

# A Survey On Spatial & Temporal Reasoning In Computer Science Using Data Mining(Stresdm)

P.Sathish Saravanan , AP-Department Of Information Technology ,  
Velammal Institute of Technology ,  
Chennai – India .

## Abstract

Nowadays Data Mining in Geo-spatial data extracted from spatio-temporal data will help us to have better prediction of spatial processes or events. Research area that started in Computer Science and information technology in the last decade and now penetrates into almost all data-rich environments like **spatio-temporal data mining**<sup>[1][3][5]</sup> here refers to the extraction of implicit knowledge, spatial and temporal relationships, or other patterns not explicitly stored in spatio-temporal databases. It is a subfield of data mining and knowledge discovery in databases.

This paper provides an overview of spatial and temporal reasoning in computer science and analyze a few research areas and their issues about the spatio-temporal importance. We have also listed a few techniques and comments for developing useful applications for future usage of **spatio-temporal data mining in computer science specialization**<sup>[1]</sup>.

**Keywords:** Spatio-Temporal Data Mining, Data Mining Process and Tasks, Spatio-Temporal Database, Constraint Networks and Qualitative Temporal, Spatio-Temporal Reasoning and Modelling, Region Connection Calculus, Geospatial Data Repositories, Algorithms Analyzation (Path-Consistency & Back-Track)

## 1. Introduction

Qualitative spatial-temporal reasoning which is based on qualitative abstractions of temporal and spatial aspects of the common-sense background knowledge on which our human perspective of physical reality is based. We analyze **spatiotemporal database**<sup>[6]</sup> is a database that manages both space and time information.

- 1. Tracking of moving objects, which typically can occupy only a single position at a given time.
- 2. A Database of wireless communication networks, which may exist only for a short timespan within a geographic region.
- 3. An index of species in a given geographic region, where over time additional species may be introduced or existing species migrate or die out.

Spatio-temporal databases are an extension of spatial databases. A spatio-temporal database embodies spatial, temporal, and spatio-temporal database concepts, and captures spatial and temporal aspects of data and deals with

- geometry changing over time
- location of objects moving over invariant geometry

### 1.1. Data Mining Techniques used in Spatialization & Temporalization

Data Mining includes spatio-temporal databases, machine learning systems, statistics, geographic visualization, and information theory.

- Spatial and Temporal relationships exist among spatial entities at various levels.
- The spatial relations, both metric (such as distance) and non-metric (such as topology, directions, shape, etc.), and temporal relations (such as before or after) may be explicit or implicit in the geographic databases.
- Spatial and temporal dependency and heterogeneity are intrinsic characteristic of spatio-temporal databases.
- Scale effect in space and time is a challenging research issue in geographic analysis.

### 1.2. Data Representation in Spatial-Temporal

Four broad categories of temporality within data are classified in a review of temporal knowledge discovery

- Like **Static**<sup>[13]</sup> -Time has to be traced by external information such as database construction,
- Like **Sequences**<sup>[13]</sup> -Ordered list of events, reveals relationships such as before and after, or the richer relationships described as meets, overlaps, contemporary,
- **Timestamped**<sup>[14]</sup> -a timed sequence of static data taken at more or less regular intervals,
- Fully **temporal**<sup>[14]</sup> -Integrated spatio-temporal data, e.g. via events, processes.

## 2. Constraints used in Qualitative Temporal

Using the steps to accomplished this model QT<sup>[14]</sup>

- the selection of a model that allows a trade off(**Time & Space Tradeoff**)between efficiency and representation
- a preprocessing step for adapting the input to the model
- a data mining algorithm able to deal with the properties provided by the model for generating a representative output

## 3. Representing Temporal Information

Temporal reasoning is an essential part of many artificial intelligence tasks. To develop a temporal reasoning component that is useful across applications such as planning and scheduling. Most of these calculi can be formalized as abstract relation algebras, such that reasoning can be carried out at a symbolic level. For computing solutions of a constraint network, a **Path-Consistency Algorithm**<sup>[8]</sup> and **Back-Tracking algorithm**<sup>[9]</sup> are important for two fundamental tasks: For that determining whether the temporal information is consistent, and, if so, finding one or more scenarios that are consistent with the temporal information.

### 3.1 Analization with Constraint Networks

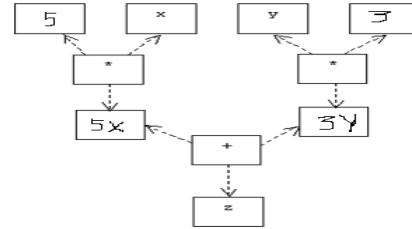
A **constraint network**<sup>[10]</sup> represents a mathematical relationship between several variables, and is able to compute the value of any one of these variables given the values of all the others. There differed in 2 ways in the nodes constraint are cells and constraints. Cells represent variables (read-only cells represent constants) and constraints represent primitive mathematical relationships such as  $z = x + y$  and  $z = x * y$ . The neighbors of a constraint are the cells that it constrains. The neighbors of a cell are the constraints that constrain it.

A number can be saved to or erased from a cell. In either case, the cell's neighboring constraints will be notified. Upon notification that a neighboring cell has been erased, a constraint erases all of its neighbors. Upon notification that a number has been stored in a neighboring cell, a constraint attempts to compute new values for its remaining neighbors. In either case, cells updated by a constraint notify the other constraints they are connected to. In this way cell updates cascade through the network.

For assume a, b, and c are cells connected to an addition constraint. If a number is stored in a, and if b already holds a number, then the adder constraint stores  $a + b$  in c. However, if b doesn't hold a value, but c does, then the adder constraint stores  $c / a$  in b. Contrast this "bi-directional" behavior with the "uni-directional" behavior of the ordinary assignment statement:

- $c = a + b$ ; which only assigns  $a + b$  to c

- $z = 5x + 3y$  Representation of this constraint network consist of 3 constraints & 7 cells:



**Fig.1 Constraint Network for Complex Relationship**

In this network the cells labeled 5 and 3 are constant or read-only cells. The cells labeled 5x and 3y hold  $5 * x$  and  $3 * y$ . The nodes labeled + and \* are constraints.

### 3.1.1 Analyze CONNNET constraint network

For an example CONNNET is a constraint network and executes the following console level commands:

#### help

```
help (displays this message)
quit (terminates session)
set NAME VALUE (update a variable cell)
addVar NAME (add a variable cell)
addConst NAME VALUE (add a constant cell)
display NAME (display a cell)
undisplay NAME (undisplay a cell)
addAdder ARG1 ARG2 RESULT (add an adder constraint)
addMult ARG1 ARG2 RESULT (add a multiplier constraint)
clear (get rid of all cells & constraints)
forgetAll (forget all cell values)
forget NAME (forget a cell value)
```

Let an example the program CONNNET to solve the equation:  $z = 3x + 4y$

Before that we create 3 cells representing the variables x, y, and z: **addVar x, addVar y, addVar z**

Next, we create 2 cells that hold the constants 3 and 4, as well as 2 cells to hold the intermediate values of 3x and 4y:

**addConst 3 3, addConst 4 4, addVar 3x, addVar 4y**

We create three constraints. The first asserts that 4y holds the product of 4 and y, the second asserts that 3x holds the product of 3 and x, and the third asserts that z holds the sum of 3x and 4y:

**addMult y 4 4y, addMult x 3 3x, addAdder 3x 4y z**

usually cells are updated quietly. If we want a cell to display itself when it is updated, we must explicitly ask it to do so. In our example, we are only interested in the values held by x, y, and z:

**display x, display y, display z**



```

s ← first(P,c)
while s ≠ Λ do
    bt(s)
    s ← next(P,s)
    
```

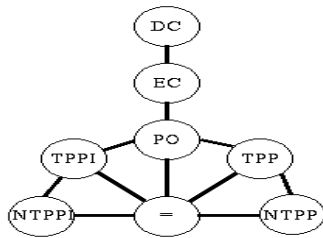
**4. Spatio-Temporal Reasoning and Modelling**

It encompasses the work done or in progress in the following broad sub-areas:

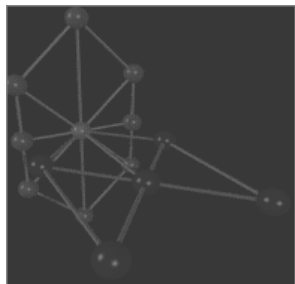
- (1)Spatio-Temporal Continuity,(2)Qualitative Modelling,(3)Spatio-Temporal Ontology

**4.1 Spatio-Temporal Continuity**

To formalize the intuitive notion of **spatio-temporal continuity**<sup>[11]</sup> for a qualitative theory of motion for describing how the relationships between regions change with time is necessary for modelling real-world processes. The generation of continuity network that demonstrates the allowed transitions between spatial relationships for the basic eight relation of RCC logic. This is shown in the figure on the left. Similar networks for extended forms of RCC have also been developed. The image on the right is actually a 3-dimensional depiction of a continuity network for an extended 15 relation form of RCC.



**Fig.5 Spatial relationship Diagram**



**Fig.6 A 3-dimensional depiction of a continuity network**

**4.2 Qualitative Modelling**

An important application of **RCC**<sup>[12]</sup> has been to produce a Qualitative Spatial Simulation program (QSSIM) for modelling physical and biological

systems.QSSIM is then able to generate successive 'environments' of the stages of the device's operation.

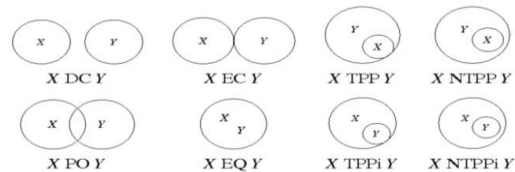
**4.3 Spatio-Temporal Ontology**

Taking space-time as primitive, we are working on an axiomatic theory for physical objects. We introduce and characterize by means of logical axioms the basic ontological distinctions needed to reason about physical objects for this **ontology**<sup>[12]</sup>.

The most important spatial-temporal calculus is region connection calculi(RCC), spatio temporal constraint calculus (STCC), and qualitative trajectory calculus (QTC) allows for reasoning about moving objects.

**5. Region Connection Calculus**

The **region connection calculus (RCC)**<sup>[12]</sup> serves for qualitative spatial representation and reasoning. RCC abstractly describes regions (in Euclidean space, or in a topological space) by their possible relations to each other. RCC8 consists of 8 basic relations that are possible between two regions:Like disconnected (DC),externally connected (EC),equal (EQ),partially overlapping (PO),tangential proper part (TPP),tangential proper part inverse (TPPi),non-tangential proper part (NTPP),non-tangential proper part inverse (NTPPi)



**Fig.7 Diagrammatic View of Relations, Combinations for Proper Part (PP) is the union of TPP and NTPP**

The composition table of RCC8 are as follows: "\*" denotes the universal relation.

		DC,	DC,	DC,	DC,			
		EC,	EC,	EC,	EC,			
<b>DC</b>	*	PO,	PO,	PO,	PO,	DC	DC	DC
		TPP,	TPP,	TPP,	TPP,			
		NTPP	NTPP	NTPP	NTPP			

**Tab.1 The Composition Table of RCC8**

**Examples**

The RCC8 calculus can be used for reasoning about spatial configurations. Consider the following example: two houses are connected via a road. Each house is

located on an own property. The first house possibly touches the boundary of the property; the second one surely does not. What can we infer about the relation of the second property to the road?

The spatial configuration can be formalized in RCC8 as the following constraint network:

```
house1 DC house2,house1 {TPP, NTPP}
property1,house1 {DC, EC} property2,house1 EC
road,house2 { DC, EC } property1,house2 NTPP
property2,house2 EC road,property1 { DC, EC }
property2,road { DC, EC, TPP, TPPI, PO, EQ, NTPP,
NTPPi } property1,road { DC, EC, TPP, TPPI, PO,
EQ, NTPP, NTPPi } property2
```

Using the RCC8 composition table and the path-consistency algorithm, we can refine the network in the following way:

```
road { PO, EC } property1,road { PO, TPP } property2
```

That is the road either overlaps with the second property, or is even (tangential) part of it.

## 6. Data Mining

**Data mining**<sup>[13]</sup> (the analysis step of the "Knowledge Discovery in Databases" process, or KDD), a field at the intersection of computer science and statistics, is the process that attempts to discover patterns in large data sets. It utilizes methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The Goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

### 6.1 Mining Process

The **Knowledge Discovery in Databases (KDD)**<sup>[13]</sup> process is commonly defined with the stages:

**(1) Selection (2) Pre-processing (3) Transformation (4) Data Mining (5) Interpretation/Evaluation.**

KDD process simplified knowledge discovery as

(1) pre-processing, (2) data mining, and (3) results validation.

#### 6.1.1 Pre-processing

Data mining can only uncover patterns actually present in the data, the target dataset must be large enough to contain these patterns while remaining concise enough to be mined within an acceptable time limit. A common source for data is a data mart or data warehouse. Pre-processing is essential to analyze the multivariate datasets before data mining. The target set is then cleaned. Data cleaning removes the observations containing noise and those with missing data.

#### 6.1.2 Data Mining Tasks

Data mining involves six common classes of tasks:

- **Deviation detection(OR) Anomaly detection** – The identification of unusual data records, that might be interesting or data errors and require further investigation.
- **Dependency modeling(OR)Association rule learning** – Searches for relationships between variables.
- **Clustering** – is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.
- **Classification** – is the task of generalizing known structure to apply to new data. For example, an e-mail program might attempt to classify an e-mail as "legitimate" or as "spam".
- **Regression** – Attempts to find a function which models the data with the least error.
- **Summarization** – providing a more compact representation of the data set, including visualization and report generation.

#### 6.1.3 Results Validation

The missing information about non-classification tasks in data mining, it only covers machine learning.

It is common for the data mining algorithms to find patterns in the training set which are not present in the general data set. This is called overfitting. To overcome this, the evaluation uses a test set of data on which the data mining algorithm was not trained.

## 7. Spatial Data Mining

Spatial data mining is the application of data mining methods to spatial data. The end objective of spatial data mining is to find patterns in data with respect to geography. So far, data mining and **Geographic Information Systems (GIS)**<sup>[2][4][5][6]</sup> have existed as two separate technologies, each with its own methods, traditions, and approaches to visualization and data analysis. The results for those organizations are:

- offices requiring analysis or dissemination of geo-referenced statistical data
- public health services searching for explanations of disease clustering



- environmental agencies assessing the impact of changing land-use patterns on climate change
- geo-marketing companies doing customer segmentation based on spatial location.

### 7.1 Challenges in Geospatial Data Repositories

**Geographic Information Systems (GIS)**<sup>[2][5][6]</sup> datasets are often splintered into feature and attribute components that are conventionally archived in hybrid data management systems. Algorithmic requirements differ substantially for relational (attribute) data management and for topological (feature) data management.

#### 7.1.1 Research Areas for Geospatial Data Repositories

- Developing and supporting geographic data warehouses (GDW's)
- Better spatio-temporal representations in geographic knowledge discovery
- Geographic knowledge discovery using diverse data types
- And also used in Sensor data mining, Visual data mining, Music data mining

### 7.2 Pattern Mining

Pattern mining is a data mining method that involves finding existing patterns in data. In this context patterns often means association rules. The "**Pattern-based data mining**"<sup>[15]</sup> looks for patterns (including anomalous data patterns) that might be associated with terrorist activity these patterns might be regarded as small signals in a large ocean of noise.

### 8. Conclusion & Future Work

In this paper, we propose various computational activities to obtain Qualitative spatial-temporal reasoning with Computer Science. Spatio-Temporal data mining is a promising research frontier for developing applications on Computational techniques in both Geographical Information Systems and Data Mining. It's not sufficient with manual data analyzation in spatio-temporal data mining with computer science Arena. So that we provide useful tools and technology for improving spatio-temporal activities in computational devices with increasing computerization in many fields, these days vast amounts of data are routinely collected. Here which we have listed few issues on which research can be done in future. Since many research challenges exist so scope of exploration in this field is quite vast. Hence there is always a need for efficient algorithms. Future efforts will be needed to refine the parameters in this approach. Incrementally update or approximate computing techniques can be applied to improve the computational efficiency with time space trade off system.

### References

- [1] Gennady Andrienko & Donato Malerba & Michael May & Maguelonne Teisseire "Mining spatio-temporal data".
- [2] Battenfield, B., Gahegan, M., Miller, H., Yuan, M. (2001), Geospatial data mining and Knowledge Discovery. UCGIS white paper on Emergent Research Themes.
- [3] Koperski, K., Adhikary, J., and Han, J. (1996), Spatial Data Mining: Progress and challenges. Proceedings of the ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery. Montreal, Canada: 55-70.
- [4] Miller, H.J. and Han, J. (2001), Geographic data mining and knowledge discovery: an overview. In Miller, H.J. and Han, J. (eds) Geographic data mining and knowledge discovery. London, New York : Taylor & Francis, 3-32.
- [5] Roddick, J.F. and Lees, B.G. (2001), Paradigms for spatial and spatio-temporal data mining. In H.G. Miller and J. Han (eds), Geographic Data Mining and Knowledge Discovery. London: Taylor & Francis.
- [6] Tamas Abraham and John F. Roddick, Survey of Spatio-Temporal Databases, *GeoInformatica* 3:1, 61-99 (1999).
- [7] Xiaobai Yao-Research Issues in Spatio-temporal Data Mining A white paper submitted to the University Consortium for Geographic Information Science (UCGIS) workshop on Geospatial Visualization and Knowledge Discovery, Lansdowne, Virginia, Nov. 18-20, 2003.
- [8] Moninder Singh-"Efficient Path Consistency Algorithms for Constraint Satisfaction Problems" in University of Pennsylvania.
- [9] Gurari, Eitan (1999). Backtracking algorithms "CIS 680:Datastructures:Chapter19:BacktrackingAlgorithms.
- [10] Rina Dechter, Itay Meiri and Judea Pearl-"Temporal Constraint networks" from Computer science Department, Haifa 32000, Israel & USA NOV-1989, 1990.
- [11] Brandon Bennett, and Derek R. Magee and Anthony G. Cohn and David C. Hogg "Using Spatio-Temporal Continuity Constraints to Enhance Visual Tracking of Moving Objects".
- [12] Randell, D. A., Cui, Z. and Cohn, A. G.: A spatial logic based on regions and connection, Proc. 3rd Int. Conf. on Knowledge Representation and Reasoning, Morgan Kaufmann, San Mateo, pp. 165-176, 1992.

[13] Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02-5, Two Crows Corporation, 10500 Falls Road.

[14] A. Cohn, and S. Hazarika. Qualitative Spatial Representation and Reasoning: An Overview". In Lecture Notes in Computer Science, Springer Berlin/Heidelberg, Qualitative Spatial Reasoning with Topological Information, Vol 2293/2002 , pp. 179-182, 2001.

[15] Taneli Mielikäinen "Summarization Techniques for Pattern Collections in Data Mining" Department of Computer Science Series of Publications A Report A-2005-1.

IJERT