



A FUZZY APPROACH TO CATEGORIZE THE ATTRIBUTES IN HEALTH CARE STRUCTURED DATA

Naresh Patel K M¹, Bindu D P², Deepa A L³, Deepika A R⁴, Jaina H Shah⁵,

¹Assistant Professor, Dept of CSE, BIET College, Davangere.

^{2,3,4,5}Final year student, Dept of CSE, BIET College, Davangere

ABSTRACT

Healthcare performs a key role in the health sector in today's era. While gathering healthcare data, the problems will appear in the classification of data. Structured data conforms to a datamodel, which has a well-defined structure, follows a consistent order and can be easily accessed and used by a person or a computer program. Fuzzy logic comes into picture when healthcare structured data is used to classify sensitive and non-sensitive attributes. Fuzzy logic is capable of handling the partial truth, where truth value may range between completely true and completely false. In Boolean algebra, 1 represents truth and 0 represents false, while fuzzy logic is capable to accept partial truth or partial false(intermediate values).

Key-Words:Healthcare, fuzzy logic, structured data.

INTRODUCTION

Online diagnosis is becoming popular day by day. In today's world people are so busy, that they hardly have enough time to visit a doctor. So they can use the service of this online diagnosis system residing home or office and have an idea about the disease. After that they can consult the specialist doctor if it is necessary or serious based on the grading of diseases. Structured data is the data which conforms to a data model, has well defined structure, follows a consistent order and can be easily accessed and used by person or computer program. Structured data is usually stored in well-defined schema such as Databases. It is in a form of tabular columns and rows that clearly define its attributes. So, that its elements can be made addressable for more effective processing and analysis. The term fuzzy refers to things

which are not clear or are vague. In the real world many times we encounter a situation when we can't determine whether the state is true or false, their fuzzy logic provides a very valuable flexibility for reasoning. In this way, we can consider the inaccuracies and uncertainties of any situation. In fuzzy mathematics, fuzzy logic is a form of many-valued logic in which the truth values of variables may be any real number between 0 and 1 both inclusive. It is employed to handle the concept of partial truth, where the truth value may range between completely true and completely false. By contrast, in Boolean logic, the truth values of variables may only be the integer values 0 or 1.

LITERATURE SURVEY

[1]Guney. Fuzzy Logic (FL) is a method of computing that follows human cognitive ability. FL acts like the way of decision making in humans which based on degrees of truth instead of Boolean logic YES and NO. He highlighted that human decision making includes a lot of possibilities between YES and NO, such as possibly, certainly, almost.

[2]Torres and Nieto. Explained diagnosis of the disease involves several levels of uncertainty and inaccuracy. A single disease may appear in many forms based on the patient, and with different intensities. A single symptom may correspond to different diseases. The description of disease entities uses linguistic terms that are also not exact and vague. To deal with inaccuracy and uncertainty, fuzzy logic introduces fractional truth values, between YES and NO.

[3]Warren et al. Presented a decision support system for automating the application of clinical practice guidelines based on fuzzy method. The study shows that the test report yields likelihood

estimates rather than confirmation of presence or absence of disease and in fuzzy method, likelihood estimates can be handled as membership values and used as such in fuzzy inference model. Thus the study shows that the fuzzy logic has great contribution in health industry.

[4]K.V.Kanimozhi1, Dr. M. Venkatesan. Structured data is a data included in relational database system. Examples are database tables, objects, tags, reports, indexes etc. Being structured and highly organized it can be managed by SQL and its multiple variations developed by IBM, ADO.net, ODBC and many RDBMS support, Due to explicit semantics and structure efficient search is possible for focused content by simple and straightforward search engine.

[5] R. Vaarandi.He discussed the state-of-the-art algorithms for information extraction from structured data. We divided the algorithms into relationship-based, which used feature-driven approaches, template-based, that used data-driven approaches and deep learning based technique.

DATASET COLLECTION

Dataset in this project is a collection of information of individual patients which is used as the main database. This dataset is acquired from the Local hospitals patient record. These dataset contains following attributes:

PID, Name, Age, Sex, Zip code, City, Disease

PID	NAME	SEX	AGE	CITY	ZIPCODE	DISEASE
1953726	Rakesh	M	28	Harihar	577598	Fever
1953727	Sakkamma	F	34	Davangere	577003	Anaemia
1953841	Basavaraj	M	55	Halibathi	577566	Lipoma
1953862	Uma	F	33	Rannebenur	581115	IHD
1953900	Janamma	F	78	Harihar	577599	Vertigo
1953906	Chandrappa	M	57	Chitradurga	577501	Malaria
1953729	Shobha dev	F	30	Haveri	581110	Fever

Example records of Patient dataset

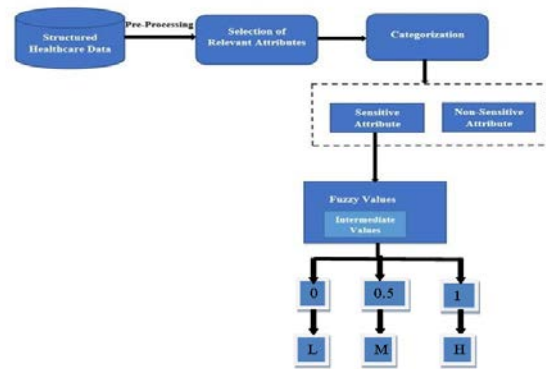
METHODOLOGY

Structured Data: Structured data is the data which conforms to a data model, has a well define structure, follows a consistent order and can be easily accessed and used by a person or a computer program. Structured data is usually stored in well-defined schemas such as Databases. It is generally in tabular form with column and rows that clearly define its attributes.

Pre-processing: The real-world data is generally, incomplete where there is lack of attributes values, containing of only aggregate data. After pre-processing is done, noise and outliers are removed. It includes the tasks like data cleaning, integration, transformation, reduction, discretization.

Extraction: Extraction is a process that involves retravel of data from various sources and store those results in the data warehouse.

Categorization:Categorization is the process through which ideas and objects are recognized, differentiated, classified, and understood. The word “categorization” implies that objects are sorted into categories, usually for some specific purpose.



Proposed block diagram to categorize the attributes in healthcarestructured data by using fuzzy logic.

ALGORITHM

This project is implemented using FCS algorithm.

1. Set $k = [N/2.4]$ (the nearest integer to $N/2.4$) and compute:

$$a = \sum_{i=1}^k x_i \mu_A(x_i) + \sum_{i=k+1}^N x_i \mu_A(x_i)$$

$$b = \sum_{i=1}^k \mu_A(x_i) + \sum_{i=k+1}^N \mu_A(x_i)$$

Compute $c' = a/b$

2. Find $k' \in [1, N - 1]$ such that $xk' \leq c' \leq xk' + 1$
3. Check if $k' = k$. If yes, stop and set $c' = cl$, and $k = L$. If no, go to Step 4.

4. Compute $s = \text{sign}(k' - k)$ and:

$$a' = a + s \sum_{i=\min(k,k')+1}^{\max(k,k')} x_i [\mu A(x_i) - \mu A(x_{i-1})]$$

$$b' = b + s \sum_{i=\min(k,k')+1}^{\max(k,k')} x_i [\mu A(x_i) - \mu A(x_{i-1})]$$

Compute $c''(k') = a'/b'$

5. Set $c' = c''(k')$, $a = a'$, $b = b'$ and $k = k'$ and go to Step 2

END if

6. if $K \leq 0.5$ Fuzzy value= Low

else if $0.6 < K \leq 0.7$ Fuzzy value=Medium

else Fuzzy value= High

End if

return K

RESULTS AND SCREENSHOTS

Disease Categorization using Fuzzy Method

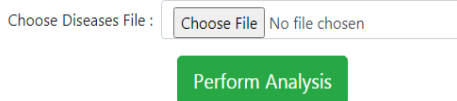


Fig: Front end web application of our FLS

The home page web application of our FLS. It consists of file choose option for uploading dataset file from our system. After uploading we need to press perform analysis button to get output csv file containing disease grading values for each disease. It also contains Enter disease name and severity level fields for getting grading and value along with severity graphs manually

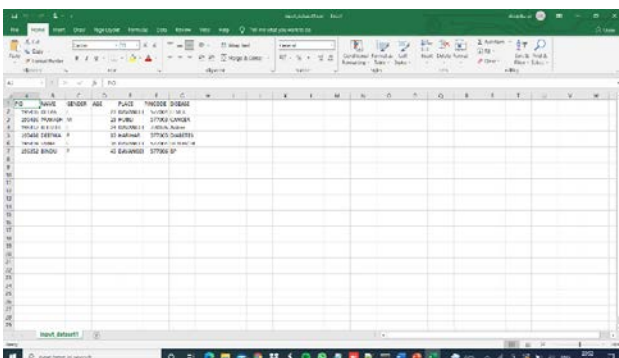


Fig: Input Dataset for FLS

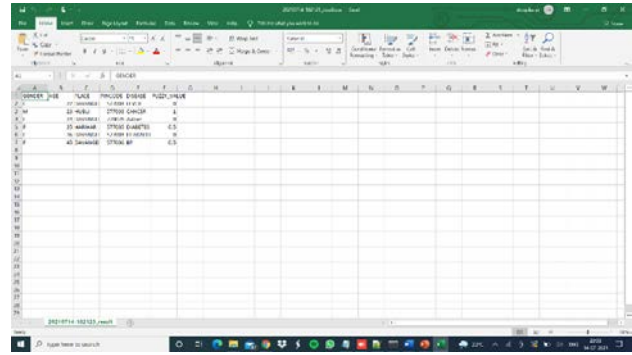


Fig: Output of FLS

CONCLUSION

Overall, the project was successful in developing the FLS for its purposes. However, the main concerns remain about the accuracy of the results of this Fuzzy Logic System, as it was developed using incomplete and relatively inaccurate data. With higher quality data a more effective system with a larger rule base and more accurate membership functions should be possible. In terms of its basic goals, this FLS is capable of taking a user's input and evaluating whether the severity of disease is high, medium and low by assigning the grade values. However, it is inaccurate, and will generally not be of huge use to the general public at this stage. Further work is also needed to refine the rule base so that it can provide accurate values.

REFERENCES

- [1.]Guney (2019) "Fuzzy Sets for Knowledge Representation and Inferencing," Proc. IEEE FUZZ Conference, IEEE World Congress on Computational Intelligence.
- [2.] Torres and Nieto (2019) "An interval fuzzy logic toolbox for control applications," Proc. IEEE FUZZ Conference, pp.
- [3.] Singh (2018) "Research about Fuzzy logic and software developers" IEEE FUZZ Conference
- [4.] Szolovits (2017) "According to his principle, in fuzzy logic each declarative sentence is required to be either true or false" IEEE FUZZ Conference.
- [5.] R Bělohlávek, JW Dauben, GJ Klir (2017) "Fundamental principles of classic logic" IEEE FUZZ Conference
- [6.] Warren et al (2017) Presented a decision support system for automating the application of clinical practice guidelines based on fuzzy method. IEEE World Congress.
- [7.] Mustafa Demetgul, Osman Ulkir (2016) Fuzzy logic is an extension of two valued

logic or Boolean logic. IEEE FUZZ Conference.

[8.] R. Vaarandi (2016)He discussed the state-of-the-art algorithms for information extraction from structured data. IEEE International Conference.

[9.] K.V.Kanimozhi1, Dr. M. Venkatesan (2015)Structured data is a data included in relational database system. IEEE International Conference.