Hybrid Recommendation System Based Upon Network

Major Project submitted in partial fulfillment of the

Requirements for the award of degree of

Master of Technology

In

Information Systems

Submitted By:

Ankit Kanojia

(2K12/ISY/05)

Under the Guidance of:

Prof. Dr. O.P. Verma

(IT Department)

Department of Information Technology

Delhi Technological University

Delhi-110042

(2012-2014)

CERTIFICATE

This is to certify that **Mr. Ankit Kanojia (2K12/ISY/05)** has carried out the major project titled **"Hybrid Recommendation System Based Upon Network"** as a partial requirement for the award of Master of Technology degree in Information Systems by Delhi Technological University.

The major project is a bonafide piece of work carried out and completed under my supervision and guidance during the academic session **2012-2014**.

The matter contained in this report has not been submitted elsewhere for the award of any other degree.

(Project Guide)

Prof. Dr. O.P. Verma

Head of Department

Department of Information Technology

Delhi Technological University

Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

I express my gratitude to my major project guide **Prof. Dr. O.P. Verma, HOD, IT Dept.** Delhi Technological University, for the valuable support and guidance he provided in making this project. It is my pleasure to record my sincere thanks to my respected guide for his constructive criticism and insight without which the project would not have shaped as it has.

I humbly extend my words of gratitude to other faculty members of this department for providing their valuable help and time whenever it was required. I am thankful to all teaching or non-teaching staff of DTU and my fellows, who have helped me in the completion of this thesis report.

Ankit Kanojia

Roll No 2K12/ISY/05

M.Tech (Information Systems)

Department of IT

E-mail: ankitkanojia89@gmail.com

List of Figure

Table of content

Chapter 1

Introduction to recommender system

1.1 Preview

In today world large number of people uses internet. People always connected to internet through their gadgets (smartphone, pads and laptops). Due to advancement in communication technology, internet available at low cost which lead people always online or connected to internet. So activity such as online shopping, searching, social interaction etc. increase over internet with exponential rate. As a result more online data is available about users. E-commerce industry uses these available data to suggest their items to customers. They develops recommendation system. These recommendation systems increase their sells. Also provide new platform to customer for comparing and refer similar items. Now days thousands of products available in market. Most of people confused between products, which products best or most appropriate for them. Recommendation systems also helps them to compare products based upon similarity and dissimilarity. Enable customer to select best appropriate product according to their requirement. Recommendation system also able to handle ambiguous query posted by user and varied user profile. Therefore, help in reducing information load by considered only selected items based upon input query. They targeted to deliver specific information to the specific user. Recommendation process basically based on classification and clustering methods. They are usually hidden from end user, delivered them high quality information which enhance user satisfaction and performance of system.

1.2 Definition of Recommendation System

Recommendation systems goal to predict level of likelihood of user towards certain item, with aim to suggest items, which may they like from set of items they have not considered so far[1].

1.3 User Outline

Recommenders can be seen as extension of customer relationship management (CRM) systems. They exploit all the available information of the customer by referring customer history through cookies, by integrating customer social site profile e.g. Facebook and Twitter etc. and through registration form data. They also track user insight behavior. Based upon all available information about the customer, they provide personalized insight to the customer. In the sense recommenders act as a personalized tool. Recommenders heart maintains user profile which includes demographic information belongs the customer (e.g. location, family status, job, age, and sex) and customer behavior tracking(both online and in physical world, through expense tracking, loyalty program etc.).In many cases partial profile is available which can be integrated through customer social site profile and customer cookies. So, richer the profile, the more precise and accurate recommendation will be.

While calculating result recommenders not only exploit current user profile but also exploit profile of other users(similar or dissimilar) as group or clusters of user. The extension of the contribution of each factor depends upon the recommendation approach adopted.

1.4 Types of Recommendation Systems

Recommendation Techniques can be classified in following main categories:

- Content-based recommendation technique: exploits information and features of the items, with the aim of recommending to user a set of items that have some similarity or relationship with items they brought or examined in detail during the current or past online activity.
- Collaborating Filtering: exploits similarities between users (i.e. commonalities between user profile) for suggesting items. It is based upon the assumption that people with the same opinion on an item are more likely to have same opinion on the other items too. Therefore, it is possible to predict the taste of a user by collecting preferences from other similar users.
- Knowledge based recommendation: considered facts to make recommendation about the user requests and liking.

A hybrid approach can be constructed which is combination of at least two above mention approaches in order to eliminate the problems of an individual approach used separately e.g. we can combine content based and collaborative filtering algorithm in such a way that they produce individual list of recommendations that can be merge to make final recommendation list from which we can recommend top elements . Experiments show they actually slightly improve performance of both[1].

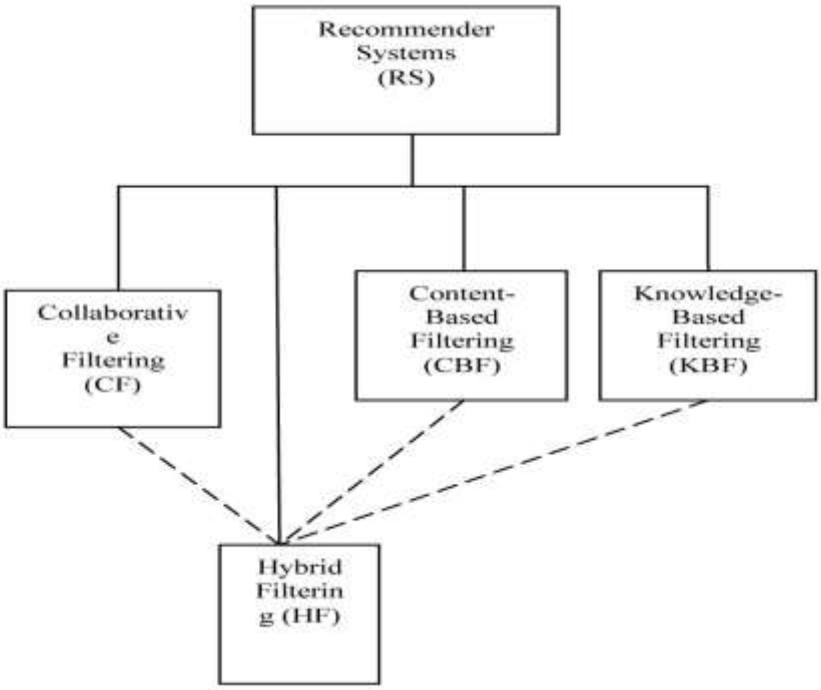

Fig 1.1: Classification of Recommender Systems

1.4.1 Content- Based Recommendation Techniques

The main content based technique include following for building recommenders:

- Classifiers: can be used to build recommendation solutions, as the build classes of items, which can then be recommend to user. They can also use to classify items basis on search query and suggest items related search query . They can also use collectively with other recommendation approach.
- Category-Based Recommendation: This technique assumes that each item belongs to more than one category. Once a customer picks an item, the system determine possible categories of interested domain based on the past customer activities over items(or over categories), and then select the top items from considered categories, which may can recommended. This approach requires that the customer activity be tracked and that the recommendation be somehow personalized.
- Search-Based Recommendation: Starting from a search query and the corresponding result set, search-based recommendation selects and highlights some of these results based on general, non-personalized ranking, e.g. on the number of sales, number of views, number of views. This method is very simple and only requires some basic

statistics knowledge on the items; however, it can hardly be considered a recommendation at all, as items are just retrieved in a specific order.

- Semantics-Based Recommendation: this extension of category-based recommendation because it switches categories of items with complex domain model for describing the semantics of the items (e.g., ontologies, vocabularies, and any other kind of define conceptual model) and matches these descriptions with semantic model of customer.
- Information Filtering: This technique exploit syntactical knowledge on the item and/or semantic information about their kinds or categories; the available information span from unstructured, organized in typed attributes. When a customer declares interest in an item, the system also recommends those item that are more alike to it, where similarity calculated in a way that depends on the type of information. This solution does not need user tracking, but it suffers from two main problems:

  1.) The recommendation only depends on the description of item and therefore will not change over the time to reflect the user's change of interest.

  2.) It does not work when the items are not well described.

1.4.1.1 Advantage

- Recommend new, rare, or unpopular items because the matching and prediction are based on item descriptions and not on statistics of usage or preferences.
- Able to recommend item to users with unique tastes, again because no large statistics on preferences are needed.
- They are extremely interested because they are able to provide explanations of why items where recommended by considering content features of item that makes an item to be recommended.
- They do not suffer with typical problems of collaboration techniques, such as cold start, sparsity and first-rated problems.

1.4.1.2 Disadvantage

- They required to convert content into meaningful and measurable features; and they need to represent users' tastes as a learnable function of these content.
- Unable to exploit quality judgments of other users(unless they are explicitly included in the content feature).

1.4.1.3 Metrics to measure similarity/ dissimilarity in content-based techniques

In content-based technique, we construct feature vector or attribute profile of item. The feature vector holds numerical or nominal values that represents definite feature of the item like price, size, color etc. A range of metrics available to compute similarity/ dissimilarity between items. Following metrics can be used:

- Euclidean Distance: is used to evaluate the distance between two points in multidimensional space, which can be any kind of distance we can evaluate through ruler. Considered the point in area as (p_1, p_2, p_3, p_4,........) and (q_1,q_2,q_3,q_4,…..), then formula for Euclidean distance defined below:

$$\sqrt{(p_1-q_1)^2+(p_2-q_2)^2+\dots+(p_n-q_n)^2} = \sqrt{\sum_{i=1}^{n}(p_i-q_i)^2}$$

Euclidean Distance

- Pearson Correlation: is used to evaluate how two variables are highly correlated. It predict values between 1 and -1, where 1 means that variables are highly correlated, 0 denotes no correlation exist, and -1 means they are highly inversely correlated.

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{N}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{N}\right)}}$$

Pearson Correlation

- Cosine Similarity: It is used to evaluate similarity between two vectors of an inner product that evaluate cosine angle between them. The cosine 0 degree represent 1, and it has value less than 1 for any other angle. It is just represent alignment and not magnitude: two vectors with the same alignment have a cosine similarity of 1, two vectors at right angle(90 degree) alignment has a similarity of zero, and two vector diametrically aligned opposed have similarity of -1, independent of their magnitude[2].

Given two vector of features, A and B, the cosine similarity, is represented using a dot product and magnitude as :

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

Cosine Similarity

- Tanimoto Coefficient: It is used evaluate similarity between two groups. It can also be used to evaluate likeness between two items based on lists of features e.g. if we considered two group of alphabets , G_1 and G_2, where:
  G_1= { A, B, C, B,D}
  G_2={ A, A, B}
  Then the overlapping(joint) group, which let G_T, is { A , B}. The Tanimoto coefficient is defined below, where N_a is number of elements in G_1, N_b is number of elements in G_2, and N_c is number of elements in G_T, the overlapping. In considered case the Tanimoto coefficient : 2/(5+3-2)=2/5=0.33

$$T = \frac{N_c}{N_a+N_b-N_c}$$

Tanimoto Coefficient

- Gini Impurity: It is used to evaluate how impure a set or group is. If we considered a group of elements, define as [a,a,a,a,b,b,c,c,c], then Gini impurity evaluate the probability that we would be incorrect if we selected one element and randomly predict its class. If the set contains only one type of element e.g. a's, we would always predict a and never be incorrect, so the set considered as completely pure.

$$I_G(i) = 1 - \sum_{j=1}^{m} f(i,j)^2 = \sum_{j \neq k} f(i,j) f(i,k)$$

Gini Impurity

- Entropy: It is taken from part of the information theory and it is used evaluate the amount of disorder in a set or group. Informally defined, entropy represent how surprising a randomly chosen element from group is. If the complete group contain only one element e.g. E, then entropy would be 0. We would never surprise on getting element E. The formula define below:

$$H(X) = \sum_{i=1}^{n} p(x_i) \log_2 \left(\frac{1}{p(x_i)} \right) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i)$$

Entropy

1.4.2 Collaborating Filtering Technique

A collaborating filtering technique usually work by exploring a large set of human population and find a small subset of humans which have similar taste or attitude. In this technique user rating database, of many items as possible, need to be maintained. The methods find those users whose rating strongly correlate with the current user, and then recommend items which are highly rated by similar users.

Considered an example:

Customers vs. **Items:**	**Item1**	**Item2**	**Item3**	**Item4**	**Item5**	**Item6**
Customer 1		X	X			
Customer 2	X			X	X	
Customer 3	X				X	X
Customer 4		X	X	X		
Customer 5	X				X	X
Customer 6		X	X			X
Current customer	X				X	

Table1.1 Purchase Matrix: customers and items they purchased in the past

In the above example, where an X represents the fact user brought an item. To recommend an item to current user; we find similar customers. In this example user 2,3,5 are similar (they all brought Item1 and Item5). Therefore , the system will suggest other items they brought. In particular, it will strongly suggest Item4(purchased by two customer) and also Item6(purchased by two customer).

In general , the more a customer is similar to the current one, the higher is his weight in the prediction and As the number of similar customer that brought a specific item increases, rank of the item also incremented in the recommendation.

1.4.2.1 Advantage

- Able to delivered very relevant recommendations.
- Extremely powerful and effective approach.

1.4.2.2 Disadvantage

- Need to maintain huge database, which requires lot of resource.
- Time consuming approach: database is dynamic in nature need to change with time.
- Cold Start: is a problem which occur at the starting of the system lifespan. When the machine is switched on, it doesn't contain any record of user interaction or rating. This make impossible for system to evaluate any recommendation.
- First Rater: When new items comes in market domain that have never purchased then it would never be recommended by above technique.
- New User: when new customers come, who never purchased anything cannot receive any recommendation suggestion because they cannot be linked to other customers.
- Data Sparsity: In general, the purchase table(as well as any rating table or click table) is very sparsely occupied: the typical situation is that, upon a warehouse which contain tens millions of products, each customer may have purchased only few items at most, may have visited only few times, and may have explicitly rated or commented on just one or two. This make an effective implementation more difficult.

To address above problem, clustering techniques can be applied over users, thus aggregating customers that have similar behavior and hence labeling and keeping the common behavior through a representative set of select actions. This tremendously reduce the size of the customer set that considered for analysis.

1.4.2.3 Metrics for Collaborating filtering technique

- Weighted Mean: can be define as type of average that has a weight for every performance result being averaged. It can be used in numerical predictions based on similarity scores(similarity score can be calculated through above content based metrics, on the user.). Let x_1, x_2, ….,x_n are the performance result and w_1, w_2,…,w_n are the weights. The weighted mean define as:

  .

$$\bar{x} = \frac{w_1x_1 + w_2x_2 + \ldots + w_nx_n}{w_1 + w_2 + \ldots + w_n}$$

Weighted Mean

- Variance: It is used evaluate divergence of list of number varies from the mean(average) value. Informally, it is used to evaluate how large the divergence of every number from the mean, formula shown below:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

Variance

- Clustering technique : defined in next section.

1.4.3 Knowledge Based Recommendation System

Knowledge based approaches are comprise factual knowledge, they contain knowledge about how a particular item can satisfied a certain need. Therefore, it reason about relationship between user requirements and possible recommendation or suggestions. The user outline can be used for construct of knowledge structure that supports this inference.(Fuzzy inference approach can be used.)[3]

1.5 Challenges Of Recommendation Techniques

- Data sparsity
- Cold Start

- Scalability: When the number of existing users and items population grows tremendously, the traditional system undergoes from fatal scalability problems, leads acute decline in performance which goes beyond acceptable level.
- Synonyms: It occurs when number of the indistinguishable or very similar items have a many name and recommender systems unable to reveal this hidden association, then considered these products differently.
- Grey Sheep and Black Sheep: Problem occurs when a customer whose views are constantly unable to correlate in agreement or disagreement with any certain group of people. Thus customer unable to take advantage of Recommender System and known as Grey Sheep. This problem also acutely increase error rate in collaborative filtering recommender system. Customer who correlate with very limited people or none can be considered as Black sheep. This scenario makes challenging situation for recommendation system.
- Fraud: Now days recommender Systems are gradually adopted by e-commercial websites. Recommender System provide financial benefits to the vendors and service provider. Unethical competition between vendors have started them to involve in different kinds of fraud in order to con the system to get their advantage. They have tried to inflate the perceived attractiveness of their own commodities(push attacks) or reduce the ratings of their competitors (nuke attacks). These attacks are also known as shilling attacks[4].

1.6 Problem Statement

In the present research work, efforts are made to proposed an efficient approach to develop a recommender system, which can fulfill above mention challenges. It essential to reduce space and time complexity of recommendation system. Also need to focus on accuracy as well as dynamic behavior of customers. Different techniques of classification and clustering can be adopted to fulfill recommendation system criteria. In current approach, we use MLP- Neural Network which act as a classifier, takes attribute of book(publisher and title) and user demographic information(age, sex, location) as input and classify books according to rating(1 to 5), where rating 1 and 2 represent rejected book, rating 3 represent average and rating 3,4 represent accepted book for current user. By defining rating threshold, books above rating threshold can be recommended to user.

Thus, the problem taken for this research work is " Hybrid Book Recommendation System Based Upon (neural) network".

1.7 Performance Measure For Recommender Systems

Recommender Systems performance can be measure by comparing predicted recommendation rating with a known user rating. These systems are commonly measured using predictive accuracy metrics, where the predicted ratings are directly compared to actually user ratings. The commonly used metrics are Mean Absolute Error(MAE) and Root Mean Error(RME) described below[4]:

$$MAE = \frac{\sum | P_{u,i} - R_{u,i} |}{N}$$

$$RME = \sqrt{\frac{\left(P_{u,i} - R_{u,i}\right)^2}{N}}$$

Fig 1.2 Evaluation Metrics

Where, $P_{u,i}$ is the predicted rating for u on item i,

$R_{u,i}$ is the actual rating

N is the total number of ratings in the test set

Predictive accuracy metrics treat all items equally.

1.8 Literature Review

In today world millions of people are connected to internet, they are heavily dependent on it for day to day activities such as shopping, information access, referring surveys, education and many more. Over internet enormous amount of data available, to perform such activities over internet user need to search exhaustively, even after exhaustive search some time user unable to achieve their activity goals. Here recommender systems play significant role, which recommend information(or may be item) to every user personally by referring user details, behavior and input query. In other words, we can say that recommender system can handle problem of information overload. The basic function of recommender system was to provide recommendation to the user based on items or information related to their interest and in some cases, to provide guesses or rating to each item which the user may prefer[5]

In real life scenario whenever we plan to perform any new task e.g. to buy a new car, selecting of insurance policy, admission to college etc. We always go for recommendation initially from friends, relatives or teachers etc. So, recommendation can be define as one of the primary activity, which help us in resolving problem or current activity task. Now, days this recommendation task is automate with help of technological research. Basically recommender technique includes information categorization, information clustering, information summary and information filtering. They also makes user to enter keywords or query that characterize their interest. In many cases recommender system acts similar to normal academic search engine, which retrieve information based on input query. Therefore, we can say both search engine and recommender system are highly related. As a result, any advanced research in search engine field, can also be applied to recommender systems. Further related fields are: citation recommender for patents[6], educational recommender systems[7], expert search[8], academic alerting service[9], venue recommendation[10].

C. Lee Giles et al. presented the first research paper on recommender system as part of their CiteSeer Project. CiteSeer understand how to parse citations, identify the context of citations in the body of articles[11]. In this project similarity is measured between different research paper through following mechanisms such as word vector, LikeIt and citation similarity. To measure similarity weighting scheme is TF-IDF.

For comprehensive review we categorize recommender systems in following groups according to their implementation and techniques:

1.) Stereotyping

It is one of the oldest technique for recommender systems. It was proposed by Elaine Rich in the recommendation system known as "Grundy", A system is described that bunds models of its user, with the aid of stereotypes, and then exploits those models to guide it in its tasks, suggesting novels that people may find interesting.[12]. In this author defined facets, which are based on a stereotype as collection of psychology characteristics e.g. Grundy assume female user are independent, romantic, perseverance, caring, loving , negative interest in Piety, liberal. Based on this assumption it recommend books, which are manually classified to suit the facets. Rich recommendation suffers from problem by considering all female have negative interest in piety but in real scenario exception occur. Another problem by which Stereotyping suffer from is a labour-intensive task and allow only limited personalization[13]. But some researcher of Stereotypes argues that once it is created the it required little computation and performs quite well. Some recommender systems customized search according to gender[14].

2.) Content based filtering

It is predominant recommendation approach. Features of items are used in this approach, which are typically words. The majority of researcher used plain words. Jiang et at used subject, which utilized only those words and combinations of words that occurred as social tags on CiteULike[15].

Some recommender systems non textual features i.e. writing style, layout information, XMLtags[16]. Author Giles et al. used citations[5].

In this not only features of items are considered but user information also considered equally, even feedback from user given important. Author Mohammand Hamidi Esfahani et al. considered both item features and user's properties for recommendation, they uses both features for construction of cluster using c-Means clustering methods, which used form extract weighted Map for recommendation purpose[17]. Author also used metadata for recommendation purpose.

Most of author used machine learning techniques for content based filtering. Most popular used technique is SVM(Support Vector Machine). Author Jianying Mai et al. used Neural Network, machine learning algorithm for evaluate pre-classification results[18]. Some researcher also applied stemming and removed stop words.

Some researcher considered fact that influence of certain field have more dominance than other filed e.g. A word occurs in the abstract of paper is more dominance over word present in the body text. Nascimento et al. considered this fact, and weighted terms from the title three times stronger than terms from the body - text, and text from the abstract twice as strong[19].

3.) Collaborative Filtering

The term " Collaborative Filtering " was first coined by Doug Terry, David Goldberg et al. in 1992, they were driven by idea that " information filtering can be more effective when human are involved in the filtering Process[20][21]. In this approach author designed a system known as Tapestry that allowed people to mark document as interested or uninterested, then used this information to filter documents for other people. Now days this recommender technique is frequently applied.

Paul Resnick et al., after two year proposed that similar minded people like similar items. Two user considered as similar minded, when they like or positive rated same item. CF also offers following advantages such as Collaborative offers content independent, count on connection i.e. rating, human provide explicitly rating which provide real time estimation of item, collaborative filtering support serendipitous recommendations because it is relies on user similarities.

Yang et al. proposed explicit rating approach allowed user to rate their research paper, but user were too " too lazy to provide rating"[15]. To overcome this problem Yang et al. inferred implicit ratings from the number of pages user read[15].

4.) Item Centric

In this technique, affinity between items is calculated and those item which are related to the item purchased by user or have connection with the user are recommended. To evaluate affinity between items, correlation based upon occurrence of items is determined. Higher correlation indicated higher affinity. Basic in this technique association rules are developed. E-commerce website adopt this technique e.g. amazon, which said "customer who brought this item also brought…", amazon analyses , which items are frequently brought together.

5.) Graph Based

In this technique graph is generated based upon fact that object are integrally connected. Basically many research form graph based upon citation for recommendation purpose[22][23][24]. Sometime research considered parameters such as author, user/customer, venues and the publication of the year.

Once the graph matrix, graph metrics can be used to select best candidates for recommendation.

6.) Global Relevance

In this items which have highest global relevance are recommend, relevance is usually considered by popularity of item over internet e.g. a book always recommended whose author highest average rating. Popular rating metrics are PageRank, Hits.

7.) Hybrid Recommender System

This technique represent hybrid approach in which previously defined approach are combined to increase accuracy. It take advantage of both technique and reduce their individual demerit. It is novel way enhance precision of the system. Mohammand Hamidi Esfahani et al. proposed hybrid approach using c-means clustering[17].

Chapter 2

Classification and Clustering Techniques

2.1 Preview

In recommender systems information overhead cost can be address through machine learning techniques that organize and categorize large amount of data. The two major methods are Classification and Clustering. Classification is a supervised technique that allocate a grade two each data point by performing an initial training phase over set of human marked data and then a subsequent testing phase which applies the classification of remaining data elements whereas Clustering is an unsupervised technique that does not require any priori information: data are grouped into classes on the basis of some similarity measure metric, which compute similarity between instances, in such a way that object belongs to one cluster are very alike(compactness property) and object belongs to different clusters are unlike (separateness property). Similarity can be calculated in different ways mentioned in previous chapter, based upon data norms namely numeric, mixed data, categorical, or textual.

2.2 Supervised learning Vs. Unsupervised learning

- Supervised learning assumes the availability of a training set of correctly identified observations. Classification problems are often modelled as supervised learning problems, with techniques such as naïve Bayes, regression, decision trees, support vector machine and neural network.
- Unsupervised learning does not require any priori information but involves grouping data into classes based upon evaluate characteristic similarity between instances. An example of unsupervised learning is clustering, which can be performed through partition (e.g. k-means) or hierarchical approach.

2.3 Classification

Classification is the problem of assigning an object (item or observation) to one or more categories (subpopulations). The individual items or observations are characterized by some quantifiable properties, called features; these can be categorical, ordinal, or numerical. An algorithm that implements classification is known as a classifier. Some algorithms work only on discrete data, while other also works on continuous values (e.g. real number).

Classification is based on a training set of data containing observations for which the category is known a priori, as provided by a human analyzer; therefore classification is a supervised learning technique.

2.3.1 Bayesian Classifier

It is also known as Probabilistic Classifier, which is established on the Bayes theorem of strong (naïve) independence assumption P (Y| X) =P (X|Y) P(Y)/ P(X). In the classification context, the probability to be estimated is the probability of an object belonging to a class, given a number n of its features.

The naïve Bayes classifier can be defined by combining the naïve Bayes probability model with a decision rule. The classification function simply assigns the elements with feature value f_1, f_2, $f_{3, \ldots\ldots}$, f_n to the most probable class:

$$\text{classify}(f_1, \ldots, f_n) = \text{argmax}\, P(C = c) \prod_{i=1}^{n} P(F_i = f_i | C = c)$$

Bayesian Classifier

2.3.1.1 Strengths

- It can be trained fast and can be queried with large dataset. In these classification of items just refer to a mathematical manipulation of the probabilities of these item attributes. So, less training data is needed.
- Support incremental training- in this every single new member of training dataset can be used to upgrade the probabilities without using any of the old training dataset. This support of incremental training is very crucial for dynamic application, which require constantly training on new items or data point that come in, has to be updated rapidly.
- In naïve Bayesian, it is relative simple to interpret and understood what the classifier actually learned. Because probability associated with each attribute is stored, we can simply observe at database at any time and check which attribute are best in division. It can possibly to use with another applications or used as a beginning point for another applications.

2.3.1.2 Weakness

- Naïve Bayesian unable to deal with consequences that varies on combination attributes or inability to learn interaction between features.
- If a category and attribute value never occur together in the training set, the corresponding probability estimate will be zero.

2.3.2 Regression Classifiers

- Linear regression classifier model: In this we interested in a random variable Y i.e. associated to a number of independent variables x_1, x_2,…,x_n. The aim is to construct a prediction equation that expresses Y as a function of these independent variables. Then, we can measure the independent variables, substitute these values into prediction equation and obtain the prediction for Y.

$$Y= \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_{n-1} x_{n-1} + \beta_n x_n + \varepsilon$$

The values of β_i and ε are estimated based on observed data

In case when random variable Y and independent variables are not linearly interrelated then logistic and polynomial regression can be used.

- Logistic regression: In this random variable Y and the independent variables are not linear but logistic; this suitable for binomially or multinomially distributed data i.e. categorical data which can assume two (binomial case) or more (multinomial case) possible values. Logistic regression is based on the logistic function shown below, which has the useful property of taking in input any value from real number and producing in output values between 0 and 1.

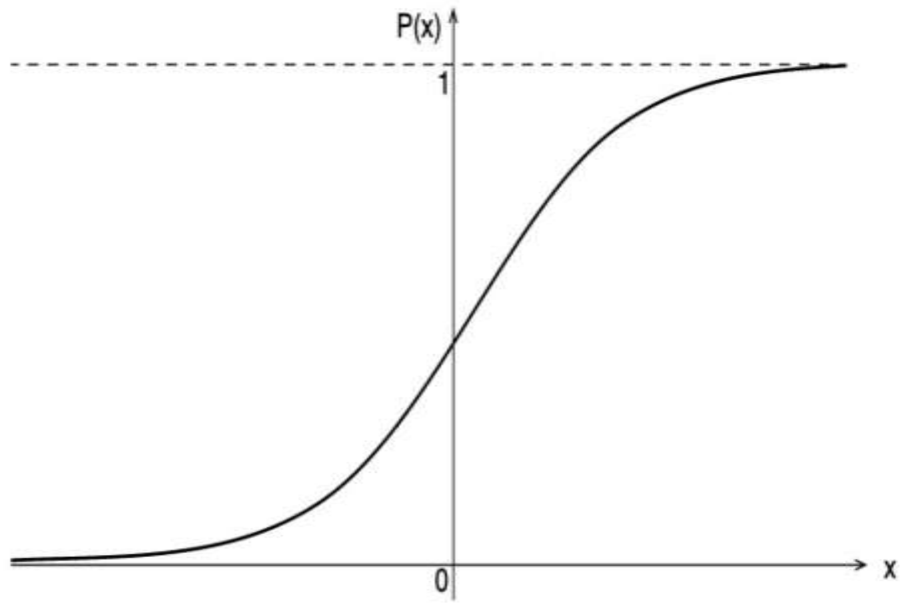

Fig 2.1 Logistic function representation

2.3.2.1 Strength

- Logistic regression has a simple probabilistic interpretation and allows models to be updated to reflect new data easily (unlike e.g., decision trees or support vector machine).
- Logistic regression approaches are also known as maximum entropy (MaxEnt) techniques, because they are based on the principle that the probability distribution.
- Logistic regression can be applied to problem like predicting the presence or absence of a disease given the characteristics of patients, or predicting the outcomes of political elections.
- Regression model able to learn from all the previous significance judgments including judgments form various queries.

2.3.2.2 Weakness

- Regression models basically based on heuristic features in the initial place; it required lots of experimental computation in order to determine a set of useful features.

2.3.3 Decision tree Classifier

A decision tree can be considered as flow-chart like tree structure, where each inner node represents assessment on an feature, each branch of tree denotes an result of the assessment, and leaf nodes denotes division or class distribution. They can be used for classification by creating a predictive model that forecasts the value of a targeted variable based upon various input variables (attribute).

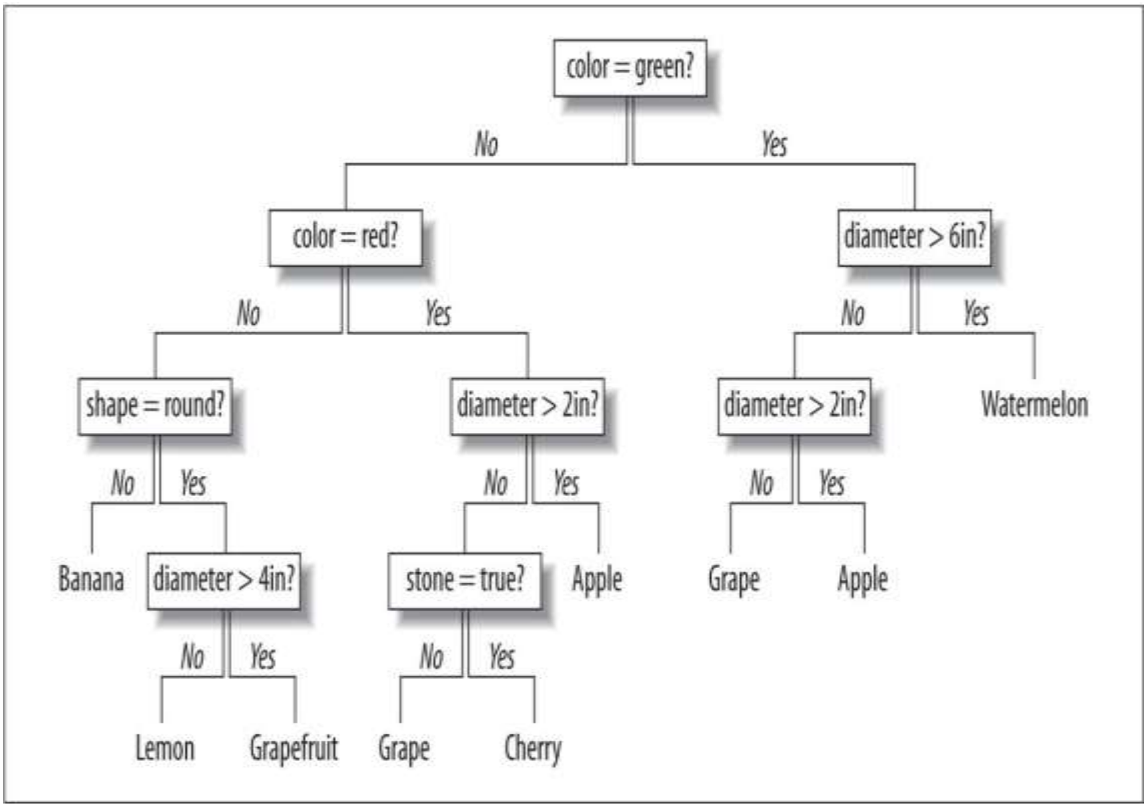

Fig. 2.2 Decision Tree example

From above diagram, we can clearly understand how a decision tree tackle task of classification of new item. Classification task begin from the topmost node of the tree, it evaluate the item against the node's condition-if condition meets with item then it follows the Yes branch of tree; otherwise it follow the No branch of tree. This process is continue until an terminate point is reached, which predict category of new item.

A classic decision tree algorithm, which learns a decision tree from a group of labeled data samples, which exploit the concept of information entropy by evaluating uncertainty of a random variable: the basic idea is that at each decision point, the action maximizing the information again (and therefore minimizing the entropy) associated with the value of the random variable should be taken. In a decision problem, this means associating each decision point with the most discriminative “question “available. Therefore, the concept of entropy measure how much effective a division is:

- P(i) = frequency(outcome) = count(outcome) / count(total rows)
- Entropy = sum of p(i) * log(p(i)) for all outcomes

A lower value of entropy with in a group indicates that the group is typically identical items(homogenous), and a value 0 indicated that it contain only one type of elements. The entropy for each group is used to evaluate the information gain, which is defined as:

- Weight1 = size of subset1/ size of original set
- Weight2 = size of subset2/ size of original set
- Gain = entropy(original) – weight * entropy($group_1$) – weight * entropy($group_2$)

So for each feasible distribution, the information gain is evaluated and used to determine the distribution variable. Once the distribution variable has been select, the first node of the decision tree can be created.

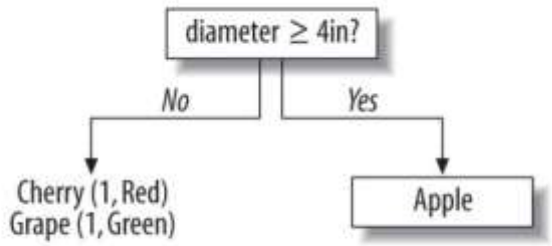

Fig2.3 Fruit decision tree root node

From above figure, we can understand distribution criteria, the data which not able to qualify the criteria is gone down to the No branch of tree whereas data that qualify the criteria gone down to Yes branch of tree. Since in above case Yes branch has just only one possible outcome, it becomes a terminating point. The No branch still contain a mixture, so it can be further divide by using exactly same method that was used to select top node. In this considered case, color is the best variable for division of data. This process continue until there is no information gain for division of data on a given branch

2.3.3.1 Strength

- Decision trees are good at handling feature interaction and are insensitive to outliers and to linear separability i.e. when data points belonging to different classes can be distinguished by mean of a linear function.
- In decision tree important factors present at top of the tree, which makes decision tree easy to interpret.
- Decision tree can also use when numeric data provided as inputs, since by locating the division link that increases information gain. The ability of combining categorical and numerical data is beneficial in case of many different divisions problems-somewhat that traditional statistical method regression has trouble in doing.

2.3.3.2 Weakness

- They do not support online learning, meaning that tree must be rebuilt if new training samples come in or does not support incremental training.
- They easily over fit the training data i.e. they adhere too much to the specificities of training data and do not perform well when applied to previously unseen data. This problem can be solved by adopting evolutionary ensemble methods, like random forests or boosted trees, which increase further complexity.
- Decision tree are not suitable for making prediction for problem which have numeric solutions.

2.3.4 Support Vector Machine

The goal of support vector machine is very simple: given a set of instances belonging to two classes (e.g. POS and NEG) represented as vectors in a d-dimensional space, find an optimal hyperplane separating two classes. The term " vector" here refers to the fact that finding such a hyperplane implies finding the instances of each class that minimize distance from hyperplane from both "sides" of the hyperplane itself; these are called support vector.

Basically an SVM construct a predictive model by finding the division line between two groups e.g. considered a graph in which we plot a set of value for height versus speed and the best position for each person, we get a graph like one shown below. Front-court players are shown as Xs and back-courts are shown in O's. Graph represents few lines that separate the data into two categories[25].

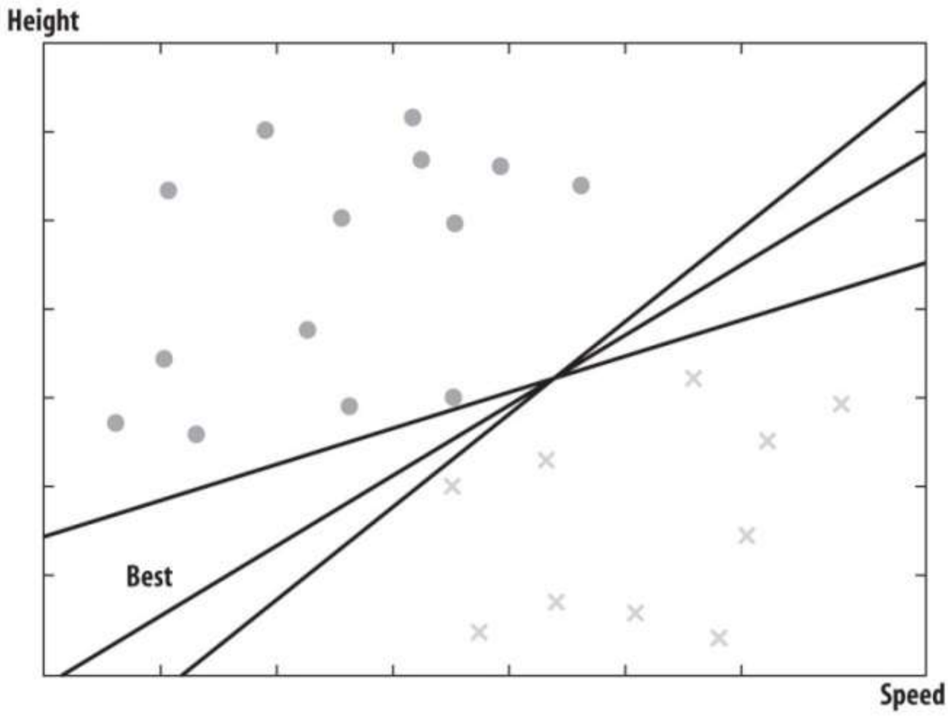

Fig2.4 Plot of basketball and dividing lines

A support vector machine locate the line that split the data appropriately. The line name as "best" in above graph represent the best support vector, which able categories data perfectly. After locating the dividing line, classification of new items is just refer as plotting them on the graph and check on which side of the line they lie. In this technique doesn't require to go back to training data for classification of new points once the dividing line has been found, so classification process is very quick.

Sometimes data is inseparable in current dimension space by linear support vector machine for such case we can took advantage of kernel trick, which transform current data into multi dimension space by applying different function to axis variables. By transformation we can able classify the data elements. This is also known as polynomial transformation, it able transform data along different axis.

2.3.4.2 Strength

- Due to convergence properties, SVMs are best technique for supervised classification.
- SVM are very rapid in classification of new observations by simply refer a point lie on which side of dividing line, training also very fast.
- By converting unqualified(categorical) inputs to numerical, we can make SVM work with the combination of various data.

2.3.4.3 Weakness

- The parameter and best kernel transformation function will be different for every problem(dataset) , need to find in every case. Iterative(or looping) approach with possible data values helps improve current problem, but it require to have a large database to do cross-validation.
- SVM suitable to problems in which we have huge amount of data available whereas other technique such as decision tree can still produce valuable information with the small dataset.
- SVMs behaves as black box technique, it is very challenging to understood how SVM performing classification task because of transformation of data into multi dimension space.

2.3.5 Neural Network Classifier

Define in next section

2.4 Clustering

Clustering is the technique in which we categories data into clusters or groups so items within a cluster are highly similar(compactness Property) in comparison to one another, but very different to objects in other clusters(separateness property). Evaluation criteria in clustering basically refer to measure similarity and dissimilarity between clusters. Clustering is an unsupervised technique because it doesn't require any prior knowledge and training data for objects. They also not perform prediction activity. In clustering, it is not always required to know how many cluster are needed and what parameter play major role forming clusters in advance.

Clustering process can be defined as follow:

Fig 2.5 Steps and outputs of the clustering process

1. Data Processing: conversion of raw data to standard data i.e. data matrix through a set of high representation and discriminant features.

2. Similarity Function Selection: Defines (dis)similarity evaluation criteria i.e. how data set objects must be compared. Similarity function and Dataset plays crucial role in overall quality of clustering process. (some similarity Function is defined in section 1.4.1.3)

3. Cluster analysis: is the process of categories a set of physical or abstract items into classes of similar objects using clustering algorithm and chosen similarity function.

4. Cluster Validation: refers to process in which we evaluate produce clusters. Validity indexes used evaluate quality of cluster.

Clustering can be state as optimization problem in which aim is to locate data objects or items into cluster best possible way. Two main categories of clustering can be define which are:

- Hard Clustering: In this clustering technique each item or data point just belongs to one cluster.
- Soft Clustering(Fuzzy Clustering): In this clustering technique each data point belongs to different categories, with a certain degree of membership that represents the prospect of belonging to that cluster.

2.4.1 Some Major Clustering Methods

In this section we define only some clustering technique, for further details and more technique reader must refer to bibliographic[26][27].

Selection of particular clustering algorithm refers to which type of data is available and purpose of clustering and type of application. If clustering technique used for exploitation of data then different clustering techniques can be apply for analysis of data.

1.) Hierarchical clustering methods

It receive data objects, creates hierarchical decomposition of that data objects.

Based upon hierarchical decomposition process hierarchical decomposition can be further divides as follow :

- Agglomerative Approach
- Divisive

Agglomerative Approach: also known as bottom up approach, in this initially each object represent distinct group. It successively merge groups which are similar or close to one another, until all of the groups combined into one(the top most level of hierarchy), or until termination condition occurs.

Divisive Approach: also known as top-down approach, considered all data objects in one cluster initially, it is highly iterative process, in every single iteration cluster break up into smaller cluster, until each cluster contain highly similar data objects or termination condition occurs.

Strength

- ➢ Simplest way to form clusters of data objects
- ➢ Easier to understand
- ➢ Because of iterative approach, same codes execute again and again, which makes quite simple to implement.
- ➢ Computation cost is quite low.
- ➢ Create trees of items for analysis.

Weakness

- ➢ In this after merge or split step, we can’t undone that step. In other words, we can’t regain previous stage.
- ➢ Unable to correct invalid decisions.

Following approaches can be used to improve the quality of hierarchical clustering :

- Perform careful analysis of objects internal interconnectivity at each hierarchical partitioning.
- Integrate hierarchical agglomeration and iterative relocation by first using hierarchical agglomerative and then refining the result using iterative relocation.

2.) Partitioning Clustering Methods

Considered a dataset of n objects, a partitioning clustering method creates k clusters of the data. It actually separate the data into distinct clusters . In this technique before running clustering algorithm, we need to define how many clusters needed. It classifies data into k groups, where each group need to satisfy following requirements:

- At least one object must be present in each group.
- In this each object must be associated to exactly one group.

By initially providing k, denoted number of partition for creation, a partition method creates an primary partitioning. It then uses an iterative relocation approach that to improve the partitioning or clustering by moving objects from one group to another.[26]

Popular heuristic model for partitioning clustering methods:

- K-means: In this clustering algorithm each cluster denoted by its mean value.
- K- medoids: In this clustering algorithm each cluster is denoted by one of the objects located at near the centre of the cluster.

3.) Fuzzy c-mean clustering

In this method one data object is belongs to different clusters through the degree of membership which represent likelihood to different clusters. The FCM algorithm is used for grouping or clustering of data x_1, x_2, ….,x_n into c cluster, the result in the form of fuzzy membership u_{ij} and cluster centroid c_j . FCM has goal to minimize the objective function shown below:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2 \quad , \quad 1 \le m < \infty$$

Objective Function

Where , m represent real number greater than 1. $U_{i,j}$ is the degree of membership of x_i in the cluster j, x_i is the i^{th} of d-dimensional measured data, c_j is the d-dimensional center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center[28].

FCM clustering is evaluate membership $u_{i,j}$ and cluster center c_j through an iterative process which repeatedly update both values by using below define equation and has aim to optimize objective function define above.

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}}, \quad c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

This iteration process will terminate when $\max_{ij}\{\|u_{ij}^{(k+1)} - u_{ij}^{k}|\} < e$, where' e' represent as stopping criteria between 0 and 1, whereas k considered as iteration steps. This procedure converges to a local minimum or a saddle point of J_m .

The algorithm composed of the following steps:

1.) Initialize U =[u_{ij}] matrix , $U^{(0)}$
2.) At k-step: calculate the centers vectors $C^{(k)}=[c_j]$ with $U^{(k)}$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

3.) Update $U^{(k)}$, $U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}}$$

4.) If $\| U^{(k+1)} - U^{(k)} \| < e$ then STOP; otherwise return to step 2

Chapter 3

Neural Network

3.1 A brief review of neural network

Neural Network is basically based on the structure and function of neurons present in human beings. This field is originally generated by psychologists and neurobiologist. A neural network contains multiple sets of connected layers, each layer contains many processing units(neurons), which take inputs and produce outputs, where each connection has a weight linked with it. The output from one group of neurons are forwarded to the next layer through the connection. Neural network basically act as black box, which required two phase such as Training Phase and Testing Phase. During training phase neural network adjust its connection weights through iterative process. So, it able to classify and predict the correct class label of the inputs samples. During testing phase neural network classify data using updated weights.

There are many different kinds of neural network developed by researchers, in this section we only discuss about multilayer perception neural network.

3.2 Multilayer Perception Neural Network(MLP)

A Multilayer Perception Neural Network is define as feed forward neural network that associate set of input value to set of appropriate output values. A MLP Neural Network contain many layers of nodes in a directional graph, in which each layer is connected with next layer except input layer. Each node called as neuron(or processing unit) which contain an activation function. MLP Neural Network is a supervised learning technique based on backpropagation algorithm for training network. MLP is a modification of the standard linear perceptron and distinguish data that are not linear separable[29].

3.2.1 Basic MLP neural network structure shown below:

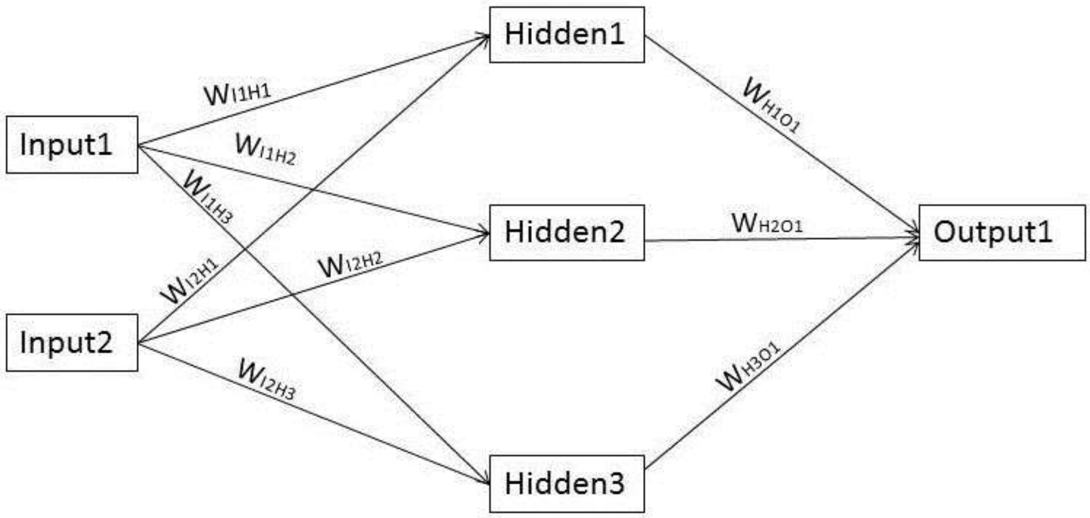

Fig 3.1 Basic MLP neural network structure

In above structure, we divide neural network into three layers architecture such as:

Input Layer: contain input values to neural network, Input1 and Input2 represent two different inputs to neural network, these values must be numerical in nature.

Hidden Layer: contain neurons acts as processing unit, in this layer Hidden1, Hidden2, Hidden3 are three neurons. More than one hidden layer can be possible and any number of neuron possible depend upon neural network topology.

Output Layer: also contain neurons act as processing unit, in this layer output1 is a neuron. More than one neuron possible depend upon neural network topology.

Where, W_{I1HI}, W_{I1H2}, W_{I1H3}, W_{I2HI}, W_{I2H2}, W_{I2H3} represent connection weights between input layer and hidden layer.

W_{H1O1}, W_{H2O1}, W_{H3O1} represent connection weights between hidden layer and output layer.

Weights can be initialized randomly or uniformly, depends on context of application within range of 0.0 to 1.0.

3.2.2 Inside Neuron

Below image represent neuron structure present in a hidden or output unit j: The inputs to unit j are outputs from the previous layer. These inputs are multiplied by their corresponding weights in order to get a weighted sum, which is added to the bias associated with unit j. A nonlinear activation function can be applied to net input, function can be sigmoid or logistic.

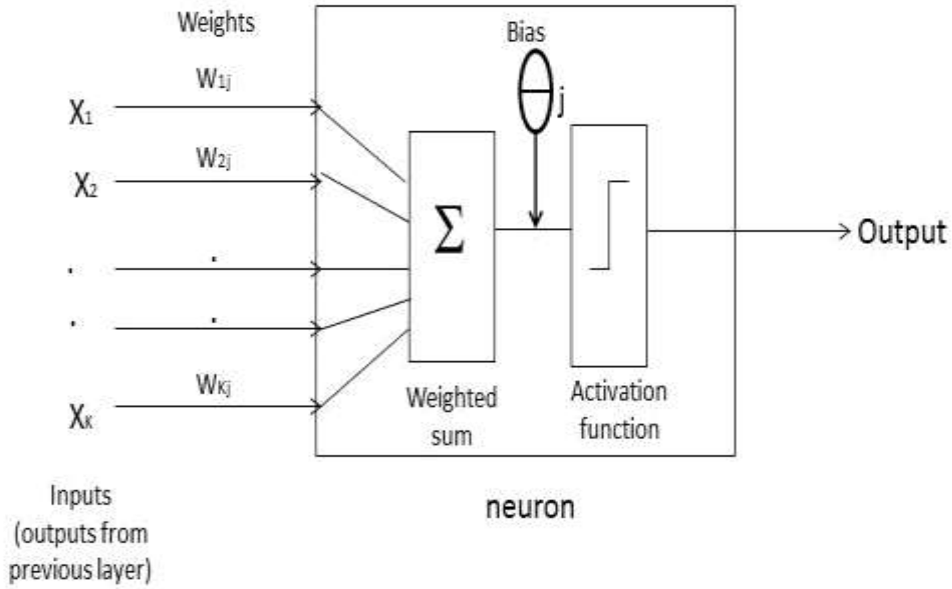

Fig 3.2 Neuron structure

3.2.2.1 Activation Function

Activation function can be considered as a squashing function, since it associate a large input domain onto the smaller range of 0 to 1. The logistic function is non- linear in nature and differential function, which allows the backpropagation algorithm to solve classification problems that is linearly inseparable. Similarly Hyperbolic function also.

- Sigmoid Function:

$$f(x) = \frac{1}{1+e^{-x}}$$

Sigmoid function Formula

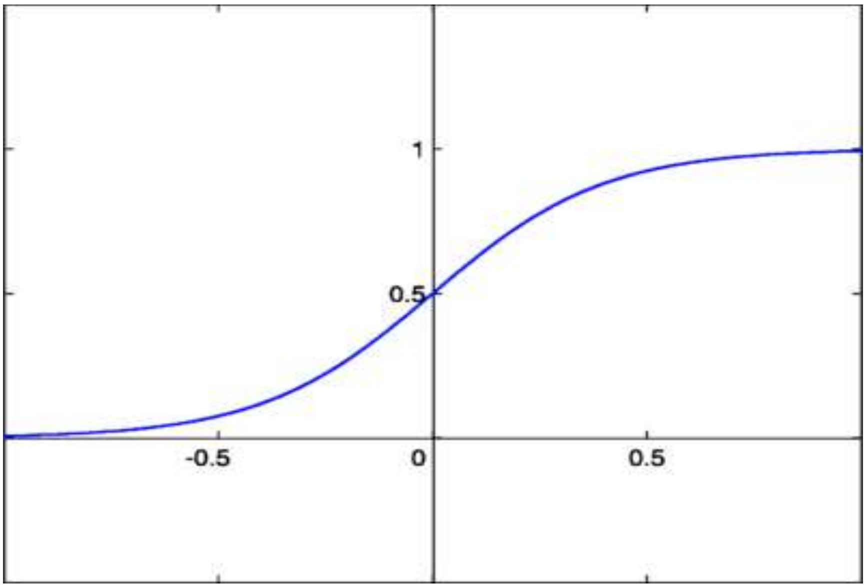

Fig3.3 Sigmoid function graph

- Hyperbolic Tangent:

$$f(x) = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1}$$

Hyperbolic Tangent function formula

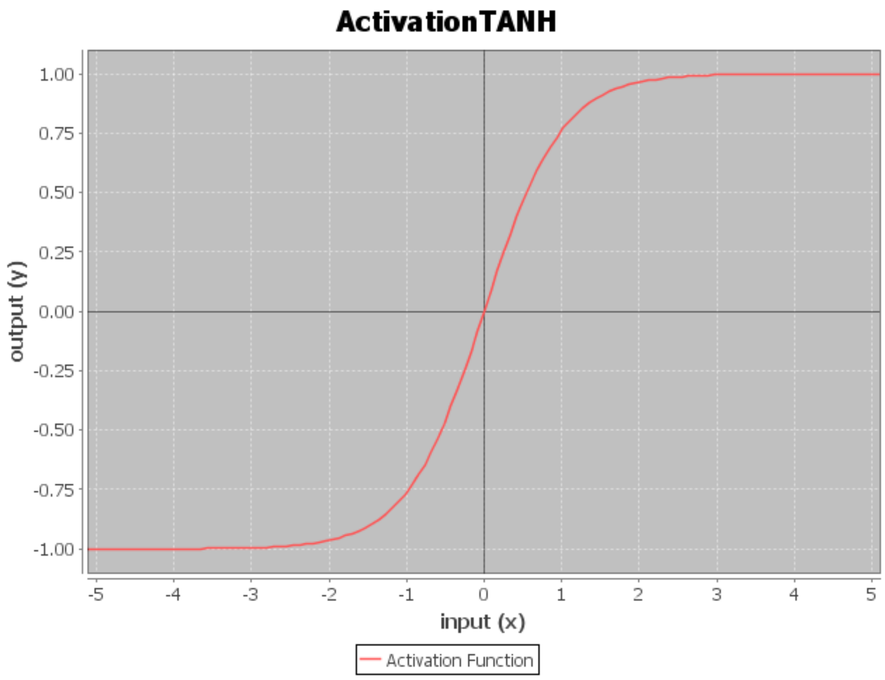

Fig 3.4 Hyperbolic Tangent Function Graph

3.2.3 MLP neural network algorithm

a.) **Backpropagation**: It is learn through iterative processing by a set of training samples, each sample applied to network for prediction by initializing connection weights once, network's prediction compared with actual known class label. For each training sample, the weights are modified or updated so as to minimize the mean squared error between the network's prediction and the actual class. These modifications are made in the " backwards" direction i.e. from the output layer, through each hidden layer down to the first hidden layer(hence the name backpropagation). Below define algorithm[26].

Algorithm: Backpropagation, Neural network learning for classification, using the backpropagation algorithm.

Input: The training samples, samples; the learning rate, l; a multilayer feed-forward network, network.

Output: A neural network trained to classify the samples.

Method: A neural network trained to classify the samples.

(1) Initialize all weights and biases in network;

(2) **while** terminating condition is not satisfied {

(3) **for** each training sample X in samples{

(4) // Propagate the inputs forward:

(5) **for** each hidden or output layer unit j {

(6) $I_j = \Sigma_i w_{ij}O_i + \theta_j$; // compute the net input of unit j with respect to the previous layer, i

(7) $O_j = 1/ 1+e^{-1j}$;} // compute the output of each unit j using activation function

(8) // Backpropagate the errors:

(9) **for** each unit j in the output layer

(10) $Err_j = O_j (1 - O_j) (T_j – O_j)$; // compute the error

(11) **for** each unit j in the hidden layers, from the last to the first hidden layer

(12) $Err_j = O_j (1- O_j) \Sigma_k Err_k w_{jk}$; // compute the error with respect to the next higher layer, k

(13) **for** each weight w_{ij} in network {

(14) $\Delta w_{ij} = (l)Err_j O_i$; // weight increment

(15) $W_{ij} = w_{ij} + \Delta w_{ij}$; } // weight update

(16) **for** each bais θ_j in network {

(17) $\Delta\theta_j =(l)Err_j$; //bias increment

(18) $\theta_j = \theta_j + \Delta\theta_j$; }// bias update

(19) }}

In neural network, connection weights are set with random number (e.g. ranging from -0.0 to 1.0 or -0.5 to 0.5) or may be uniform number (ranging from 0.0 to 0 1). The learning rate l initialized with a constant value between 0.0 and 1.0. Backpropagation algorithm learns through technique of gradient descent aim to find set of weights that can solve the given classification problem so as to minimize the mean squared distance between the network's class prediction and the actual class label of the samples. The learning rate helps to avoid getting stuck at a local minimum in decision space (i.e., where weights appear to converge, but are not the optimum solution) and encourages finding the global minimum. If we considered learning rate too small, then learning will occur at a very slow pace. If considered learning rate is too large, then oscillation between inadequate solutions may occur. A rule of thumb is to set the learning rate to 1/t, where 't' is the number of iterations through the training set so far. In above algorithm weights are updating and biases after the presentation of each sample. This is referred to as case updating..

b.) **Feed Forward:** in this input is given to each neuron layer by layer, the output from one set of neuron is feed into another layer neuron. In this way we get output after processing all layers. In this none of the weights cycle back to an input unit or to an output unit of a previous layer. This part of algorithm used for classification of test samples.

Algorithm: Feed Forward

Input: The testing samples, samples; the learning rate, l;

Output: samples classified

Method: Classified test samples

(1) **for** each training sample X in samples{

(2) // Propagate the inputs forward:

(3) **for** each hidden or output layer unit j {

(4) $I_j = \Sigma_i \, w_{ij}O_i + \theta_j$; // compute the net input of unit j with respect to the previous layer, i

(5) $O_j = 1/\ 1+e^{-1j}$;} // compute the output of each unit j using activation function

3.2.4 Strength

- Provide high tolerance against noisy data.
- Able to classify unseen data i.e. for which they are not trained.
- Neural network able to handle complex non-linear functions and able to predict association between combination of inputs.
- It also support incremental learning and doesn't require much storage space; only need to store connection weights.
- It doesn't require to store or keep training data, which means neural network can handle continuous stream training data applications.
- Several approaches are proposed which can extract rule from trained neural network.

3.2.5 Weakness

- It required long training time.

- Neural network acts as a black box which contains hundreds of nodes and thousands of connections between them, which makes difficult to interpret how neural network come up to answers. It is hard for humans to interpreted the symbolic meaning behind the learning weights.
- Parameters for neural networks such as decisive rules for considering the t`raining rate and network topology are quite difficult to decide for a particular problem. They are highly experimental based. Lot of experimentation required to select best candidate solution. If we keep high training rate than neural network might overgeneralize on outliers(noise data) , and if we keep it low then neural network never learn to classify on the given input data.

Chapter 4

Proposed Approach

4.1 Overview

The main area of interest of this research is here, in this we proposed an approach to develop a recommender system, which based on machine learning technique such as neural network. Initially neural network learn from previous available data that is Training dataset, after learning phase neural network classify and predict rating of books. Predicted rating is in range of 1 to 5 and specified meaning associate with it e.g. rating 1 represent highly rejected book for current customer whereas rating 5 highly accepted book for current customer.

Based on predicted rating and considered threshold rating, books are recommended to current customer. The purpose this recommender system is to accept all the challenges of recommendation techniques(such as Data sparsity, Cold start, Scalability etc. define in section 1.5) and provide innovative solution to all accepted challenges.

4.2 Appropriateness of Neural Network

In this defined approach we considered neural network for classification purpose over other available classifiers because of following reasons:

- Neural Network support incremental training as compared to other techniques such as SVM, decision tree etc. In this domain every year thousands of book got published, which have different characteristic and due to advancement in accessibility of internet more customer purchase books online, each customer have different characteristic associated with them. Therefore, to deal with current scenario we need considered a classifier, which support incremental training.
- To deal with consequences of scalability (when degree of existing user and items increased tremendously) such as noisy data. We need a technique, which is highly tolerance to noisy data. Neural Network provide high tolerance against noisy data.
- Sometime classifier need to classify data, which is unseen (unseen refers here data which is highly diverse in nature from data used in training set). Neural Network able to classify appropriately unseen data as compared to other technique such as decision tree, SVM, regression classifier etc. Recommender System uses different combination

of parameters for recommendation, which can lead to unseen combination as input to classifier.

- In current approach, we considered different parameters for recommendation e.g. publisher, title , demographic information(such as age , sex, location), which required to predict dependencies between combination of inputs. Neural Network can predict dependencies between combination of inputs efficiently and also have ability to handle complex non-linear function.
- Neural Network offers less space complexity only need to store connection weights as compared to other technique such as SVM, which require to store complete training data set.

4.3 Structure of Dataset

In the following approach, we construct a dataset which consist of following tables:

- Book Table: which contain all information associated with books such as ISBN Number, Book title, Book Authors, Book Publisher.
- Customer Table: which consist detail of customer, who previously purchased book from system. Customer table contain following details: customer id(auto generated), age, sex, country, IP address number.
- Rating Table: contain explicit rating for books previously purchased by customer. Rating table contain following details: ISBN Number, customer id, book rating.
- IP table: which is used to convert IP address in unique IP number and also used to obtain current user country name. This table contain following attributes: ID, Beginning IP, Ending IP, Two Country Code, Three Country Code, Country Name.

From below E-R Diagram, we can understand relationship between tables defined above. It's clearly understandable one customer can by more than one book and one particular book can be purchased by more than one customer. When customer purchased a book, he/ she can label rating of purchased book.

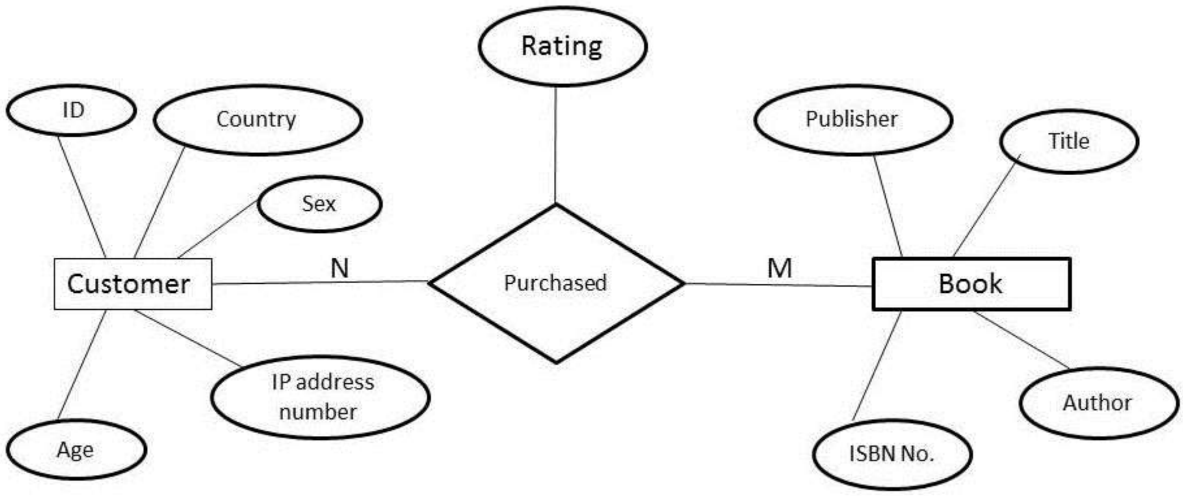

Fig 4.1 E-R Diagram of dataset

4.4 Procedure to translate IP address into IP Number

Every visitor to site have a unique IP address, which can be easily obtained. It can be used for tracing. IP address can also identify current location of visitor or at least it's Service Provide in the world. Sometimes difficulty arise when visitor behind a proxy of some sort, in this case we can only obtain IP address of proxy server.

Method to convert IP Address into IP Number:

As we know Internet Protocol(IPv4) divides IP address into four sub-blocks. Each block has separate weight associated with it, which are powered by 256.

The starting IP number and the finishing IP number are compute by using following formula:

IP number = 16777216*a + 65536*a + 256*c + 1*d

Or

IP number= 256^3*a + 256^2 *b + 256^1*c + 256^0*d

Where : IP address = a.b.c.d

e.g. considered IP Address : 205.109.50.7, then its IP number can be calculated as follow:

IP Address: 205.109.50.7 so, a= 205, b= 109, c= 50, d= 7

IP Number =16777216*205+65536*109+256*50+1*7

= 3439329280+7143424+12800+7

= 3446485511

Method to convert IP Number into IP Address:

a = integer (IP Number / 16777216) % 256

b = integer (IP Number / 65536) % 256

c = integer (IP Number / 256) % 256

d = integer (IP Number) % 256

where, % denotes modulus operator and integer used to type cast return result part of division into integer.

Country name can be obtained from the IP number by using IP-country database. Through IP-country database we can tally unique record that has the IP Number that fits between the starting IP number and the finishing IP number e.g. the IP address: 205.109.50.7 is corresponding to the IP Number: 3446485511. It is associated to the following record in the database because it is between the starting and the finishing of the IP Number.

From the IP-country dataset, the country name is Malaysia and the country code is MY

From above procedure we convert IP Address into IP Number, which can be provide as input to neural network for prediction purpose. For recommendation to current customer, we obtain it IP address and translate it into IP number. Use this IP number to recommend him/her.

4.5 Hierarchy of ISBN Number

In late 1960's, publishers got understood that they required a unique uniform number to recognize all the books that were being published throughout the world. In 1970, the 10 digit an ISBN format is developed by the International Organization for Standardization(ISO) and was published as international standard ISO 2108.Every book, including new edition of older book, was issued ISBN number[30].

An ISBN is a "structured" number, different part associated with different meaning(similar to telephone codes). Different associated part are separated by hyphen and space(hyphen are preferred, but not necessary). The ISBN is 13- digits long if assigned on or after 1 January 2007, and 10 digit long if assigned before 2007. An International Standard Book Number consists of four parts(if it is a 10 digit ISBN) or 5 parts (for a 13 digit ISBN)[31]:

1.) For a 13-digit ISBN, a prefix element – a *GSI prefix*: so far 978 or 979 have been made available by GSI
2.) The registration group element,(language sharing country group, individual country or territory)
3.) The registrant element
4.) The publication element
5.) A checksum character or check digit

The 13 digit ISBN separates its part(prefix element, registration group, registrant, publication and check digit) with either a hyphen or space. Other than the prefix and the check digit, no part of the ISBN has a fixed number of digits.

The 10 digit ISBN also separated its parts(registration group, registrant, publication and check digit) with either hyphen or space e.g.

The parts of a 10-digit ISBN and the corresponding EAN-13 and barcode. Note the different check digits in each. The part of the EAN-13 labeled "EAN" is the Bookland country code.

Fig 4.2 Hierarchy of ISBN Number

Registration group Identifier

This part contain 1 to 5 digit which identifies a country, region or language area participating in ISBN system within a single prefix element(i.e. one of 978 or 979). The single digit group identifier in the 978 prefix element are:

Group Number: 0 or 1 for English speaking countries

Group Number: 2 for French speaking countries

Group Number: 3 for Germany speaking country

Group Number: 4 for Japanese speaking country

e.g. Group number 99942 is Sudan (Africa)

Book published in infrequent language usually have long Group Number.

Registrant Element

Registrant element also refer as publisher, which may contain up to seven digits. The publisher identifier directs particular identification of publication house and its address. If publisher issue title number from their initial collection then they need to provide additional publisher identifier.

Title Identifier

The third part is used to identify specific edition of specific publisher, it may contain up to 6 digits. If ISBN have less than 13 digits, then 0's are prefix to make number 13 digit.

Check Digit

A check digit is used for error detection, which represent by single digit.

e.g. Considered an ISBN number :81 - 8147 – 049 - 4

To calculate ISBN check digit, we need to multiply the first digit by 10, the second digit by 9, the third digit by 8, ..., the ninth digit by 2 and sum up all these numbers. The check digit is the number have to sum to this total to get up to a multiple of 11. For ISBN number given above we compute check digit in the following way:

$$8\times10 + 1\times9 + 8\times8 + 1\times7 + 4\times6 + 7\times5 + 0\times4 + 4\times3 + 9\times2 =$$

$$80 + 9 + 64 + 7 + 24 + 35 + 0 + 12 + 18 = 249.$$

As we know 249 is between 242(22×11) and 253(23×11). We need to add up 4 to 249 in order to get 253. So, check digit number is 4.

4.6 Neural Network Input-Output parameters

In defined approach, we considered a neural network which has 5 inputs and 5 outputs, one hidden layer and one output layer. In our approach we only considered books which are belong to English speaking countries. So, Registration group Identifier can't be act as appropriate candidate for input to neural network for classification and prediction.

Five inputs to neural network are:

Input		Data Type	Length		Range	
			Min	Max	Min	Max
Registered Element (part of ISBN)		Integer	1	7	1	9999999
Title Identifier (part of ISBN)		Integer	1	6	1	999999
Age		Integer	1	3	10	100
Unique IP Number		Integer	1	16	1	4294967265
Gender	Male	Char	1	1	-	-
	Female	Char	1	1	-	-

Table 4.1 Input parameters

Gender value is converted from character to integer value for gender differentiability and classification and prediction purpose. In defined approach, we converted as follow:

- Male: 1
- Female: 1000

Five output represent rating predicted by neural network:

Rating	Associated Meaning
1	Highly Rejected
2	Rejected
3	Marginal Accepted
4	Accepted
5	Highly Accepted

Table 4.2 Output parameters

Rating values during training phase :

Rating	**Corresponding value** (for error evaluation)
1	1,0,0,0,0
2	0,1,0,0,0
3	0,0,1,0,0
4	0,0,0,1,0
5	0,0,0,0,1

Table 4.3 Rating values during training phase

Above define corresponding value is used evaluate error during training phase and errors are backpropagate to adjust weights of neural network for prediction and classification.

Output rating during testing phase:

In order to select rating for testing phase, we obtain index of maximum value occurred at output e.g. 5 outputs of neural network:

Values:	0.95619	0.0119020	0.022318	0.0014514	0.12561
Index:	1	2	3	4	5

Table 4.4 Output rating during testing phase

Maximum value: 0.95619

Maximum value index: 1

So, consider rating of current sample as : 1.

Below diagram represent inputs and outputs parameters to neural network:

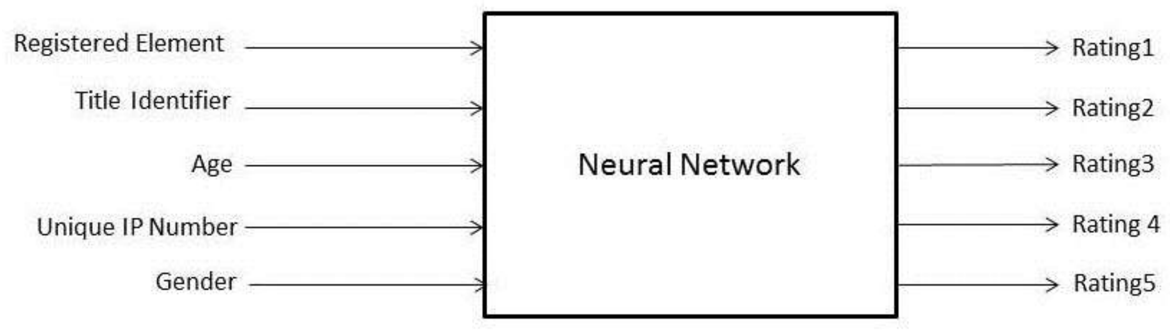

Fig 4.3 Neural Network input and output parameters

4.7 Neural Network functional parameters

Functional Parameter	Data Type	Range	
		Min	Max
Number of neuron in hidden layer	Integer	5	20
Momentum	Real	0	1
Epoch	Integer	1	1000
Error Threshold	Real	0	0.5

Table 4.5 Neural Network function parameters

4.8 Activation Function in Defined Approach

In this approach we considered two activation function, which are :

- Sigmoid Function
- Hyperbolic Tangent Function

4.9 A Personalized Hybrid Book Recommender System Algorithm

The approach consists of following steps:

Initially we divide complete data set into two different set: Training Set for learning phase and Test Set for subsequent performance evaluation of the during testing phase e.g. considered data available form 10 user who at least rate two books, we divide data set as the Training Set (7/10) and the Testing Set (3/10).

// Training Phase include following Step

Step 1: Obtain Training input dataset

Step 2: Obtain Training output dataset // rating of corresponding customers whose data previously retrieved

Step 3: Perform data transformation // rating to corresponding value define in table 4.3 and map non numeric data to numeric

Step 4: Load data for training phase

Step 5: Initialize Neural Network parameters

Step 6: Implement Backpropagation algorithm of MLP Neural Network (defined in section 3.2.3.a) // Train neural network

Step 7: If (Error < Threshold)

Go to step 8 // Testing Phase

Else repeat step 6 and 7

// Testing Phase include following step:

Step 8: Obtain Testing input dataset

Step 9: Obtain Expected ratings // corresponding to testing input dataset

Step 10: : Load data for testing phase

Step 11: Implement Feed-Forward algorithm with updated weights (defined in section 3.2.3.b)

Step 12: Obtain Predicted Rating // max_value_index(Rating1, Rating2, Rating3, Rating4, Rating5)

Repeat Step 11 and 12 for each testing sample

Step 13: Evaluate Mean Absolute Error // among predicted rating and expected rating(defined in section 1.7)

The block diagram of the proposed Personalized Hybrid Book Recommender System procedure are shown below:

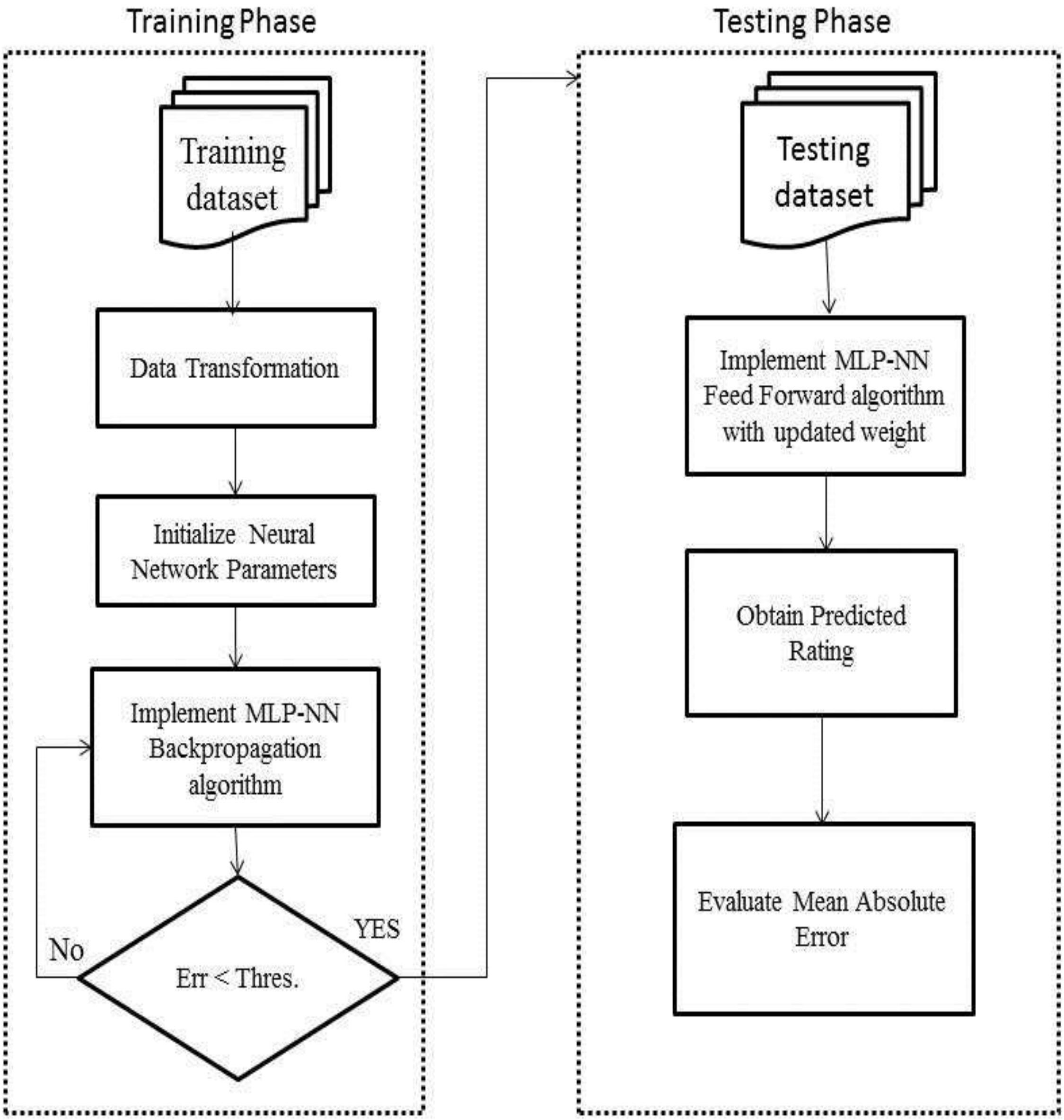

Fig 4.4 A Personalized Hybrid Book Recommender System Algorithm

Chapter 5

Experiments Results

The proposed method 'Hybrid Recommender System Based Upon (Neural) Network ' has aim to predict rating of books using MLP-Neural Network. In following approach, we evaluate performance of MLP-neural network in matlab. In order to select best MLP-Neural Network parameters for optimal results, we evaluated defined approach by differing parameter values(such as activation function and number of neurons at hidden layer) and estimated the results. The outstanding MLP- Neural Network is one which classified data and predict accurate rating of book. We considered index value of maximum output as rating, output value which is nearer to 1 and index values whose output nearer to 0 are rejected. Such neural network can be used for predict rating of books. Firstly, we trained MLP-Neural Network with training dataset for accurate prediction, training would stop when error threshold criteria meet. After training phase completion, we applied testing dataset to Network for performance evaluation of Network. As a result of testing phase, we attempt to determine optimal parameters for accurate prediction.

5.1 Expected Rating of testing dataset

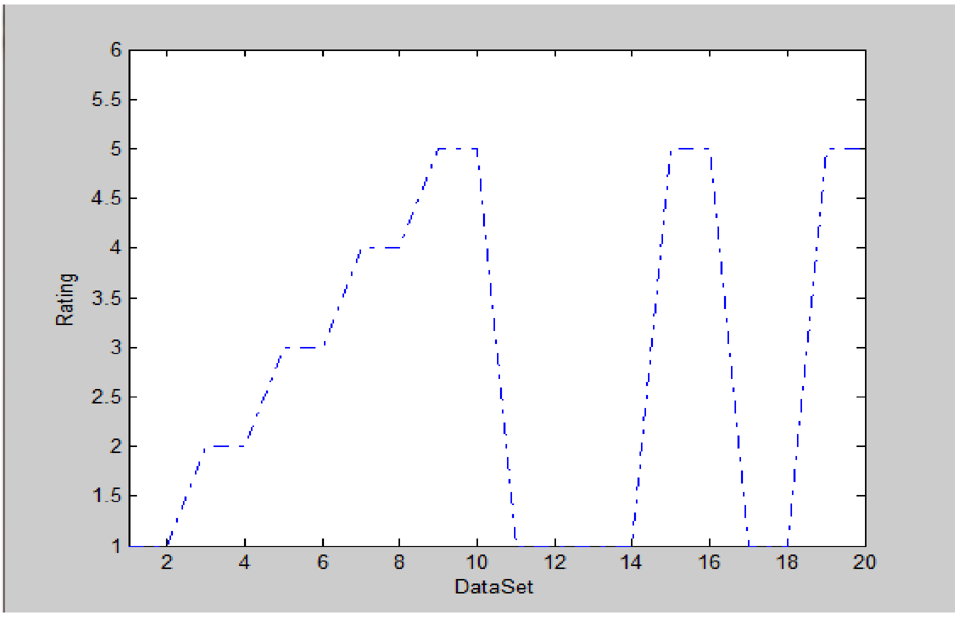

Fig 5.1 Expected rating of testing data

Figure 5.1 shows graph which represent expected rating of testing dataset. We considered 20 samples in testing dataset. Y-axis represent output rating of books and X-axis represent samples.

Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Expected Rating	1	1	2	2	3	3	4	4	5	5	1	1	1	1	5	5	1	1	5	5

Table 5.1 Sample and expected rating

5.2 MLP-Neural Network performance with various activation function

Activation function is one of significant parameter of MLP-Neural Network, play significant role in Network performance. By changing activation function at hidden layer and output layer, further accuracy can be achieved.

- Hidden Layer activation function: Sigmoid

  Output Layer Activation function: Sigmoid

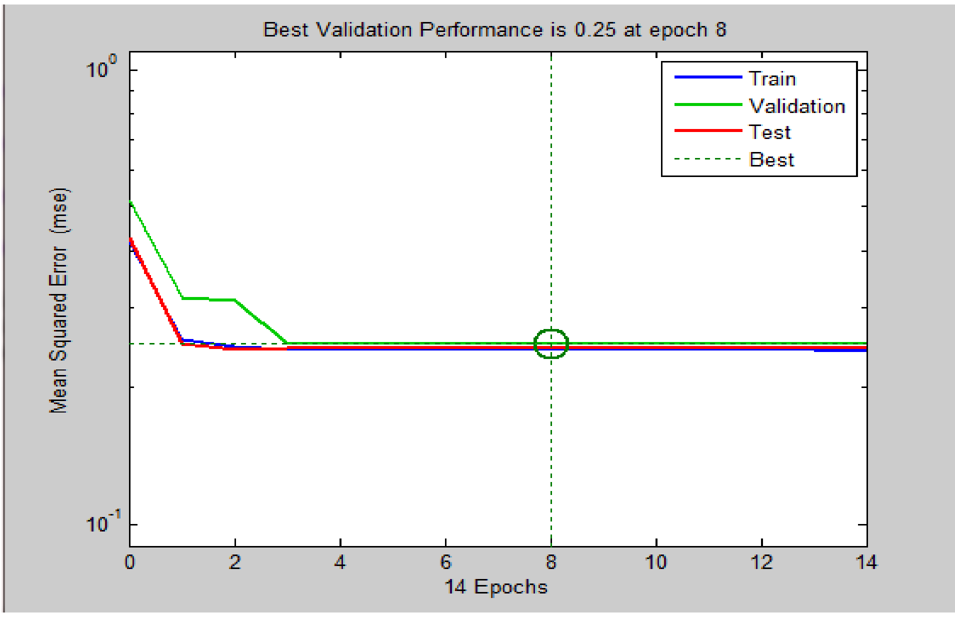

Fig 5.2 Performance graph(Activation Functions: Sigmoid, Sigmoid)

Figure 5.2 represent performance graph

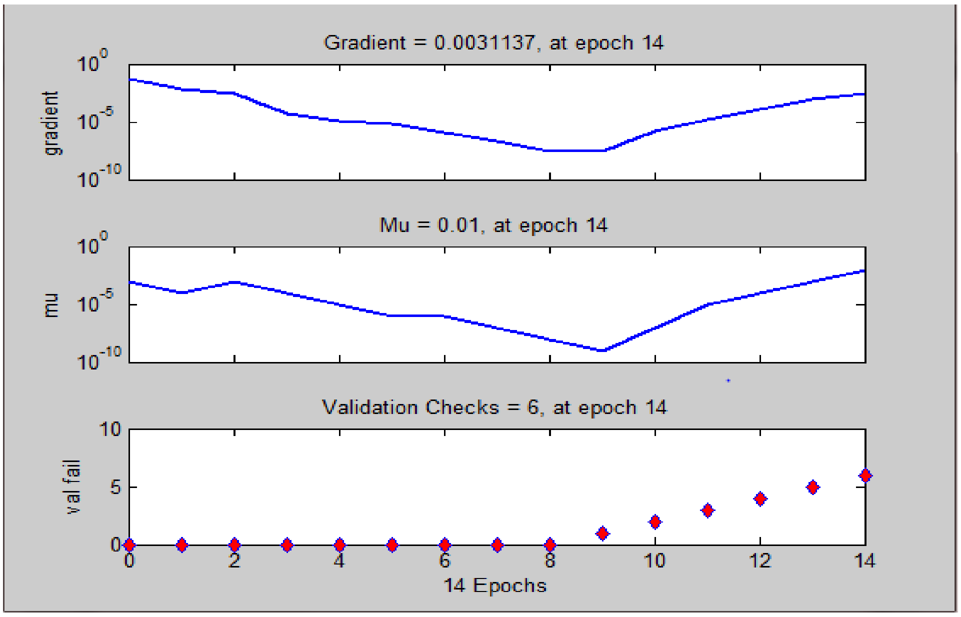

Fig 5.3 Training State graph(Activation Functions: Sigmoid, Sigmoid)

Figure 5.3 represent training state graph

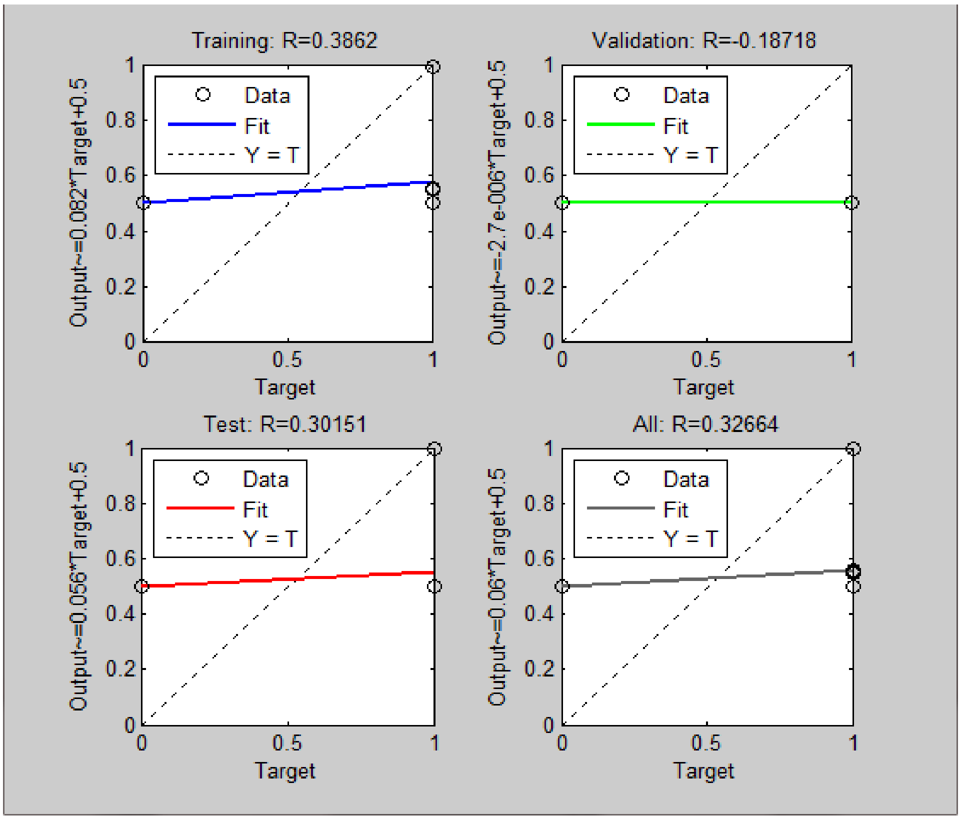

Fig 5.4 Regression graph(Activation Functions: Sigmoid, Sigmoid)

Figure 5.4 represent regression graph

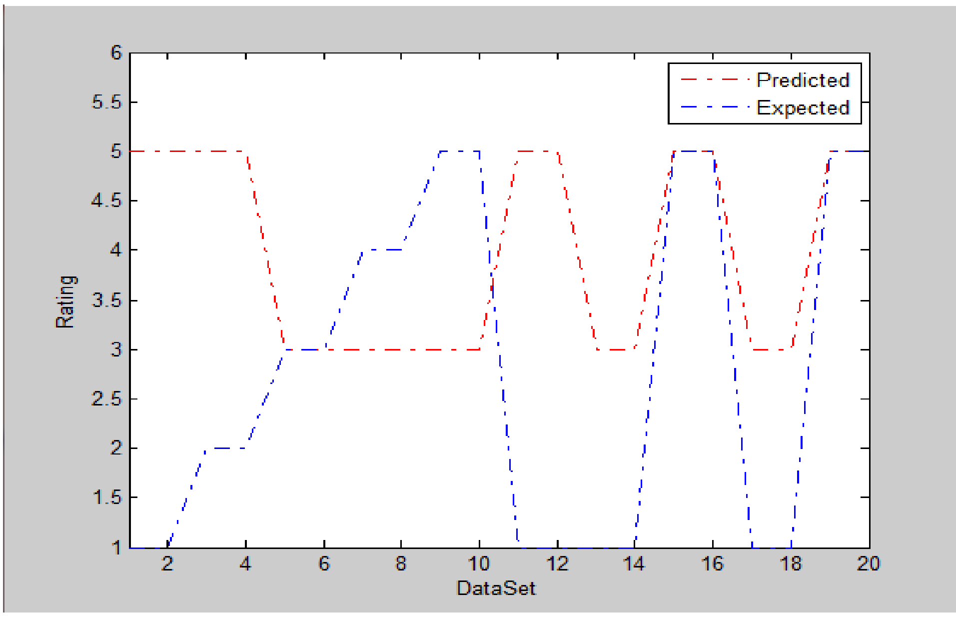

Fig 5.5 Predicted vs. Expected rating (Activation Functions: Sigmoid, Sigmoid)

Figure 5.5 represent MLP-Neural Network predicted rating and expected rating

Mean Absolute Error(MAE) on Expected rate and Predicted rate is : 1.8

- Hidden Layer activation function: Hyperbolic Tangent

  Output Layer activation function: Sigmoid

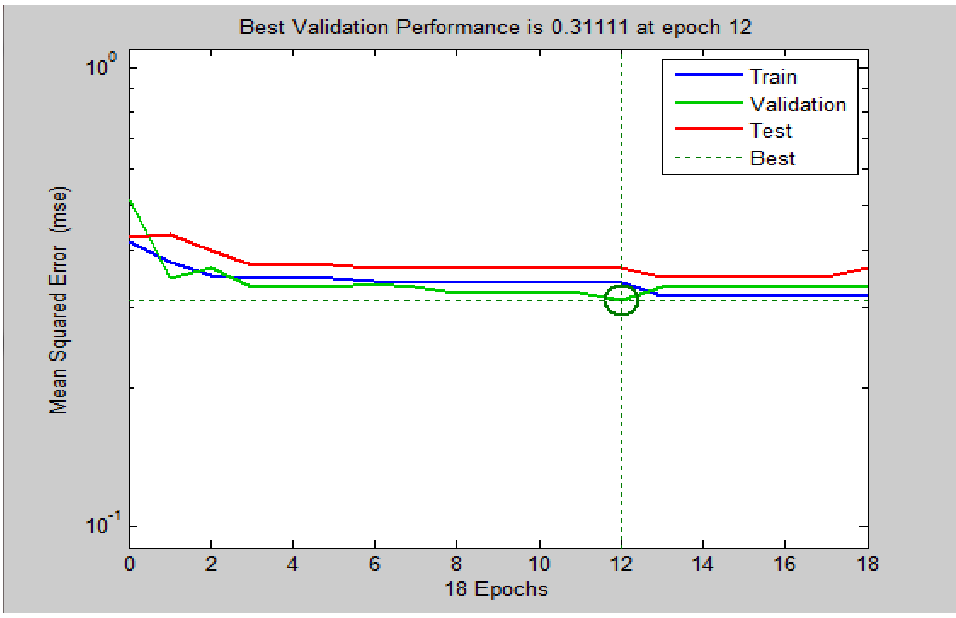

Fig.5.6 Performance graph(Activation Functions: Hyperbolic Tangent, Sigmoid)

Figure 5.6 represent performance graph

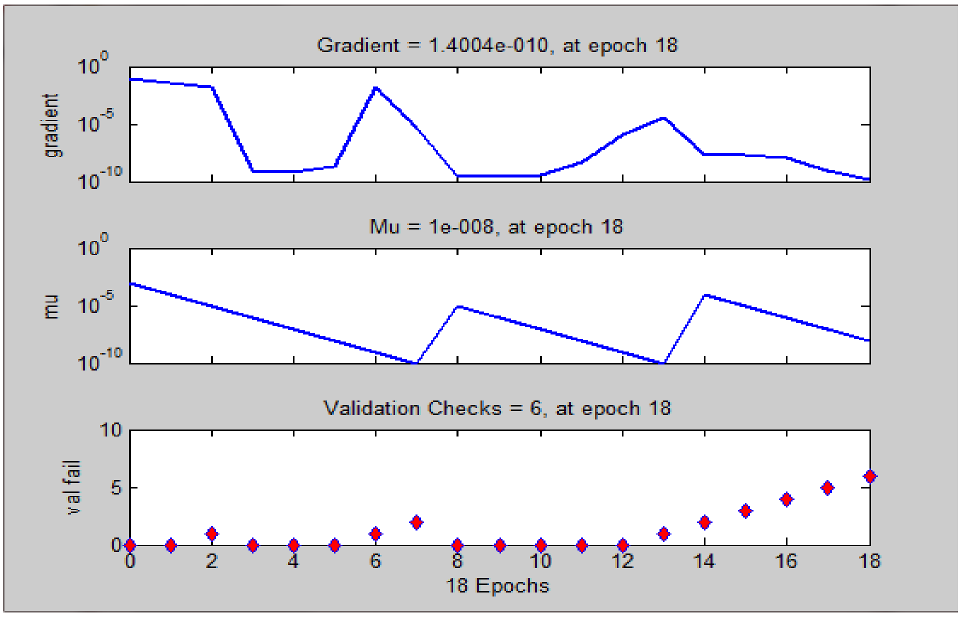

Fig 5.7 Training state graph(Activation Functions: Hyperbolic Tangent, Sigmoid)

Figure 5.7 represent Training state graph

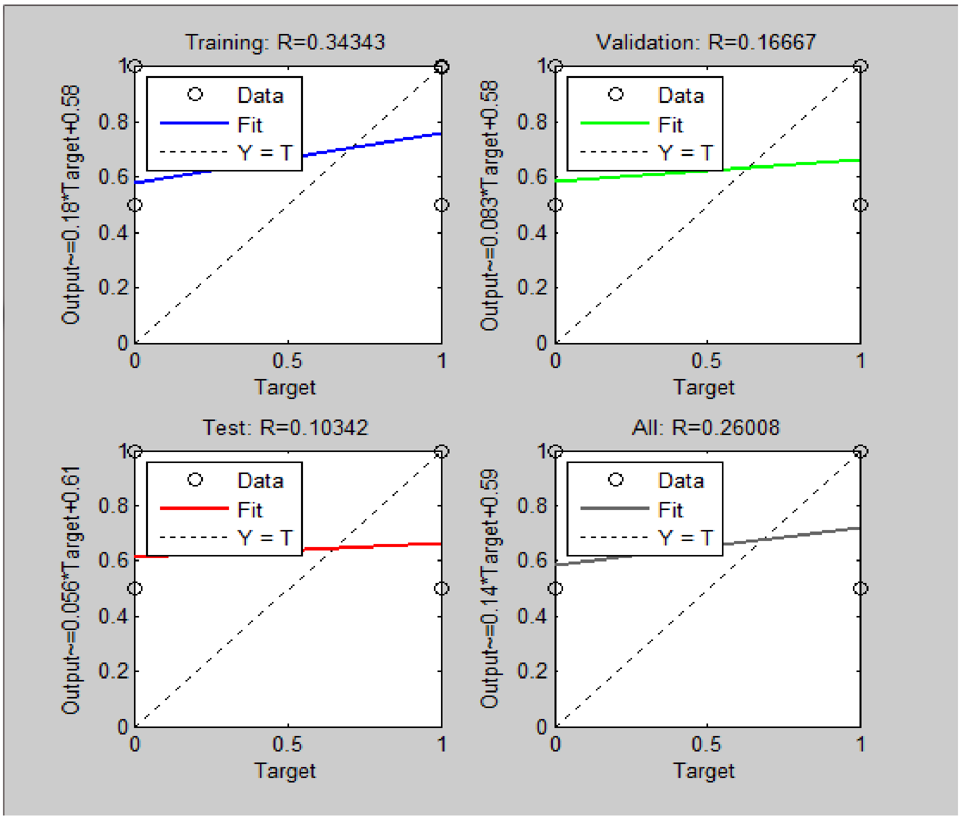

Fig 5.8 Regression graph(Activation Functions: Hyperbolic Tangent, Sigmoid)

Figure 5.8 represent regression graph

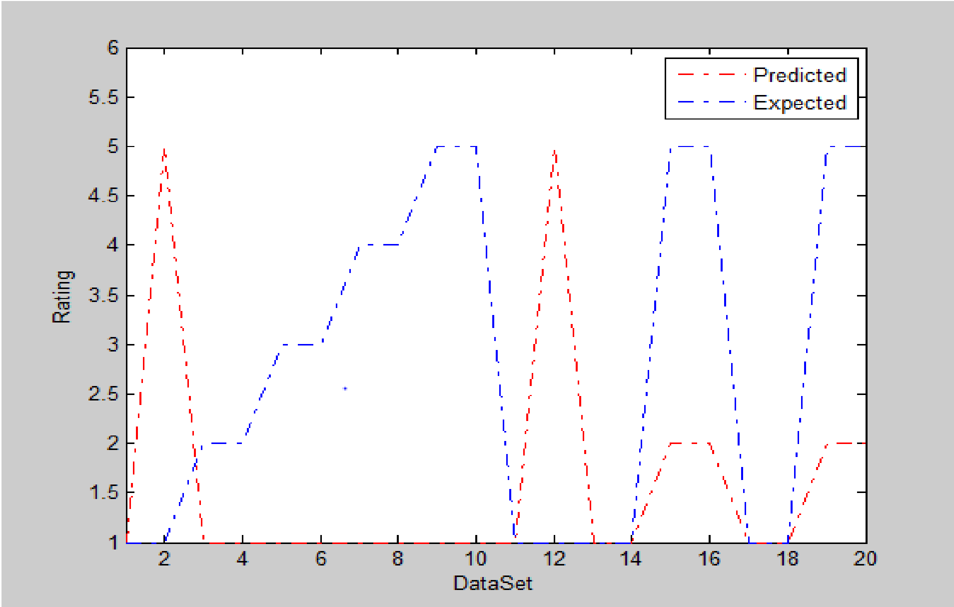

Fig 5.9 Predicted vs. Expected rating(Activation Functions: Hyperbolic Tangent, Sigmoid)

Figure 5.9 represent MLP-Neural Network predicted rating and expected rating

Mean Absolute Error(MAE) on Expected rate and Predicted rate is : 2

- Hidden Layer activation function: Sigmoid

  Output Layer activation function: Hyperbolic Tangent

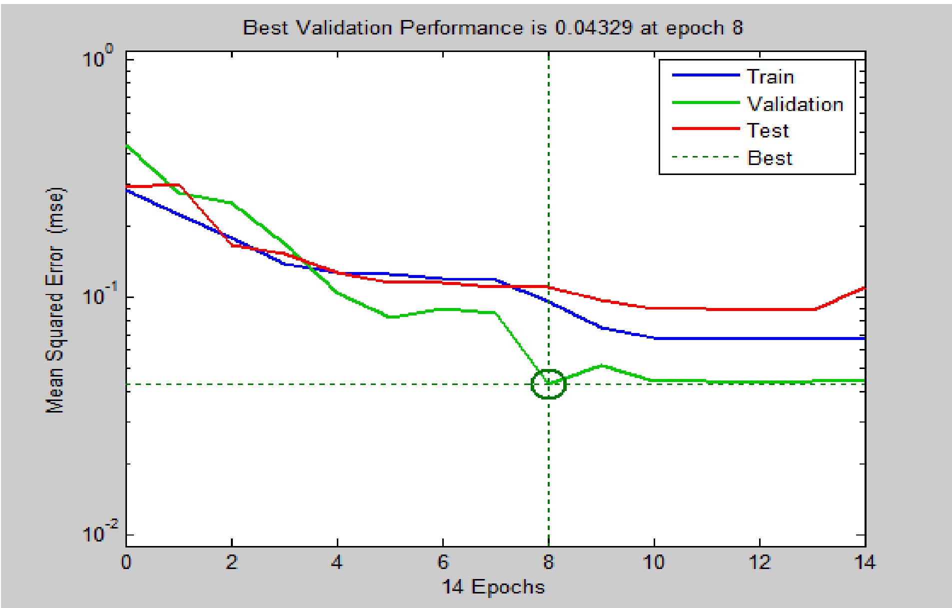

Fig 5.10 Performance graph(Activation Functions: Sigmoid, Hyperbolic Tangent)

Figure 5.10 represent performance graph

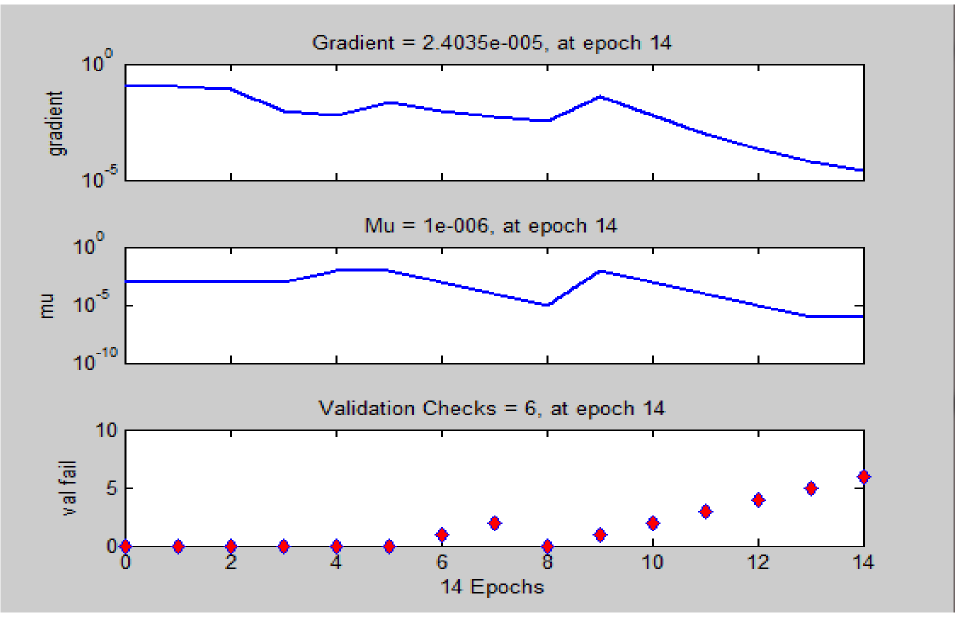

Fig 5.11 Training state graph(Activation Functions: Sigmoid, Hyperbolic Tangent)

Figure 5.11 represent training state graph

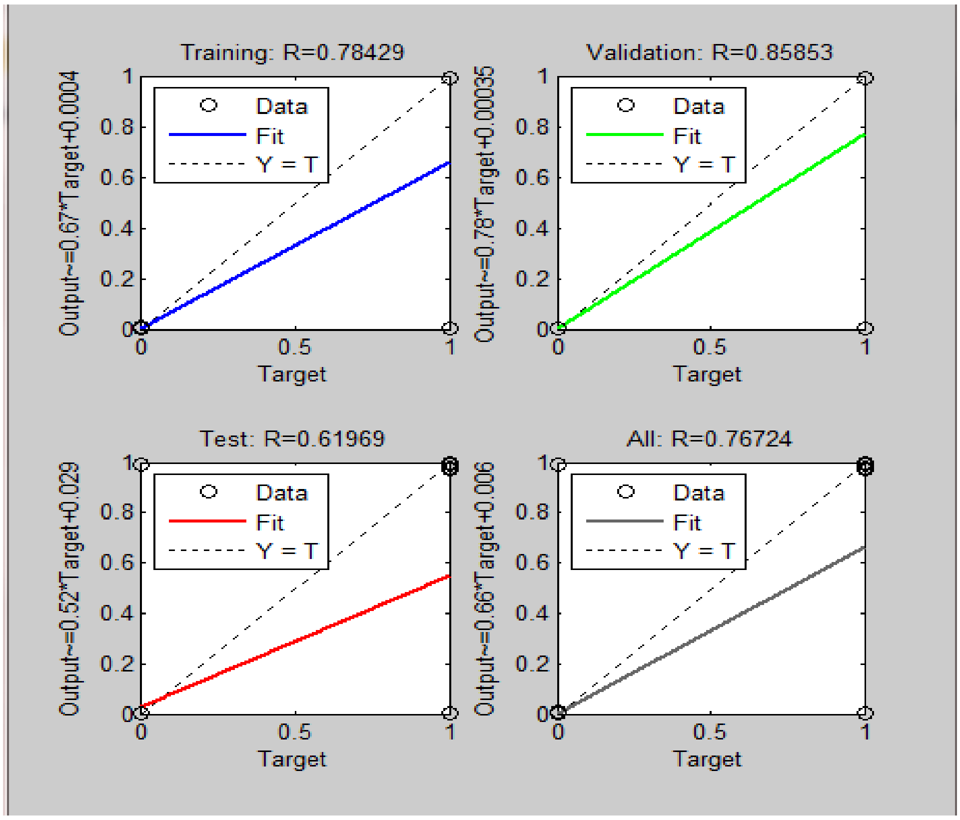

Fig 5.12 Regression graph(Activation Functions: Sigmoid, Hyperbolic Tangent)

Figure 5.12 represent regression graph

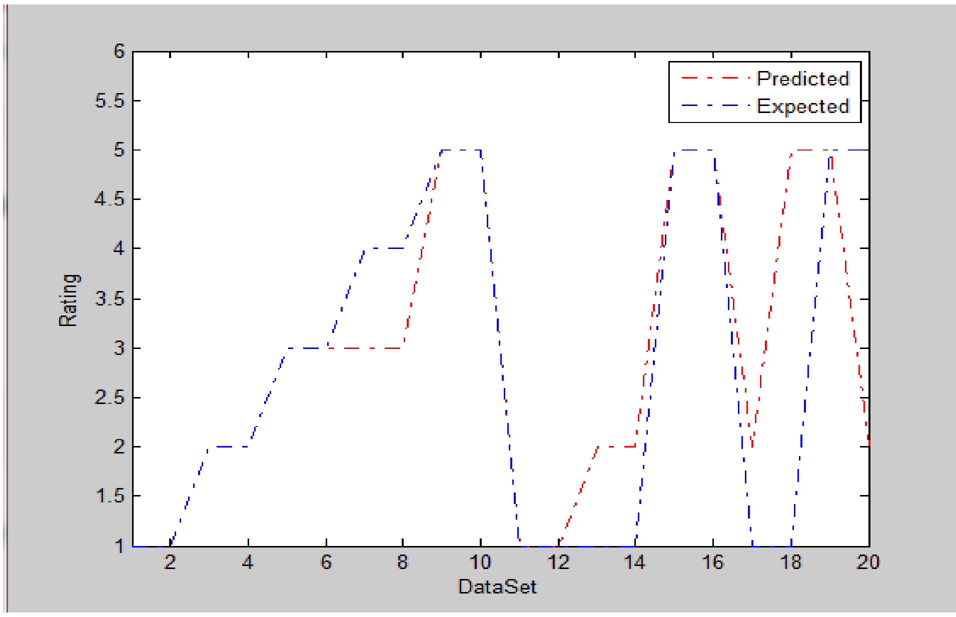

Fig 5.13 Predicted vs. Expected rating(Activation Functions: Sigmoid, Hyperbolic Tangent)

Figure 5.13 represent MLP-Neural Network predicted rating and expected rating

Mean Absolute Error(MAE) on Expected rate and Predicted rate is : 0.600

- Hidden Layer activation function: Hyperbolic Tangent

  Output Layer activation function: Hyperbolic Tangent

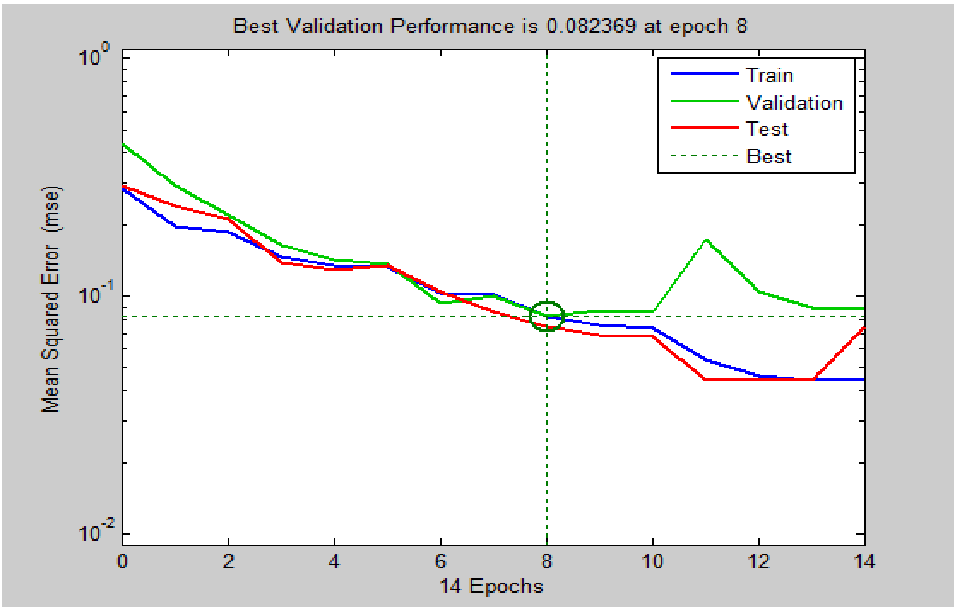

Fig 5.14 Performance graph(Activation Functions: Hyperbolic Tangent, Hyperbolic Tangent)

Figure 5.14 represent performance graph

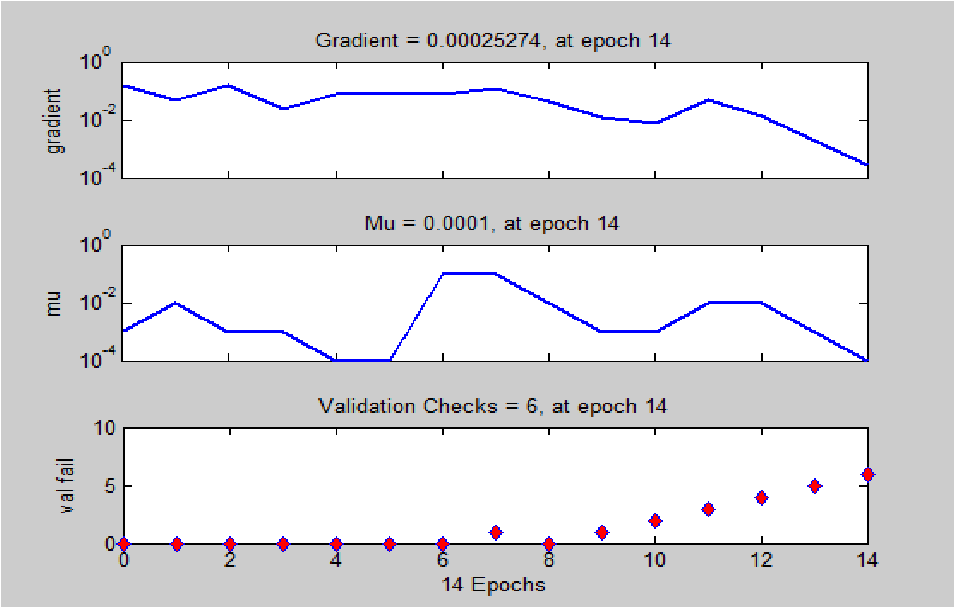

Fig 5.15 Training state graph(Activation Functions: Hyperbolic Tangent, Hyperbolic Tangent)

Figure 5.15 represent training state graph

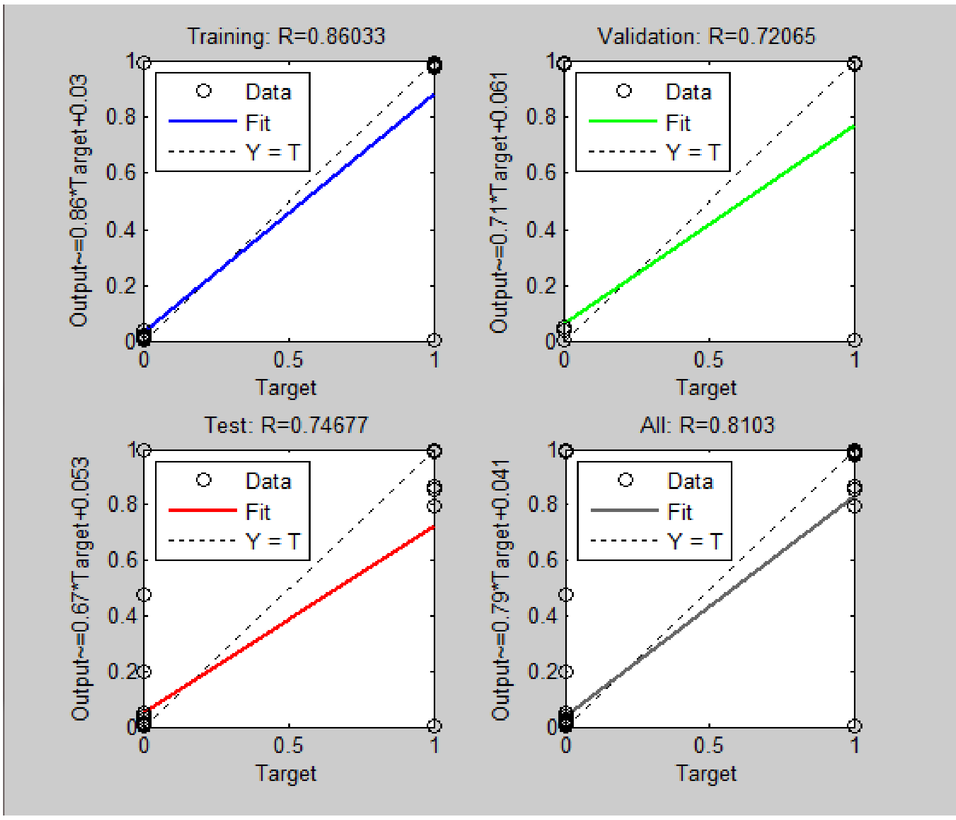

Fig 5.16 Regression graph(Activation Functions: Hyperbolic Tangent, Hyperbolic Tangent)

Figure 5.16 represent regression graph

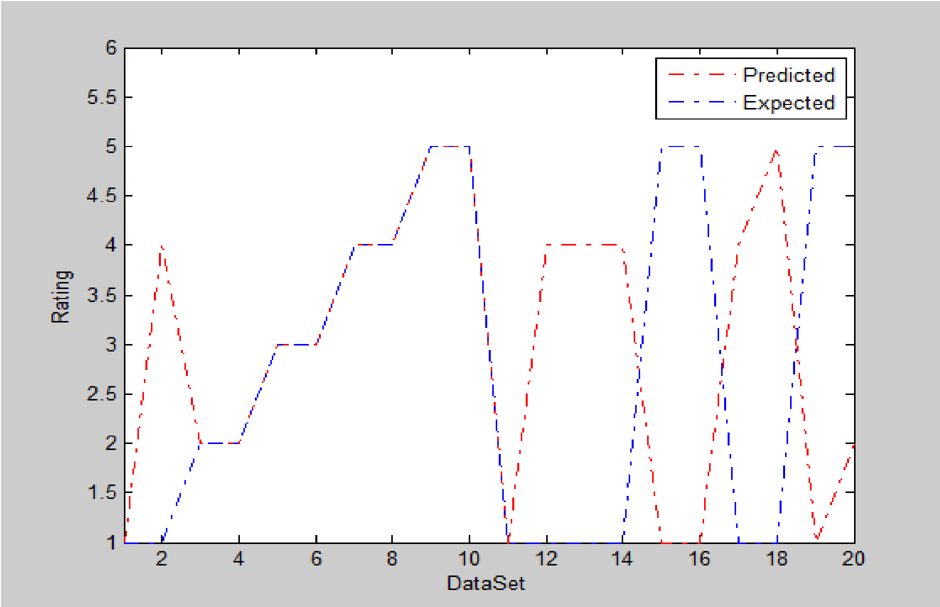

Fig 5.17 Predicted vs. Expected rating(Activation Functions: Hyperbolic Tangent, Tangent)

Figure 5.17 represent MLP-Neural Network predicted rating and expected rating

Mean Absolute Error(MAE) on Expected rate and Predicted rate is : 1.7

5.2.1 MLP Neural Network performance summary for various functions:

Below table represent performance of various activation function implemented at Hidden Layer and Output Layer.

Serial No.	Activation Function		MAE on rating
	Hidden Layer	Output Layer	
1	Sigmoid	Sigmoid	1.75
2	Hyperbolic Tangent	Sigmoid	1.7
3	Sigmoid	Hyperbolic Tangent	0.600
4	Hyperbolic Tangent	Hyperbolic Tangent	1.5

Table 5.2 Summary of MLP Neural Network performance for various MLP functions

From table 5.2, results shown that activation function Sigmoid at Hidden Layer and Hyperbolic Tangent function at Output Layer delivers the optimal results. Whereas all other activation functions combination delivers nominal results. Therefore, in the subsequent experiment activation function Sigmoid at Hidden Layer and Hyperbolic Tangent at Output Layer were used.

5.3 MLP-Neural Network performance by varying numbers of neuron at hidden layer

As we know each neurons play significant role in classification and prediction in MLP-Neural Network. Performance of Network is highly influence by number of neuron present at hidden layer. In this section, we investigate performance of Network by considering different numbers of neuron at hidden layer. If we considered number of neuron to high then time complexity of Network increase significantly where as if we considered number of neuron quite low then Network unable to classify and predict. Therefore, we evaluate performance with in defined range of number of neurons.

- Number of neuron at hidden layer: 5

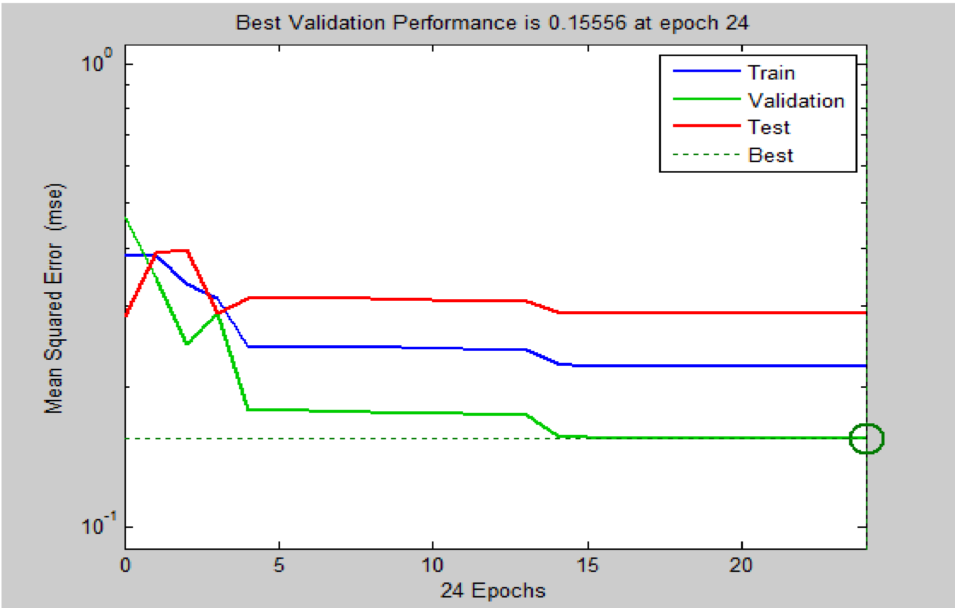

Fig 5.18 Performance graph(No. of neurons:5)

Figure 5.18 represents performance graph

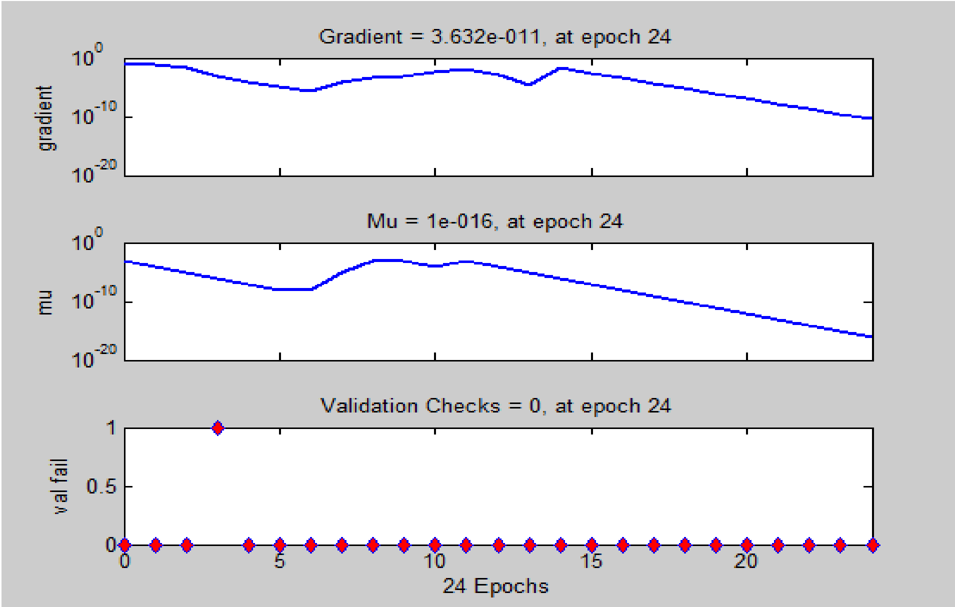

Fig 5.19 Training state graph(No. of neurons:5)

Figure 5.19 represents training state graph

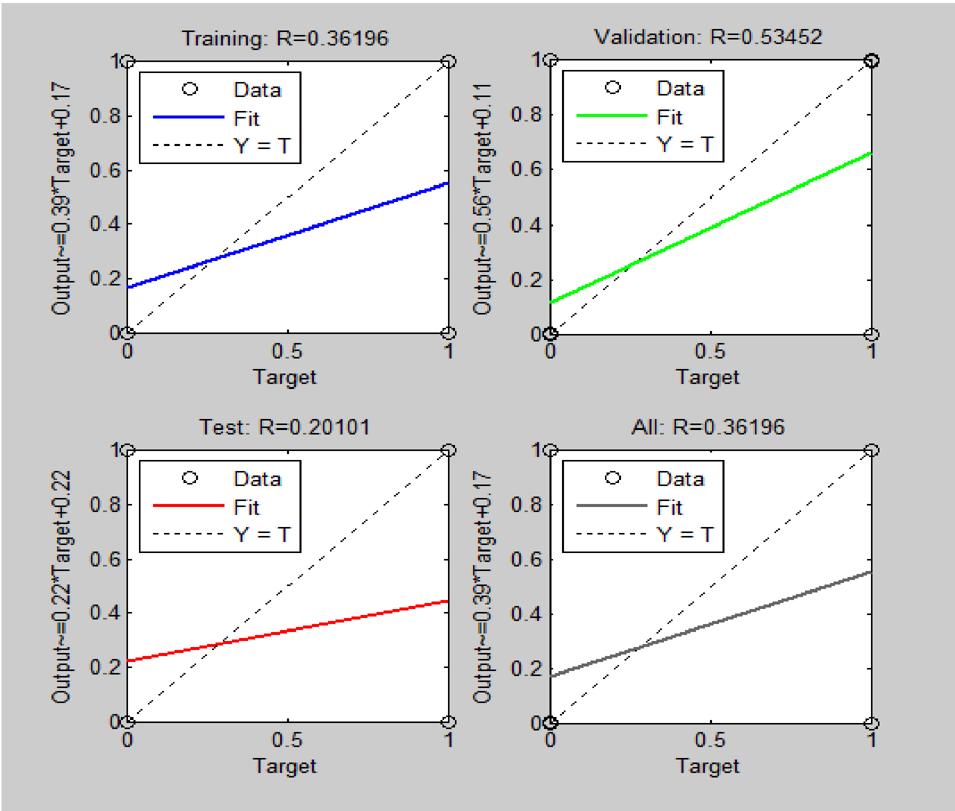

Fig 5.20 Regression graph(No. of neurons:5)

Figure 5.20 represents Regression graph

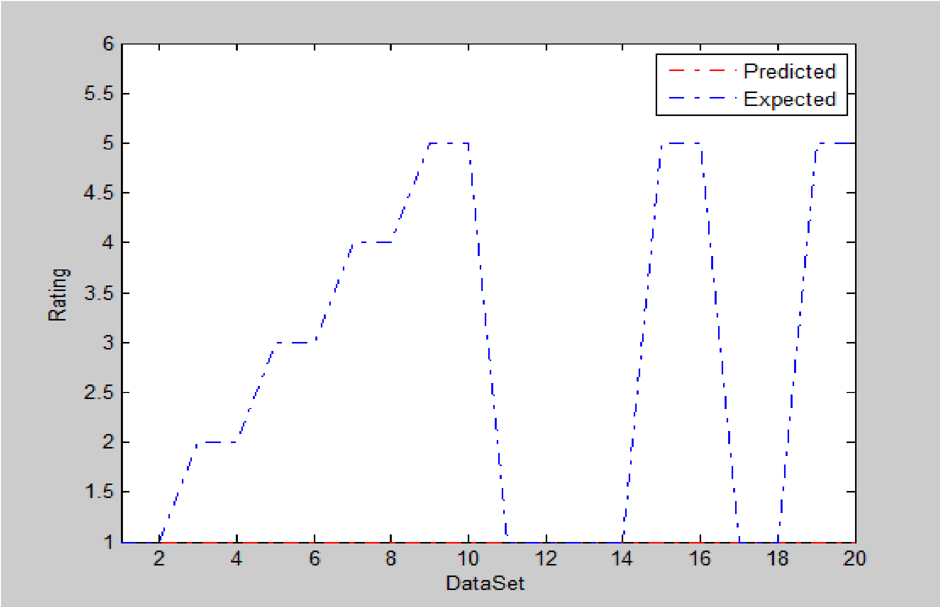

Fig 5.21 Predicted vs. Expected rating(No. of neurons:5)

Figure 5.21 represent MLP-Neural Network predicted rating and expected rating

Mean Absolute Error(MAE) on Expected rate and Predicted rate is : 1.8

- Number of neuron at hidden layer: 10

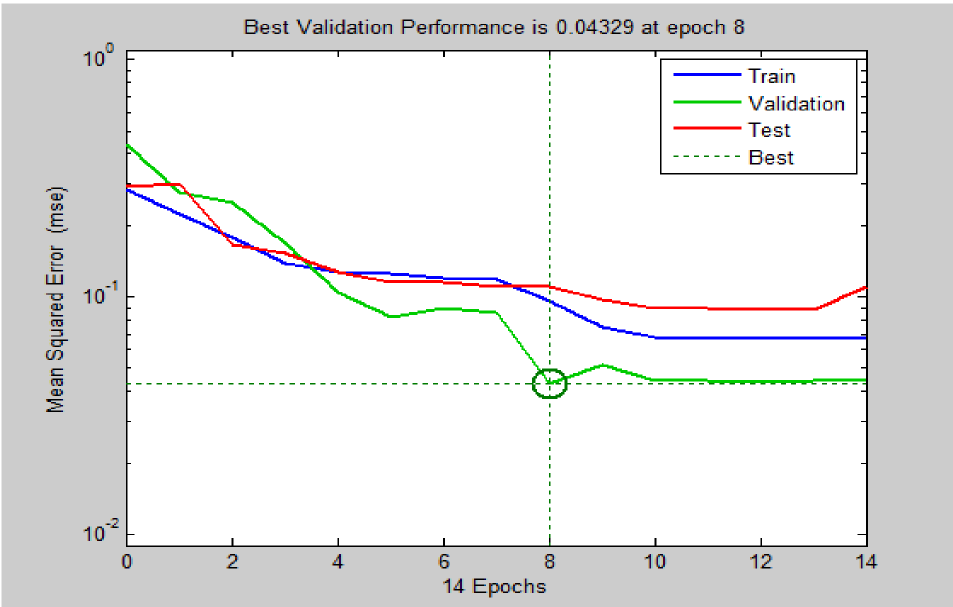

Fig 5.22 Performance graph(No. of neurons:10)

Figure 5.22 represent performance graph

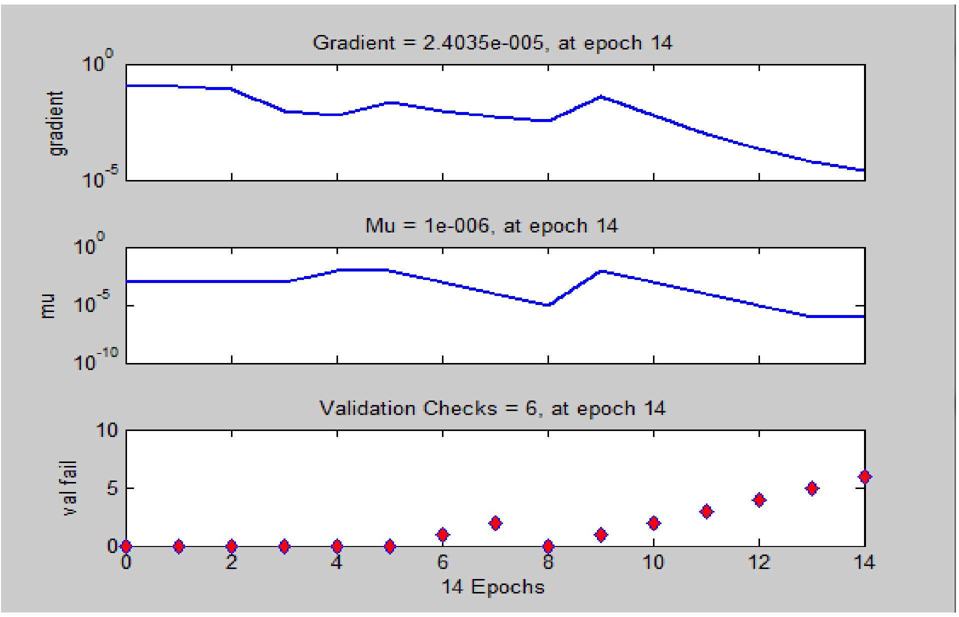

Fig 5.23 Training state graph(No. of neurons:10)

Figure 5.23 represent training state graph

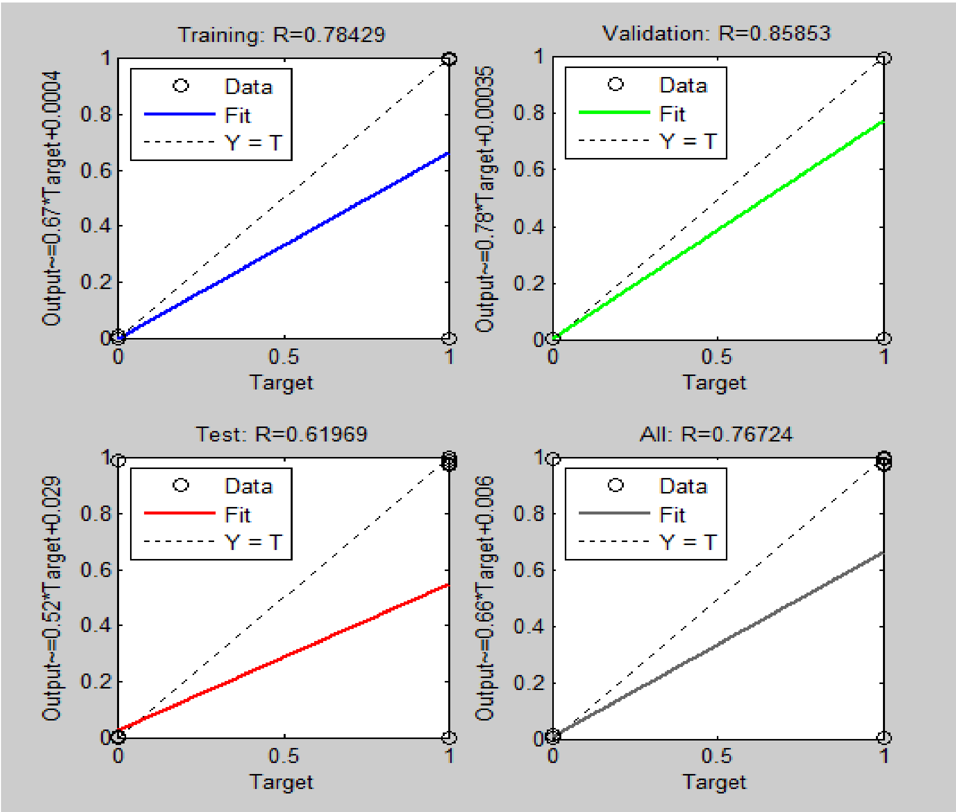

Fig 5.24 Regression graph(No. of neurons:10)

Figure 5.24 represent regression graph

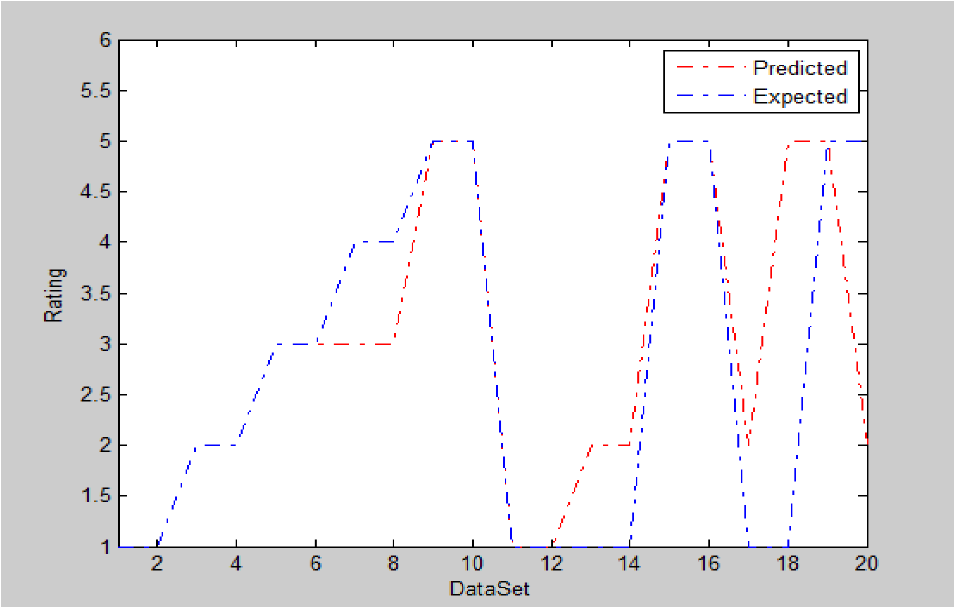

Fig 5.25 Predicted vs. Expected rating(No. of neurons:10)

Figure 5.25 represent MLP-Neural Network predicted rating and expected rating

Mean Absolute Error(MAE) on Expected rate and Predicted rate is : 0.600

- Number of neuron at hidden layer: 15

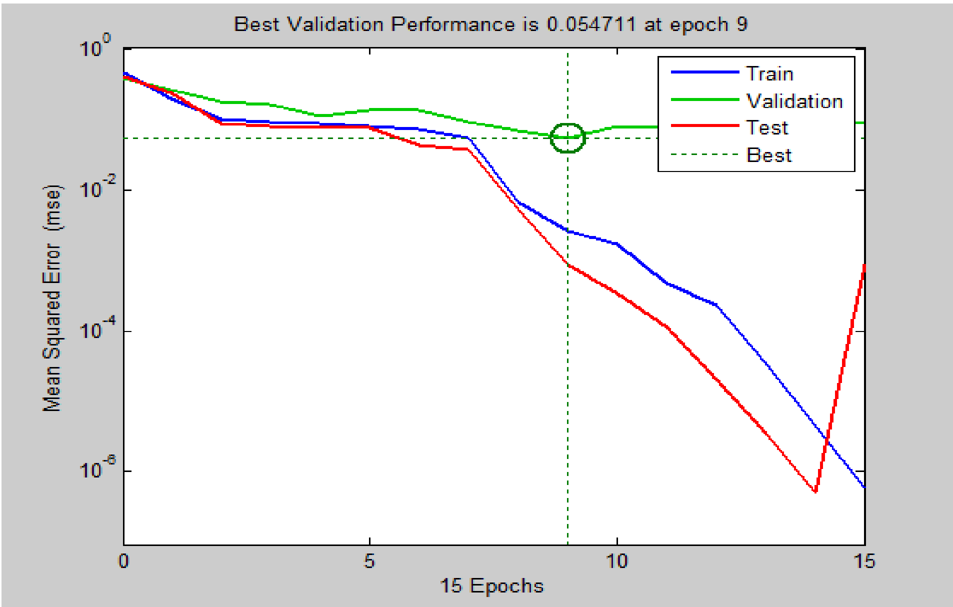

Fig 5.26 Performance graph(No. of neurons:15)

Figure 5.26 represents performance graph

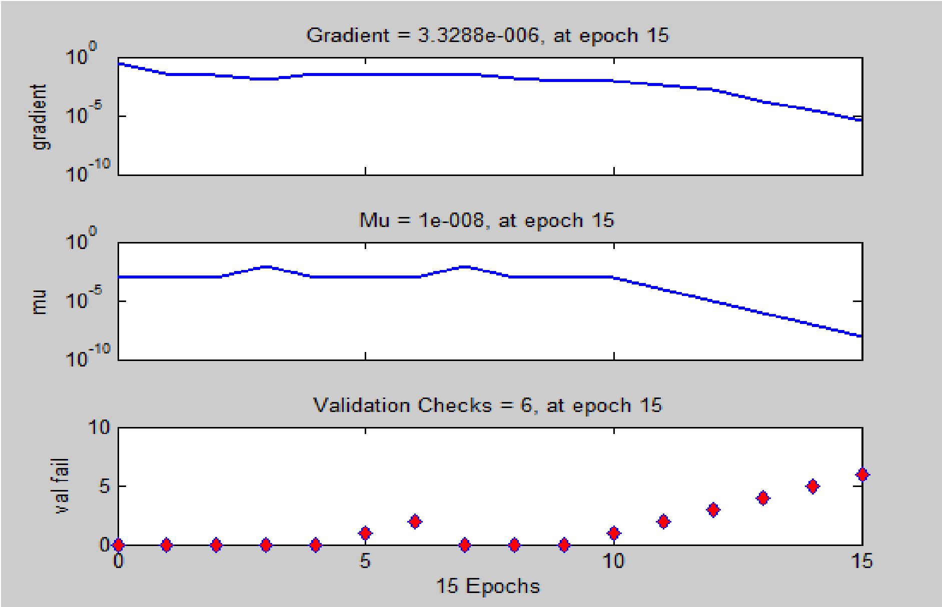

Fig 5.27 Training state graph(No. of neurons:15)

Figure 5.27 represents training state graph

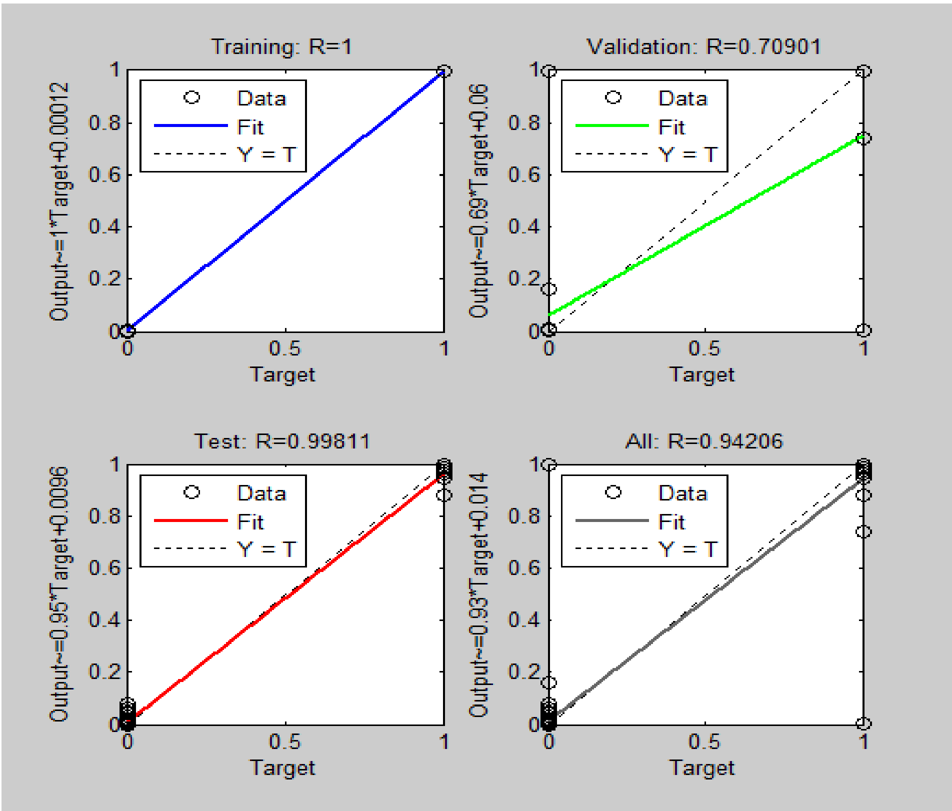

Fig 5.28 Regression graph(No. of neurons:15)

Figure 5.28 represents regression graph

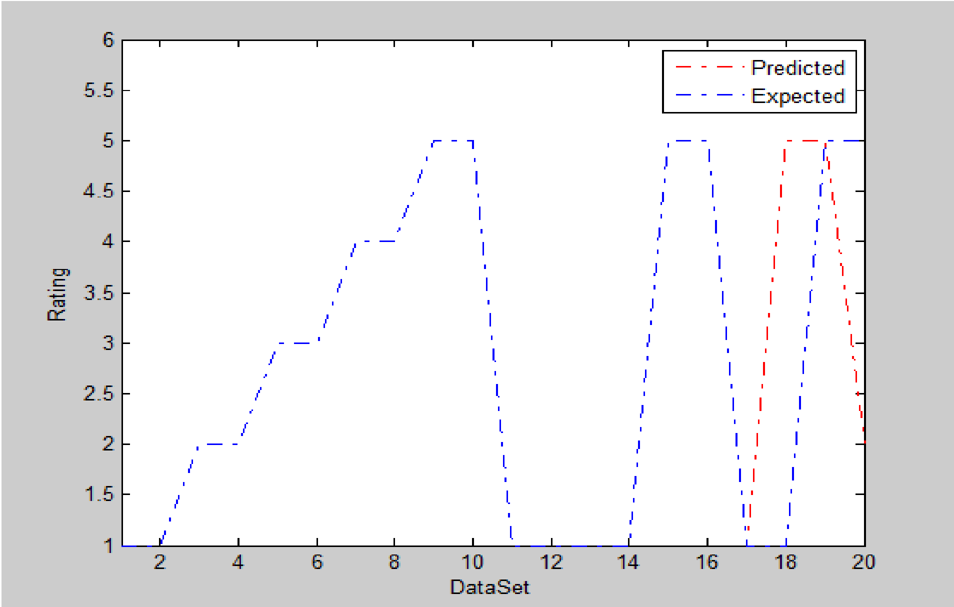

Fig 5.29 Predicted vs. Expected rating(No. of neurons:15)

Figure 5.29 represent MLP-Neural Network predicted rating and expected rating

Mean Absolute Error(MAE) on Expected rate and Predicted rate is : 0.35

- Number of neuron at hidden layer: 20

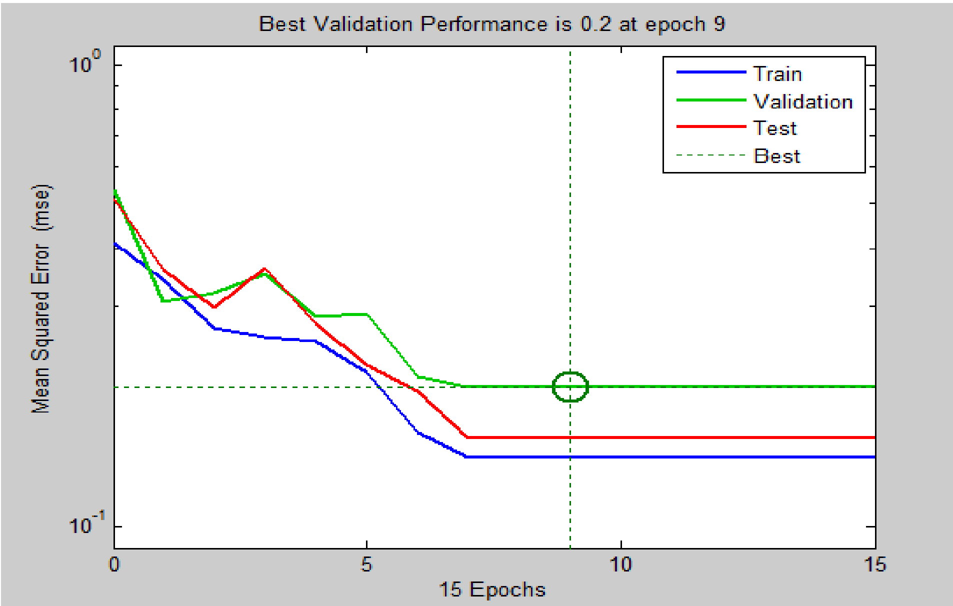

Fig 5.30 Performance graph(No. of neurons:15)

Figure 5.30 represents performance graph

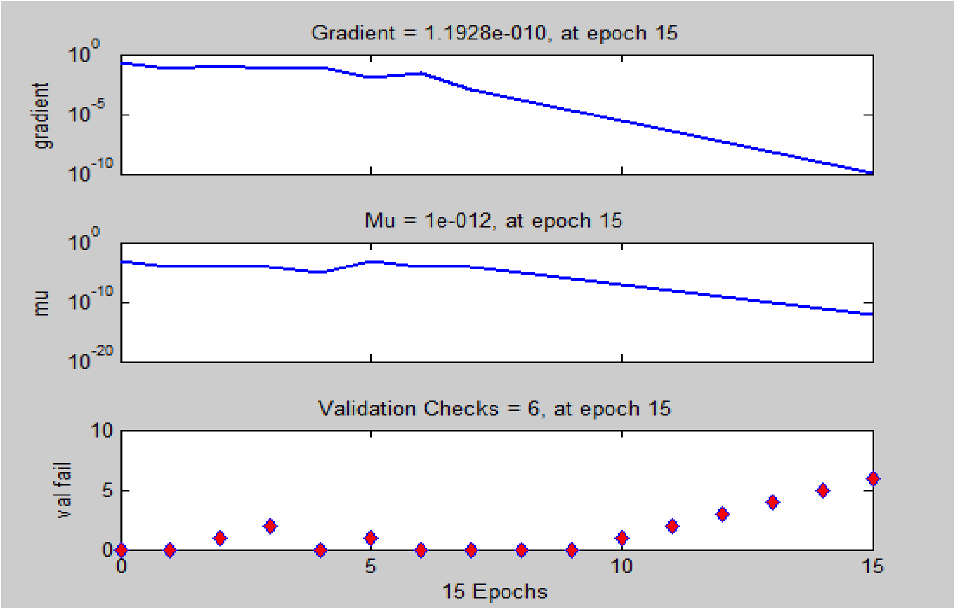

Fig 5.31 Training state graph(No. of neurons:15)

Figure 5.31 represents Training state graph

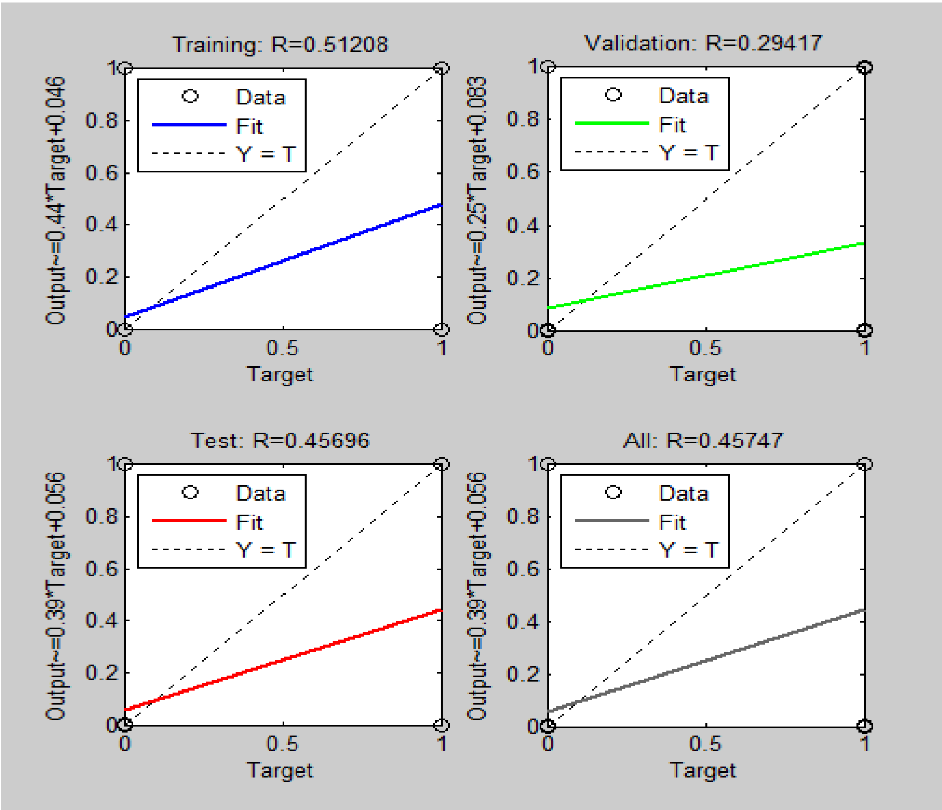

Fig 5.32 Regression graph(No. of neurons:15)

Figure 5.32 represent regression graph

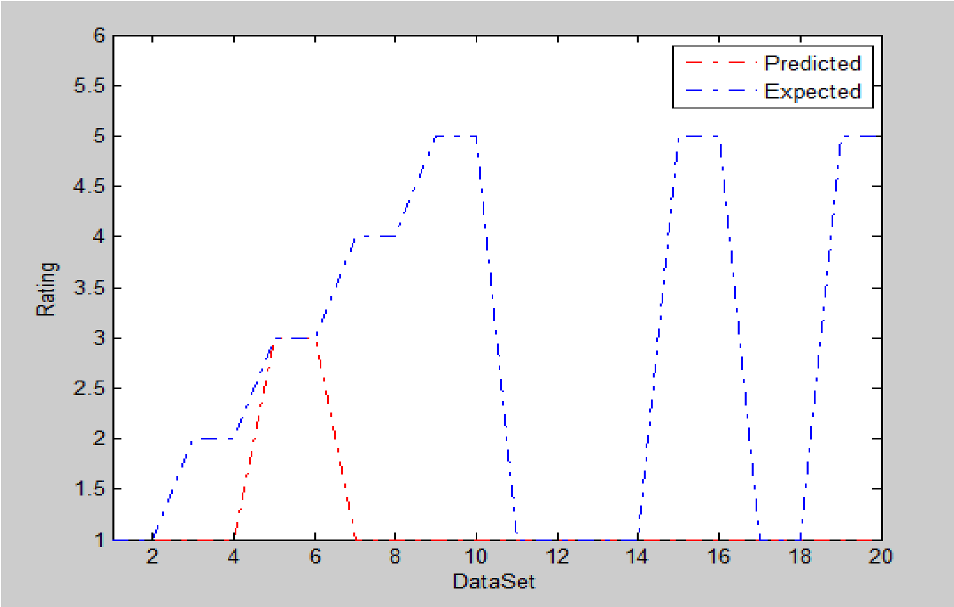

Fig 5.33 Predicted vs. Expected rating(No. of neurons:15)

Figure 5.33 represent MLP-Neural Network predicted rating and expected rating

Mean Absolute Error(MAE) on Expected rate and Predicted rate is : 1.6

5.3.1 MLP Neural Network performance summary for different number of neuron present at hidden layer

Number of neuron at Hidden Layer	MAE on rating
5	1.8
10	0.60
15	0.34
20	1.6

Table 5.3 Summary of MLP Neural Network performance for different number of neurons at hidden layer

From table 5.3, result shown that when number of neurons at hidden layer is 15 then MLP-Neural Network provides optimal results. Therefore, 15 neurons can be used for optimum prediction and has a better chance to stabilize at the most optimal global solution.

Chapter 6

Conclusion

The current work has been focused on personalized recommendation to customer by using book attributes as well as customer attributes. The proposed approach efficiently rate books. By considering book attributes as well as customer attributes together in process of recommendation serendipity factor has significantly increased. For further usage of our approach, by apply K-means on customer attribute e.g. age. We can cluster all existing customer according to considered attribute. By analysing clusters, we can easily determine from which cluster current customer belong; whom we have to recommend books. By obtaining details of the books purchased by all existing customer from cluster, which associated with current customer . We can use these books attributes and current customer attribute to evaluate books rating and can recommend books to him as a result of collaborating filtering and content based recommendation. Current customer attributes can be easily obtained by cookies, integrated profile and registration information or if in any case system not able to obtain current customer attribute, we can use default values which based on analysis.

Our purposed system effectively handle recommendation technique challenges such as:

- Frist Rater: Our system can effectively rate any new book, no prior knowledge of review associated with new book required. New book attribute and current customer attribute are sufficient for rating.
- Dynamic Nature: Our system support incremental training, which can easily adopt dynamic nature of customer.
- Data sparsity: This problem can be easily resolve by combine our proposed approach with K-means clustering for recommendation.
- New User: System can able to recommend new user just by obtaining its attribute.
- Scalability: In our approach, no need to save training set because Neural Network support incremental training. So as number of books and existing users increase, it doesn't affect much on acceptable performance.
- Synonyms: Problem is resolved by using Unique ID (ISBN) in our approach.

- Grey Sheep and Black sheep: problem can resolve by considering suitable customer attribute for k-means clustering, which collaborate similar customers together for recommendation.
- Fraud : shilling attacks can be avoided by using security parameters such as Logins, firewalls and Object oriented programming approach with our define methodology, which implements data hiding, encapsulation, inheritance, polymorphism etc.
- Cold start: can be resolve by training neural network on popular books initially, after construction of initial database, neural network weights can updated according customer rating pattern.

The future scope of the work is to explore other items such as clothes, accessories, gadgets for personalized recommendation to individual customer. We can also use clustering technique to retrieve important training samples, which have significant effect on training phase of neural network. So, process of classification and prediction can be improved further.

References

[1] Stefano Ceri, Alessandro Bozzon, Marco Brambilla, Emanuele Della Valle, Piero Fraternali, Silvia Quarteroni, "Web Information Retriveal", Data-Centric Systems and Applications 2013, pp 39-56,Springer 2013

[2] http://en.wikipedia.org/wiki/Cosine_similarity

[3] F.Ricci,"Mobile Recommender System." IT & Tourism, vol 12, pp 205-231,2010

[4] P.Melville and V.Sindhwani,"Recommender /systems", in Encyclopedia of Machine Learning, S. Verlag, ED.,ed Berlin:Springer, 2010, pp.1-9

[5] Asela Gunawardana and Guy Shani. 2009. A Survey of Accuracy Evaluation Metrics of Recommendation Tasks. Journal of Machine Learning. Volume 10 (December 2009), pp. 293 5 – 2962.

[6] S. Oh, Z. Lei, W.-C. Lee, P. Mitra, and J. Yen, “CV-PCR: a context-guided value-driven framework for patent citation recommendation,” in Proceedings of the 22nd ACM international conference on Conference on information & knowledge management, 2013, pp. 2291–2296.

[7] P. Brusilovsky, R. Farzan, and J. Ahn, “Comprehensive personalized information access in an educational digital library,” in Digital Libraries, 2005. JCDL’05. Proceedings of the 5th ACM/IEEE-CS Joint Conference on, 2005, pp. 9–18.

[8] S. Das, P. Mitra, and C. L. Giles, “Similar Researcher Search’ in Academic Environments,” in Proceedings of the JCDL’12, 2012, pp. 167–170.

[9] D. Faensen, L. Faultstich, H. Schweppe, A. Hinze, and A. Steidinger, “Hermes: a notification service for digital libraries,” in Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries, 2001, pp. 373–380.

[10] Z. Yang and B. D. Davison, “Venue Recommendation: Submitting your Paper with Style,” in Machine Learning and Applications (ICMLA), 2012 11th International Conference on, 2012, vol. 1, pp. 681–686.

[11] CiteSeer: An Automatic Citation Indexing System C. Lee Giles, Kurt D. Bollacker, Steve Lawrence NEC Research Institute, 4 Independence Way, Princeton, NJ 08540

[12] User Modeling via Stereotypes by ELAINE RICH

[13] M. Barla, “Towards social-based user modeling and personalization,” Information Sciences and Technologies Bulletin of the ACM Slovakia, vol. 3, pp. 52–60, 2011.

[14] I. Weber and C. Castillo, “The demographics of web search,” in Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval, 2010, pp. 523–530.

[15] Y. Jiang, A. Jia, Y. Feng, and D. Zhao, “Recommending academic papers via users’ reading purposes,” in Proceedings of the sixth ACM conference on Recommender systems, 2012, pp. 241–244.

[16] D. Buttler, “A short survey of document structure similarity algorithms,” in Proceedings of the 5th International Conference on Internet Computing, 2004.

[17] Mohammand Hamidi Esfahani,Farid Khosh Alhan,"New Hybrid Recommendation system Based On c-means Clustering Method", in proceedings 5th conference on Information and Knowledge Technology,2013

[18] Jianying Mai,Yongjian Fan, Yanguang Shen," A Neural Networks-based Clustering Collaborative Filtering Algorithm on E-commerce Recommendation System", in Proceeding of the International Conference on web Information Systems and Mining, 2009

[19] C. Nascimento, A. H. Laender, A. S. da Silva, and M. A. Gonçalves, “A source independent framework for research paper recommendation,” in Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries, 2011, pp. 297–306

[20] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, “[Using collaborative filtering to weave an information Tapestry],” Communications of the ACM, vol. 35, no. 12, pp. 61–70, 1992.

[21] D. A. Maltz, “Distributing information for collaborative filtering on Usenet net news,” 1994.

[22] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, “GroupLens: an open architecture for collaborative filtering of netnews,” in Proceedings of the 1994 ACM conference on Computer supported cooperative work, 1994, pp. 175–186.

[23] M. Baez, D. Mirylenka, and C. Parra, “Understanding and supporting search for scholarly knowledge,” in Proceeding of the 7th European Computer Science Summit, 2011, pp. 1–8.

[24] Y. Lai and J. Zeng, “A cross-language personalized recommendation model in digital libraries,” The Electronic Library, vol. 31, no. 3, pp. 164–277, 2013.

[25] OReilly.Programming.Collective.Intelligence.Aug.2007 by Toby Segaran

[26] Data mining concepts and techniques, 2010 by Jiawei Han and Micheline Kamber

[27] Text Mining Classification, Clustering, and Applications by Ashok N. Srivastava Mehran Sahami

[28] http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/cmeans.html

[29] http://en.wikipedia.org/wiki/Neural_Network

[30] http://www- math.ucdenver .edu/~wcher owi /jcorner /isbn.html

[31] http://en.wikipedia.org/wiki/ISBN