

# TRAJECTORY DATA FUZZY MODELING : AMBULANCES MANAGEMENT USE CASE

Eya Errajhi

Department of Computer Sciences, ISG-University of Tunis, Tunisia

## **ABSTRACT**

*Data captured through mobile devices and sensors represent valuable information for organizations. This collected information comes in huge volume and usually carry uncertain data. Due to this quality issue difficulties occur in analyzing the trajectory data warehouse. Moreover, the interpretation of the analysis can vary depending on the background of the user and this will make it difficult to fulfill the analytical needs of an enterprise. In this paper, we will show the benefits of fuzzy logic in solving the challenges related to mobility data by integrating fuzzy concepts into the conceptual and the logical model. We use the ambulance management use case to illustrate our contributions.*

## **KEYWORDS**

*Trajectory data warehouse, fuzzy concept, moving object, fuzzy modeling*

## **1. INTRODUCTION**

Data captured through mobile devices and sensors represent valuable information for organizations. This collected information, which are sets of time-stamped positions called trajectory data, are the foundation for decision making. Trajectory data are spatial and temporal data related to moving objects and comes with huge volume. These data are normally available as sample points. Trajectory data play a fundamental role to a number of applications, such as traffic control, transportation management, animal migration, and tourism.

As a TDW often provides the base of information for taking business decision, it is crucial that it is able to challenge fuzzy information. Today's TDW architectures only hardly handle imprecise and qualitative data. In fact, TDW generate a large volume of mobility data and it increases every day. These amounts of data received for processing turns into challenging tasks for administration and analysis because it is fundamental to turn these huge sets of mobility data into useful knowledge to understand objects behaviors.

Next to the problem of high quantity, there is a quality issue because data received for analysis are often incomplete, vague or uncertain. For example, positions can not be determined accurately, i.e. the recorded location of a moving object does not represent its precise location. As an example, the location of an object can be recorded as an area instead of point. Due to these unreliable reading difficulties occur in analyzing and querying the TDW.

Uncertainty sources may be one of the following: (1) Imperfect observation of the real world, (2) Incomplete representation language, (3) Ignorance, laziness or inefficiency, (4) measurement errors, and (5) sampling error. Measurement error is the result of inaccurate instruments, while sampling error occurs because the system only captures the continuous movement of an object periodically, bringing uncertainty between two consecutive observations [1].

In addition to the qualitative issues of the data entered in the TDW, the interpretation of the analysis can vary depending on the background of the user. Moreover, most analysis that data

warehousing systems perform is numerically oriented such as counting the total sales of products within a certain period. Although the numbers give some indications of the sales behavior, their real meaning is missing. There are several approaches discussed in the past to integrate fuzzy concepts in classic data warehouse, spatial data warehouse, spatio-temporal data warehouse and OLAP cubes. Nevertheless, most of these approaches have been developed for a certain field of application and they often imply restrictions on how fuzzy concepts can be applied. Or they substitute the original values of the data warehouse with information from the fuzzy concept. On the other hand and as far as we know, there is no work that integrates fuzzy concepts in TDW, either in the conceptual model or in the logical model and therefore, there is a strong urge to address this issue. Fuzzy sets provide mathematical meanings to natural language and are therefore able to handle the imprecision of the natural language. Accordingly, the integration of fuzzy logic on TDW technologies might improve the analysis of data, maintain data processing efficiency and control the amount of uncertainty and might lead to better decision making. That is why it may be useful to construct a Fuzzy Trajectory Data Warehouse (FTDW) that can directly support the analysis of fuzzy data. In order to address the problems stated in the problematic, our main goal is to achieve complex OLAP analysis, querying the multidimensional model in flexible ways and giving a meaning to TDWs when fuzzy data are involved and especially when users want to ask questions in natural language.

For those reasons, a trajectory data fuzzy modeling will be presented using StarUML. Next, a Trajectory Data Warehouse Fuzzy modeling with the integration of fuzzy logic in dimensions and facts and its implementation in a snowflake schema will be realized. To prove the feasibility of the FTDW and the benefits of integrating fuzzy logic concepts, a FTDW analysis will be realized and a complex OLAP queries in natural language will be achieved using a real-world case study which is the ambulance service.

The rest of this paper is organized as follows. Section 2 gives a survey on fuzzy trajectory data models. Our trajectory data fuzzy modeling with its implementation using the ambulance management use case is described in section 3. Section 5 includes the implementation of our FTD using starUML. The last section concludes this paper and describes our future work.

## **2. STATE OF THE ART**

Numerous efforts to integrate fuzzy logic into data warehousing process have been reported. In this related work, several existing approaches are discussed.

In order to deal with complex objects and imprecise and uncertain information in conceptual data modeling, one needs fuzzy extension to conceptual data models, which allow imprecise and uncertain information to be represented and manipulated at a conceptual level. There exist several studies on fuzzy data models including [2, 3, 4] and a recent survey of these models is presented in [5].

Galindo et al. [6] extend the Entity-Relationship Model to incorporate fuzziness was presented, where fuzzy elements as entities, attributes and associations are represented graphically. This extension defines three levels of fuzziness:

1. Entities, attributes and associations could be fuzzy; in other words, could have membership grades to the E-R model.
2. Fuzzy occurrences of entities and associations.
3. Fuzzy values of the attributes, expressed through a distribution of possibility.

In [7] three different types of fuzziness are addressed at the attribute level. The first type is called incompleteness and corresponds to the case when the value of an attribute is specified as a range value. The second one which is called null type is used to represent the cases when the value of

an attribute is not known, does not exist, or there is no information on whether a value exists or not. The last type, which is called fuzzy type, corresponds to the cases when the value of an attribute is vaguely specified. Fuzziness at the object/class level is used to represent imprecise information due to the existence of a partial membership of an object to its class. If the boundary of a class is uncertain, then the objects of this class may be a member of the class with a membership degree between 0 and 1. The relevance of an attribute to an object can also be used to represent imprecise information at the object/class level. Class/subclass level fuzziness refers to the existence of a partial membership of a class to its superclass(es). This type of uncertainty indicates that the fuzziness occurs at the class inheritance hierarchy since we might not be able to construct a class hierarchy precisely.

### 3. TRAJECTORY DATA FUZZY MODELING

In our ambulance use case, each moving object is equipped with a sensor and has a trajectory which is composed of at least two milestones that represent the beginning and the ending of a moving object's trajectory. The milestone is a record which consists of position and time recording which are extracted from GPS data. Each attribute chosen to be fuzzy is associated with a fuzzy class which contains its membership degrees of the linguistic variable to each of its linguistic terms. Our following methodology can be used in many use cases using any other representation of a trajectory.

#### 3.1 General trajectory data conceptual model

To incorporate fuzzy concepts into the conceptual model, a five step methodology should be followed:

##### 3.1.1 Identify the linguistic variables

The first step is to identify the attributes that contain values that are aimed to be represented fuzzily. A FTD contains at least one linguistic variable. Zadeh in [8] introduced the concept of linguistic variable as "a variable whose values are not numbers but words or sentences in a natural or artificial language". For example, the ambulance speed is a linguistic variable.

**Definition 1** [8]: A linguistic variable is characterized by a quintuple (H,T(H),U,G,M) in which H is the name of the variable; T(H) (or simply T) denotes the term set of H, i.e., the set of names of linguistic values of H, with each value being a fuzzy variable denoted generically by X and ranging across a universe of discourse U which is associated with the base variable u; G is a syntactic rule (which usually takes the form of a grammar) for generating the names of values of H; and M is a semantic rule for associating its meaning with each H, M(X), which is a fuzzy subset of U.

##### 3.1.2 Identify the set of linguistics terms

The second step is to determine the linguistics terms, i.e. to determine how the linguistic variable should be presented. This step is repeated for all the linguistic variables. For example, the set of linguistics terms for the attribute speed are: {*slow, average, fast*}.

##### 3.1.3 Define the membership function

In this step a membership function is calculated for each linguistic term. The membership function is represented by  $\mu$  and has a scale of values between 0 and 1.

**Definition 2** [9]: If X is a collection of objects denoted generically by x, then a fuzzy set A in X is a set of ordered pairs:

$$A = \{ (x, \mu_A(x)) \mid x \in X \}$$

$\mu_A(x)$  is called the membership function (generalized characteristic function) which maps  $X$  to the membership space  $M$ . Its range is the subset of nonnegative real numbers whose supremum is finite.

**Definition 3:** A membership function for a fuzzy set  $A$  on the universe of discourse  $X$  is defined as  $\mu_A: X \rightarrow [0,1]$ , where each element of  $X$  is mapped to a value between 0 and 1. This value, called membership value or degree of membership, quantifies the grade of membership of the element in  $X$  to the fuzzy set  $A$ .

For example, the membership functions for each linguistic term of the target attribute speed are  $\mu_{slow}$ ,  $\mu_{average}$ ,  $\mu_{fast}$ . The membership functions can have different shapes. Some of the most frequently used shapes include triangular, trapezoidal, and Gaussian shapes.

### 3.1.4 Add the fuzzy membership classes

In this step a fuzzy membership class should be created for each linguistic variable. This class contains the primary key of the class and the membership degrees of the linguistic variable to each of its linguistic terms. For example a fuzzy membership table named fuzzySpeed is added for the linguistic variable speed.

### 3.1.5 Relate the classes

The last step is to relate the fuzzy membership class with the class that contains the linguistic variable. The two classes are related in a 1-1 relation. For example the class trajectory that contains the fuzzy attribute speed is related to the class fuzzySpeed.

The UML class diagram of a general trajectory data is represented in figure 1. It is composed of crisp classes and fuzzy classes and it represents the associations between them.

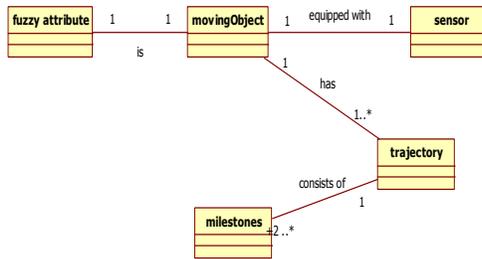


Figure 1: General trajectory data fuzzy conceptual model

The next phase after designing the general trajectory data is the application of our methodology to ambulance management use case.

## 3.2 Ambulance trajectory data conceptual fuzzy model

In order to see the effectiveness of our proposed model, we applied the proposed methodology for ambulance management use case. In the first and second step, the linguistic variables and their linguistic terms should be defined, which are:

- The attribute speed with linguistics terms: *slow, average, fast*.
- The attribute duration with the linguistics terms: *short, moderate, long*.
- The attribute age with the linguistics terms: *young, middle age, old*.
- The attribute state with the linguistics terms are: *medium, serious, mortal*.

In the third step, a membership function should be calculated for each linguistic term of the identified fuzzy attributes.

The membership function  $\mu_{\text{short}}(\text{duration})$  has an R-function shape and is presented analytically below:

$$\mu_{\text{short}}(\text{duration}) = \begin{cases} 1 & \text{if duration} \leq 10 \\ 0 & \text{if duration} \geq 20 \\ \frac{20 - \text{duration}}{20 - 10} & \text{if } 10 < \text{duration} < 20 \end{cases}$$

The term moderate has a triangular membership function which is presented analytically below:

$$\mu_{\text{moderate}}(\text{duration}) = \begin{cases} 0 & \text{if duration} \leq 10 \\ 0 & \text{if duration} \geq 40 \\ 1 & \text{if } 20 \leq \text{duration} \leq 30 \\ \frac{\text{duration} - 10}{20 - 10} & \text{if } 10 < \text{duration} \leq 20 \\ \frac{40 - \text{duration}}{40 - 30} & \text{if } 30 \leq \text{duration} < 40 \end{cases}$$

The term long has an L-function shape (figure 2) with parameters  $a = b = +\infty$  and is a special case of the trapezoidal function. The analytical representation of this latter linguistic term is:

$$\mu_{\text{long}}(\text{duration}) = \begin{cases} 0 & \text{if duration} \leq 30 \\ 1 & \text{if duration} \geq 40 \\ \frac{\text{duration} - 30}{40 - 30} & \text{if } 30 < \text{duration} < 40 \end{cases}$$

The following figure shows the graphical representation of each linguistic term of our linguistic value duration.

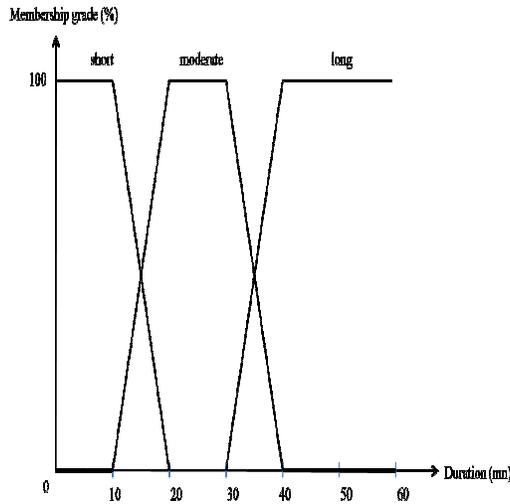


Figure 2: Membership function for linguistic variable duration

The following formulas present the membership functions for each linguistic term of the linguistic variable speed with its analytical representation:

$$\mu_{\text{slow}}(\text{speed}) = \begin{cases} 1 & \text{if speed} \leq 80 \\ 0 & \text{if speed} \geq 100 \\ \frac{100 - \text{speed}}{100 - 80} & \text{else} \end{cases}$$

$$\mu_{\text{average}}(\text{speed}) = \begin{cases} 0 & \text{if speed} \leq 80 \\ 0 & \text{if speed} \geq 100 \\ \frac{\text{speed}-80}{90-80} & \text{if } 80 < \text{speed} \leq 90 \\ \frac{100-\text{speed}}{100-90} & \text{if } 90 \leq \text{speed} < 100 \end{cases}$$

$$\mu_{\text{fast}}(\text{speed}) = \begin{cases} 0 & \text{if speed} \leq 80 \\ 1 & \text{if speed} \geq 100 \\ \frac{\text{speed}-80}{100-80} & \text{else} \end{cases}$$

The following figure shows the graphical representation of each linguistic term of the linguistic value speed.

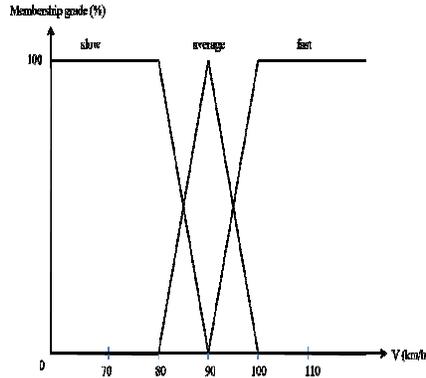


Figure 3: Membership function for linguistic variable speed

The following formulas present the membership functions for each linguistic term of the linguistic variable state with its analytical representation:

$$\mu_{\text{medium}}(\text{state}) = \begin{cases} 1 & \text{if state} \leq 2 \\ 0 & \text{if state} \geq 4 \\ \frac{4-\text{state}}{4-2} & \text{if } 2 < \text{state} < 4 \end{cases}$$

$$\mu_{\text{serious}}(\text{state}) = \begin{cases} 0 & \text{if state} \leq 2 \\ 0 & \text{if state} \geq 8 \\ 1 & \text{if } 4 \leq \text{state} \leq 6 \\ \frac{\text{state}-2}{4-2} & \text{if } 2 < \text{state} < 4 \\ \frac{8-\text{state}}{8-6} & \text{if } 6 < \text{state} < 8 \end{cases}$$

$$\mu_{\text{mortal}}(\text{state}) = \begin{cases} 0 & \text{if state} \leq 6 \\ 1 & \text{if state} \geq 8 \\ \frac{\text{state}-6}{8-6} & \text{if } 6 < \text{state} < 8 \end{cases}$$

The following figure shows the graphical representation of each linguistic term of the linguistic value state.

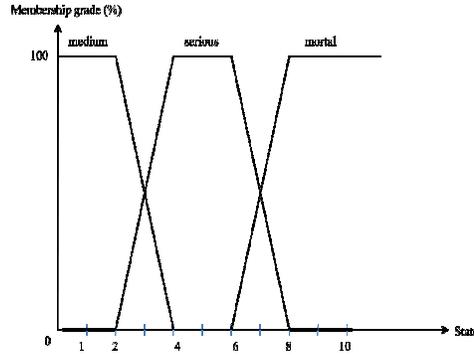


Figure 4: Membership function for linguistic variable state

The following formulas present the membership functions for each linguistic term of the linguistic variable age with its analytical representation:

$$\mu_{\text{young}}(\text{age}) = \begin{cases} 1 & \text{if } \text{age} \leq 25 \\ 0 & \text{if } \text{age} \geq 40 \\ \frac{40-\text{age}}{40-25} & \text{if } 25 < \text{age} < 40 \end{cases}$$

$$\mu_{\text{middle age}}(\text{age}) = \begin{cases} 0 & \text{if } \text{age} \leq 25 \\ 0 & \text{if } \text{age} \geq 65 \\ 1 & \text{if } 40 \leq \text{age} \leq 50 \\ \frac{\text{age}-25}{40-25} & \text{if } 25 < \text{age} < 40 \\ \frac{65-\text{age}}{65-50} & \text{if } 50 < \text{age} < 65 \end{cases}$$

$$\mu_{\text{old}}(\text{age}) = \begin{cases} 0 & \text{if } \text{age} \leq 50 \\ 1 & \text{if } \text{age} \geq 65 \\ \frac{\text{age}-50}{65-50} & \text{if } 50 < \text{age} < 65 \end{cases}$$

The following figure shows the graphical representation of each linguistic term of the linguistic value age.

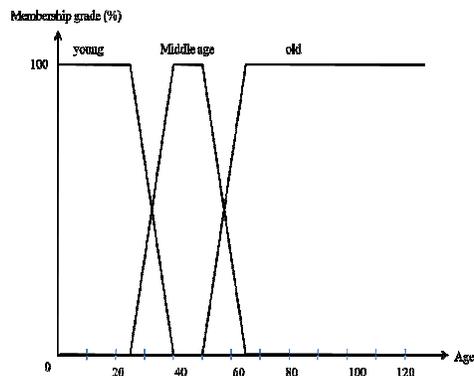


Figure 5: Membership function for linguistic variable age

The creation of the fuzzy membership class is realized in this fourth step. After creating our membership classes, each class should be related, in this last step, to the class that contains the corresponding linguistic variable. The following figures present the association between the fuzzy classes and the crisp classes.

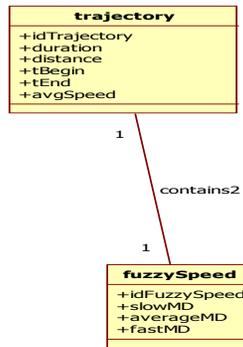


Figure 6: Linguistic values associated to the attribute speed

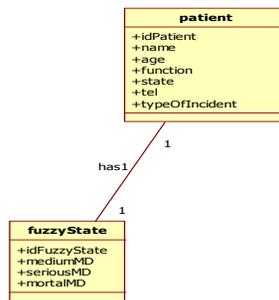


Figure 7: Linguistic values associated to the attribute state

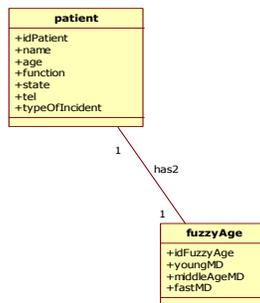


Figure 8: Linguistic values associated to the attribute age

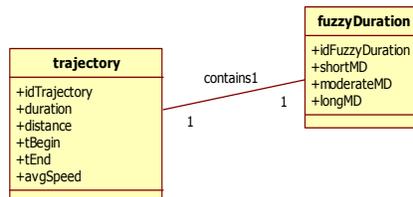


Figure 9: Linguistic values associated to the attribute duration

After applying our methodology, the fuzzy conceptual modeling of trajectory data applied on ambulance management use case is presented in Figure 10 as a UML class diagram.

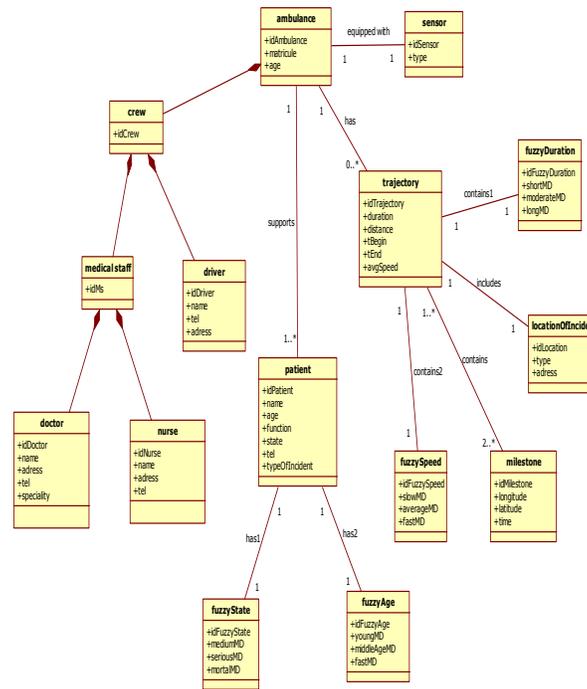


Figure 10: Ambulance trajectory data fuzzy conceptual model

Each component in this diagram is described as follows:

**Ambulance:** The ambulance class is associated with sensor and patient classes. It is composed of a crew including a medical staff, and a driver. Each ambulance has its own trajectory which is composed by a set of milestones.

**Trajectory:** the trajectory class is a travel in the space and in the time. It is composed by a list of milestones. It has several attribute like the time of begin, the time of end and the traveled distance.

**Milestone:** the milestone class contains a record of the position and the time of the ambulance trajectory which are extracted from GPS data.

**Sensor:** the ambulance is tracked every where to know its position. This is performed thanks to a sensor like GPS.

**Crew:** the crew class describes the persons involved in the ambulance including the driver and the medical staff.

**Driver:** The driver class describes the person who is driving the ambulance.

**Medical staff:** The medical staff could include a doctor or more, and a nurse or more depending on the incident

**Patient:** the patient class describes the person treated by the medical staff in the moving ambulance.

**Doctor:** the doctor class describes the person who is curing the patient.

**Nurse:** the Nurse class describes the person who is working in the moving ambulance.

#### 4. CONCLUSION AND FUTURE WORKS

In this article, we presented our contribution which is the integration of fuzzy logic in the conceptual model of a trajectory data and its implementation to ambulance management domain. The fuzzy logic was integrated as a fuzzy class that contains the membership degrees of the linguistic variable to each of its linguistic values. With such relationship, all the attributes get a semantic meaning in the trajectory data. The fuzziness was integrated at the attribute and class level. The proposed model is generic in the sense that it could easily be adapted to any domain and it is easy to apply. Among the advantages of our contribution, it is not limited for a specific application scope. Also, the proposed idea can be applied not only on trajectory data but also on a 'classic' data as well as a spatial or a spatio-temporal data.

For future work, we plan to apply the fuzzy concept to real world trajectory data to strengthen the validity of the fuzzy data warehouse concept because our application is based on a virtual ambulance trajectory data.

#### REFERENCES

- [1] D. Pfoser and C.S.Jensen, "Capturing the uncertainty of moving-objects representations", In Proc. of the SSDBM Conference, pp. 123-132, 1999.
- [2] M. Koyuncu and A.Yazici, "IFOOD: An Intelligent Fuzzy Object-Oriented Database Architecture", IEEE Transactions on Knowledge and Data Engineering, vol. 15, no. 5, 2003.
- [3] N. Marin, O.Pons and M. A.Vila, "A Strategy for Adding Fuzzy Types to an ObjectOriented Database System", Int. Journal of Intelligent Systems, vol. 16, no. 7, pp 863-880, 2001.
- [4] J. M. Medina, O. Pons and M. A.Vila, "GEFRED: A Generalized Model of Fuzzy Relational Databases", Information Sciences, vol. 76, pp. 1-2, 1994.
- [5] Z. M.Ma and L. Yan, "A Literature Overview of Fuzzy Database Models", International Journal of Information Science and Engineering, vol. 24, no. 1, pp 189-202, 2008.
- [6] J. Galindo, A.Urrutia and M.Piattini, "Fuzzy Databases: Modeling, Design and Implementation", Idea Group Publishing, 2006.
- [7] A. Yazici, Q.Zhu and N.Sun, "Semantic Data Modeling of Spatiotemporal Database Applications", Int. Journal of Intelligent Systems, vol. 16, no. 7, pp 881-904. 2001.
- [8] L. A. Zadeh, "The Concept of a Linguistic Variable and its Application to Approximate Reasoning - Part I", Information Science, No.8, pp.199-249, 1975.
- [9] H.J. Zimmermann, "Fuzzy Set theory", Advanced Review, Germany, vol.2, 2010.