Reliability of a Dataset using Z Number

Asst. Prof. Rama Dan[1], Sourav Prakash Sunny[2], Harsh Kumar Upadhayay[3], Utkarsh Sao[4], Piyush Gautam[5]

[1,2,3,4,5]*Dept. of ISE, MVJ College of Engineering, Bangalore, India*

Abstract **- In large measure, science and engineering dwell in the world of measurements and numbers. In this world, a basic question which arises is: How reliable are the numbers which we deal with? This question plays a particularly important role in decision analysis, planning, economics, risk assessment, design and process analysis. A Z number, Z=(A,B) has two components, The first component A is a restriction (constraint) on the values which a real valued uncertain variable, X, is allowed to take. The second component B is a measure of reliability (certainty) of the first component. The possibility of using the novel concept of Z numbers is introduced for the first time. This concept improves the computation time in handling uncertainty of the dataset with respect to fuzzy numbers. Uncertainty is quite relevant in real-time data. Due to lack of uncertainty, decision-based system would not be able to provide correct information. Z numbers thus enables to convert data to natural language. Data will be collecting from different domains such as tide energy, solar energy, food security and performing calculations to check the precision of the reliability and the uncertainty of the predicted data. This accuracy of the measure of the certainty or the unreliability of data can then be expressed in natural language, which can be understood commonly by all. To elaborate, the data classified above a certain threshold can be clubbed to say that the particular corresponding part can be termed as 'very highly probable', below a certain threshold can be clubbed to say that it can be termed as 'very lowly probable' and in between can be termed as 'highly probable', 'moderately probable' and so on. Hence, we aim to build a system with a graphical user interface that will provide correct information based on this measure of uncertainty computed using Z numbers.**

1.INTRODUCTION

In the tangible world, the data is not perfect, most of the time. This imperfection in data is mainly due to the unreliability of available or given data. The other aspects of data being imperfect, include the nature of data, the uncertainty of data and the correctness or precision of the data. Much information upon which decisions are made is uncertain and hence requires more carefulness. Thus, the measure of reliability comes out to be a very important factor in the analysis of data. There are various systems used for decision making. One of the decision-making theories which deals with imprecise information described in Natural Language (NL), is proposed in.

Apart from this, classical fuzzy set has been applied for decision making. Fuzzy sets were introduced independently by Lotfi A. Zadeh and Dieter Klaua in 1965 as an extension of the classical notion of set theory [1]. In classical set theory, the membership of elements in a set is assessed in binary terms according to a bivalent condition — an element either belongs or does not belong to the set. By contrast, fuzzy set theory permits the gradual assessment of the membership of elements in a set; this is described with the aid of a membership function valued in the real unit interval [0, 1].

A fuzzy set A is defined on a universe X may be given as: $A = \{\langle x, \mu A(x)\rangle \mid x \in X\}$ where $\mu A: X \rightarrow [0, 1]$ is the membership function A [1]. The membership value $\mu A(x)$ describes the degree of belongingness of $x \in X$ in A. The issue with fuzzy sets is that reliability of the information is not well being taken into consideration. Hence, Z-number are taken into consideration. The concept of a Z-number is intended to provide a basis for computation with numbers which are not totally reliable. Z-number can describe both the constraint and the reliability, and therefore, is used to resolve the issue of reliability [2][4].

A Z-number, Z, has two components, Z = (A, B) [1][2]. The first component, A, is a restriction (constraint) on the values which a real-valued uncertain variable is allowed to take. The second component, B, is a measure of reliability (certainty) of the first component. Typically, A and B are described in a natural language [1].

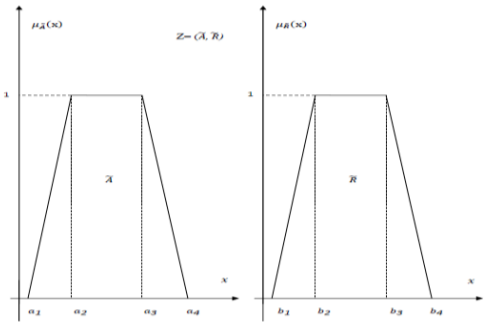

Fig 1: Z Number representation

2. OBJECTIVES

Uncertainty is quite evident in the real world where much of the decision-based information requires more carefulness due to the lack of the uncertainty. As a result, it is hard to formalize the ability to make rational decisions to be uncertain, inexact and/or incomplete information. The purpose of the project is that it is possible to classify the data by using the z-numbers technique in. Therefore, our aim is to build a system that handles the unreliability in the predicted data. This unreliability can be handled using the concept of Z numbers as suggested in.

The issue is that the reliability of information is not precisely taken into account. The reliability, among the most significant qualitative attributes of decision-relevant information, is not there in most of the cases. Decision based systems, hence would not be able to provide correct information. Therefore, Z numbers take into account, the reliability, while performing the classification of the dataset and help converting data to natural language. Hence, the problem statement can be framed as: Computation of Z numbers for unreliable data using the Tidal energy dataset as a case study

3. RELATED WORK

The Existing system present uses adaptive neuro-fuzzy inference system (ANFIS) and fuzzy logic. They are equivalent to fuzzy inference systems, with adaptive networks as mentioned in [7], thereby named as Adaptive Neuro Fuzzy Inference System (ANFIS), organically blending fuzzy logic and neural network.

A Sugeno fuzzy model with inputs *x*, *y* and an output *z*, shall comprise of:

- If *A*1 defined by *x* and *B*1 defined by *y*, then

$f1 = p1x + q1y + r1$

- Similarly, if *A*2 defined by *x*, *B*2 defined by *y*, then

$f2 = p2x + q2y + r2$

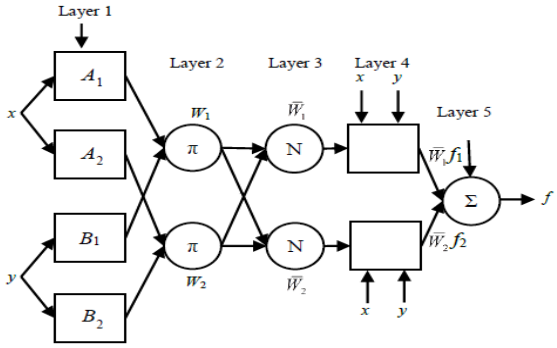

Fig 2: ANFIS Architecture

Drawbacks of the Existing System

The Existing system uses ANFIS and Fuzzy logic which increases computation time and no Graphical user interface has been provided which will enable the user to find the unreliability.

4. METHODOLOGIES

In this section we will discuss the methodology that we are using to avoid accidents block diagram of the proposed system is shown below

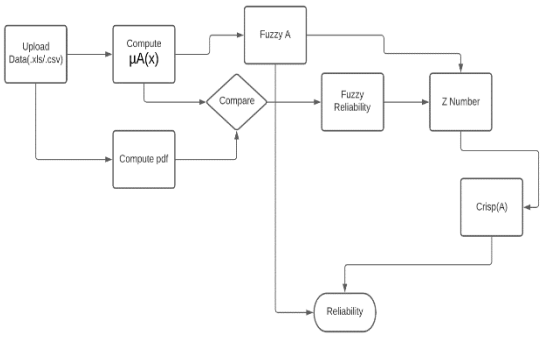

Figure 3: Block diagram of our proposed system

4.1 Block Diagram Description

This project mainly contains five modules

4.1.1 Data Uploading and Preprocessing: The dataset is uploaded to the interface in form of an excel file. The dataset could be one day's data (in this case) or one month data. It consists of the Intensity of sun at given time intervals of 10 minutes. To elaborate, the intensity at time 0 could be starting at 8:00 a.m., the subsequent Intensity at time 10 would correspond to 8:10 a.m., at time 20 to 8:20 a.m., at time 30 to 8:30 a.m., and so on till the last value.

4.1.2 FUZZYFICATION: The Project uses the initial concept of Fuzzy Numbers to derive a more elaborate Z Number. The first step towards this process is Fuzzification [3][4]. The Data which as mentioned above has Intensity with respect to time. So, we need to make a model first to calculate uncertainty corresponding to data. We thus plot a Triangular Membership Function taking Time as X axis and Intensity as Y Axis. The Triangular Membership data acts as a preliminary Fuzzy Number [3].

4.1.3 Z NUMBER: Now we convert need a Probability Density Function for our Input data before we get our Z number [5]. For getting a Probability Density Function we used an inbuilt python directory Scipy.data.norm.pdf (). Our Fuzzy Data now can be converted into Z Number. The Dot Product of PDF and Triangular MF gives our Z Number [3][5].

4.1.4 DE-FUZZIFICATION: The Z number has all the information needed for predicting uncertainty. But it's unreadable at the moment. The Z number is converted thus to a Crisp Number. The Crisp Number is then converted to Natural Language for easier understanding [6].

4.1.5 Graphical user interface: The graphical user interface is a form of user interface that allows users to interact with electronic devices through graphical icons and audio indicator such as primary notation, instead of text-based user interfaces, typed command labels or text navigation.

The graphical user interface in this project is developed using the Django web interface. Django is a Python-based free and open-source web framework that follows the model–template–views architectural pattern. It is maintained by the Django Software Foundation

5. SYSTEM DESIGN AND WORKING:

5.1 Flow Diagram

The flow diagram depicts the design of the project. The flow diagram for this project has the following phases:

- Input the dataset.
- Plot the input data, i.e., the given dataset.
- Plot the membership function.
- Calculate the reliability.
- Display the output.

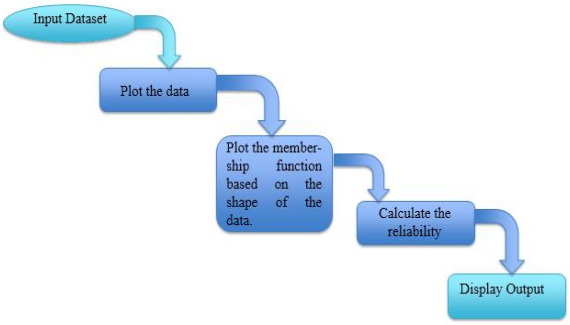

Figure 4: Flow diagram

5.1.1 Input the dataset

The dataset is uploaded to the interface in form of an excel file. The dataset could be one day's data (in this case) or one month data. It consists of the speed of the tide in m/s at given time intervals of 5 minutes. To elaborate, the speed of the tide at time 0 could be starting at 8:00 a.m., the subsequent speed at time 5 would correspond to 8:05 a.m., at time 10 to 8:10 a.m., at time 20 to 8:20 a.m., and so on till the last value.

5.1.2 Plot the input data

The dataset is plotted into a graph. The graph could be possibly plotted as line or scatter or any other graph. For simpler identification of the type of function and for ease in overlapping of the graph obtained in the later phase, the graph is plotted as a scatter plot. A scatter plot (also called a scatterplot, scatter graph, scatter chart, scattergram,or scatter diagram) Cartesian coordinates to display values for typically two variables for a set of data. The shape of the data is taken into account based on which, the membership function will be plotted.

5.1.3 Plot the membership function

Depending upon the shape that the data takes, be it a triangle or a trapezoid, we plot the triangular or trapezoidal membership function. In case of a triangular membership function, the minimum and the maximum values are determined from the data and a triangle is plotted taking the three points to be the two minimum values and one maximum value.

After plotting the triangle, the corresponding y values on the triangle are calculated by using the x values

from the original data whose values are obtained using the equation of line:
(y-y1)/(y2-y1) = (x-x1)/(x2-x1)
The y values obtained help in plotting the membership function. The plot for the data and the membership function are plotted so that they overlap each other [3][4].

5.1.4 Calculate the reliability

The points for the triangular membership function on the graph are calculated with the given x (time) values. The difference between the original data values and the triangular membership function is then obtained. The difference, thus obtained gives us the reliability values for each time interval [2].

5.1.4 Display the output

The dataset is then classified into three classes namely 'not likely', 'less likely' and 'more likely' according to the calculated reliability values. If the reliability falls between 0 and 0.3, the time interval is classified as 'not likely'. If the reliability falls between 0.3 and 0.65, the time interval is classified as 'less likely' and if the reliability is greater than 0.65 and less than 1, the data or the time interval is classified as 'more likely'.

6. EXPERIMENTAL RESULTS

The User Interface of the final project is built on Django web interface. The front page is designed to be a modern home page which serves as an index to locate the predictor. The predictor allows you to upload the file which you want to check for reliability. Upon clicking Upload Button we can browse the file we are looking for, the file can be an Excel file or a CSV file.

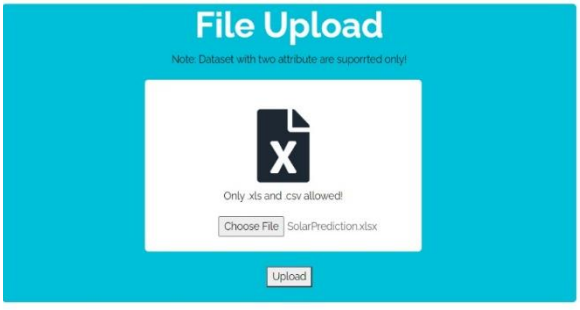

Figure 5: Dataset uploading page

After Uploading Successfully, you get a popup message as Upload Successful. Upon Successfully uploading file we can Predict the Uncertainty by clicking Predict button. The predict button performs computations on the backend and then dumps the final output. It also returns a Graph with Input data in a Scattered Graph and a Triangular Membership Function.

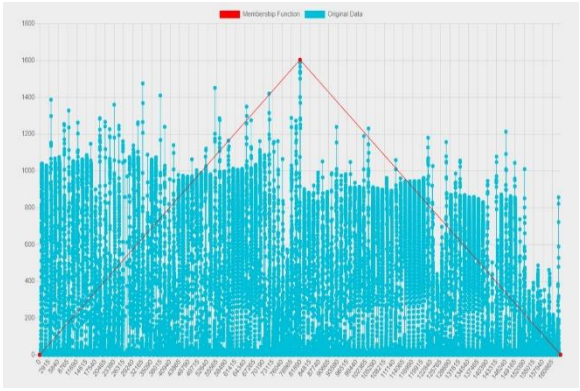

Figure 6: Resultant graph

User can also download the entire result in excel format as the output of the project is the classification of the dataset based on the calculated reliability that is obtained with the help of the membership function. The output is in the form of natural language, which is easily understandable.

time	radiation	result
0	1.21	likely
5	1.21	likely
10	1.23	less likely
15	1.21	likely
20	1.17	more likely
25	1.21	likely
30	1.2	likely
35	1.24	less likely
40	1.23	less likely
45	1.21	likely
50	1.23	less likely
55	1.21	likely
60	1.22	less likely
65	1.21	likely
70	1.23	less likely
75	1.22	less likely
80	1.21	likely
85	1.22	less likely
90	1.22	less likely
95	1.2	likely

Figure 7: Resultant Dataset

7. CONCLUSION

During the course of this project, a better understanding of the z-numbers has been gained come up to the surface. Various methods for implementation of the classification of data sets were able to come into light. Based on the references provided, the Existing system present uses adaptive network based fuzzy inference system and fuzzy logic. The existing system does not produce the output directly and does not consist of a graphical user interface. The proposed system is a system that handles the unreliability in data. The unreliability can be handled using the concept of Z numbers.

The project can be used in many applications provided. The scope of the project is suitable for all the data sets that follow the concept of triangular or trapezoidal membership function as suggested in. The system hence developed performs classification based on the given dataset, say wind data and produces results by classifying as more likely, less likely and so on, based on the reliability calculated by the membership function. The system provides output in natural language that is easily understood by all. The output produced can be helpful in various applications such as predicting the possibility of installation of windmills or solar panels, etc.

REFERENCES

[1] L. A. Zadeh, A Note on Z-numbers, Information Sciences, 181 (2011) 2923-2932

[2] Aliev, R. A., Alizadeh, A. V., & Huseynov, O. H. (2017).“An introduction to the arithmetic of Z-numbers by using horizontal membership functions”. Procedia Computer Science, 120, 349–356.

[3] Ahmed A.Valieva , Tarlan S. Abdullayeva , Akif V. Alizadehb , Nigar E. Adilovac. ‘Comparison of measures of specificity of Z-numbers’. ICSCCW 24-25, 2017.

[4] Derar Atallah Talal Mraiziq (2015), ‘Computation and Reasoning with Z-Numbers’, Rahib H. Abiyev, Kaan Uyar, Umit Ilhan, Elbrus Imanov, and Esmira Abiyeva (2018), Estimation of Food Security Risk Level Using Z-Number-Based Fuzzy System.

[5] Ahmed A.Valieva , Tarlan S. Abdullayeva , Akif V. Alizadehb , Nigar E. Adilovac. ‘Comparison of measures of specificity of Z-numbers’. ICSCCW 24-25, 2017.

[6] Bingyi Kang, Daijun Wei, Ya Li, Yong Deng (2012).“A method of converting Z-number to classical fuzzy number”. Journal of Information & Computational Science 9(3): 703-709

[7] Andrey (2017, May). Solar Radiation Prediction, Version 1. Retrieved 20 June 2021 from ‘https://www. kaggle.com/dronio/SolarEnergy’