**Ramaiah Institute of Technology**

(An Autonomous Institute, Affiliated to VTU)

MSR Nagar, MSRIT post, Bangalore-54

A Dissertation Report on

US Fatal Road accidents Using Fuzzy Set Approach

Submitted by

Aishwarya Sitharam 1MS15CS149

Shweta Baindur 1MS15CS122

Shrisha Hallale 1MS15CS121

*Bachelor of Engineering in Computer Science & Engineering*

Under the guidance of

Sowmya B.J

Assistant Professor

Computer Science Department

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**M.S.RAMAIAH INSTITUTE OF TECHNOLOGY**

**(Autonomous Institute, Affiliated to VTU)**

**BANGALORE-560054**

www.msrit.edu, **2017**

**Introduction**

Many real-world events that are encountered daily are perceived to be vague or ill-defined rather than being a discrete probabilistic value. For example, when solving a real-life problem, the role of "either ... or" may be replaced by that of "both ... and." In image processing, the gray level of an image element is defined as both black and white to a certain degree of the two attributes, respectively, where the fullness of being black or white is only a special situation. The underlying logic of fuzzy-set theory is that it allows an event to belong to more than one sample space where sharp boundaries between spaces are ill-defined.

For example, when a question is asked whether it will rain tomorrow, an answer is given to the likelihood of raining normally with a degree of confidence unless occurrence of the event is totally ensured. If saying that the chance of rain tomorrow is 40%, the chance of no rain is implied to be 60%. However, in daily living, a crisp estimate for such an event is unusual. Some linguistic expressions such as "It is very likely to rain," and "the chance of rain is about 40%," are preferred.

The US fatal road accident data set for automobiles has been chosen here for analysis. Data is for automobiles where the right passenger seat was occupied, with one observation for each such passenger. Observations for vehicles where the most harmful event was a fire or explosion or immersion or gas inhalation, or where someone fell or jumped from the vehicle, were not included. The dataset excludes large trucks, pickup trucks, vans and buses.

The dataset was cleaned and normalized before using it for analysis. The obtained plots are thereby analyzed to draw inferences based on the attributes chosen.

**Data Set description:**

**Source of Dataset**

The US Fatal Road Accident dataset was taken from the US-FARS website.

**Attributes Description**

* State: This data element identifies the state in which the crash occurred. Ranges from 1 to 51 according to the Geographic Location Codes (GCL).
* Age: Corresponds to the age of the passenger.
* Injury: Indicates the severity of the passenger’s injury with ranges from 0 to 4 where 0 corresponds to no injury and 4 corresponds to fatal injury.
* Restraint: Restraint equipment in use by the occupant at the time of crash.

0: None Used, 1: Shoulder Belt, 2: Lap Belt, 3: Lap and Shoulder Belt, 4: Child Safety Seat.

* Impact: Area on this vehicle that was most damaged during an event in the crash. Ranges from 1 to 12 clock points.
* Model year: This data element identifies the manufacturer's model year of the vehicle.
* Driver injury: Ranges from 0 to 4 where 0 corresponds to no injury and 4 corresponds to fatal injury.
* Year: Year the accident occurred.

**Data Set size in terms of Bytes and Number of Tuples:**

The size of the dataset is 2,461 KB that is 2,520,064 bytes consisting of 8 attributes and 78150 tuples.

**Inferences drawn**:

1. Inimpact v/s driver injury.

The severity of the driver’s injury can be predicted based on which part of the vehicle was most damaged in the accident.

1. Inimpact and restraint v/s injury of passenger.

Based on which part of the vehicle was most damaged, and whether the front seat passenger was wearing a seatbelt a prediction can be made on the severity of the injury of the passenger.

1. Model year v/s driver injury.

The severity of the injury to the driver which occurred can be predicted based on which year the vehicle was manufactured in.

**Algorithm Description**

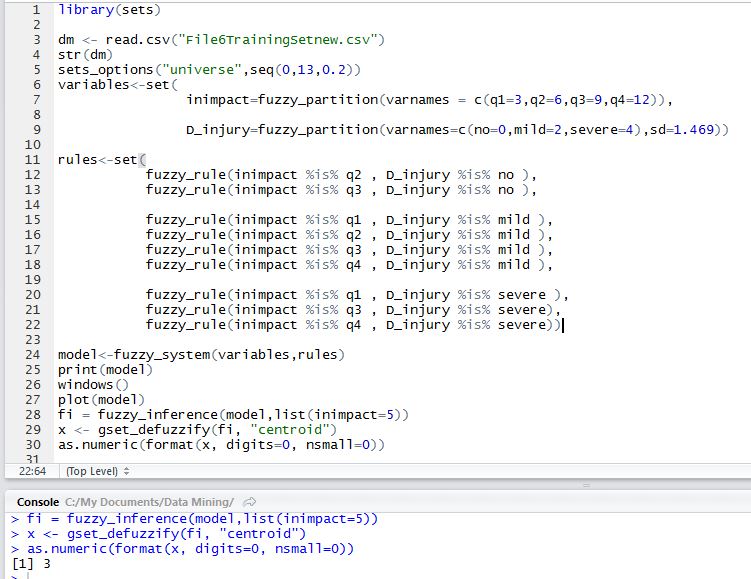
Fuzzy Set Theory is also called Possibility Theory. This theory allows us to work at a high level of abstraction. It also provides us the means for dealing with imprecise measurement of data.

Fuzzy sets generalize classical sets, but the membership functions take a value between 0 and 1. In fuzzy set theory, classical bivalent sets are usually called crisp sets. The fuzzy set theory can be used in a wide range of domains in which information is incomplete or imprecise, such as bioinformatics

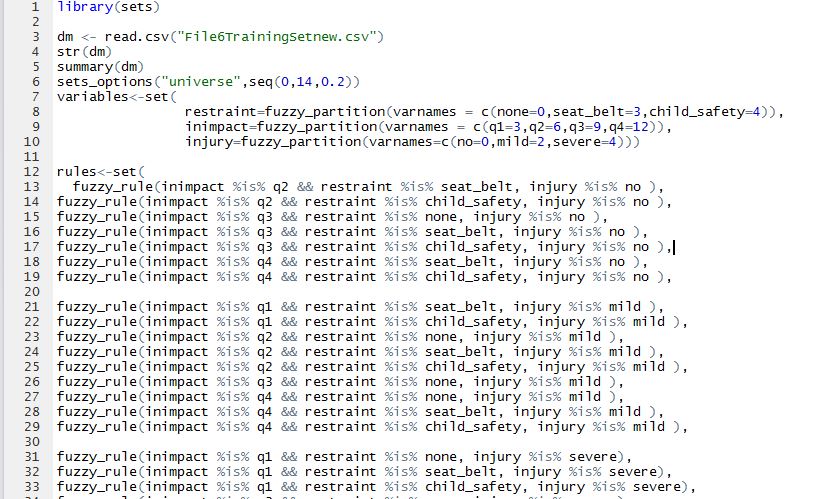
Fuzzy logic is a form of many-valued logic or probabilistic logic it deals with reasoning that is approximate rather than fixed and exact. In contrast with traditional logic theory, where binary sets have two-valued logic, true or false, fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Fuzzy logic has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely false.

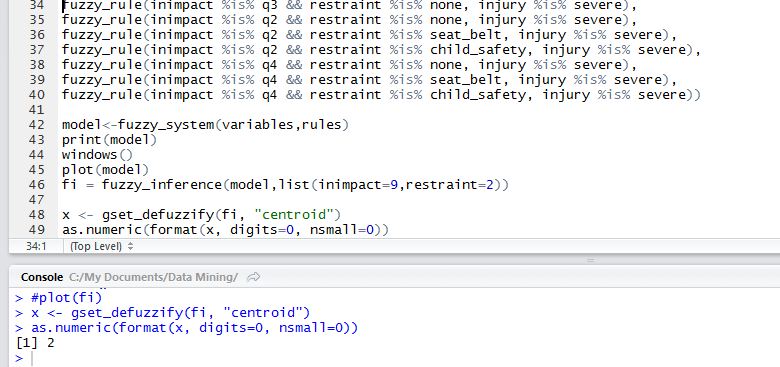
**Code**

**Inimpact v/s driver injury:**

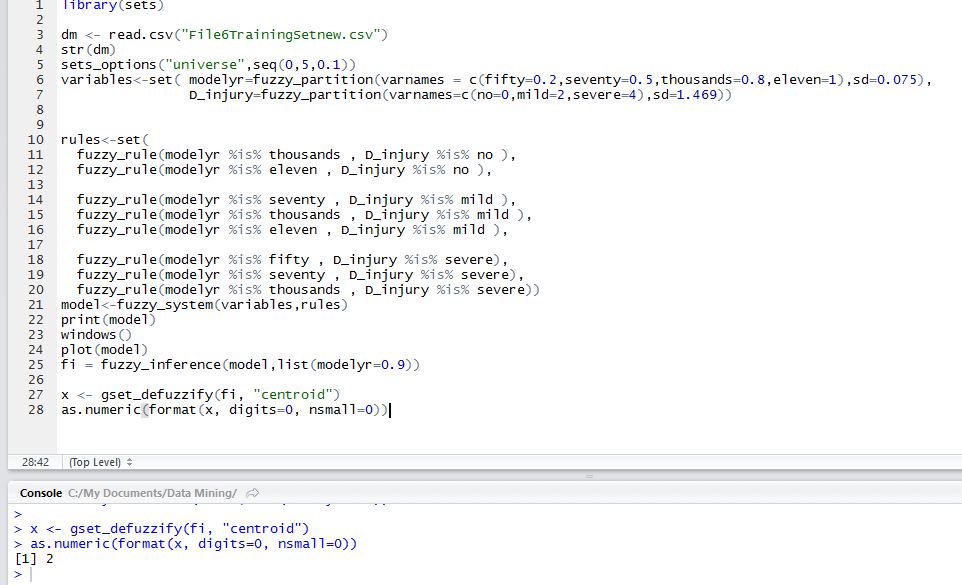


**Inimpact and restraint v/s injury of passenger:**

****

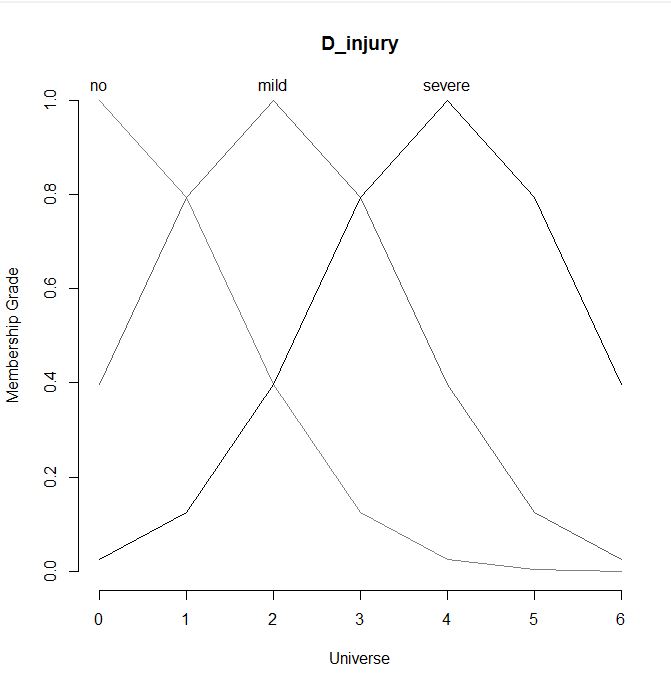
****

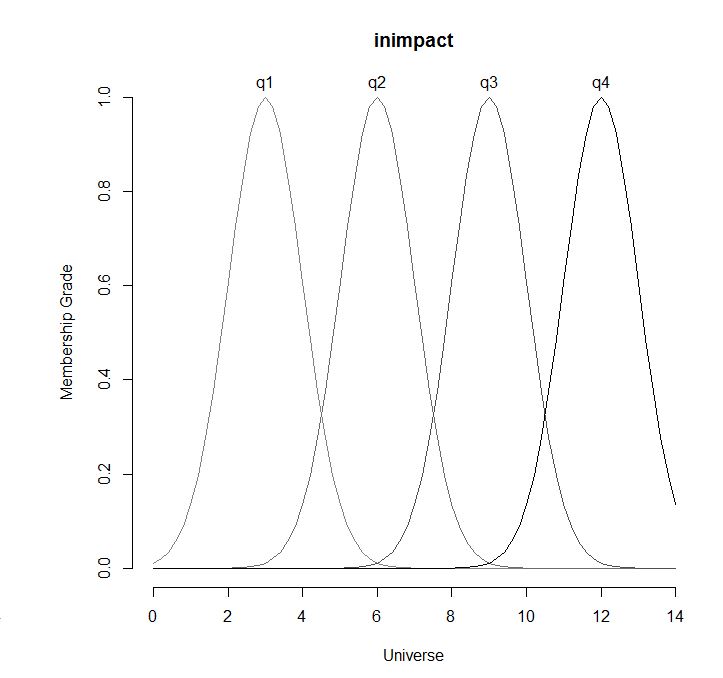
**Model year v/s driver injury:**

****

**RESULT SNAPSHOTS:**

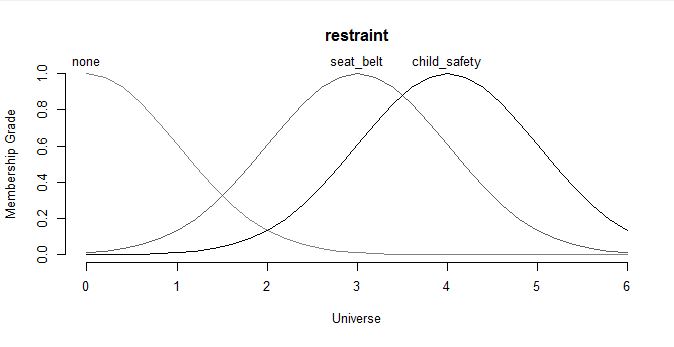
**Inimpact v/s driver injury:**

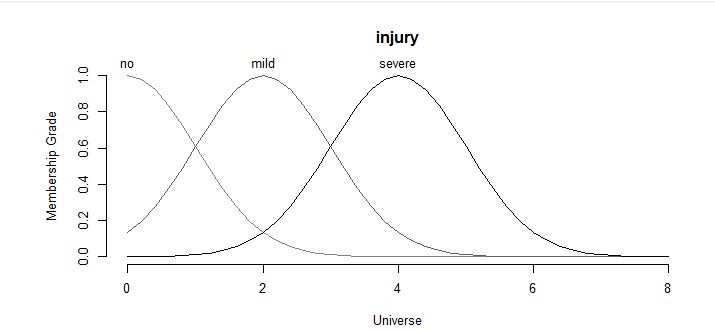
****

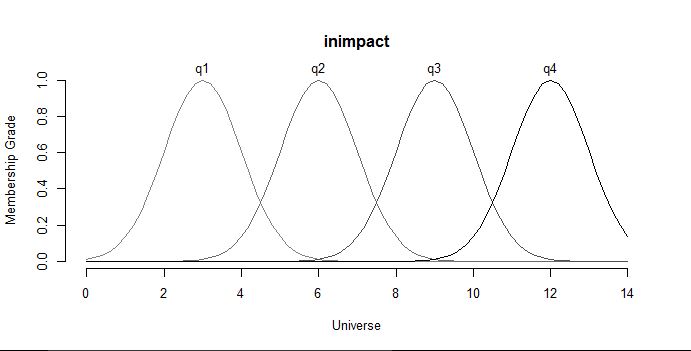
****

The above graphs showcase the fuzzy values for the seriousness of the driver’s injury as well as which area of the vehicle had the most damage done to it. In the first graph, consider the D\_injury to be 2, as this 2 is present in all three member functions that is, no, mild and severe, this 2 is known as a fuzzy value for the D\_injury graph. Similarly for inimpact, here 4 is a fuzzy value as it is present in both quadrant 1(q1) and quadrant 2 (q2). After obtaining this graph using the fuzzy algorithm, we defuzzify the values; this means that after training the dataset by classifying the values into their particular groups, we obtain a crisp output value for the particular input. For example, to predict the D\_injury for a particular inimpact, such as 5, we obtain the seriousness of the driver injury to be 2 which is mild.

**Inimpact and restraint v/s injury of passenger:**

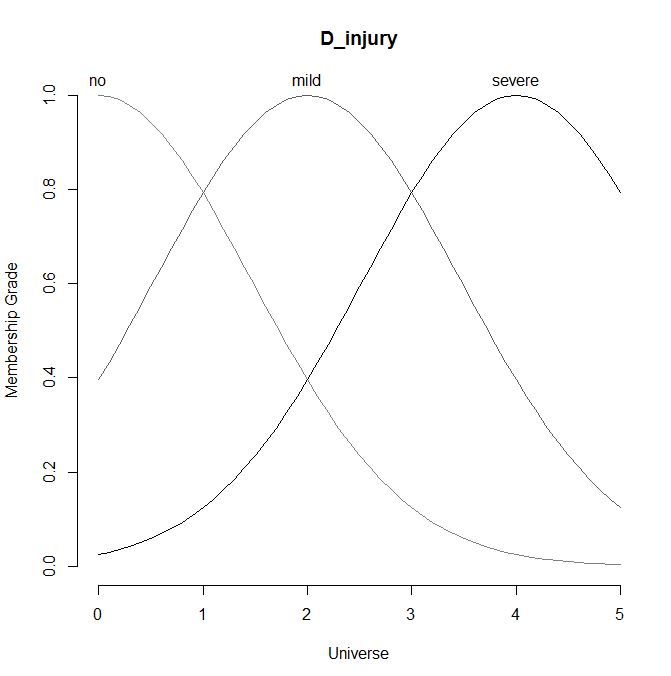
****

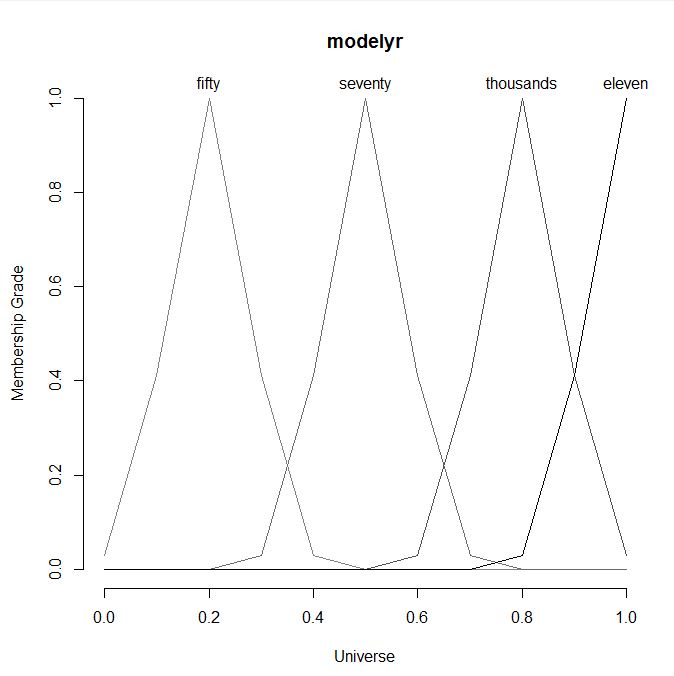
****

****

For the above 3 fuzzy graphs, we classify the attribute inimpact and which type of restraint the front seat passenger was wearing in regards to the seriousness of the passengers injury. After obtaining the membership functions in regards to the graph and classifying inimpact and restraint based on the fuzzy rules to obtain the injury of the passenger, we defuzzified the values.

For example, in regards to this inference, if want to defuzzify the values for inimpact as 9 and restraint as 2, we obtain the seriousness of the passengers injury as 2.

**Model year v/s driver injury: **

****

The above two graphs show the inference between the model year of the vehicle and the seriousness of the drivers’ injury. This means that we classify the drivers’ injury based on the model year using the given fuzzy rules. Using these rules we can defuzzify the values to give us our required crisp output value. For example, if we want to predict how serious the drivers’ injury will be when given a particular model year such as 0.9 (0.9 is the year obtained after applying min max normalization) where 0.9 is also in the year between the 2000s and 2011s, we can find that the drivers’ injury is mild, which is represented as 2.

**Implementation:**

A fuzzy logic system (FLS) can be defined as the nonlinear mapping of an input data set to a scalar output data. A Fuzzy logic consists of four main parts: fuzzifier, rules, inference engine, and defuzzifier.

The process of fuzzy logic is explained in: First, a crisp set of input data are gathered and converted to a fuzzy set using fuzzy linguistic variables, fuzzy linguistic terms and membership functions. This step is known as fuzzification. An inference is then made based on a set of rules. Finally, the resulting fuzzy output is mapped to a crisp output using the membership functions, in the defuzzification step.

1. Define the linguistic variables and terms (initialization)

2. Construct the membership functions (initialization)

3. Construct the rule base (initialization)

4. Convert crisp input data to fuzzy values using the membership function (fuzzification)

5. Evaluate the rules in the rule base (inference)

6. Combine the results of each rule (inference)

7. Convert the output data to non-fuzzy values (defuzzification)

Membership function (MF) is a function that specifies the degree to which a given input belongs to a set. Membership functions are used in the fuzzification and defuzzification steps of a Fuzzy logic system to map the non-fuzzy input values to fuzzy linguistic terms and vice versa.

The process of fuzzy logic:

A crisp set of input data are gathered and converted to a fuzzy set using fuzzy linguistic variables, fuzzy linguistic terms and membership functions. This step is known as fuzzification.

Fuzzification - The process of generating membership values for a fuzzy variable using membership functions.

An inference is made based on a set of rules.

Fuzzy Rules: In fuzzy expert systems, linguistic variables are used in fuzzy rules. For example:

IF wind is strong THEN sailing is good.

The fuzzy output is mapped to a crisp output using the membership functions, in the defuzzification step. Defuzzification is the process of transforming a fuzzy output of a fuzzy inference system into a crisp output. This is the purpose of the defuzzifier component of a FLS. Defuzzification is performed according to the membership function of the output variable.

**SOCIAL IMPACT:**

Road traffic accidents are a serious problem worldwide. The consequences of road traffic accidents primarily affect the direct participants of road traffic accidents and their families. The direct participants of road traffic accidents particularly suffer health consequences.

We have tried to understand what kind of parameters affect the seriousness of the accidents. One type of parameter that we have analyzed is the model year in regards to how serious the driver was injured. We have found that with vehicles that are of an older model, the seriousness of the injury to the driver was comparatively more than with vehicles having a recent manufacturing date.

Another parameter that was analyzed was which type of seatbelt along with which area had the most damage caused to the vehicle and how it in turn affected the seriousness of the front seat passengers injury.