A Neuro- Fuzzy Approach for Formulating Survey and Managing recorded Information in Data Warehouses

Rajdev Tiwari
ABES Institute of Technology
Campus-II, Vijay Nagar
Ghaziabad, U.P, (INDIA)

Manu Pratap Singh
Institute of Computer and Information Science
Dr. B.R.Ambedkar University, Khandari,
Agra U.P.,(INDIA)

ABSTRACT
The major activity of Business Intelligence (BI) is to dig out the various trends and patterns from variety of authentic sources that helps the managers to take appropriate decision in framing the policies for business plans accordingly. Surveys are considered to be an essential part of BI. Surveys conducted amongst different or same groups by different team may yield conflicting reports. Moreover the recorded answers during the surveys may even contain a lot of vagueness in it.
This paper suggests and implements a neuro-fuzzy approach for processing and storing the vague information captured during the surveys. This approach shall help the personnel involve in BI to get the more appropriate analysis based on human like reasoning, out of the Data Warehouse (DW) as compared to the DWs based on the crisp values only.

General Terms: Data warehouse, Questionnaire Design, Fuzzy Information.

Keywords: Neuro-Fuzzy, Fuzzy Information, Fuzzy sources, Data Warehouse, Questionnaire, Business Intelligence, Surveys.

1. INTRODUCTION AND RELATED WORK
Data warehouses and data marts were conceived as a means to facilitate the compilation of regular reports on the status of the business by continuously collecting, cleaning, and summarizing the core data of the organization. Data warehouses provide a clean and organized source of data for data mining.
Surveys are perhaps the most common and expensive source of data because they capture information by direct interaction with customers. Surveys collect data through different communication channels with customers, such as mail, email, interviews, questionnaires, and forms on websites. There are many anecdotes about the accuracy and validity of the data collected from the different forms of surveys. However, they all share the following two common features. First, the number of customers who participate in the survey is usually limited because of the cost and the number of customers willing to participate. Secondly, the questions asked in the survey can be designed to directly address the objective of the planned model. For example, if the objective is to market new products, the survey would ask customers about their preferences in these products, whether they would buy them, and what price they would pay for them. More detailed description is found in [1]. Human Response to a question with multiple alternatives can be imprecise or fuzzy for various reasons. Part of this fuzziness can be attributed to the lack of sincerity and clarity in the thought process of the respondent. A formal measure of this aspect of fuzziness is formulated by [2].
In most countries, there are commercially available databases that contain detailed information on consumers of different products and services. The most common type is demographic databases based on a national census, where the general demographic profile of each residential area is surveyed and summarized. It is believed that data obtained from such data warehouses is usually clean and information-rich.
The quality and authenticity of the surveys strongly depends upon the manner in which it is conducted. In fact followings are a few of the points to be taken care:

- Whether the survey has recorded answers correctly without any ambiguity or vagueness.
- Was it conducted amongst different groups by the same team?
- Whether the questionnaire was designed keeping open-ended answers and closed-ended answers both in view?
- Was the questionnaire same for different area/ group?
- Whether the questionnaire was designed keeping two-valued answers and multi-valued answers both in view?

1.1 Questionnaire Design
A questionnaire is a data collection method that a respondent fills in written format. Questionnaire surveys are an important part of marketing and customer relationship management. The use of questionnaires is even popular in schools to collect students' opinions of teaching performance as mentioned in [3] where a technique for mining fuzzy association rules from questionnaire data is discussed in detail.
According to [4], the types of questions in the questionnaires can be roughly classified into two categories, open-ended and closed-ended questions. Accordingly, the answers to these two types of questions constitute open questionnaire data and closed questionnaire data, respectively.
To collect open questionnaire data, structured questions must be supplemented with open-ended questions to get respondents to answer the problems in their own words. In [5] a responsive design of paradata (process data) is defined to guide mid-survey decisions affecting the non-response, measurement and sampling variance properties of resulting statistics of a survey. In [6], a

concept based questionnaire system is presented to improve the search quality of web-based search engines.

Apart from asking users to answer or fill out any types of questionnaires, [7] presented an initial design for assessing the user's product-specific knowledge using the proposed innovative method for detecting it directly in a non-intrusive way. Their method is based on customer's on-line navigation behaviors by analyzing their navigation patterns through pre-trained artificial neural networks. In [8] it is proposed to have a fuzzy based method to identify rules to get the subjective requirements of the wheel chair users by evaluating some questionnaire data.

For instance, let us discuss the questionnaire used by NSSO [9] for determination of 'Access to Improved Drinking Water' in India. The two variables used by them are:

Var1: 'Whether the availability of water is sufficient throughout the year?'

Var2: 'Facility of drinking water.'

The answers recorded for both the questions are of 'yes' or 'no' type. More details can be found in [10]. With these kinds of responses it is determined whether a particular household has adequate facility of drinking water or not. Measuring adequacy by means of 'yes' or 'no' types of answers only does not seem to be appropriate. In fact more appropriate answers that can be expected from the people in this case are 'more or less sufficient', 'all most sufficient', 'nearly sufficient', absolutely sufficient' and may more. In general answers recorded during surveys may be full of vagueness. It is essential in that case not to avoid such vague answers neither restricting them to two-valued only(i.e. 'yes' and 'no' types) rather capturing them in its original form to the best possible extent. Even questions framed for surveys meant for desert areas might not be equally relevant for other areas.

1.2 Fuzzy logic

Fuzzy logic has already proven its worth to be used as tool to deal with real life problems that are full of ambiguity, imprecision and vagueness.

In written sources, we can find a large number of papers dealing with Fuzzy logic theory, which was first introduced by Lotfi A. Zadeh in 1965 [11]. According to [12], Fuzzy sets can provide solutions to a broad range of problems of control, pattern classification, reasoning, planning, and computer vision. Their work bridges the gap between theory and practice. An advanced introduction to the problems of fuzzy modeling and to fuzzy control, one of its most important aspects is covered in [13]. He covered the topics including applications in control, diagnostics, decision making, optimization, pattern recognition and robotics. [14] covers the basics leading to: fuzzy clustering, fuzzy pattern recognition, fuzzy database, fuzzy image processing, soft computing, fuzzy applications in operations research, fuzzy decision making, fuzzy rule based systems, fuzzy systems modeling, fuzzy mathematic. A firm mathematical basis for the calculus of fuzzy concepts necessary for designing intelligent systems and a solid background for readers to pursue further studies and real-world applications is provided in [15]. The original interpretation of fuzzy sets arises from a generalization of the classic concepts of a subset extended to embrace the description of "vague" and "imprecise" notions. This generalization is made considering that the membership of an element to a set becomes a "fuzzy" or "vague" concept. In the case of some elements, it may not be clear if they belong to a set or not. Then, their "membership degree" of the element to the set, and it takes a value in the interval [0,1] by agreement. Using classic logic, it is only possible to deal with information that is totally true or totally false; it is not possible to handle information inherent to a problem that is imprecise or incomplete, but this type of information contains data that would allow a better solution to the problem. In classic logic, '0' represents membership of an element to a set if it does not belong to and by '1' if it does, having a set {0, 1}. On the other hand, in fuzzy logic, this set is extended to the interval [0, 1]. Therefore, it could be said that fuzzy logic is an extension of the classic systems. Fuzzy logic is the logic behind approximate reasoning instead of exact reasoning. Its importance lies in the fact that many types of human reasoning, particularly the reasoning based on common sense, are by nature approximate. Note the great potential that the use of membership degrees represents by allowing something qualitative (fuzzy) to be expressed quantitatively by means of the membership degree.

Since surveys are one of the costliest tools to collect data, sometimes individual organizations restrict themselves to focus on limited geographical areas or on small groups for collecting the data. However, if we device a mechanism which is capable of giving predictions about the responses on the basis of some observations and past experiences without actually collecting them, it shall help the whole BI community to mine and analyze the market trends and so.

1.3 Artificial Neural Networks

In fact Artificial Neural Networks (ANN) has become a popular tool for classification and predictions. The field of neural networks was originally kindled by psychologists and neurobiologists who sought to develop and test computational analogues of neurons. Roughly speaking, a neural network is a set of connected input/output units in which each connection has a weight associated with it. During the learning phase, the network learns by adjusting the weights so as to be able to predict the correct class label of the input tuples. A bank loans officer needs analysis of his data in order to learn which loan applicants are "safe" and which are "risky" for the bank, A medical researcher wants to analyze breast cancer data in order to predict which one of three specific treatments a patient should receive. In each of these examples, the data analysis task is classification, where a model or classifier is constructed to predict categorical labels, such as "safe" or "risky" for the loan application data; "treatment A," "treatment B," or "treatment C" for the medical data respectively. These categories can be represented by discrete values also like 1,2,3,4 or 5.

Advantages of neural networks, however, include their high tolerance of noisy data as well as their ability to classify patterns on which they have not been trained. They can be used when you may have little knowledge of the relationships between attributes and classes. They are well-suited for continuous-valued inputs and outputs. They have been successful on a wide array of real-world data, including handwritten character recognition, pathology and laboratory medicine, and training a computer to pronounce English text. Neural network algorithms are inherently parallel; parallelization techniques can be used to speed up the computation process. In addition, several techniques have recently been developed for the extraction of rules from trained neural networks. These factors contribute toward the usefulness of neural networks for classification and prediction. [16], proposed a neural-network-based fuzzy system (NNFS) which is a self-organizing neural-network that can partition the input spaces in a flexible way, based on the distribution of the training data in order to reduce the number of rules without any loss of modeling accuracy. A new approach of neuro-fuzzy learning algorithm for tuning fuzzy rules by using training input–output data, based on the gradient descent

method is developed in [17]. Three approaches of implementing and constructing the fuzzy neural networks, neural networks based fuzzy logic, and an improved genetic algorithm, which is a special neural networks learning algorithm are discussed in [18]. Issues related to design of Neural networks, neuro-fuzzy systems and their learning capabilities with their applications have been the prime target of researchers working in the field of Artificial Intelligence.

This paper proposes and simulates a Fuzzy-Neuro approach for formulating questionnaires, conducting surveys and finally preparing reports to populate the Data Warehouse. This paper exploits the capability of fuzzy-neuro (i.e. the integration of fuzzy logic with Neural Networks) system to capture the more realistic data from the questionnaire and to train the systems with patterns identified, for the purpose of creation of summaries and reports. The paper is organized as follows;

Firstly, it describes the proposed approach which includes design of questionnaire and types of responses expected, fuzzification and defuzzication of compiled questionnaire data, use of ANN for classification and prediction of overall responses and preparing reports from the analyzed data. Secondly, it describes the simulation of the proposed method with all its relevant details. The paper concludes with the conclusion section followed by future directions and references used.

2. PROPOSED METHOD

2.1 Questionnaire Design for Surveys

The way in which questionnaire is laid out is very important to its success as an instrument to capture accurate data from the survey. Ian Brace [19] has thoroughly covered the issues laying out for interviewer-administered questionnaires, for paper self-completion questionnaires and for online questionnaires in his book on questionnaire design.

Let us take the example of demographic surveys of a country. Housing conditions of the people seems to be a good indicator of how developed a country is. Improving the living conditions of people through better provision of access to basic necessities such as drinking water, sanitation and adequate living space is one of the important developmental goals. Like NSSO in India, other countries also conduct demographic surveys for different areas, periodically. The content and coverage of data collected in these surveys has only changed marginally over the years and the basic concepts, definitions and procedures have remained the same. The sample survey considered and reformulated in this paper is based on the indicator variables usually taken by NSSO India for the assessment of housing conditions. Indicator variables used here are durability of housing structure, over-crowding, access to various facilities like drinking water, sanitation, electricity etc.

In the proposed formulation of surveys, surveyors are now supposed to record the ranked answers on a 5-point scale instead of recording ‘yes’ or ‘no’ types of answers specially for those questions where the answers may not be given in ‘yes’ or ‘no’ only.

For instance to assess the “Adequacy of drinking water” etc., respondents are asked to give the ranking on a 5-point scale. The questionnaire format for data collection shall now look like as shown in the table 1.

Table 1. Sample of questionnaire

S.No	Indicator Variables	1	2	3	4	5
1	Adequacy of access to drinking water				√	
2	Adequacy of sanitation facilities					√
3	Adequacy of power supply				√	
4	Durability of dwellings			√		
5	Overcrowded		√			
6	…					

Collected data is then compiled on area basis and is represented as fuzzy sets, details of which are covered in the forthcoming section. As far as other design specifications for surveys are concerned these may be any one of the standards available suitable for the conditions. Various specifications, concepts, definitions and guidelines for the design and conduction of surveys are covered in [20].

2.2 Fuzzyfication and defuzzification of compiled data

Collected data will not convey the values of the indicator variables in terms of ‘yes or ‘no’ rather, it will give a rating about the adequacy of either of the amenities. Taking the sample size of survey as 1000 and collecting the responses according to their rank, the complied data will now look like as shown in table 2.

Table 2. Compiled data

S.No	Indicator Variables	1	2	3	4	5
1	Adequacy of access to drinking water	90	110	140	260	400
2	Adequacy of sanitation facilities	100	150	230	250	270
3	Adequacy of power supply	300	230	190	150	130
4	Durability of dwellings	110	160	220	240	290
5	Overcrowded	350	270	160	120	100
6	…	…	…	…	…	…

Fuzzification of the compiled data shall now be done to find out degree of belongingness of individual response in terms of ranks to the fuzzy sets corresponding to every indicator variables. To represent the compiled data as fuzzy sets some identified fuzzy sets and their denotations are: Adequate access to drinking water $\left(\widetilde{A}\right)$, Adequate sanitation facilities $\left(\widetilde{B}\right)$, Adequate power supply $\left(\widetilde{C}\right)$, Durable dwellings $\left(\widetilde{D}\right)$, Overcrowded $\left(\widetilde{E}\right)$. To find out the memberships of individual ranks to the fuzzy sets, Gamma function is chosen for the purpose, which is shown in figure 1.

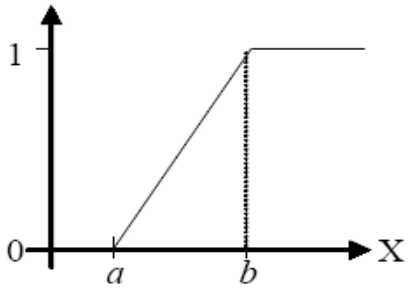

Figure1. Gamma Function (membership function)

So the whole data collected during survey is now represented by foresaid fuzzy sets as follows:

$$\widetilde{A} = \left\{\frac{0.16}{1} + \frac{0.2}{2} + \frac{0.27}{3} + \frac{0.51}{4} + \frac{0.8}{5}\right\}$$

$$\widetilde{B} = \left\{\frac{0.18}{1} + \frac{0.29}{2} + \frac{0.45}{3} + \frac{0.49}{4} + \frac{0.53}{5}\right\}$$

$$\widetilde{C} = \left\{\frac{0.59}{1} + \frac{0.45}{2} + \frac{0.37}{3} + \frac{0.29}{4} + \frac{0.24}{5}\right\}$$

$$\widetilde{D} = \left\{\frac{0.2}{1} + \frac{0.31}{2} + \frac{0.43}{3} + \frac{0.47}{4} + \frac{0.57}{5}\right\}$$

$$\widetilde{E} = \left\{\frac{0.69}{1} + \frac{0.53}{2} + \frac{0.31}{3} + \frac{0.22}{4} + \frac{0.18}{5}\right\}$$

Defuzzification is need because reports are supposed to contain crisp values ultimately. Weighted-Average method of defuzzification is preferred over others for the data considered for simulation in this paper. Weighting of each membership function of fuzzy set by its height is done in this method values of qualifying elements, as shown in the figure 2, are computed by $\frac{x1.h1 + x2.h2}{x1 + x2}$.

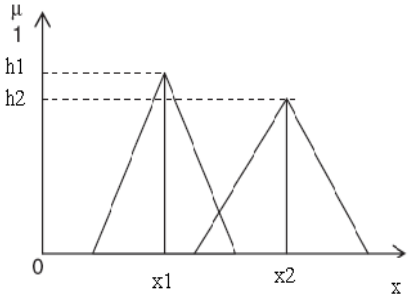

Figure 2. Weighted-Average method of defuzzification

2.3 Use of ANN for classification and prediction

Defuzzified values corresponding to every indicator variable and their corresponding cumulative out puts are given to the neural networks to train them. Once the ANN has been trained it shall now be used for predicting cumulative responses based on a number of indicator variables. Before training the neurons it is required to design an appropriate topology for ANN and choose the appropriate learning algorithm for training them.

In the proposed method Multilayer feed forward neural network (as shown in figure 3) trained by back propagation algorithm is used.

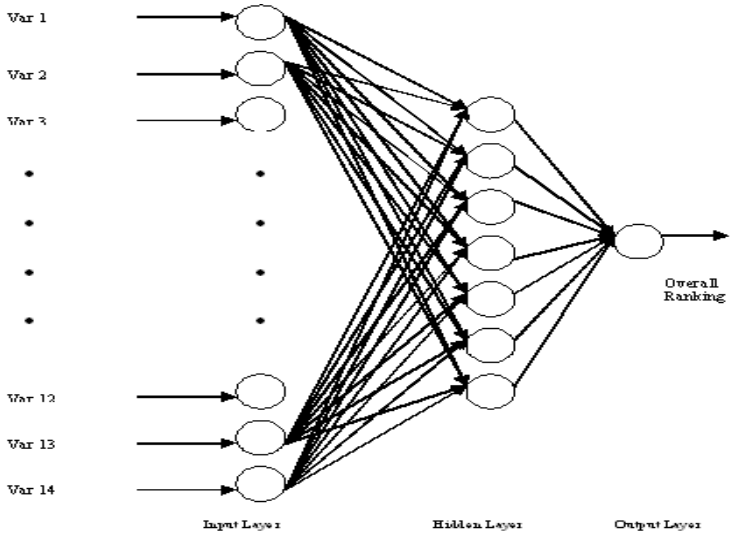

Figure 3. ANN Architecture

2.4 Report preparation

A sample report in terms of percentage is shown in the table 3. Similar reports can be made on the basis of rankings also shall be discussed in the simulation section. These reports are now used to populate the Data Warehouse. Sample report shown in the following table is one of the instances in demographic data warehouse created and maintained for the assessment of living conditions in India.

However, in the simulation of proposed approach the real data is taken from the questionnaire of feedback system of an academic institution.

Table 3. Sample Report on adequacy(%) of drinking water

	Rural		Urban	
State	**Drinking water facility**			
	Inadequate	**Adequate**	**Inadequate**	**Adequate**
Jammu &Kashmir	13.5	39.2	2.7	81.6
Himachal Pradesh	14.2	29.4	1.6	83.1
Punjab	2.9	81.1	0.3	90.1
Chandigarh	--	65.7	1.3	91.6
Uttaranchal	9.3	43.4	0	90
Haryana	10	39.5	3.3	79.8
Delhi	10.6	32.3	6.8	73
Rajasthan	11.7	21.9	5.3	66.9
Uttar Pradesh	1.5	48.4	0.9	78
Bihar	2	51.2	2.9	75.3
Sikkim	6.4	47.6	0.3	85.8
Arunachal	6.9	33.9	0.7	85.6
Nagaland	39.8	10.9	25.5	38.6
Manipur	14	10.9	17	28.8

3. SIMULATION DETAILS AND RESULT DISCUSSION

Data used for simulation is taken from the questionnaire of an academic institution. Institutions use the questionnaire for taking feedbacks regarding teaching qualities of individual teacher for all the subjects taught by him/her in every semester. In this questionnaire there are 14 variables/ parameters. Students are suppose rate the individual teacher for all 14 variable on 5-point scale. The variables are as follows;

Var1: Course Planning
Var2: Complete coverage of syllabus
Var3: Pace of coverage
Var4: Speech audible & clear
Var5: Use of simple language
Var6: Variety of teaching aids / methods
Var7: Use of practical examples
Var8: Ability to control the class
Var9: Encouragement to class participation
Var10: Punctuality in class
Var11: Lectures held regularly
Var12: Approachability of faculty
Var13: Fairness of evaluation
Var14: Response to student's questions

Students are also suppose to give the over all rating again on 5-point scale which can be taken as final assessment. Linguistic grades in 5-point scale are; Outstanding, Excellent, very good, good and satisfactory, obviously in decreasing order from 5 to 1.

Simulation is done on 50 samples of 20 students each resulting in 1000 responses containing 14 variables each. Responses of two such samples are shown in the table 4 and their rank-wise compilation is shown in table 5. It is worth mentioning that tables shown here contain data corresponding to two samples only. Responses corresponding to individual questions are fuzzified using Gamma function and mapped to 5 fuzzy sets per indicator variable as shown in table 6.

Table 4: Responses of questionnaire

Response No./ Variable	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11	Var12	Var13	Var14	Overall Assessment
R1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
R2	4	4	4	5	5	5	4	5	5	5	5	4	5	5	4
R3	5	4	4	5	4	4	5	4	5	5	5	4	4	5	5
R4	3	4	2	5	5	5	3	2	2	2	2	5	3	2	4
R5	3	4	4	5	5	5	4	4	5	5	4	4	4	5	5
R6	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
R7	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5
R8	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5
R9	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5
R10	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
R11	1	2	2	2	2	2	1	2	2	4	3	1	4	4	2
R12	4	4	4	5	5	4	3	4	3	4	3	4	5	4	3
R13	4	3	4	5	5	5	5	4	5	5	5	4	5	5	5
R14	5	5	5	5	5	4	4	5	5	5	5	5	5	5	5
R15	2	3	4	4	4	5	5	4	4	4	4	4	4	4	4
R16	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
R17	4	3	3	4	5	4	4	4	5	5	5	5	5	4	5
R18	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5
R19	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
R20	5	5	4	5	5	4	4	5	5	4	4	4	4	5	5
R21	2	2	2	2	2	2	1	2	2	3	3	2	2	2	2
R22	4	3	4	4	5	4	5	4	5	5	5	5	4	4	4
R23	4	4	4	4	4	4	3	4	3	4	4	3	4	4	4
R24	3	4	3	5	5	3	2	4	3	4	3	5	5	5	5
R25	5	5	5	5	5	4	4	5	4	4	4	5	5	5	5
R26	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
R27	1	3	2	3	3	3	2	4	4	3	3	3	3	4	4
R28	4	4	4	4	1	4	4	4	4	4	5	4	4	4	4
R29	3	3	4	4	3	3	4	4	4	4	4	3	4	3	4
R30	2	2	3	3	1	4	3	1	1	2	3	2	3	1	1
R31	3	3	3	4	4	3	3	4	4	4	4	5	5	5	5
R32	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
R33	4	4	4	3	4	4	3	4	5	4	5	5	5	5	5
R34	2	3	2	5	5	3	1	3	5	5	5	5	5	5	4
R35	2	2	2	2	2	1	1	2	2	2	4	4	3	2	2
R36	2	2	2	1	2	1	1	1	1	2	2	3	3	1	1
R37	2	2	2	2	2	2	1	2	2	2	3	2	2	2	2
R38	3	3	4	4	3	5	4	5	4	4	4	4	4	4	4
R39	2	2	4	4	4	2	2	4	2	4	4	4	2	4	3
R40	4	4	4	5	5	4	4	5	4	5	4	5	5	4	4

Table 5. Compiled Data for above two samples

Rank														
						Sample 1								
5	10	6	8	15	15	12	10	10	14	13	12	10	11	13
4	4	8	7	2	2	5	5	6	1	4	3	7	6	4
3	2	3	1	0	0	0	2	0	1	0	2	0	1	0
2	3	1	4	1	3	1	2	2	4	1	3	0	2	1
1	1	2	0	2	0	2	1	2	0	2	0	3	0	2
						Sample 2								
5	3	3	3	6	7	3	3	5	5	5	6	9	8	7
4	5	5	8	7	4	7	5	9	7	9	8	4	5	7
3	4	6	3	3	3	5	4	1	2	2	5	4	4	1
2	7	6	6	3	4	3	3	3	4	4	1	3	3	3
1	1	0	0	1	2	2	5	2	2	0	0	0	0	2

Table 6. Fuzzified data

						Sample 1									
Outstanding	0.5	0.1	0.3	1	1	0.7	0.5	0.5	0.9	0.8	0.7	0.5	0.6	0.8	0.8
Excellent	0	0.3	0.2	0	0	0	0	0.1	0	0	0	0.2	0.1	0	0
Very good	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Good	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Satisfactory	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
						Sample 2									
Outstanding	0	0	0	0.1	0.2	0	0	0	0	0	0.1	0.4	0.3	0.2	0.1
Excellent	0	0	0.3	0.2	0	0.2	0	0.4	0.2	0.4	0.3	0	0	0.2	0.3
Very good	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0
Good	0.2	0.1	0.1	0	0	0	0	0	0	0	0	0	0	0	0
Satisfactory	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Defuzzified cumulative responses are then computed using weighted- average method and are normalized using normc() function of MATLAB 7.0 before giving them as input to the neural network. The cumulative over all responses are given as target output to ANN and then it is trained by using Backpropogation algorithm in MATLAB. Selection of MATLAB for simulation is purely an arbitrary choice as per our convenience and its availability. Parameters for ANN used are listed in the table 7.

Prediction capability of the trained ANN is then tested for 25 samples, comparison of predicted output and actual output is shown in the figure 4. Finally reports are prepared for individual teacher for all the subjects taught by him/her, a sample of report prepared is shown in table 8.

Table 7. ANN Specifications

SNo.	Parameters	Value
1	Number of neurons in input layer	14
2	Number of neurons in hidden layer	7
3	Number of neurons in output layer	1
4	Number of Epochs	100
5	Learning rate	.001

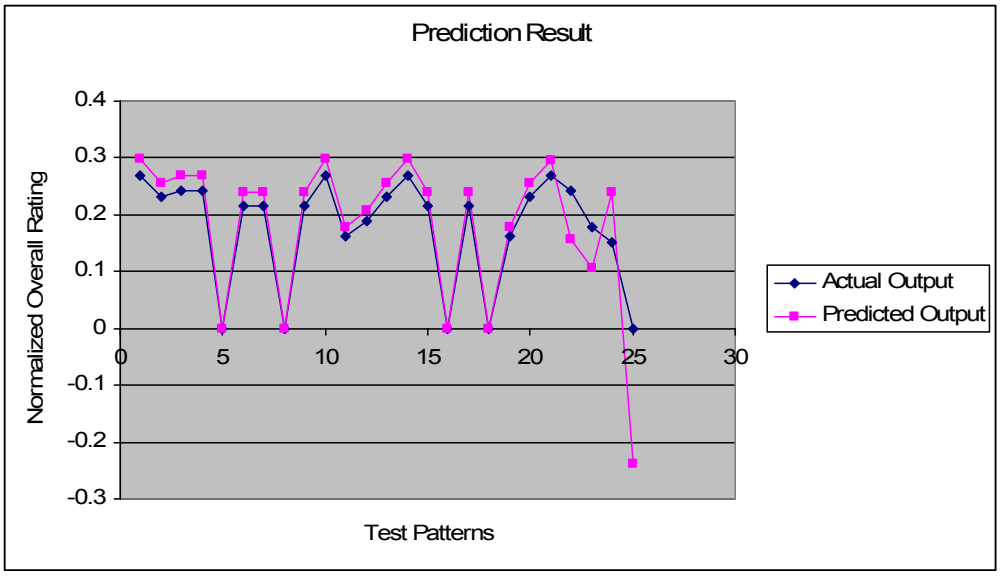

Figure 4. Prediction Result

Table 8. A Report on teaching quality

Academic Information System			
Academic Feedback (2010-11 Odd Semester)			
Teacher: Mr. X			
	Parameters/Subjects	Sub 1	Sub 2
	Teaching		
1	Course Planning	4.37	2.27
2	Complete coverage of syllabus	4.36	2.75
3	Pace of coverage	4.5.0	2.12
4	Speech audible & clear	4.50	2.25
5	Use of simple language	4.00	2.25
6	Variety of teaching aids / methods	4.50	2.00
7	Use of practical examples	4.38	2.30
	Study environment		
1	Ability to control the class	4.00	2.12
2	Encouragement to class participation	4.50	2.50
3	Punctuality in class	4.50	2.38
4	Lectures held regularly	4.00	2.12
	Genaral Professional Attitude		
1	Approachablity of faculty	4.00	2.25
2	Fairness of evaluation	4.50	2.62
3	Response to students' questions	4.08	2.40
	Overall assesment		
1	Overall assesment	4.00	2.33

4. CONCLUSION

The simulation of the proposed method itself advocates about the goodness of it. This method for formulating the surveys and creating the reports to populate the Data Warehouses by capturing fuzzy information have following advantages over the traditional methods:

- It does neither ignore the vagueness in the answers nor forcefully make them crisp.

- Conflicting or ambiguous reports submitted buy surveyors does not raise any issue.
- Capable of generating the reports even with partially available inputs.
- Alleviate us from looking for every piece of data used in analysis.
- Reports generated are more realistic.
- Reduces the cost of conducting surveys.

The only limitation of this approach that could be observed is a lot of mathematical computation involved in fuzzification and defuzzification process. But as compared to the traditional statistical method it is not be a big deal.

5. FUTURE DIRECTIONS

Further improvement in this approach may be achieved by identifying fuzzy association rules from the questionnaire data and training the ANN using such rules. In cases, where we have very large set of variables, feature selection can also be applied before training and prediction of ANN. Details of feature section is found in [21].

6. REFERENCES

[1] Mamdouh Refaat (2007). Data Preparation for Data Mining Using SAS", Morgan Kaufmann Publishers , pp.44-47

[2] Das, Shubhabrata(2006). On measuring imprecision in human response due to respondent and attribute and its utility in questionnaire design. In International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems, v 14, n 2, p 155-173.

[3] Chen, Yen-Liang ;Weng, Cheng-Hsiung (2009). Mining fuzzy association rules from questionnaire data. In Knowledge-Based Systems, v 22, n 1, p 46-56.

[4] Marshall, G. (2005). The purpose, design and administration of a questionnaire for data collection, Radiography 11 (2) 131–136.

[5] Groves, Robert M. & Heeringa, Steven G.(2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. J.R.Statist.Soc.A, 169(3), 1-19

[6] Nikravesh, Masoud (2008). Concept-based search and questionnaire systems. In Soft Computing, v 12, n 3, p 301-314, Special issue on BISCSE 2005 'Forging the Frontiers' Part II.

[7] Chang, S.E., Changchien, S.W., Huang, R.H. (2006). Assessing users' product-specific knowledge for personalization in electronic commerce, Expert Systems with Applications 30 (4) 682–693

[8] Lafuente, Ruben ; Page, Alvaro; Sanchez-Lacuesta, Javier; Tortosa, Lourdes (1998). Application of fuzzy logic techniques for the qualitative interpretation of preferences in a collective questionnaire for users of wheelchairs. In Journal of Rehabilitation Research and Development, v 35, n 1, p 91-107.

[9] NSSO (2001). Concepts and Definitions used by NSSO, Golden Jubilee Publication .

[10] Mohanan, P. C. & Sourav Chakrabortty (2008). Inter-State Comparisons of Housing conditions – A study based on NSS 58th round. In Journal of National Sample Survey Organization, 94 Issue, Vol. XXVIII No. 3&4.

[11] Zadeh, L.A. (1965) Fuzzy sets. In Information and control, 8, pp. 338-353.

[12] Pedrycz, W., & Gomide, F. (1998). An introduction to fuzzy sets: Analysis and design (A Bradford Book). The MIT Press.

[13] Piegat, A. (2001). Fuzzy modeling and control. In Physica-Verlag (Studies in Fuzziness and Soft Computing).

[14] Buckley, J.J., & Eslami, E.(2002). An Introduction to fuzzy logic and fuzzy sets (advances in soft computing). Physica-Verlang Heidelberg.

[15] Nguyen, H.T., & Walker, E.A.(2005). A first course in fuzzy logic (3rd ed.). Chapman & Hall/CRC.

[16] Wang, Y. & Rong, G. (1999). A self-organizing neural-network-based fuzzy system. In Fuzzy Sets and Systems.

[17] Shi, Y. & Mizumoto, M. (2000). A new approach of neuro-fuzzy learning algorithm for tuning fuzzy rules. In Fuzzy Sets and Systems.

[18] Yang, Y., Xu, X. & Zang, W. (2000). Design neural networks based fuzzy logic. In Fuzzy Sets and Systems.

[19] Ian Brace (2004). Questionnaire Design,. In Kogan Page,.

[20] Groves, Robert, M.; Benson, Grant; Mosher, William; Rosenbaum, Jennifer; Granda, Peter; Axinn, William; Lepkowski, James; Chandra, Anjani.(2005) "Plan and Operation of Cycle 6 of the National Survey of Family Growth". Hyattsville, MD: National Center for Health Statistics, Vital Health Statistics, 1(42).

[21] Rajdev Tiwari & Manu Pratap Singh (2010). Correlation-based Attribute Selection using Genetic Algorithm. In International Journal of Computer Applications (0975 – 8887) Volume 4– No.8