
	A	B	C	TID
C1	10	20	30	1
C1	15	20	20	3
C1	20	20	10	5
C1	25	20	5	6
C1	30	25	5	8
C1	35	25	5	9
C1	40	25	5	11

The above D depicts the seven changes made to the collocation sequence C1 during different time stamps, The collocation sequence C1: {A→B→C} has occurred with various weights of features as follows

C1 → {(A=10)→(B=20)→(C=30)}
C1 → {(A=15)→(B=20)→(C=20)}
C1 → {(A=20)→(B=20)→(C=10)}
C1 → {(A=25)→(B=20)→(C=5)}
C1 → {(A=30)→(B=25)→(C=5)}
C1 → {(A=35)→(B=25)→(C=5)}
C1 → {(A=40)→(B=25)→(C=5)}

From the above sequences, by observation, the event A has an incremental change, the event B has been feebly incremented, and the event C has been decremented.

6. CHANGE DETECTION



Identification of the nature of the feature values of A, B, C in various time stamp value. (Output generated by DATAPLOT.)

6.1 RUDIMENTARY APPROACH

The change detection methods based on statistical approach are quite sophisticated in nature. However, the applied method had to be flexible and easy-to-implement to new operating conditions. Therefore, detecting changes on the basis of the data without any a priori assumptions was an important issue. It would reduce the uncertainties involved with the a priori estimates as well as initial tuning requirement. On the basis of these criteria, three different change detection methods were chosen for further work in this experiment, namely:

1. Difference Method
2. Possibilistic Approach
3. Model Based Approach

Out of all three methods, the difference method is flexible and efficient according to [11, 12], where it was calculated on the basis of the feature values of sequences. However, the changes to be detected were slower than those of consecutive parameter values in this case.

Therefore, the difference method was modified as follows:

$$\Delta_i = \left| x_i - x_{i-n} \right| \rightarrow (1)$$

Where Δ_i is the difference, x_i is the current value of the parameter and x_{i-n} is the lagged value.
A change notion is observed if

$$\Delta_i \geq ThVal.$$

Where *ThVal* is a threshold value.

The difference is calculated when a new measurement is received. For each variable, the lag n and the threshold value *ThVal* must be predefined. The absolute value in (1) enables the use of threshold of value for both positive and negative changes.

6.2 MARTINGALE AND EXCHANGEABILITY EMENDATAION

In probability theory, a (discrete-time) martingale is a discrete-time stochastic process (i.e., a sequence of random variables) X1, X2, X3, ... that satisfies the identity

$$E(X_{n+1} \mid X_1, \ldots, X_n) = X_n$$

i.e., the conditional expected value of the next observation, given all of the past observations, is equal to the last observation. As is frequent in probability theory, the term was adopted from the language of gambling.

In the context of experiment, $\{Z_i: 1 \leq i < \infty\}$ be a sequence of random variables that represent a set of functions which results values in various quantum for a feature. As the change occurred to the feature is positive and realistic the result of the stochastic process is a positive number including zero. A finite sequence of random variables $Z_1, \ldots, Z_n$ is *exchangeable* if the joint distribution $p(Z_1,\ldots,Z_n)$ is invariant under any permutation of the indices of the random variables. A martingale sequence of random variables $\{M_i: 0 \leq i < \infty\}$ such that M_n is a measurable function of $Z_1, \ldots, Z_n$ for all n = 0, 1, … and the conditional expectation of M_{n+1} given $M_0,\ldots,M_n$ is equal to M_n, i.e.,

$$E(M_{n+1} \mid M_1,\ldots,M_n) = M_n \rightarrow (1)$$

The idea of testing exchangeability online using the martingales holds importance in statistics. After observing a new data point, a learner outputs a positive martingale value reflecting the strength of the evidence found against the null hypothesis of data exchangeability.

Consider a set of labeled examples of this experiment (features of a collocation) $Z = \{z_1, \ldots z_{n-1}\} = \{(x_1,y_1\}, \ldots, (x_{n-1},y_{n-1}))$ where x_i is an object that describes the feature values and $y_i \in \{-1,1\}$ as its corresponding label, when holding the feature or not, for i = 1, 2, … , n – 1. Assuming that new labeled example, z_n, is observed, testing exchangeability for the sequence of examples $z_1, z_2, \ldots, z_n$ consists of two main steps.

A. EXTRACT A p-VALUE p_n FOR THE SET $Z \cup \{z_n\}$ FROM THE STRANGENESS MEASURE DEDUCED FROM A CLASSIFIER.

The randomized p-value of a set $Z \cup \{z_n\}$ is defined as

$$V(Z \cup \{z_n\}, \theta_n) = (\#\{i: \alpha_i > \alpha_n\} + \theta_n \#\{i: \alpha_i = \alpha_n\})/n \rightarrow (2)$$

Where α_i is the strangeness measure for z_i, i = 1, 2, … , n and θ_n is randomly chosen from [0,1]. The strangeness measure is a way of scoring how a data point of this context as feature is different from the rest.

The p-values $p_1, p_2 \ldots$ output by the synthesized randomly p-value function V are distributed uniformly in [0,1], provided that the input examples $z_1, z_2, \ldots$ re generated by an exchangeable probability distribution in the input space. This property of output p-values no longer holds when the exchangeability condition is not satisfied.

B. CONSTRUCT THE RANDOMIZED POWER MARTINGALE

A family of martingales, indexed by $\varepsilon \in [0,1]$ and referred to as the randomized power martingale, is defined as

$$M_n^{(\varepsilon)} = \prod_{i=1}^{n} (\varepsilon p_i^{\varepsilon-1}) \rightarrow (3)$$

where the p_is are the p-values output by the randomized p-value function V, with the initial martingale $M_0^{(\varepsilon)}=1$. We note, $M_n^{(\varepsilon)} = \varepsilon p_n^{\varepsilon-1} M_{n-1}^{(\varepsilon)}$. Hence it is not necessary to store the previous p-values. In our experiment, we consider $\varepsilon = 0.92$, which is within the desirable range where the martingale value is more sensitive to a violation of the exchangeability condition. When $\theta_n = 1$, the p-value function V is deterministic, the martingale constructed is also deterministic.

6.3 TESTING FOR CHANGE DETECTION

By using martingale value and martingale function leads to efficient statistical instrumentation for change detection. An effective change-detecting algorithm requires that (i) the mean delay time between a true change point and its detection be minimal, (ii) the number of miss detections be minimal, and (iii) collocation objects and their respective data be handled efficiently. A martingale exchangeability framework can be supported to succeed into the change detection exercise as follows: When a new collocation occurrence is observed, hypothesis testing can be made using a change function to detect whether a change could be made or not. In this framework, the exercise to detecting changes is carried out in two fold; (i) observing a martingale value and (ii) a martingale difference.

This approach is a one-pass incremental algorithm. The base-classifier is first assumed to be the first incidence of the collocation from the entire collocation-sequence.

Extract a randomized p-value a change threshold from the assumed collocation sequence for a feature z, which is identified to be the strangeness deduced from the classifier. The martingale value is a constant added to the p-value in the hypothesis. Consider α as the significant level to stabilize the p-value inflation. As the p-value is influenced by the martingale value, without a significant α it is very difficult to identify the change levels. The strangeness measure is a way of scoring how a feature is different from the rest. Assuming a sequence containing events of various collocations represented as objects; slice the sequence into two called S1 and S2, such that the changes of S1 and S2 parts of collocation sub-sequences have been considered as C1 and C2 respectively, where C1 is not equal to C2. Switching the feature z_i from S2 to a position in S1 will make the feature stands out in S1. The exchangeability condition is a necessary condition for a conceptually stable collocation sequence. The absence of this exchangeability may lead to identification of many changes.

When a change occurrence is identified, the distribution of the feature with respect to the p-value is normal and distributed uniformly among all the collocations, when a change is not observed the distribution of the feature with respect to the p-value will be null, and no kind of distribution is identified in the set of values for the feature as all the values are quantified to some equivalent value. On inflating the p-value by the martingale value and continuing the process of sampling the sub-sequences with higher-granularity we can identify the changes.

6.4 CHANGE-STAMP

The set of initial p-value, the martingale-value and the number of times the p-value is inflated becomes as a structure to define the proper change-stamp. The precision is a probabilistic measure to identify the detected changes. The precision in the change stamp is calculated as the number of correct detections divided by the number of detections observed. The recall is the actual inference of change that is available among the collocation objects of a sequence. The recall is calculated by dividing the number of correct detections by the true number of changes available naturally in the collocation sequence. The delay time also plays an important role to reside in the change stamp that designates the number of time units from true change point to the detected change point.

6.5 ALGORITHM

The prerequisites for the algorithm implementation in this experiment play a vital role in achieving the result. The base classifier which is the first identification of the collocation objects in the sequence is a mandatory parameter. The selection of the feature on which the change detection is performed is very important, the sampling or the selection of the feature is to be done according to the real-time application of the knowledge analysts. The p-value choice should be made carefully based on the V described in the section 6.2 (A) eqn. (1).

```
Algorithm
Classifier: V;
Randomly Synthesized: p:
Feature: z
Significance: α
Sequence: S1, S2
Change: C1,C2
V=classify_feature();
while(END)
do
        p=strange_measure();
        if(p>α) then
                // end of change detection
                // terminate loop
        slice_sequence(S,V)
                // slice the sequences into S1 and S
                // according to classifier V
        z=read_features();
        while(all values of feature z)
        detect_change(S1,S2);
                // changes with regard to p in S1, S2
                // are C1, C2
        if(C1!=C2) then
                change_stamp(C1,C2);
                // record the change
end of while loop
```

7 EXPERIMENTAL RESULTS

In the rudimentary approach, the $\Delta_i = | x_i - x_{i-n}|$ holds good for the series of feature values which exhibit normal distribution. But this is not always natural because, in the real time applications, the difference or the mean of the series will not be uniform among all the entire series. There may be change that identifies the increase or decrease in the value of a candidate feature. The absolute number of changes cannot be inferred by the rudimentary algorithm proposed in the section 6.1. Of course, the algorithm can explain the overall mean of the series that will identify the inference of the change in the feature of a particular class of collocation. Classification of a collocation or identifying the classifier of a collocation is a very important activity, which is not done in the rudimentary approach. Identifying the classifier helps know the end user understand the number of types of collocation objects required to work in the experiment.

At the outset, the experimenter can define various classes of collocation and form collocation sequence with temporal significance. In the section 6.2, algorithm and method, considering the classifier the change detection is made for every class of collocation sequences. The algorithm in 6.2 explains slicing of the collocation sequence into two subsequences, where it is quite possible for an experimenter to find the relevant change between common set of collocation sequences that belong to same class, rather for different. Finding the change between two different collocation subsequences that belong to two different classes will be irrelevant to any kind of application areas.

The approach used in the 6.2 uses martingale preliminaries, which are the basic identifiers that observe change in the various experiments of a stochastic process, which helps us in identifying the collocation object the result of spatial data mining at a time stamp, and a common set of collocation objects as a collocation sequence. Where all the collocation objects of the collocation sequence contain the objects containing the similar properties, at least common set of features, however their weights differ. In the context of the difference of the weights of the features the change of the collocation is identified between the collocation objects of a collocation sequence.

8 CHANGE-STAMP NOTATION IN DATA STRUCTURE

The result of the algorithm is inferred to be as the change stamp which precisely explains the nature of change that is prevailing among the collocation objects of a sequence. The number of iterations performed for various classifiers will bring out various change stamps for the same feature and that can be analyzed in common as a cardinality of change. The same process may be continued with respect to all the features that are application relevant and the nature of the change for each feature can be known and finally a generic notation for identifying the collocation object with the nature of change for each feature can be described, which the ultimate goal of the project is help the GIS analyst to conform the collocation completeness.

$$C1 \rightarrow \{A^{\uparrow}, B^{\uparrow}, C^{\downarrow}\}$$

Thus the spatio-temporal object (*5.1*) is defined as:

- $ST_1 = (C1, d, t)$;
- d = difference factor, which is also called as change stamp.
- t = is the timestamp.

The timestamp for the spatio temporal object is a temporal-difference factor, since the spatio temporal object is collocation, which is not a physical object of the spatial domain.

9 COMPLEXITY

The algorithmic complexity for the rudimentary approach is almost *O(N.f(N))*, where the difference is calculated for each two subsequent feature value of a series and finally the gross difference may only be found out, which will heal the actual difference details in the course. The algorithm proposed in the 6.2 contains only *O(f(N))* where a particular change is identified about a class of collocation sequences, which is more application relevant. This may also lead to the $O(f(\log_n N))$, since all the collocation objects are not interfaced by the algorithm which does not fall into the purview of the class, so it is almost a 50-50 standard for a complete preprocessed set of collocation sequence generated with a definite even number of classes.

10 CONCLUSION

In this paper we have identified the need of the data structure to represent the collocation, the spatial knowledge. The temporal study is made on the set of such structures, and a spatio-temporal coincidence has been developed with the definitions given before proceeding into the experiment. The rudimentary approach for finding the change has been analyzed and designed for comparison. A statistical approach of *martingale framework* is used to be apt in this side of finding the difference in the series of feature values. The comparison is made with regard to the functionality and the result what both the approaches infer. The complexity analysis has been made and it is identified that, the proposed framework of algorithm satisfies the expected results with sufficient time constraint and proved to be application relevant. Finally the change inference is identified with the cardinality measurements and an attempt to quote against each feature is made while storing the collocation structure.

It is very essential to understand the basic structure of the knowledge whatever is mined from the mining processes, to analyze and find the complete meaning of the knowledge and apply to the real world applications. The pompous achievement is the idea of storing the knowledge with its time-varying natures to enable the analyst in finding the alarming solutions in the spatial applications.

NOTE: This article is an extension for which we submitted in “Data structure for Spatial Knowledge “, Int. Conference. On. Rural Geo-Informatics (ICORG)-July 2006, Hyderabad, India.

11 REFERENCES

[1] Shashi Shekar, Yan Huang , Hui Xiong,”Discovering Colocation Patterns from Spatial Data Sets : A General Approach ,IEEE –KDE-Vol 16,No 12,DEC 2004.

[2] Shashi Shekar,” Mining confident co-location rules without a support threshold “, Symposium on Applied Computing Proceedings of the 2003 ACM symposium on Applied computing, Melbourne, Florida, session: Data mining, Pages: 497 - 501, Year of Publication: 2003.

[3] Shashi Shekar,” Fast mining of Collocation Rules”, Conference on Knowledge Discovery in Data Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining ,Seattle, WA, USA ,P: 384 - 393 , 2004

[4]M. Nagabhushana Rao , A. Ramamohan Reddy, Dr.P.Govindarajulu, “Spatial Knowledge System Frame Work Through Self-Adaptive Modeling “, Int. Journal .of Computer Science and Network Security, Vol.6,No 8, P-128-136,Aug-2006.

[5] M.Nagabhushana Rao[M.N.RAO] ,Dr .P. Govindarajulu, ” Spatial knowledge for disaster Identification ”, International Conference Systemics, Cybernetics and Informatics, Hyderabad , India ,vol.1.P-610-616 ,Jan 2006.

[6]M.NagabhushanaRao ,A.V.Sreeharsha,Dr.P.Govindarajulu, “Data Structure for Spatial Knowledge“, International Conference On Rural Geo-Informatics-(ICORG)July 2006, Hyderabad, India.

[7] Sanjay Chawla, Florian Verhein, “Mining Spatio-Temporal Association Rules, Sources, Sinks, Stationary Regions and Thoroughfares in Object Mobility Databases”, Lecture Notes, Springer Berlin/Heidelberg, Computer Science, P: 187-201, Mar-2006.

[8] Mennis and Liu, “Mining Association Rules in Spatio-Temporal Data: An Analysis of Urban Socioeconomic and Land Cover Change ”, *Transactions in GIS* **9** (1), 5-17, Black Well Synergy .

[9] Roddick, Myra , Spiliopoulou , ” A Survey of Temporal Knowledge Discovery Paradigms and Methods ”, Knowledge and Data Engineering, IEEE Transactions on, 2002 - ieeexplore.ieee.org.

[10] Yannis Theodoridis , Jefferson R.O. Silva, Mario A.Nascimento, “On the Generation of Spatiotemporal Datasets”, Proceedings of the 6th International Symposium on Advances in Spatial Databases,P: 147 - 164 , Year :1999, ISBN:3-540-66247-2, Springerverlag.

[11] Mohammed Zaki, “Efficient enumeration of frequent sequences ” , 7th International Conference on Information and Knowledge , 1998 , http://www.datamining.org.tw/doc/similar/CIKM98_ps.pdf .

[12] Marko Paavola, Mika Ruusunen and Mika Pirttimaa, *Some change detection and time-series forecasting algorithms,* Report A No 26, March 2005, University of Oulu, Control Engineering Laboratory

[13] Dmitry B. Rokhlin, “Martingale selection theorem for a stochastic sequence with relatively open convex values”, 2006, Mathematics Subject Classification. 60G42, arXiv:math.PR/0602587, v1. 26 Feb 2006

12 ABOUT AUTHORS

1. **Mr. M.Nagabhushana Rao**, Professor in Department of computer science, C.R.Engineering ,Tirupati is also a Research Student in Dept. of Computer Science, S.V.University, Tirupati. Completed his BE in Computer Science from Amaravathi University, Completed MS in Software Systems from BITS, Pilani.

2. **Prof. P.Govindarajulu**, Professor in Department of computer science at S.V.University, Tirupati, has completed M.Tech in Computer Science from IIT Madras (Chennai), PhD from IIT Bombay (Mumbai). His area of research is Databases, Data mining, Image Processing ,Software Engineering.