International Journal of Artificial Intelligence and Neural Networks
Volume 6 : Issue 1 [ISSN 2250-3749]
Publication Date : 31 August, 2016

Regression Based Fuzzy Logic Classification Approach for Non Linear Data Set in Health Care System

Dr. Manish Pandey, Gurinderjit Kaur, Sachin Chauhan, Jagbir Gill

Abstract –Data processing, popularly known as data discovery in giant information, allows companies and organizations to form calculated choices by grouping, accumulating, analyzing and accessing company information. Authors propose a sturdy metaphysics based mostly third-dimensional information deposit and mining approach to deal with the problems of organizing, reportage and documenting polygenic disease cases as well as causalities. data processing procedures, during which map and information views representational process similarity and comparison of attributes extracted from warehouses, square measure utilized in this studies, for understanding the ailments supported gender, age, geography, food habits and hereditary traits. Statistic statement takes the past values of a statistic and uses them to forecast the longer term values. Fuzzy regression strategies have normally been wont to develop shopper preferences models that correlate the engineering characteristics with shopper preferences relating to a replacement product; the patron preference models offer a platform, wherever by product developers will decide the engineering characteristics so as to satisfy shopper preferences before developing the merchandise. Recent analysis shows that these fuzzy regression strategies area units normally won't to model client preferences. We tend to propose a Testing the strength of Exponential Regression Model over regression toward the mean Model.

Keywords – Health-care management systems, Fuzzy regression, Data mining, Forecasting, Fuzzy membership function.

I – Introduction

Mining is a methodology of isolating gaining from gigantic content reports. [7]. Along these lines information mining procedures, for example, grouping, affiliation and bunching are for the most part used to remove the covered up, beforehand inconspicuous learning from voluminous of databases. Of the different information examination method arrangement is a directed machine learning system which makes forecasts about future class cases by mapping cases of testing information to the predefined class marks which is gain from the supplied examples of classes with class names. There are a few models in characterizations, for example, probabilistic model, developmental algorithmic model and so forth.

Characterization includes of foreseeing a definite lead to read of a given data. To anticipate the result, the calculation forms a preparation set containing a meeting of traits and also the separate result, usually known as objective or forecast characteristic. The calculation tries to search out connections between the traits that may create it conceivable to foresee the result. Next the calculation is given data settled not seen your time recently, known as forecast set, that contains identical arrangement of properties, other than the expectation property – not nonetheless illustrious. The calculation investigations the information and produces an expectation. The expectation exactitude characterizes however "great" the calculation is. Order procedure is provided for making ready a additional intensive smorgasbord of knowledge than relapse and is developing in fame.

Arrangement is an alternate method than grouping. Order is like grouping in that it likewise sections client records into unmistakable portions called classes. Yet, not at all like bunching, an arrangement examination obliges that the end-client/examiner know early how classes are characterized. It is vital that every record in the dataset used to manufacture the classifier as of now have a worth for the credit used to characterize classes. Since every record has a worth for the credit used to characterize the classes, and on the grounds that the end-client settles on the ascribe to utilize, arrangement is a great deal less exploratory than grouping. The target of a classifier is not to investigate the information to find intriguing fragments, but instead to choose how new records ought to be characterized. Characterization schedules in information mining additionally utilize a mixture of algorithms.

Fuzzy rule-based systems (FRBSs) square measures acknowledge ways inside soft computing, supported fuzzy ideas to handle advanced real-world issues. They need become a robust methodology to tackle numerous issues like uncertainty, impreciseness, and non-linearity [7]. They're usually used for identification, classification, and regression tasks. FRBSs are deployed in a very variety of engineering and science areas. FRBSs are referred to as fuzzy abstract thought systems or just fuzzy systems. once applied to specific tasks, they additionally could receive specific names like fuzzy associative recollections or fuzzy controllers. They're supported the fuzzy pure mathematics that aims at representing the information of human specialists in an exceedingly set of fuzzy IF-THEN rules. Rather than victimization crisp sets as in classical rules, fuzzy rules use fuzzy sets. Rules were ab initio derived from human specialists through information engineering processes.

II – Review of literature:

Shastri et.al [1] propose a strong metaphysics primarily based three-d knowledge deposit and mining approach to handle the problems of organizing, news and documenting polygenic disorder cases, as well as causalities. data processing procedures, at intervals that map and data views depiction similarity and comparison of attributes extracted from warehouses, unit utilized during this studies, for understanding the ailments supported gender, age, geography, food habits and hereditary traits. Besides data image, data interpretation is planned for wealthy diagnosis, ensuing prescription and applicable medication. Higuchi et al. [2] describes AN analysis technique supported fuzzy set for health scrutiny knowledge. This technique converts health info into fuzzy degree to manage a variable info analysis. The obtained fuzzy degree is taken under consideration as associate attribute price in interval [0, I]. The degree shows a normality of health condition. Throughout this study, fuzzy membership functions are created from commonplace divisions of reference interval of health medical info. As associate example, they calculated fuzzy degrees and ill health index from Japanese health medical info. throughout this result, they confirmed that the obtained ill health index corresponded with medical established theories.

Chowdary et al. [14] built up another system for choice tree for arrangement of information utilizing an information structure called Peano Count Tree (P-tree) which improves the productivity and adaptability. They apply Data Smoothing and Attribute Relevance methods alongside a classifier. Test results demonstrate that the P-tree strategy is altogether quicker than existing characterization systems and the favoured technique for mining on information to be arranged. Kishana et al. [15] concentrates on visual information digging applications for improving business choices. The product based framework is executed as a completely robotized and sufficiently shrewd to produce into results of every business exchange.

.

III - Fuzzy Based Regression Analysis:

Profit foretelling ways embrace quantitative and qualitative foretelling ways. Quantitative prediction includes statistic analysis prediction and correlation analysis prediction. Statistic ways applied to the industries and firms, that the long run sales trends and history area unit consistent like food, energy, electricity, drugs and alternative basic industries. Stable and defensive corporations use moving averages, exponential smoothing and line analysis tools to predict. Correlation analysis forecast apply for the trade and firms, like building materials and business trade, of that external or internal factors have an effect on their sales, and that they believe the innovation-based industries and innovative corporations, like the knowledge trade. The factors as independent variables and sales or profits because the variable quantity, victimization single or variable multivariate analysis, to determine relationship multiple statistic regression between earnings and also the factors. The ability to forecast time series accurately is essential in a wide range of domains such as weather forecasting, electric power demand forecasting, earthquake forecasting, and financial market forecasting. Because of the fact that these time series are affected by a multitude of interrelating macroscopic and microscopic variables, the underlying models that generate these time series are nonlinear and extremely complex.

Formally, a time series, Xt, is a sequence of data points measured in equidistance time intervals, defined by:

$$X_t = \{x_t \in R : t = 1,2,3, \dots\dots, n\}$$

Where t is the time index and n is the number of samples or observations. The aim of forecasting is to provide an algorithm that allows, with a certain level of confidence, the future values of the time series given by X_t+k where $k \in \mathbb{N}^+$ represents the prediction horizon of k steps ahead. According to chaos theory, forecasting accuracy exponentially grades with increase in the prediction horizon. For instance, although we can forecast tomorrow's weather with a certain degree of accuracy, it is very difficult to forecast next year's weather with the same degree of accuracy. Therefore, long-range time series forecasting is very challenging. Choice of time lags to represent time series is also a very important process in time series forecasting [1-2]. Real world time series are generated by nonlinear dynamical systems with an astronomical number of input variables. Such systems are extremely sensitive to initial conditions.

Time series occur in various domains in great number and heterogeneity. In general, a statistic s may be represented as a sequence (x1, x2, x3 ..., xn) containing n information points xi. These information points will comprises real numbers, for instance of the river level or the voltage of an EEG derivation [8] measured at sure usually equal points in time; or additional advanced, they'll be extremely three-dimensional, e.g. the time of the transaction, a customer ID and bought items. Because of the fact that these time series are affected by a multitude of interrelating macroscopic and microscopic variables, the underlying models that generate these time series are nonlinear and extremely complex. Therefore, it is computationally infeasible to develop full-scale models with the present computing technology. Fully shaped applied mathematics models for random simulation functions, therefore on generate various versions of the statistic, representing what would possibly happen over non-specific time-periods within the future. Simple or totally shaped applied mathematics models to explain the probably outcome of the statistic within the immediate future, given data of the foremost recent outcomes (forecasting).

A. Modelling the Causal Time Series

With multiple regressions, we are able to use quite one predictor. It's continuously best, however, that's to use as few variables as predictors as necessary to urge a fairly correct forecast. The forecast takes the form:

Y = β0 + β 1X1 + β 2X2 + . . .+ β nXn, Where β zero is that the intercept, β 1, β 2, . . . β n are coefficients representing the contribution of the freelance variables X1, X2,..., Xn. Statistical management limits are calculated in an exceedingly manner kind of like different internal control limit charts, however, the residual variance are used.

B. Modelling Seasonality and Trend

Seasonality could be a pattern that repeats for every amount. As an example annual seasonal pattern features a cycle that's twelve periods long, if the periods are months, or four periods long if the periods are quarters. We'd like to urge AN estimate of the seasonal index for every month, or different periods, like quarter, week, etc, reckoning on the information handiness.

The formula for computing seasonal factors is:

Si = Di/D, Where:

Si = the seasonal index for i^{th} amount

Di = the common values of i^{th} amount

D = grand average

i = the i^{th} seasonal amount of the cycle.

C. Simulation methodology

Modelling and Simulation:

In this problem we consider a medical dataset containing observations of patients in 12 months of the continuous dependent variable Y and independent variable X. Table 1 shows the medical dataset of patients over three years and forth year predicted value through Time Series Analysis. Let y_j denote the value of the variable y for observations, j (j = 1… N) Where N is number of months in a year, and let x_i be the observed value of the independent variable x for observation of number of patients. Suppose we have constants 'a' and 'b' for an exponential function

$y_j = a\, e^{(a+b)x}$ (1)

Where number of patients 'y' depends on the exponential function of 'x' with constants 'a' and 'b'. Now taking logarithmic analysis in both sides for the calculation of the two constants.

$\log_e y = \log_e a\, e^{(a+b)x}$ (2)

$\log_e y = \log_e a + (a+b)\, x$ (3)

$\log_e y = \log_e a + (a+b)x$ (4)

Now consider Y = loge y ; A = loge a, m = a+b and X = x – M , Where M is the mean of N observations.

We get the equation; Y = A + mX (which is a linear equation of exponential values of a dependant variable X)

To calculate the values of A and m we have the following equations:

$A = \frac{\sum Y - b \sum X}{n}$ (5) and $m = \frac{n \sum X\, Y - \sum X \sum Y}{n \sum X2 - (\sum X)2}$ (6)

Now X = x – M, X = x – 6.5 (6.5 is the mean value of 12 observations of x)

$\sum X = \sum x - \sum 6.5$ We get,

$\sum X = 0$

Putting $\sum X = 0$ in above values of constants A and m we get,

$A = \frac{\sum Y}{n}$ (7) and

$m = \frac{\sum X\, Y}{\sum X2}$ (8)

For the year 2013, after putting the values of X, Y and n we get the values of the constants:

m = 30.223 and A = 562.08, Taking antilog of A we get,

a = antilog A = 278.8 and b = -212.14

For the year 2014, after putting the values of X, Y and n we get the values of the constants:

m = 15.97 and A = 543.9 Taking antilog of A we get,

a = antilog A = 228.14 and b =-212.14

For the year 2015, after putting the values of X, Y and n we get the values of the constants:

m = 20.50 and A = 606.25Taking antilog of A we get,

a = antilog A = 408.3 and b =-407.8

For the predicted year 2016, after putting the values of X, Y and n we get the values of the constants:

m = 39.2 and A = 606.25 Taking antilog of A we get,

a = antilog A = 742.4 and b = -703.2

Result Analysis

Calculation of line equation for the month 2013:

After applying the following linear regression formula of slope and constant for the line equation, we will get the value of m and b. Where n = 12 (for 12 months in a year), after applying the formula for the slope of the line and the constant b for the line equation y = mx + b, we get: m = 30.223, a = 278.8, b = -245.6, A = 562.08

Now for the plot of the graph for x = month of the year, y = variable from line equation after applying the value of slope (m) and constant (b), and y1 = number of patients in the month. We get y1 as Table 1.

Application of linear regression technique for the calculation of line equation for the month 2014:

After applying the following linear regression formula of slope and constant for the line equation, we will get the value of m and b: m = 15.97, a = 228.14, b = -212.14, A = 543.9

Where n = 12 (for 12 months in a year) Now for the plot of the graph for x = month of the year, y = variable from line equation after applying the value of slope (m) and constant (b), and y2 = number of patients in the month. We get y2 as Table 1.

Application of linear regression technique for the calculation of line equation for the month 2015:

After applying the following linear regression formula of slope and constant for the line equation, we will get the value of m and b: m = 20.50, a = 428.3, b = -407.8, A = 606.25

Where n = 12 (for 12 months in a year), Now for the plot of the graph for x = month of the year, y = variable from line equation after applying the value of slope (m) and constant (b), and y3 = number of patients in the month. Y3 as Table 1.

Application of linear regression technique for the calculation of line equation for the month 2016:

After applying the following linear regression formula of slope and constant for the line equation, we will get the value of m and b. Where n = 12 (for 12 months in a year): m = 39.2, a = 742.4, b =-703.2, A = 661. Now for the plot of the graph for x = month of the year, y = variable from line equation after applying the value of slope (m) and constant (b), and y1 = number of patients in the month. We get y4 as Table 1.**Comparison of slopes for four year:**

X	y1	y2	y3	y4
1	373.6	270.1	527.2	109.9
2	505.4	317.0	647.1	162.7
3	683.7	371.9	794.4	240.8
4	925.0	436.3	975.2	356.5
5	1251.4	511.9	1197.1	527.7
6	1693.0	600.6	1469.5	781.2
7	2290.4	704.6	1804.0	1156.4
8	3098.7	826.7	2214.5	1711.8
9	4192.2	970.0	2718.4	2534.0
10	5671.5	1138.0	3337.1	3751.1
11	7672.9	1335.2	4096.5	5552.6
12	10380.5	1566.6	5028.7	8219.0

Table 1: Comparison of slops for four years

We get the different slope for different years

Year	m
2013	30.22
2014	15.97
2015	20.50
2016	39.2

Table 2: Slope of four different years

On comparing the four different slopes of four years we got that the predicted slope of the fourth year is about the average of the three year slopes hence we can say that the fourth year prediction is the good quality prediction for the patient data. The statistical reports in the following pages shows the various reports and analysis of patient data that can be represented using figure 6.1

IV - Conclusion

Inherent in the collection of data taken over time is some form of random variation. There exist methods for reducing of cancelling the effect due to random variation. Widely used techniques are smoothing. This technique, when properly applied, reveals more clearly the underlying trends. However, the data is not properly managed. As a result of this, majority of out-patients do not have full medical record. With this situation, the physician's time is wasted since they have to collect this information again and in addition, it becomes very difficult for them to keep track of the patients. This reduces the ability to carry out high quality clinical research in the hospitals, and compromises the continuity of healthcare as well as the quality of healthcare delivery in the hospital. A Data Mart has been designed to collect, store, organize and retrieve the medical information of patients. A simple way of detecting trend in seasonal data is to take averages over a certain period. If these averages change with time we can say that there is evidence of a trend in the series.

References:

[1] Kit Yan Chan, Member, IEEE, Hak Keung Lam, Senior Member, IEEE, Tharam S. Dillon, Life Fellow, IEEE, and Sai Ho Ling, Senior Member, IEE "A Stepwise-Based Fuzzy Regression Procedure for Developing Customer reference Models in New Product Development" IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 23, NO. 5, OCTOBER 2015

[2]H. Tanaka, S. Vejima, and K. Asai, "Linear regression analysis with fuzzy model," IEEE Trans. Syst., Man, Cybern., vol. SMC-12, pp. 903–907, 1982.

[3] Celikyilmaz A. and Turksen B., Fuzzy functions with support vector machines, Information Sciences, vol. 177, pp. 5163-5177, 2007.

[4] Chen S.P. and Dang J.F. A variable spread fuzzy linear regression model with higher explanatory power and forecasting accuracy, Information Sciences, vol. 178, pp. 3973-3988, 2008. [5] Chen X.B. and Ke H., Effect of fluid properties on dispensing processes for electronic packaging, IEEE Transactions on Electronic Packaging Manufacturing, vol. 29, no. 2, pp. 75-82, 2006.

[5] Kim H.K., Yoon J.H. and Li Y., Asymptotic properties of least squares estimation with fuzzy observations, Information Sciences, vol. 178, pp. 439-451, 2008.

[6] Takagi T. and Sugeno M., Fuzzy identification of systems and its application to modeling and control, IEEE Transactions on Sys.
tems, Man and Cybernetics, vol. 15, no. 1, pp. 116-132, 1985.

[7] Tanaka H. and Watada J., Possibilistic linear systems and their application to the linear regression model, Fuzzy Sets and Systems, vol. 272, pp. 275-289, 1988.

[8] Stefan Kleinmann, Ralf Stetter, Praveen Kumar Kubendra Prasad" Optimization of a Pump Health Monitoring System using Fuzzy Logic", 2013 Conference on Control and Fault-Tolerant Systems (SysTol) October 9-11,2013. Nice, France.

[9] T. Schluter and S. Conrad, "TEMPUS: A Prototype System for Time ¨ Series Analysis and Prediction," in IADIS European Conf. on Data Mining 2010. IADIS Press, 2010, pp. 11–1.

[10] A.S. Chen, M.T. Leung and H. Daouk, "Application of Neural Networks to an Emerging Financial Market: Forecasting and Trading the Taiwan Stock Index," Computers and Operations Research 30, 2003, 901-923.

[11] IANOSI ENDRE "Considerations about efficient health care management systems", Proceedings of the 3rd International Conference on E-Health and Bioengineering - EHB 2011, 24th-26th November, 2011, Iaşi, Romania

[12] Endre Ianosi, V. Vacarescu “Dialysis apparatus Technical and quality aspects (in Romanian)”, Timisoara, Ed. Orizonturi Universitare, 2002, ISBN 973-8391-26-1.

[13] Lan Yu “Data Mining on Test Data of Physical Health Standard”, 978-1-4244-3894-5/09/$25.00 ©2009 IEEE.

[14] Lan Yu” Association Rules based Data Mining on Test Data of Physical Health Standard”, 2009 in International Joint Conference on Computational Sciences and Optimization.

[15] W. J. Frawley, G. Piatetsky-Shapiro and C. J. Matheus, "Knowledge Discovery in Databases: An Overview", in G. Piatetsky-Shapiro and W. J. Frawley (eds.), Knowledge Discovery in Database. AAAI/MIT Press, pp.127, 1991.

	Jan	Feb	Mar	April	May	June	July	Aug	Sept	Oct	Nov	Dec
2013	275	300	370	475	550	625	800	920	880	600	500	450
2014	250	325	400	515	679	720	910	700	650	532	456	390
2015	300	400	500	545	600	750	900	850	750	700	560	400
Mean	275	341.7	423.3	511.7	609.7	698.3	870	823.3	760	610.7	505.3	413.3
Index	0.48	0.60	0.74	0.90	1.07	1.22	1.53	1.44	1.33	1.07	0.89	0.70
Expected 2016	145	240	371	489	642	919	1373	1227	1000	750	496	281

Table 3: Number of patients in different months, for three years and forecast of fourth year, using linear regression.

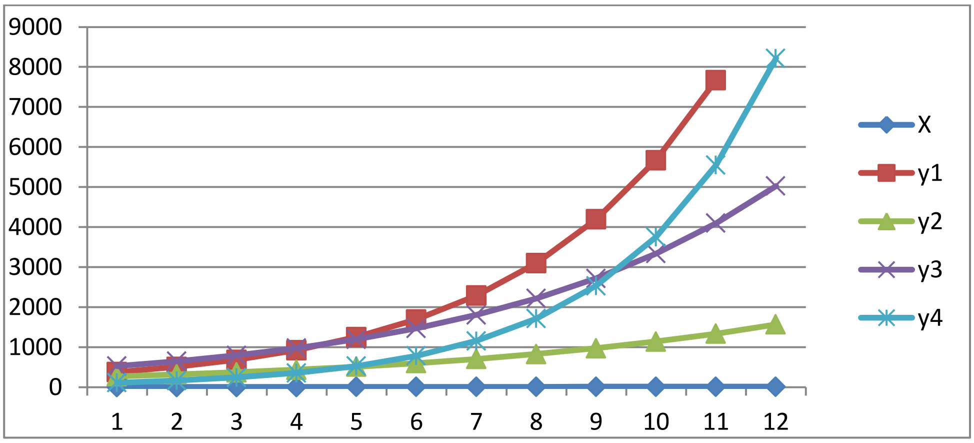

Figure 1: Comparison of slops for four years