Rajagiri Management Journal
Volume 10, Issue 2, December 2016

Spatio-Temporal Analysis of Big Data Generated from GPS Traces of Moving Objects for Personalized Customer Relationship Management

Sajimon Abraham [1] and Nishad A[2]

Abstract

Big data enables an organization to gain a much greater understanding of their user and customer base, their operations and supply chain, even their competitive or regulatory environment. Due to the rapid proliferation of online activities and the recent proliferation of connected mobile devices more customer touch points have turned electronic. Additionally temporal and moving data, in all its variant forms, might help boost efficiencies or scale down the cost of labor and other resources. The advancements in location awareness devices allows moving objects to track its movement paths with GPS traces, and users can manage their travel experiences on a web map and share travel knowledge among each other. Spatio-temporal data evaluation is vital for many applications relating to transportation infrastructure, supply chain logistics, tracking of customer tastes and personalized product advertisements. The spatio-temporal data generated from the trace of an object in motion is basically a big data to be managed in a special way and can be productively utilized for realizing personalized CRM.

Keywords: Customer Relationship Management, Moving Objects, Big Data,Spatio-Temporal Analysis.

---

[1] Assistant Professor in Computer & IT, School of Management and Business Studies. Mahatma Gandhi University, Kottayam. Mobile: 9446224240

[2] Administrative staff in Mahatma Gandhi University, Kottayam. Mobile: 8547443643

1. Introduction

The notion of big data and related application in business is gaining broad attention in recent time because of its great potential in generating business impacts (Chen and Chiang, 2012). "Big Data" is described as quantity of data over and above technology's potential to store, manage and process efficiently (Armour et al., 2013). The basic attributes of big data are identified along three important dimensions, namely volume, velocity, and variety (Zikopoulos and Eaton, 2011).

Advancement in the location centric gadgets such as smart phones, smartdriving assistants etc. generates information that has geographical and time relevance. The spatio-temporal data, one of the branches of big data management which augmented the relational data base system has thus emerged. The spatial component in the spatio-temporal data represents entities defined in geometric space. Spatio-temporal database is the augmented variant of spatial database which enables the monitoring of objects that keep on changing their place in the space.

1.1. Objective

Businesses have acquired data in the electronic form,resulting in the generation and storing up of volumes of data in the electronic repository. This warehouse of organization data (big data) becomes a potential resource for various kinds of analytical operations that will ultimately benefit the organization.

This study analyses how online marketing and customer relationship applications can benefit from the outcomes of the studies on moving objects supported by their GPS traces. The spatio-temporal details on the objects that frequently change their place of existence (customer) is recorded as trajectories in the moving object database. By applying clustering techniques on the trajectories, the customer shopping behaviour pattern can be procured which will augment the CRM analysis.

2. Relevant Studies

Data Centric Business Intelligence and Analytics Technologies are receiving wide spread acceptance. Certain discussions in this area are mentioned here.

Traditionally marketing intelligence relied on market surveys to find out consumer behavior and to improve customer relationship. For example, companies use consumer feedbacks to study customer attitudes towards company's product and services. With the help of analytic tools supported by Big data Technologies, parameters for tactics of marketing decisions like end user standpoint about a product, customer support and the like can be automatically monitored (Tan et al., 2013) by mining vast amount social media. However, while providing uncommon spaces for marketing intelligence, big data also imports challenges to practitioners and researchers. The challenges are mainly with regard to its storage, maintenance and processing (Kaisler et al.) For typical marketing intelligence tasks such as customer opinion mining, companies nowadays have innumerable options to collect data from a multiplicity of information sources like social media data, transactional data, survey data, sensor network data, etc. Based on the characteristics of collected data, different strategies can be applied to discover marketing intelligence.

Researches for reaping the advantages of advancements of social networking sites in CRM are still underway. Wide spread availability of less expensive connected devices has increased the adoption of social networking platforms for their daily updates. The situation called Social CRM (Greenberg et. al, 2009) will provide an easy board for customers to express their real experience of a purchase of good or service. Customer reviews and ratings in shopping cart sites and discussion forms (forums) gain acceptance and will be considered as unbiased opinion when compared to the taglines of the company. Edward C. Malthousea (2013) put forward a framework for the effective utilization of social networking. It examined how the vital component of CRM acquisition, maintenance, and termination fit with the data generated from social networking platforms, with special reference to the unsuspected dangers implied. Analysis models developed with the aid of a single data source may only provide limited insights, leading to potentially biased business decisions due to the dynamic changes happening in today's business. On the other hand,incorporating data derived from different sources provides a comprehensive view of the domain and leads to more specific marketing intelligence. Unfortunately, integrating big data from diverse data silos to generate marketing intelligence is not a trivial task. This prompts exploration of new methods, applications, and frameworks for efficient handling of big data in the ambience of marketing intelligence.

The advancement of mobile based communications and ubiquitous computing spread through our culture, and wireless networks sense the passage of people and vehicles, producing large volumes of mobility data. With the support of facilities such as GPS location tracking and high speed mobile networks, studies on moving object data analysis has been simulated securing widespread acceptance for robust systems for monitoring, collecting and storing of location.

All these facilitates massive depositories of spatio-temporal data (big data) tracking persons mobile activities that necessitates appropriate analytical methods capable of enabling the development of creative and position-sensitive applications. This is a situation of great freedom and challenge: on one side, obtaining these details can take one to useful knowledge, feasible mobility and intelligent transportation systems; on the other side, individual privacy is at risk, as the mobility data may reveal, if disclosed, sensitive personal information.

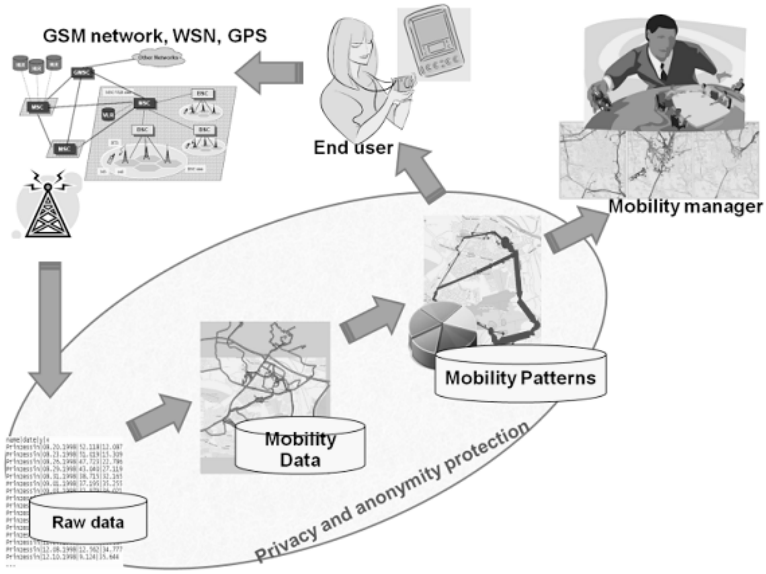
The "GeoPKDD (Geographic Privacy-aware Knowledge Discovery and Delivery)",a project backed by European Commission under the Future and emerging technologies (FET) program of the 6th Framework (FP6), has been to discover useful knowledge about human movement behavior from mobility data, while preserving the privacy of the people under observation. The various processes in this project (Giannotti and Trasarti, 2010) are shown in Figure1.

The GeoPKDD aims at improving decision-making in many mobility-related tasks, especially in tourist information system and in mining tourist mobility in the following ways.

- Forecasting traffic and public transportation systems
- Localizing new facilities and public services
- Predicting traffic-related phenomena
- Location-based marketing and advertising
- Innovative info-mobility services
- Detecting deviations in collective movement behavior.

Traditional database management systems especially RDBM Systems are not capable of managing the spatio-temporal data, which is treated as big data, generated from these kinds of devices in its sense because of its multi-dimensional nature and complex behavior. A complete replacement of existing storage and retrieval mechanism to accommodate the spatio-temporal data is not an acceptable solution; instead plug in additional features in the existing system can be pursued. Researches in the area of Moving Object Database domain have started during 1990s but recent trends in the technology demands more advancement in managing the data.

Fig. 1 : Process involved in Mobility Mining (Courtesy: http://www.geopkdd.eu)

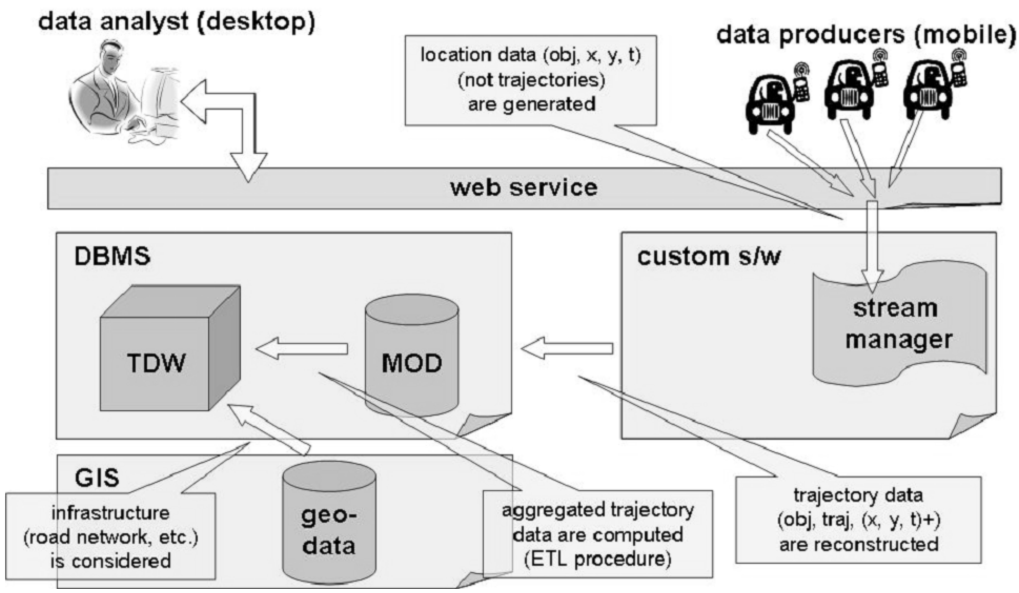


3. Definitions and Concepts

3.1. Moving Object Databases

A database management system and a trajectory warehouse have been designed around this specific form of data which are continuously moving in the two dimensional space. The design of the trajectory warehouse has been influenced by the Moving Object Databases (MOD) research community, which extends the conventional database technology for modeling, indexing and querying trajectory data. In this, the spatial and temporal dimensions are of primary concern both past and current and future positions of moving objects are of interest (Guting, 2005, Mokbel, 2004, Wolfson, 2000). The structure of a trajectory warehouse is shown in Figure2.

Fig. 2: Moving Object Database and Warehouse



3.2. Trajectory

A trajectory is a sequence of time-stamped locations, sampled from the itinerary of a moving object.A moving object trajectory is defined as a polyline in the three-dimensional space, where first two dimensions refers to the location and the third dimension to time. It may be represented as a series of points ((x1,y1,t1), (x2,y2,t2)……..(xn,yn,tn)) with t1<……… tn. For a trajectory tr, its spatial projection on the two dimensional plane is called the route of tn. The locale of object in motion is given as an implicit function of time. The entity is at position (xi,yi,) at time ti and during each period [ti,ti+1] , the object is assumed to move along a straight line from (xi,yi) to (xi+1,yi+1) at a fixed speed.

While the Trajectory Database analytical tools concentrate on presence of moving objects, mobility mining is aimed at analyzing movement. Basically mobility data mining address two different tasks: first, to define the format of special models and patterns on spatio temporal data to be extracted from trajectory data, and second, the task of designing and implementing efficient algorithms for extracting such patterns and models. The GeoPKDD(Giannotti and Trasarti, 2010) focusing on various data mining task like trajectory similarity and patterns, trajectory clustering, and trajectory classification. This paper discusses the issue related with trajectory similarity and how this method could be applied for finding similarity in customer's behaviors pattern. We use a measure for determining the similarity of customer moving vehicle on road network. Here the trajectories are made from the GPS traces of the object, recorded in the remote moving object database.

3.3. Trajectory derived from GPS Traces of Moving Objects

GPS based modern monitoring systems such as positioning and mobile phone networks have made available warehouse of data having spatial and temporal dimension by recording human mobile activities, call for suitable analytical methods, capable of enabling the development of innovative, location-aware applications. Variety of objects such as mob, wild animals, materials, eatables, data and even concepts move in increasing speeds over varying distances and in increasing quantity. Hence mobility is the key processes in our present world.Distinguishing this type of mobility patterns is a key factor for significant decision making activity in operational management domain such as fleet management, transportation modeling, urban planning, tourism, wildlife ecology, spatial epidemiology, location-based services, flight safety, personalized advertisemsent, on-line marketing and customer relationship managment.

The researchers have worked on the issue of spatio-temporal similarity of vehicle trajectories (M.F. Mokbel, 2004) considering spatial similarity as a combined version of both structural and sequence similarities measured with the dynamic programming approach. This paper is a continuation of the work mentioned above, which will use spatio-temporal similarity measure to extract knowledge of semantic location and related activities from traces of moving objects added with variety of other information, such as a list of visited websites while travel (web-click stream), user behavior patterns, environmental factors or energy aspects as large sensor data(O. Wolfson,2000).

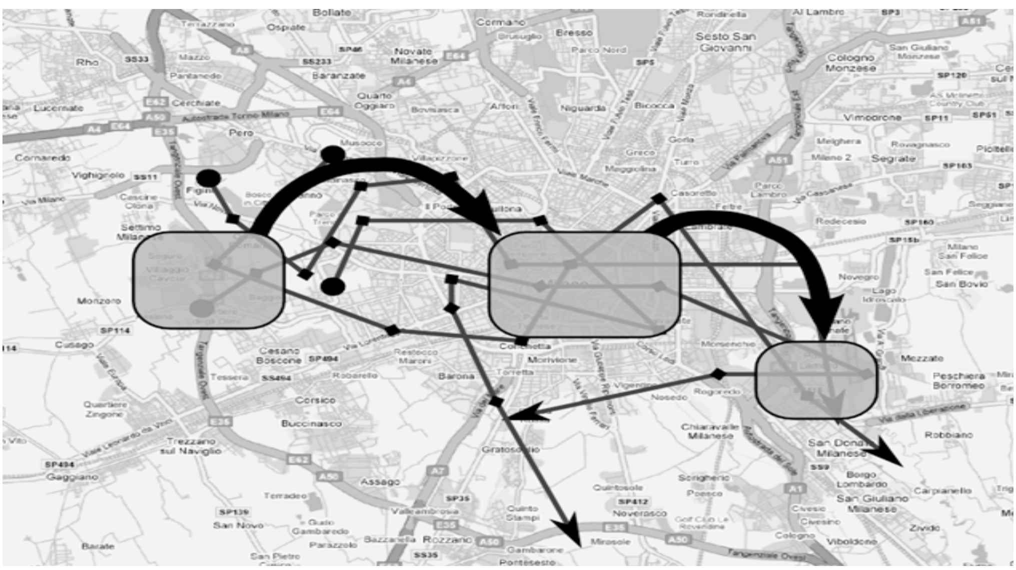
4. Design and methodology

In this paper we propose algorithmic approach over the trajectory data generated from the traces of moving objects. We have used the technique in mobility mining similarity search to predict similarity pattern of vehicle movement which will ultimately be used for customer preferences in shopping.

4.1.1.Trajectory Similarity and Pattern Mining

Trajectory pattern is a novel notion of spatio-temporal pattern, which formalizes the idea of aggregated movement behaviors. A trajectory pattern, as defined in (Giannotti, 2007), represents "a set of individual trajectories and having the property of traversing the same sequence of places. Therefore, two notions are central: (i) the regions of interest in the given space, and (ii) the typical travel time of moving objects from region to region". In this approach a trajectory pattern is a sequence of spatial regions that, on the basis of the source trajectory data, emerge as frequently visited in the order specified by the sequence; in addition, the transition between two consecutive regions in such a sequence is annotated with a typical travel time that, again, emerges from the input trajectories.

Fig. 3: A typical Trajectory Model



For example a pattern may be viewed as a typical behavior of frequently visiting places of tourists, the vehicle station and spend about two hours there before getting to the adjacent museum. A figurative illustration of a specific trajectory pattern is shown in Figure 3.

4.1.2. Trajectory Clustering

Clustering is one of the general approaches to explore and group large amounts of data, since it allows the analyst to consider set of objects rather than individual objects, which are too many. Clustering consorts objects in groups (clusters) such that the objects with in a group share some properties that do not hold (or hold much less) for the other objects. Spatial clustering builds object that are having similar spatial properties (shapes, spatial relationships in its components, etc.). Spatial clustering groups having common traits in spatial attributes (shapes, spatial relationships among components, etc.) are grouped together in spatial clustering. Simple distance-based clustering methods are not effective in separating trajectory clusters that exhibit a non-convex (non- globular) shape, as it often occurs in practice.

4.1.3. Trajectory Classification and Location Prediction

Predictive models for trajectory include a classification method for inferring the category of a trajectory, and a predictor for the consequent location forecasting of a moving object once its trajectory is given. There is strong research interest in next location prediction, in that it enables several intelligent location-based services. In the literature, this task is achieved by applying various learning methods to the trace of moving object for the purpose of creating an individual location predictor.

4.1.4. Trajectory Anonymity

In the management of personal mobility data, privacy is a major constraint in which location data allow inferences which may help an attacker to discover private information, such as individual habits and preferences. Just replacing identifiers with pseudonyms doesn’t guarantee anonymity, since location represents a property that may allow re-identification: for instance, location characteristic of house and office can soon be identified with the use of visual analytics methods, given detailed personal trajectories. Anonymization techniques provide double goals: first to decrease the probability of re-identification below an acceptable threshold, and second maintaining the analytical power of the data.

4.1.5. Visual Analytics

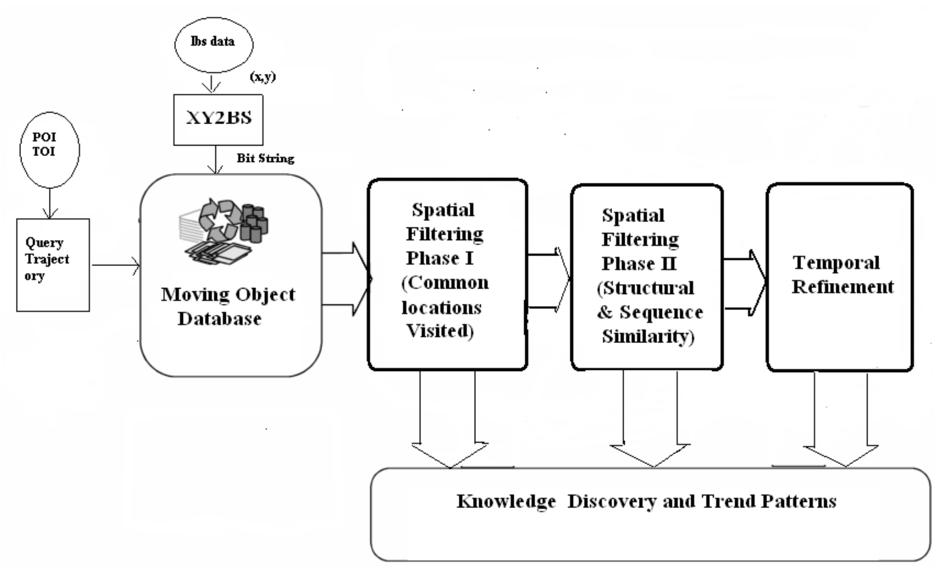
The aim of this system is to navigate through mobility data and patterns and to visually drive the analytical process (Andrienko and Andrienko, 2008). The major features include: the visualization and navigation of trajectory patterns in the spatial and temporal dimensions; the progressive refinement of trajectory clusters, through user-driven exploration and analysis of the discovered trajectory clusters (Rinzivillo et al., 2008) and the visual presentation of various measures provided by the trajectory Warehouse (Orlando, 2007).

5. Process Model

Finding the similarity pattern in the trajectory is the core theme of our research. Here we concentrate our research on spatio-temporal similarity to mine similarity in customer preferences in shopping behavior data. We introduce the similarity measures among two moving object trajectories in a constrained network incorporating concepts as discussed below and shown in Figure4.

a) Locations(Common) visited by two trajectories
b) Structural similarity of locations in the trajectories
c) Sequence Similarity of trajectories
d) Temporal refinement of trajectories

Fig.4: Similarity based Knowledge recovery Process Model



5.1. Common URL's visited by two trajectories

Let Ti and Tj be two trajectories. We introduce a measure of Spatial Similarity measure $Sim_c(T_i,T_j)$ which attempt to incorporate number of common locations in trajectories. This similarity matrix ***calculates*** the number of common locations commonly visited by two moving objects relative to the total number of locations in both trajectories.

5.2. Structural Similarity

Here by structural similarity we mean how geographically the points are close to each other. Structural similarity is measured by considering each of the four strings, separately and for each bit string, weightage is assigned from left to right in the order 1,2,3 etc. Then we compare each corresponding bit of both the token strings bit by bit from the beginning until the first pair of bits is different. The similarity is defined as the sum of the weight of those tokens which are matching divided by the sum of the total weights.

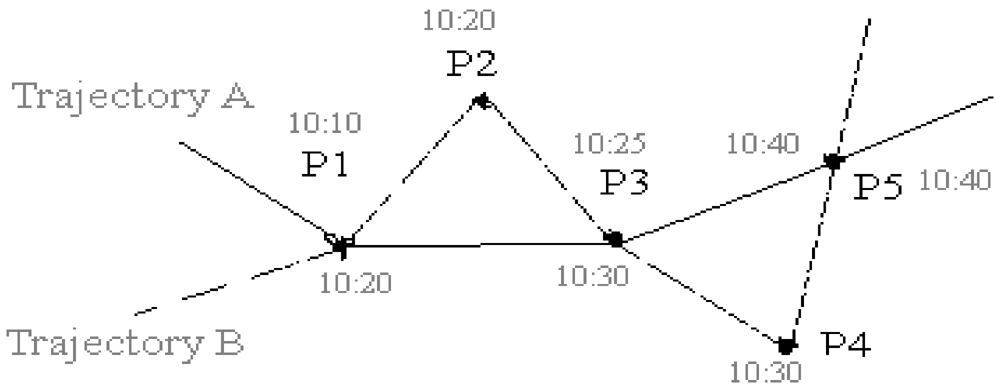
5.3. Sequence Similarity of web sessions

The mentioned spatial similarity is checking only the percentage of locations which are commonly visited by each input trajectory with the query trajectory made by points of interest. As we consider important individual locations in Point of interest, the sequence or order in which these locations visited by an input trajectory is also to be considered in calculating the actual similarity. For example in the domain of security informatics if POI contains strategic locations and then the location sequence of a moving car crosses will also have to be considered in finding how a user movement trajectory matches with the given query trajectory created with the given set of POI. Here we consider the given original trajectory data containing location and time information as a set of sequences, and apply sequence alignment method to measure similarity between trajectories.

5.4. Temporal Refinement

In real applications, the information on time associated with each trajectory is also very important in addition to location information. In order to find the similarity we have to consider the concept of space and time together. Here we are considering the temporal distance by taking the difference in web page request time as shown in Figure5.

Fig. 5: Temporal Distance of Two Trajectories



Temporal distance between Trajectory A (T_A) and Trajectory B (T_B) will be $Dist_T$ (T_A,T_B) = Differences in time at common URL's visited P1,P3,P5 = 10+5+0 = 15

6. Findings and Discussions

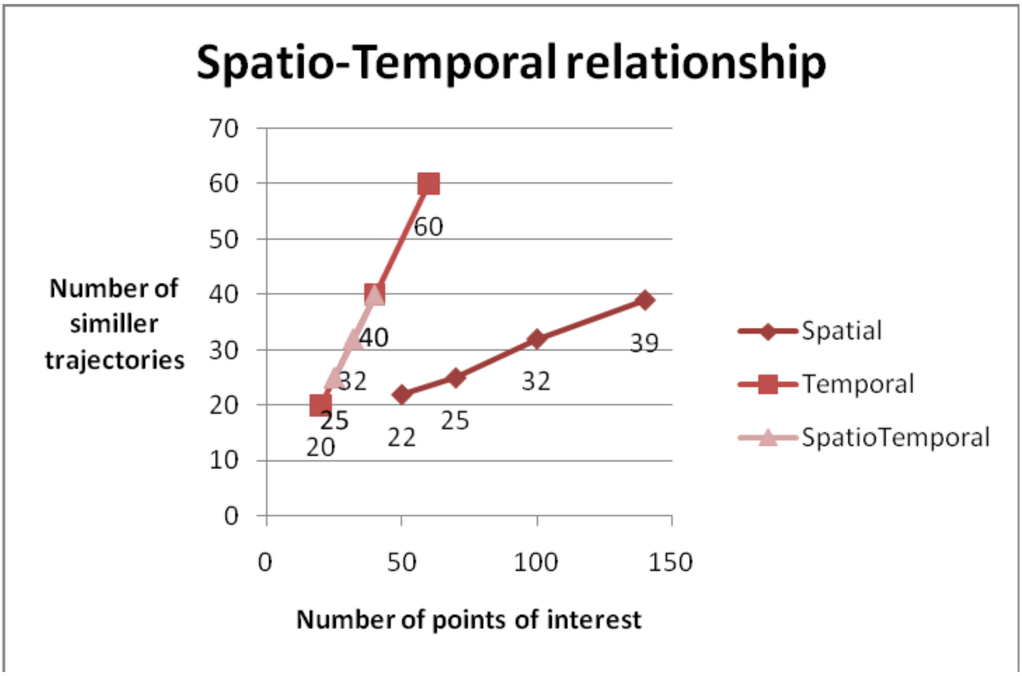
6.1. Data Set

We have taken a real-world data of customer shopping for experimentation purposes, namely a series of customer cabs that make trajectory data set. The data set consists of 576 trajectories of 150 cabs moving places around Athens metropolitan area in Greece for 30 distinct days. The format of each record is as follows:

{obj-id, traj-id, date(dd/mm/yyyy), time(hh:mm:ss), lat, lon, x, y}, where (lat, lon) is in WGS84 reference system and (x, y) is in global GGRS87 reference system.

Each of the three methods generates knowledge based on the methods proposed which can be used for mining customer shopping profiles. The results of experiments confirm that (Figure6) the third measure which is spatio-temporal based similarity search is more consistent than other two types of similarity search.

Fig.6: Comparison of Spatial, Temporal and Spatio-Temporal Similarity Search



6.2. Evaluation of Accuracy of Algorithms

In this experiment, we focus on evaluating the accuracy factors of the algorithms, i.e., the amount of false positives measure and measure of false negatives. A false negative (FN), is the error situation in not finding a pattern that does exist in the data. A false positive (FP), is the error situation of finding a "pattern" that does not exist in the data. A True Positive (TP) is the case of finding situation where a pattern does exist in the data and a True Negative (TN) is the case not finding situation where a pattern which does not exist in the data. Below formulae were used to find out measures such as sensitivity, specificity and accuracy:

Sensitivity = TP / (TP + FN)
Specificity = TN / (TN + FP)

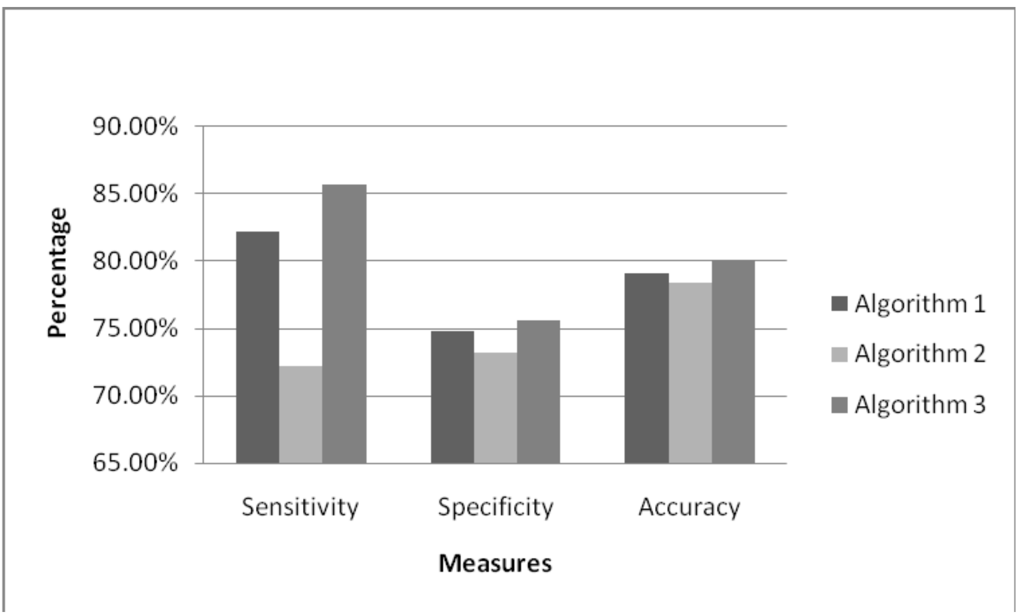
Accuracy = (TP + TN) / (TP + FP + TN + FN)

Table 1 shows sensitivity, specificity and accuracy for different similarity algorithms in our proposal. Fig 7 shows the graphical representation of these measures.

Table 1: Sensitivity, Specificity and Accuracy

Similarity Algorithms	**Sensitivity**	**Specificity**	**Accuracy**
Algorithm 1 -Spatial	82.24%	74.82%	79.11%
Algorithm 2 - Temporal	72.22%	73.16%	78.45%
Algorithm 3 – Spatio-Temporal	85.75%	75.63%	80.1%

Fig.7: Performance of Sensitivity, Specificity and Accuracy



A Receiver Operating Characteristic (ROC) space is defined by False Positive Rate and True Positive Rate which shows relative trade-off between true positive and false positive.

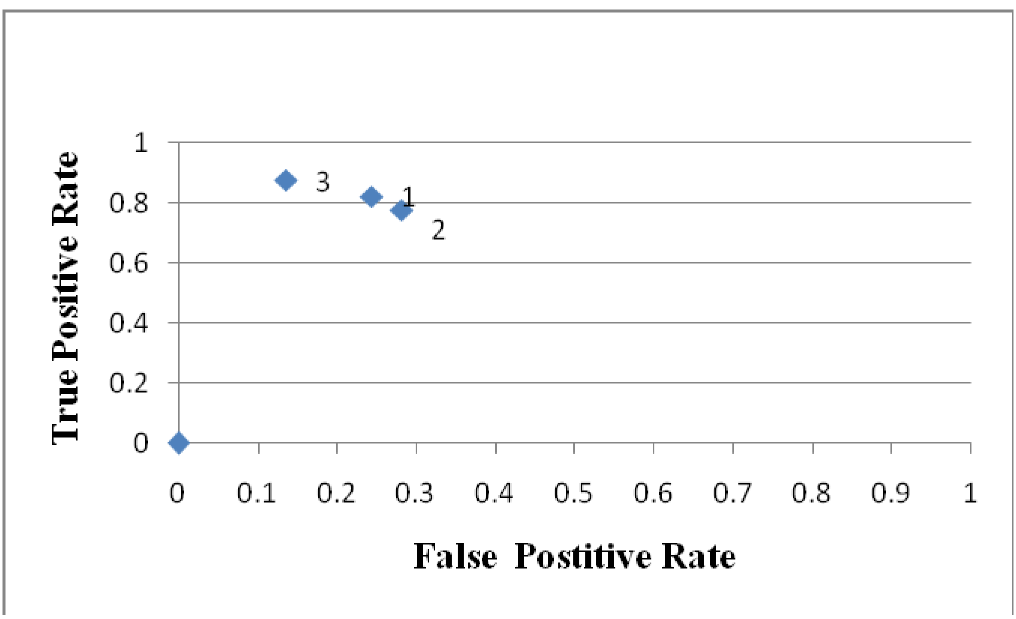
True Positive Rate = TP / (TP + FN)
False Positive Rate = FP / (FP + TN)

Table 2 shows True Positive Rate and False Positive Rate for above three Algorithms. Figure8 shows graphical format of True Positive Rate and False Positive Rate for these algorithms. The best possible prediction model will be at coordinate (0, 1) in graph on Figure8. This will represent 100% True Positive Rate and no False Positive Rate which will be ideal case. In none of these methods we could reach this ideal rate but algorithm 3 which takes the spatio-temporal measure is more nearer to the ideal point.

Table 2: True Positive Rate and False Positive Rate

Similarity Algorithms	True Positive Rate	False Positive Rate
Algorithm 1 - Spatial	0.8525	0.2433
Algorithm 2 - Temporal	0.8236	0.2512
Algorithm3–Spatio-Temporal	0.8768	0.2353

Fig.8: ROC Space Analysis



Our results shows that the revised algorithms proposed in finding spatio-temporal similarity with binary encoding scheme offers advantages over the performance of the existing algorithm in terms of search time, False classification rate and accuracy.

7. Conclusion

The similarity problem of moving object in trajectory database for objects moving on road networks has many applications in Transportation Network like ensuring tourist security, re-routing of vehicles identifying traffic congestion and re-routing etc.In this paper we have suggested a method to measure the space and time based similarity of moving object trajectories highlighting applications in mining customer mobility data in a sales and marketing environment.

As a continuation work we are planning to extend the use of spatio-temporal similarity measures in the extraction of semantic location and activity knowledge from moving object GPS traces, possibly augmented with other information such as web-sites visited by customers and customer behavior patterns.

References

Guting, R. H., Almeida, V., Ansorge, D., Behr, T., Ding, Z., Hose, T., &Telle, U. (2005, April). Secondo: An extensible dbms platform for research prototyping and teaching. In *21st International Conference on Data Engineering (ICDE'05)* (pp. 1115-1116). IEEE.

Mokbel, M. F., Xiong, X., Aref, W. G., Hambrusch, S. E., Prabhakar, S., &Hammad, M. A. (2004, August). PLACE: a query processor for handling real-time spatio-temporal data streams. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*(pp. 1377-1380). VLDB Endowment.

Wolfson, O., Cao, H., Lin, H., Trajcevski, G., Zhang, F., &Rishe, N. (2002, March). Management of dynamic location information in domino. In *International Conference on Extending Database Technology* (pp. 769-771). Springer Berlin Heidelberg.

Giannotti, F., &Trasarti, R. (2009). Data Mining and Privacy: The GeoPKDD Paradigm.

In *Proceedings of the Fourth SIAM Conference on Mathematics for Industry (MI09), San Francisco, California* (pp. 10-18).

Giannotti, F., Nanni, M., Pinelli, F., &Pedreschi, D. (2007, August). Trajectory pattern mining. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 330-339). ACM.

Andrienko, G., &Andrienko, N. (2008, September). A visual analytics approach to exploration of large amounts of movement data. In *International Conference on Advances in Visual Information Systems* (pp. 1-4). Springer Berlin Heidelberg.

Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N., &Andrienko, G. (2008). Visually driven analysis of movement data by progressive clustering. *Information Visualization*, *7*(3-4), 225-239.

Orlando, S., Orsini, R., Raffaetà, A., Roncato, A., &Silvestri, C. (2007, September). Spatio-temporal aggregations in trajectory data warehouses. In *International Conference on Data Warehousing and Knowledge Discovery* (pp. 66-77). Springer Berlin Heidelberg.

Chen, H., Chiang, R. H., &Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS quarterly*, *36*(4), 1165-1188.

Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media.

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013, January). Big data: issues and challenges moving forward. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on* (pp. 995-1004). IEEE.

Tan, W., Blake, M. B., Saleh, I., &Dustdar, S. (2013). Social-Network-Sourced Big Data Analytics. *IEEE Internet Computing*, *17*(5), 62-69.

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013, January). Big data: issues and challenges moving forward. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on* (pp. 995-1004). IEEE.

Malthouse, E. C., Haenlein, M., Skiera, B., Wege, E., & Zhang, M. (2013). Managing customer relationships in the social media era: introducing the social CRM house. *Journal of Interactive Marketing*, *27*(4), 270-280.

Greenberg, P. A. U. L. (2010). CRM at the Speed of Light: Social CRM Strategies. *Tools, and Techniques for Engaging Your Customers, New York*.