How Trajectory Data Modeling Improves Decision Making?

Noura Azaiez and Jalel Akaichi
Department of Computer Science, ISG-University of Tunis, Le Bardo, Tunisia

Keywords: Trajectory Data, Trajectory Data Source, Modeling, Trajectory Data Mart, Bottom-up Approach.

Abstract: The incredible progress witnessed in geographic information and pervasive systems equipped with positioning technologies have motivated the evolving of classic data towards mobility or trajectory data resulting from moving objects' displacements and activities. Provided trajectory data have to be extracted, transformed and loaded into a data warehouse for analysis and/or mining purposes; however, this later, qualified as traditional, is poorly suited to handle spatio-temporal data features and to exploit them, efficiently, for decision making tasks related to mobility issues. Because of this mismatch, we propose a bottom-up approach which offers the possibility to model and analyse the trajectories of moving object activities in order to improve decision making tasks by extracting pertinent knowledge and guaranteeing the coherence of provided analysis results at the lowest cost and time consuming. We illustrate our approach through a creamery trajectory decision support system.

1 INTRODUCTION

The integration of the mobility notion into the professional life is among keys that can raise organizations toward success and guarantee its durability since it offers the possibility of reducing services geographical disparities. Thus, it becomes possible to generate a large volume of Trajectory Data (TD) that reflect the movements and stops of the mobile objects in the real world. All provided TD are stored into a data model named Trajectory Data Source (TDSrc) model. However, the operational data sources are poorly suited to long-term vision and therefore seem inadequate for decision making. Towards this inadequacy, we propose an approach that offers the possibility to easily analyze trajectories of the moving objects in order to extract pertinent knowledge and guarantee the provided analysis results' coherence at the lowest cost and time consuming. We take as example the milk transfer to creameries; milk producers are asked to transfer the milk to a given creamery in a limited time while keeping its preservation; what is a little difficult for some producers given the large path between the center of collecting raw milk (creamery) and productive farms. In order to facilitate the process of the milk transfer to creameries, mobile cisterns was appeared. Those latter move towards a set of designated farms to collect raw milk and therefore to reduce the milk collection costs for producers.

In this paper, we discuss two important issues: the first one focuses on how to integrate TD in organization Information Systems (IS); the second issue expresses how can we rely on the *bottom-up* approach, which results the generation of Trajectory Data Marts (TDMs) or Trajectory Star Schemas (TSSs) directly from TDSrc, at the aim to analyze trajectories achieved by Mobile Milk Cistern for collecting raw milk, discover localizations that provide the best milk quality and therefore the main factors which are responsible for it.

To reach goals scheduled above, we propose to organize this paper as follow. Section 2 presents an overview concerning the problem of moving objects trajectories' modeling. In section 3, we discuss related works. Section 4 expresses our position. In section 5, we propose a logical model that gathers TD provided from milk mobile cisterns' trajectories. Section 6 describes our proposed approach steps. We conclude the paper in section 7.

2 STATE OF THE ART

Mobile objects are characterized with their different movements and positions into a time interval. These various positions describe the mobile object behav-

Azaiez N. and Akaichi J..
How Trajectory Data Modeling Improves Decision Making?.
DOI: 10.5220/0005558300870092
In *Proceedings of the 10th International Conference on Software Engineering and Applications* (ICSOFT-EA-2015), pages 87-92
ISBN: 978-989-758-114-4

ior changes and therefore its trajectories. In the literature, a great interest is expressed in the management of moving objects. Some works are interested only into the operational IS modeling (Spaccapietra *et al.,* 2007). While databases models are not suitable to describe the multidimensional aspects, multiple works are interested in the multidimensional modeling of mobile objects behaviour.

In the literature, there are three approaches leading to store data in a Data Warehouse (DW); the *top down* approach (Tryfona *et al.,* 1999), the *bottom up* approach (Kimball *et al.,* 1998) and the *middle out* approach (Sapia *et al.,* 1998).

To the best of our knowledge, approaches that are proposed to investigate the problem of Trajectory Data Warehouse (TDW) modeling are based only on the *top-down* approach.

Marketos *et al*. (2008) dealt with the problem of TDW building. Indeed, they proposed a framework for TDW that takes into consideration the complete flow of tasks required during a TDW development. In fact, the first step consists to apply the trajectory reconstruction process on the raw time-stamped location data in order to generate trajectories. The second step relies on the Extract-Transform-Load (ETL) process that has to plays its important role in order to feed Trajectory Data Cube. To achieve this goal, authors proposed two alternative solutions: a (index based) *cell-oriented* and a (non-index-based) *trajectory oriented ETL process*. The final step offers OLAP capabilities over the aggregated information contained in the trajectory cube model. Authors evaluated the efficiency of the proposed solutions by implementing the TDW architecture for a real-world application; an e-Courier dataset.

Arfaoui *et al.* (2011) presented a trajectory model related to the displacement of the herd, which allows building a DW containing trajectory data. These later are generated following the monitoring of herd movements. To achieve this, authors proposed two models. The first one is based on the Entity-Relation (ER) model which helps to visualize the different entities corresponding to the trajectory and its different components as well as the relation that can exist between them. The second one is based on the Spaccapietra model. Results show that the second model is more efficient for generating a TDW model since it puts into consideration spatial and temporal aspects, which is neglected by the first model.

Errajhi (2014) investigated the problem of TDW modeling using a new method which is based on a generic model that it could easily be adapted to any domain. Their work shows the benefits of fuzzy logic in solving the challenges related to TD by integrating fuzzy concepts into the conceptual and the logical model. Fuzzy sets provide mathematical meanings to natural language and are therefore able to handle the imprecision of the natural language. For that, they integrated the fuzzy logic in the TDW modeling. Then, authors' work has been implemented to ambulance management domain.

Leonardi (2014) presents a new approach that aims at designing a TDW model, having the ability to store and analyze trajectory data. Authors included the spatial and temporal notions in dimensions. Therefore, the framework collects streams of spatio-temporal observations related to the position of moving objects; this presents the first task to reconstruct TD. The next step presents the ETL phase where such reconstructed trajectories will be used in order to load the aggregated data into the proposed TDW. Authors implemented T-WAREHOUSE, a system that incorporates all the required steps for Visual Trajectory Data Warehousing, from trajectory reconstruction and ETL processing to Visual OLAP analysis on mobility data.

3 DISCUSSION

Following the study of a sample of works related to the TDSrc and TDSS modeling, we present a comparative study based on weaknesses and strengths of each research work.

As the "trajectory" term is a new concept, we note that the researchers were interested in trajectories modelling in the first place. The work of (Spaccapietra *et al.,* 2007) is among works which tried to solve the problem of TDSrc conceptual modelling. Trajectories' studies didn't stop at this level. In fact, the last work presents the basis of the work (Arfaoui *et al.*, 2011) where authors are concentrated on the modelling of the animals' trajectories. They proposed two ways of trajectory data models to build a TDW model allowing the storage of TD. The first one is based on the ER model and the second is based on the model of (Spaccapietra *et al.,* 2007).

The new notion "fuzzy", that is interposed in (Errajhi 2014), gives another manner to study the TDW modelling. This method gives a meaning to TDWs when fuzzy data are involved and especially when users want to ask questions in natural language.

The main goal that gathers the majority of works cited above is to model a generic TDW design that is able to facilitate the decision-making in different areas. To achieve this goal, we note that these works proceeded the same classic strategy; they take the *top-down* approach as the base of their works. This later is designed using a normalized data model. Atomic data, that are data at the lowest level of detail, are stored in a central repository TDW. Besides, TDMs containing data needed for specific business processes are created from TDW; this requires the TDW whole conceptual modelling as a first step; what costs expensive and therefore conflicts with the organization goal.

4 OUR POSITION

Analyzing TD requires their integration in the organization DSS. In order to guarantee the better decision making, it's important to choose the best strategy to adopt for the TDSS modelling. As it's cited in the section above, there are three approaches leading to storing data in a DW; the *top down* approach, the *bottom up* approach and the *middle out* approach. We compared these three approaches and we conclude that the *bottom up* approach is the most suitable for designing a TDSS that analyses TD since it presents the organization's information consistently and make data quickly and easily accessible; what is the key of the organization success. Unlike the *bottom up* approach, the *top down* approach is considered the heaviest one and the most expensive since it requires the whole conceptual modeling of the TDW. Besides, our TDSS modeling vision conflicts with the principle of the *middle out* approach. In fact, this latter involves, sometimes, cutting compromises that duplicate identical dimensions for practical purposes; and this leads to the analysis results disturbance.

Kimball is among authors that support the *bottom up* approach principle. He stated that the DW is nothing more than the union of all the DMs. In other word, DMs are created providing thin views into the organizational data then combined into a larger all-encompassing DW. Kimball's methodology consists to i) select business process, ii) declare the grain, iii) choose dimensions and iiii) identify facts. This methodology leads to identify star schema elements and therefore gives the birth of a set of DMs. However, these steps' scheduling seems not effectively adequate for TDMs building. In fact, identifying the grain firstly, then dimensions and finally facts can complex the process of building trajectory dimensional star schemas.

For this, we propose an approach that investigates the problem of TDSS modelling relying on the Kimball's philosophy but not Kimball's design steps. In fact, it's more coherent to extract facts from the TDSrc model and then to identify related dimensions. Thus, it becomes easy to specify granularities that can be hierarchies for these axes of subject analysis. Our methodology consists to reverse the DW design steps proposed by Kimball; this leads to facilitate the TDMs holding.

In our example, all provided TD have to be analyzed in order to get knowledge about trajectories that contains farms which product the high quality raw milk conforming to a set of selected Norms defined by the Big Creamery Manager and the producer. This gives the possibility to discover the main factors responsible for the milk quality such as the grass localization that animals are alimented from…To the best of our knowledge, no authors have designed a TDSS relying on the *bottom up* approach.

The following figure describes the global architecture of a creamery decision making system.

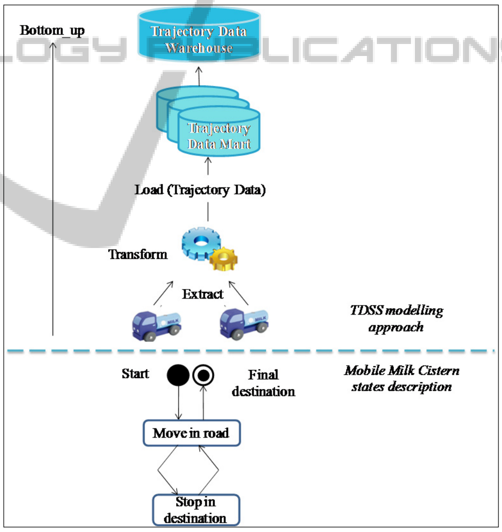

Figure 1: The creamery TDSS architecture.

5 TRAJECTORY DATA SOURCE MODELING

In the midst of the technology progress, professional success is no longer related only to the organization winning strategies but also to its ability to integrate new technologies that have become increasingly spread in all areas. For example, using mobile devices such as PDA to accomplish professional missions presents a good solution to send data concerning moving objects activities in real time.

5.1 Mobile Milk Cistern Trajectory Scenario

Creameries need regular intake of milk to ensure continuity of the process of transformation. For that, the transfer process is considered the most crucial during the production cycle of creamery products (cheese, yogurt ...). Therefore, this task's fulfillment requires the collaboration of the mission team members. Starting with the mobile cistern driver, his role is to drive according to a trajectory that is planned by the head of the mission. In fact, each cistern has its own trajectory, which is composed of a set of movements and stops. It is the "movement" state when the road traffics authorize the movement. Otherwise, or if the cistern reaches a destination (a farm) among planned destinations, it is in the "stop" state. The different information concerning the position of the mobile milk cistern are sent through wireless technology named GPS. This later is reliable to specify exactly the position of the mobile milk cistern.

The controller plays a very important role in the milk collection process. His major task is to sample the raw milk in a creamery production unit with optimum technical and hygienic conditions. This sample allows him to monitor the conditions in which the raw milk is subjected in the farm (temperature ...). Besides, the controller has to record, for each farm, the volume of the collected milk, date and time of loading the raw milk from the local cistern to the mobile cistern. Thanks to the PDA provided with the controller, data will be forwarded easily to the Big Creamery Manager.

5.2 Mobile Milk Cistern Trajectory Modeling

In this section, we focus on modeling the mobile cistern trajectory. Figure 2 presents relational TDSrc schema as a logical model.

Mobile_Milk_Cistern (<u>id_MMC</u>, #id_driver, #id_controller)
Driver (<u>id_driver</u>, name_driv, address_driv, phone_driv)
Trajectory (<u>id_traj</u>, #id_move, #id_stop, #id_mobile_cistern)
Stop (<u>id_stop</u>, duration, #id_begin, #id_end, #id_loc)
Move (<u>id_move</u>, duration, #id_begin, #id_end)
Location (<u>id_loc</u>, name)
Begin (<u>id_begin</u>, Begin_time)
End (<u>id_end</u>, End_time)
Controller (<u>id_controller</u>, name_cont, address_cont, phone_cont, #id_PDA)
Milk_local_cistern_details (<u>id_local_cistern</u>, volum, temperature, #id_farm, #id_controller)
Farm (<u>id_farm</u>, name, address, phone, #id_loc)
PDA (<u>id_PDA</u>, name)
GPS (<u>id_GPS</u>, name)
Big_creamery_manager (<u>id_Big_Creamery_Manager</u>, name_BCM, address_BCM, phone_BCM, #id_PDA)
Norms (<u>id_Norm</u>, #<u>id_local_cistern</u>, #<u>id_Big_creamery_manager</u>, #<u>id_controller</u>, Norm_Compatibility_mark)

Figure 2: TDSrc logical model.

6 TOWARDS TDM MODELING

Choosing to model a TDSS according to the principle of the *bottom-up* approach requires identifying the star schema elements directly from the TDSrc model. Collecting these elements leads to hold a set of schemas called TDMs. Those later are designed to serve a particular set of decisional analyses.

6.1 Star Schema Elements Identification

DM architecture is offered for general business people that need to move easily between marts in order to extract knowledge. Our approach is composed of three major steps for collecting the star schema elements. Firstly, we have to identify facts, then dimensions and finally hierarchies.

- ***Step 1****: Facts identification*

Whatever the kind of DW, fact tables are the most important elements constructing multidimensional star schemas. Indeed, they record measures about particular events that occur in the business. Our goal is to propose a relevant rule that is able to extract fact tables from the TDSrc model.
Rule1: In the TDSrc model, if a table T references several tables but not referenced by any table, if T contains one or several column(s) that are susceptible to be parameter(s) to calculate an additive column(s) that serves analysis and if the primary key of T contains foreign key(s), then T can be a plausible fact F.

- ***Step 2****: Dimensions identification*

In a DM, the role played by dimensions is not less important than that played by facts. Indeed, dimensions present axes of subject analysis. Their identification directly from the TDSrc model is a step that depends on the previous one since they serve to increase the level of detail of the fact table. This strong dependency facilitates the extraction of dimension tables from others.

Rule 2: If a table T is referenced by T_F (table that feeds a fact F), if T_{id} is atomic and if T contains columns that can be dimensional attributes (strong or weak), then T feeds a new dimension D for F.

- ***Step 3****: Hierarchies identification*

Dimension Hierarchies are also essential components. They are used to define data aggregation. Hierarchies group levels from general to granular; this is one of the key benefits of a DM since it enables analysts to access data quickly.

Rule 3: If a table T is referenced by table(s) that feed dimension(s) D and if it doesn't refer any table then T can feed a new level of hierarchy in each of these dimensions, the identifier of T will be a terminal parameter in multiple hierarchies and textual attributes of T become a weak attributes for the added level. Otherwise, if T refers other table T_i in the TDSrc model then the identifier of T becomes a parameter inserted between the identifier of the dimension D and the identifier T_i.

6.2 Rules Results

Applying these rules on our TDSrc model example leads to obtain a set of facts that present the analysis subjects, dimensions which are the axes of the selected subject analysis and hierarchies that are responsible to define data aggregation.

The first rule enumerates conditions that must satisfy an existing table in the TDSrc model to feed an analysis subject. In our example, the tables ***Trajectory*** and ***Norms*** satisfy those conditions. Consequently, they can be plausible facts. For the table ***Norms,*** it contains a column *Norm_Compatibility_mark*. This later verifies if the milk quality provided in the farm respects such norm defined by the farm producer and the big creamery manager. The additive columns *Farm_compatibility_rate* and the *Farm_Rank* can be calculated in function of *Norm_Compatibility_mark*. Concerning the table ***Trajectory,*** it contains columns that can be parameters to calculate the number of stops and moves and the duration of the full trajectory.

The tables ***Stop, Move, Controller, Milk_local_cistern_details, Big_creamery-_manager*** and ***Mobile_milk_cistern*** respect conditions presented in the second rule; then, they feed dimensions in TDMs.

The third rule presents conditions that must satisfy a table existing in the TDSrc model to feed a hierarchy. In our example, the tables ***Begin, End, PDA, GPS*** and ***Driver*** respect these conditions and therefore they can feed hierarchies in the star schema. The table ***Farm*** is referenced by the table ***Milk_local_cistern_details*** that feeds a dimension and it refers the table ***Location***. Consequently, the identifier of the table ***Farm*** becomes a parameter inserted between the identifier of the dimension ***Milk_local_cistern_details*** and the identifier of ***Location.*** Beside its role as a dimension in some TDMs, the table ***Controller*** can play the role of a hierarchy in others since it is referenced by a table that feeds a fact (***Norms***) and another one that feeds a dimension (***Mobile_milk_cistern***).

Arranging these elements according to a set of rules gives the birth of what is called star schema model.

6.3 Example of Mobile Milk Cistern Trajectory Star Schemas

Following the identification of fact tables, dimensions and hierarchies, it becomes possible to build a set of TDMs which their composition is optimized so that decision makers are able to look data in a unique way and this guarantee analysis results coherence. In our example, the first TDM modeled in figure 3 is extracted from our TDSrc model. It focuses on the norms compatibility analysis. This TDM model offers the possibility to discover creameries that respect Norms defined by the Big Creamery Manager and the Milk Producer and therefore localizations that provide the best milk quality.

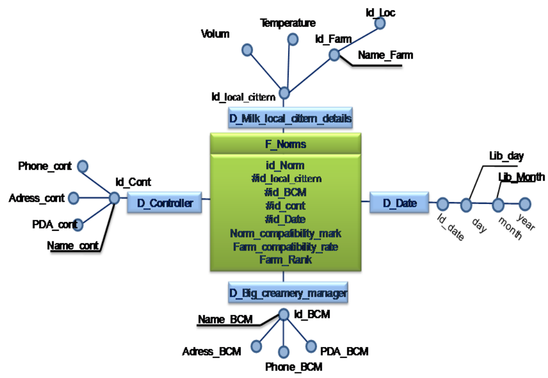

Figure 3: Norms compatibility analysis.

Figure 4 present the second TDM example. It focuses on Mobile Milk Cistern trajectory analysis. The main goal is to discover the principal factors responsible for the milk quality, such as alimentation of animals, weather...

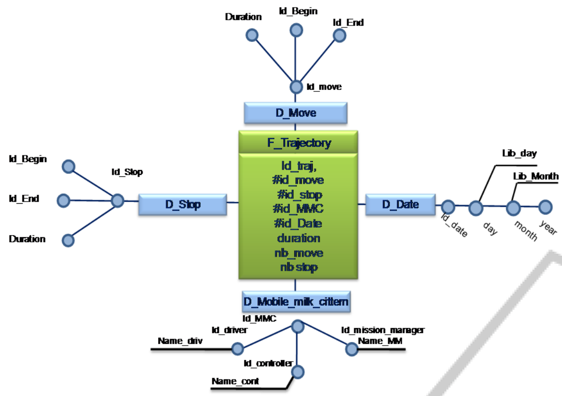

Figure 4: Mobile Milk Cistern trajectory analysis.

Each TDM modeled directly from TDSrc model focuses on a particular subject analysis. This leads to queries results coherence and therefore knowledge facility extraction. We approve that whatever the DW kind including TDW, this later is the conglomerate of all TDMs that model organization business events. Thence, once more than two TDMs are built, the fusion of these later gives the birth of a central repository of data named TDW. It's a system that collects TD organized around the organization major issues; those issues are the same that pushed building TDMs as a first step.

7 CONCLUSIONS AND PERSPECTIVE

In this paper, we presented an overview on the TDW modeling problem. We exposed some solutions proposed by different authors in recent years. Then, we expected the absence of a TDSS modeling solution that really solves the problem of organization decision making taking into account the facility that must be offered to users to extract knowledge that leads to the organization durability. Hence, we proposed an approach which deals with the TDSS modeling that is able to improve the strategy of the organization decision making. Our approach is based on the *bottom-up* approach. It consists to give the birth of a set of TDMs generated, directly, from a TDSrc model. The fusion of these TDMs leads to build the organization's complete TDW.

As the operational trajectory database is scalable due to data type and data sources heterogeneity, TDSrc model evolve over time. This, obviously, has a direct impact on TDSS. In fact, the propagation of changes performed on TDSrc towards TDSS gives the possibility to change TDMs' structures, or even to give the birth of new TDMs' models that focus on new subjects analysis, and therefore to alter the existing TDW model. Although its cruciality, the TDSS evolution problem hasn't yet received the interest that it deserves. Therefore, as perspective, we propose to deal with this problem that's still an open research issue.

REFERENCES

Arfaoui, N., Akaichi, J., 2011. Modeling Herd Trajectory Data Warehouse. *International Journal of Engineering Trends and Technology* (pp. 57-71).

Errajhi, E., 2014. Trajectory data fuzzy modeling: ambulances management use case. In *International Journal of Database Management Systems (IJDMS),* Volume 6, (pp. 1-10).

Kimball, R., Reeves, L., Ross M., Thornthwaite, W., 1998. The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses. *The book, John Wiley and sons publishing.*

Leonardi, L., 2014. A Framework for Trajectory Data Warehousing and Visual OLAP Analysis.In *doctoral thesis Ca'foscari university of Venice.*

Marketos, G., Frentzos, E., Ntoutsi, I., Pelekis, N., Raffaetà, A., Theodoridis, Y., 2008. Building real-world trajectory warehouses. In 7th *International ACM Workshop on Data Engineering for Wireless and Mobile Access(MobiDE),* (pp. 8-15), Vancouver, BC, Canada.

Sapia, C., Blaschk,a M., Höfling, G., Dinter, B., 1998. Extending the E/R Model for the multidimensional paradigm. In *Proceedings of International Workshop on Data Warehousing and Data Mining*, (pp 105-116), Germany.

Spaccapietra, S., Parent, C., Damiani, M. L., de Macedo, J. A., Porto, F., Vangenot, C., 2007. A Conceptual View on Trajectories. *Research Report*. Ecole Polytechnique Fédérale, Database Laboratory, Lausane, Switzerland.

Tryfona, N., Busborg, F., Christiansen, J.G.B., 1999. StarER: A Conceptual Model for Data Warehouse Design. In *Proceedings of the ACM 2nd International Workshop on Data warehousing and OLAP,* (pp 3-8) Missouri, USA.