Applied Mathematics and Computing
Vol. 1, No.1, (June, 2023)

Mathematical Modelling for Prediction and Growth of typhoid

Subrata Kumar Nayak

Dept. of Computer Science,
GVHSS, Sishu Ananta Mahavidyalaya,
Balipatna, Odisha, 751029, India
subratanayaksap@gmail.com

Sateesh Kumar Pradhan

Department of Computer Science
and Application, Utkal University,
Bhubaneswar, Odisha
sateesh1960@gmail.com

Sujogya Mishra

Dept. of Mathematics
Odisha University of Technology and Research
Bhubaneswar, Odisha, 751029, India
sujogya123@gmail.com

Abstract: This paper deals with analysing Typhoid data set from various medical sources. Our intention to find minimum attribute responsible for Typhoid. In this Paper we use the Rough Set [1] concept and ordinary differential equation to predict and it's growth rate.The findings are helpful in feature extraction and finding a meaningful conclusion.

Keywords: Typhoid; Rough Set Theory (RST); Soft Computing; Growth Rate; T-test.

1 Introduction

Typhoid is the oldest form of the ailment every year thousands of persons around the world suffered from Typhoid. We had gone through several medical journal and found that there was no concrete symptoms to find the exact pattern for Typhoid till date. At the present time it's symptom Often puzzle with some ordinary diseases like flu, Influenza, and some standard diseases for our work we depend data sets from various medical sources of our state. Once we collect the data we then use Correlation technique to find the dissimilar records, then apply Rough Set technique [2] on these data set to reduce the number of attributes and then verify our claim by statistical validation. The data we collected are vague, imprecise and lack of concrete information as it was collected randomly. The nature of the data was vague in nature as we collected data from various medical sources with variable age groups and from various place of our state. Rough Set is one of Soft Computing technique which deals with vagueness and impreciseness accurately. As there is no conclusive symptoms for Typhoid till date some time acute case of Typhoid may confuse the doctors with some very ordinary diseases like influenza ,flue. Our work help the doctor in diagnosis of Typhoid. We use Rough Set Technique as it is a very useful tool for analysis of vague imprecise and data with lack of information. We divide our entire work in two section in the section-1 we analyze the data collected using Rough Set Concept in the first section then in the second section we study the growth rate of Typhoid in the next decade using ordinary differential equation. Another aspect of Rough Set is to deals with the boundary to extract the hidden information better than other soft computing Techniques.

2. Literature Review

Talasila et. al [3] have discussed about prediction of diseases using RST with Recurrent Neural Network in Big Data Analytics and verify their result . This system uses Rough Set Theory and Neural Network in Bigdata to predict future diseases. Venkatesh,et. al [4] discussed Big Data Prediction Analytics Model for heart disease prediction using NB technique as BPA-NB. This scheme used a probabilistic Bayes' theorem for analysis of data, and use Hadoop-spark as tool for big data analysis to get the inside result accurately. Nilashi, et al. [5], discussed the use hybridization of Support Vector Regression (SVR), Adaptive Neuro-Fuzzy Inference System (ANFIS), Predicting Parkinson's Disease Progression accurately. Noia et al. [6], Predicting health risk problem in work place by using computational intelligence techniques. Nair et al [7], apply Big data machine engine to accurately counter diseases occurs in Real Time. Chen, et al [8] use light gradient busting machine to predict extubation failure and also feature importance analysis according to the result of Light GBM and interpret these features using shapley Additive Explanations (SHAP). Van et al. [9] presented the hierarchical analysis of machine learning algorithm for predictions of at-risk patients. Nanda et al. [15], discussed about Rough Set Approach for malicious node detection and secure data aggregation in WSN. They explained how Rough set theory is used to find the malicious node in sensor network and data packets are sending securely to the sink node.

3 Basics of Rough Set

Rough Set Theory is derived from Conventional Set theory. This theory was most widely used soft computing technique used to deals with uncertainty and vagueness. Rough Set theory has seven basic components, classified as 1. Indiscernibility 2. Upper Approximation, 3. Lower Approximation 4. Dispensable 5. Indispensable 6. Reduct [10] 7. Core

1.3.1 Indiscernibility -: Indiscernibility is define as define as IND (K) = $\{(x,y)\in U^2 \mid \forall a\in K, a(x) = a(x)\}$. It is an important concept of Rough Set Theory. Mathematically Indiscernibility implies group of attributes responsible for a decision making process if any attribute drop from this group doesn't able to produce the decision correctly. The attributes which are not Indiscernible called as Discernible.

1.3.2 Upper Approximation-: $\overline{B} = \{x \mid [x]_B \cap X \neq \varphi\}$

Where $\overline{B}$ is called as Upper Approximation, where the object interest probably present in the Set

1.3.3 Lower Approximation-: When the object of interest certainly present in the set called as Lower Approximation denoted by $\underline{B}$ and define as $\underline{B}(X) = \{x : [x]_B \subseteq X\}$

1.3.4-: Boundary Region: Boundary region define as difference between Upper and Lower Approximation

1.3.5-: Reduct: Reduct of a set define as $B' \subset B$ such that IND(B')=IND(B)

1.3.6: Core-:

$$Core = \bigcap Redut$$

2.1-: Analysis Phase-1

We had collected the data from various sources (Medical Sources). Represented the data in form of a table given below.

Representation of People suffering from Typhoid with various symptoms.

Table. 1 (Data Table)

serial	Age group	Positive Cases
1	15-25	1500
2	25-45	2500
3	>45	7000

The main objective of this paper to find minimum number of attributes (symptoms)

2.2 Analysis Phase -2

We collected data from various medical sources, in most of the cases we found that the person suffering from typhoid has the symptoms attributes <High Fever, Cough, Vomiting, Body Ache, Stomach Pain, Loss of Appetite> , we rename these attributes as <1,2,3,4,5,6> and it's values are significant and insignificant as <a, b> decision attributes rename as Positive and Negative rename as <p, q>.

Table. 3 (Rename Table)

Conditional Attributes	Reduct Name	Decision Attributes	Reduct Name	Values of Conditional attributes	Reduct Name
High Fever	1	Positive	p	Significant	a
Cough	2	Negative	q	Insignificant	b
Vomiting	3	NA	NA	NA	NA
Body Ache	4	NA	NA	NA	NA
Stomach Pain	5	NA	NA	NA	NA
Loss of Appetite	6	NA	NA	NA	NA

Here NA –Not Applicable.

After collecting the data over 10000, we apply correlation techniques and get 6 dissimilar groups (Records). All 6 groups we have are dissimilar and mutually exclusive. Represented in the form of a Table given below as:

Table. 3 (Information Table)

E	1	2	3	4	5	6	d
E_1	a	b	a	b	a	a	p
E_2	a	b	a	b	a	b	p
E_3	a	a	b	a	a	a	q
E_4	b	b	b	a	b	b	p
E_5	b	b	a	b	b	a	p
E_6	a	a	a	b	b	b	q

To find the reduct we are using Indiscernibility of each conditional attributes, by combining the conditional attributes

IND<1> = $\{\{E_1,E_2,E_3,E_6\},\{E_4,E_5\}\}$

IND<2> = $\{\{E_1,E_2,E_4,E_5\},\{E_3,E_6\}\}$

IND<3> = $\{\{E_1,E_2,E_5,E_6\},\{E_3,E_4\}\}$

IND<4> = $\{\{E_1,E_2,E_5,E_6\},\{E_3,E_4\}\}$

IND<5> = $\{\{E_1,E_2,E_3\},\{E_4,E_5,E_6\}\}$

IND<1,2> = $\{\{E_1,E_2\},\{E_3,E_6\},\{E_4,E_5\}\}$

IND<2,3> = $\{\{E_1,E_2,E_5\},\{E_3\},\{E_4\},\{E_6\}\}$

IND<3,4> = $\{\{E_1,E_2,E_5\},\{E_3\},\{E_4\},\{E_6\}\}$

IND<4,5> = $\{\{E_1,E_2\}\{E_5,E_6\},\{E_3,E_4\}\}$

IND<5,6> = $\{\{E_1,E_3\}\{E_4,E_6\},\{E_2\},\{E_5\}\}$

IND<1,3> = $\{\{E_1,E_2,E_6\},\{E_3\},\{E_4\},\{E_5\}\}$

IND<1,4> = $\{\{E_1,E_2,E_6\},\{E_3\},\{E_4\},\{E_5\}\}$

IND<1,5> = $\{\{E_1,E_2,E_3\},\{E_4,E_5\},\{E_6\}\}$

IND<1,6> = $\{\{E_1,E_3\},\{E_2,E_6\},\{E_4\},\{E_5\}\}$

IND(2,4) = $\{\{E_1,E_2,E_5\},\{E_3\},\{E_4\},\{E_6\}\}$

IND(2,5) = $\{\{E_1,E_2\},\{E_3\},\{E_4,E_5\},\{E_6\}\}$

IND(2,6) =$\{\{E_1,E_5\},\{E_3\},\{E_4,E_2\},\{E_6\}\}$

IND(3,4) $=\{\{E_1,E_2,E_5,E_6\},\{E_3,E_4\}\}$

IND(3,5) $=\{\{E_1,E_2\},\{E_3\},\{E_4\}\{E_5, E_6\}\}$

IND(3,6) $=\{\{E_1,E_5\},\{E_2,E_6\},\{E_3\},\{E_4\}\}$

IND(4,5) $=\{\{E_1,E_2\},\{E_5,E_6\},\{E_3\},\{E_4\}\}$

IND(4,6) $=\{\{E_1,E_5\},\{E_2,E_6\},\{E_3\},\{E_4\}\}$

IND(5,6) $=\{\{E_1,E_3\},\{E_2\},\{E_3\},\{E_4,E_6\}\}$

IND(1,2,3) $= \{\{E_1,E_2\},\{E_3\},\{E_4\},\{E_5\},\{E_6\}\}$

IND(1,2,4) $=\{\{E_1,E_2\},\{E_3\},\{E_4\},\{E_5\},\{E_6\}\}$

IND(1,2,5) $=\{\{E_1,E_2\},\{E_3\},\{E_4,E_5\},\{E_6\}\}$

IND(1,2,6) $=\{\{E_1\},\{E_2\},\{E_3\},\{E_4\},\{E_5\},\{E_6\}\}$

IND(2,3,4) $=\{\{E_1,E_2,E_5\},\{E_3\},\{E_4\},\{E_6\}\}$

IND(2,3,5) $=\{\{E_1,E_2\},\{E_3\},\{E_4\},\{E_6\}\}$

IND(2,3,6) $=\{\{E_1,E_5\},\{E_2\},\{E_3\},\{E_4\},\{E_6\}\}$

IND(3,4,5) $= \{\{E_1,E_2\},E_3,E_4,\{E_5,E_6\}\}$

IND(3,4,6) $= \{\{E_1,E_5\},\{E_2,E_6\},\{E_3\},\{E_4\}\}$

IND(4,5,6) $=\{\{E_1\},\{E_2\},\{E_3\},\{E_4\},\{E_5\},\{E_6\}\}$

The reduct from the above data will be attribute $\langle 1,2,6 \rangle$ and $\langle 4,5,6 \rangle$.

Core $= \bigcap \text{Reduct} = \bigcap \langle 1, 2, 6 \rangle, \langle 4, 5, 6 \rangle = 6$.

From the above equivalence classes we have two reduct set out of 6 conditional attributes, i.e., {1,2,6},{4,5,6}. Using both the reduct set we get the core i.e. 6^{th} conditional attributes (the loss of appetite) most essential attributes.

Table. 4 (Reduct Table-1)

E	1	2	6	d
E_1	a	b	a	p
E_2	a	b	b	p
E_3	a	a	a	q
E_4	b	b	b	p
E_5	b	b	a	p
E_6	a	a	b	q

from the above reduct table we have row $\{E_5,E_6\}$ somewhat provide ambiguous so we drop both the table from further analysis.

Table. 5 (Derived Reduct Table-1)

E	1	2	6	d
E_1	a	b	a	p
E_2	a	b	b	p

Now E_3, E_4 gives same ambiguous result drop both row from further analysis.

Rule generation using the reduct set-1

$$1 \rightarrow a, 2 \rightarrow b, 6 \rightarrow a > \rightarrow \text{decision is p}$$

$1 \rightarrow a, 2 \rightarrow b, 6 \rightarrow b >\rightarrow \text{decision is p}$

1,2,3,4,5,6 ,a,b,p,q is defined in the introduction section.

Rule generation using the reduct set-2.

Table. 6 (Reduct Data Table-2)

E	4	5	6	d
E_1	b	a	a	p
E_2	b	a	b	p
E_3	a	a	a	q
E_4	a	b	b	p
E_5	b	b	a	p
E_6	b	b	b	q

In reduct set-2 E_3 gives an ambiguous result.

Table. 7 (Derived Reduct Table-2)

E	4	5	6	d
E_1	b	a	a	p
E_2	b	a	b	p
E_4	a	b	b	p
E_5	b	b	a	p
E_6	b	b	b	q

Rule generated using reduct Set-2

$1.\langle 4 \rightarrow b, 5 \rightarrow a, 6 \rightarrow a \rangle \rightarrow \text{Decision is p}$

$2.\langle 4 \rightarrow b, 5 \rightarrow a, 6 \rightarrow b \rangle \rightarrow \text{Decision is p}$

$3.\langle 4 \rightarrow a, 5 \rightarrow b, 6 \rightarrow b \rangle \rightarrow \text{Decision is p}$

$4.\langle 4 \rightarrow b, 5 \rightarrow b, 6 \rightarrow a \rangle \rightarrow \text{Decision is p}$

$5.\langle 4 \rightarrow b, 5 \rightarrow b, 6 \rightarrow b \rangle \rightarrow \text{Decision is q}$

2.3 Statistical Verification of the result

Verification of the result using t-test,

Using t-test we chosen randomly from two groups of sample one Set contains 12 similar groups other one contains 10 similar groups

For t-test we apply Conventional method of treatment for Typhoid to a group of 12 people applied Modern Medical Treatment to the group of 10 people with the following data given by $\overline{X_1} = 75, \overline{X_2} = 71, s_1 = 4, s_2 = 5$ and $\sigma_1 = \sigma_2$ [8].

H_0(Null Hypothesis)-: All 6 conditional attributes are reasonable for Typhoid.

H_1(Alternate Hypothesis)-: Only two groups of reduct sets are essentially gives significant result

We can have a Confidence Interval

$$\left(\overline{X}_1 - \overline{X}_2\right) \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1^2} + \frac{1}{n_2^2}}$$

$$s_p = \sqrt{\frac{(n_1\text{-}1)s_1^2 + (n_2\text{-}1)s_2^2}{n_1 + n_2 - 2}}$$

Sp = 4.47.

So the confidence interval is given as $4 \pm t_{0.05} \times 4.47 \times 0.13$ = $4 \pm 1.725 \times 4.47 \times 0.13$.

3. Conclusion

We have considered six conditional attributes which are defined in Table-2. After complete analysis using Rough Set approach, we get three essential conditional attributes (4,5,6) and derived a set of rules using Table-6. We have considered six attributes randomly for Typhoid collected from different medical sources (Odisha) but only three attributes are essential namely, Body ache, Stomach pain and Loss of appetite along with fever (default case).

References

1 Z. Pawlak, “Rough Sets-Theoretical aspect of reasoning about data”, Kluwer Academic Publishers, 1982.

2 WHO, “Influenza update”, 359 Release, 2020.

3 Stanford Health Care, 2018.

4 R. Vashist, M. L. Garg, “Rule generation based on reduct and core: A rough set approach”, International Journal of Computer Applications,vol- 29, pp.1-4,2011.

5 S. Mishra, S. P. Mohanty, S. K. Pradhan, “Generation for student feedback system by the use of rough set theory”, IJCA, vol- 131(18), 2015.

6 E. W. Ronald, H. Myers, L. Myers Sharon, “Probability and Statistics for engineers and scientists”,8/e, 2006.

7 J. L. Devore, “Probability and Statistics for Engineering and the Sciences”,8/e, 2012.

8 V. Talasila, K. Madhubabu,M. C. Mahadasyam, N. J. Atchala, L.S. Kande, “The Prediction of Diseases Using Rough Set Theory with Recurrent Neural Network in Big Data Analytics”, International Journal of Intelligent Engineering and Systems, Vol.13, No.5, pp.10-18,2020

9 R. Venkatesh, C. Balasubramanian, and M. Kaliappan, “Development of Big Data Predictive Analytics Model for Disease Prediction using Machine Learning Technique”, Journal of Medical Systems, Vol.43, No.8, pp.272, 2019.

10 M. Nilashi, O. Ibrahim, H. Ahmadi, L. Shahmoradi, and M. Farahmand, “A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques”, Biocybernetics and Biomedical Engineering, Vol.38, No.1, pp.1-15, 2018.

11 A. Di Noia, A. Martino, P. Montanari, and A. Rizzi, “Supervised machine learning techniques and genetic optimization for occupational diseases risk prediction”, Soft Computing, pp.1- 14, 2019.

12 L. R. Nair, D. Sujala Shetty, and D. Siddhanth Shetty, “Applying spark based machine learning model on streaming big data for health status prediction”, Computers & Electrical Engineering, Vol.65, pp.393-399, 2018.

13 T. Chen, J. Xu, H. Ying, X. Chen, R. Feng, X. Fang, and J. Wu, “Prediction of Extubation Failure for Intensive Care Unit Patients Using Light Gradient Boosting Machine”, IEEE Access, Vol.7, pp.150960-150968, 2019.

14 F. Van Wyk, A. Khojandi, and R. Kamaleswaran, “Improving Prediction Performance Using Hierarchical Analysis of Real-Time Data: A Sepsis Case Study”, IEEE Journal of Biomedical and Health Informatics, Vol.23, No.3, pp.978- 986, 2019.

15 A. Nanda, A.K Rath, “Rough Set Approach for Malicious Node Detection and Secure Data Aggregation in WSN”, International Journal of Applied Engineering Research, Vol-11 (8), pp. 5536-5541, 2016.

16 N. Kausar, A. Abdullah, B. B. Samir, S. Palaniappan, B. S. AlGhamdi, and N. Dey, “Ensemble clustering algorithm with supervised classification of clinical data for early diagnosis of coronary artery disease”, Journal of Medical Imaging and Health Informatics, Vol.6, No.1, pp.78-87, 2016.