The Genetic Algorithm based Recommender Systems

Sandeep Kaur, Mini Ahuja

Research Scholar, Dept. of Computer science and applications, Guru Nanak Dev University, Amritsar, India

Asst. Professor, Dept. of Computer science and applications, Guru Nanak Dev University, Amritsar, India

ABSTRACT: The rapid growth of the internet has significantly increased the number of social web site users. At the same time there is a noteworthy increase in demand of users for particular web. A recent statistical report shows that the 70% of the web traffic suffers from congestion problem. So evaluating the actual demand of given site may improve the performance of social sites. This paper has shown that the use of hybridization of data mining techniques can be done to improve the accuracy rate further for social network recommended system. Most of the existing techniques are limited to some significant features of social networks. The use or genetic algorithm has been ignored to improve the accuracy rate further for social networks. The accuracy rate of the existing methods is found to be poor so improvement is required to make them more consistent. So in order to reduce these limitations a new technique have been proposed which will utilize genetic algorithm to enhance the prediction rate further.

KEYWORDS: Social Networks, Recommender systems,Types of Recommendation system, Genetic algorithm, Random forest algorithm

I. INTRODUCTION

Web, the fastest growing network the entire world has actually identified, has revolutionized the entire world by letting thousands of people to communicate with each other over vast distances, and across all kinds of boundaries. In these times, there are many than 1,262,032,697 Net users throughout the earth [internetworldstats.com, 2008], i.e. about 19% of the world's populace now has use of the Net, and that number is continually increasing. That large quantity of Net users accesses billions of documents (Web pages) that are interconnected by Earth Large Web (Web for short) through the Internet. Nowadays the Web types the greatest and most commonly identified information resource that is readily available and searchable. That popular use of Web as a global information process has flooded us with a significant amount of information and information that are growing tremendously whilst the individual capacity to locate and read material remains constant [1].

Recommender techniques assist customers for making choices from numerous solutions; the target of these techniques would be to calculate consumer preferences and give predictions of appropriate information. Social recommender techniques make an effort to alleviate information and interaction clog through the use of numerous techniques that ultimately present the most appropriate and attractive information to users. These personalized suggestions based on social interactions or preferences are seen as a huge chance for vendors. Indeed, a survey of online retailers in 2010 found that over half in the pipeline on implementing recommendation features on their sites. collaborative selection (CF) techniques, are the most successful; however, some important disadvantages in them have now been described ,specially in standard strategies based on nearest friend algorithms, which display serious efficiency and scalability problems. Furthermore, the great amount of evaluations needed by these techniques to be able to give specific suggestions triggers the sparsity issue when evaluations from customers are insufficient [2].

II. SOCIAL NETWORKS

A social network is identified as a social design of individuals having relationship based on casual interests e.g. friendship and integrity. Social network system centers on the design and identification of on-line social support systems for people who reveal their interests and activities or those who are interested in searching the others' interests and activities. These sites, first, are used in order to making buddies and sharing ideas among people but today they are used in order to work and data sharing. Of course as enough time moves, two company and helpful environments are mixed and modify in to substance situations. As enough time moves two helpful and company environments are mixed

and modify in to a substance one. Social networks are trusted because of the experiences, knowledge values and needs. For instance, buddies talk about eateries and cinemas. This time about social support systems contains two crucial potential benefits for marketing as well :

•The advertisement displayed through social network can be looked at as more reliable. Really, people choose to count on recommendations provided by their buddies and neighbors because human character is more interested in what a buddy acquisitions, perhaps not an as yet not known individual, and it is likely that an individual utilizes a friend's ideas and is affected by his/her actions [3].

III. RECOMMENDER SYSTEM AND THEIR TYPES

Recommender systems were introduced as a computer-based wise method to cope with the problem of information and item overload. They may be used to effortlessly provide individualized companies in many e-business domains, benefiting equally the consumer and the merchant. Recommender Techniques will benefit the consumer by making to him suggestions on goods he is assumably planning to like. At the same time frame, the business is likely to be gained by the increase of income that'll typically happen when the consumer is shown with an increase of goods he may likely find appealing. The two basic entities which come in any Recommender System are an individual (sometimes also called customer) and them (also called item in the bibliography). A consumer is someone who utilizes the Recommender System giving his opinion about numerous goods and receives suggestions about new goods from the system. The aim of advice system is to choose information whose material are many relevant to the user's curiosity from a higher level of information accessible and presenting them in appropriate means for the user. A recommendation system makes suggestions about preferred website pages to a user. Social advice techniques collect scores of website pages from many people to make suggestions to a user concerned [3]. Various types of recommender systems are as follows:

A. *Content-Based System:* Content-based process proposes data by evaluating the available assets with individual pursuits, and the important thing issues will be the extraction of individual pursuits and the calculation of similarity. The system has functions such as for instance simplicity and performance.

B. *Collaborative Filtering:* Collaborative filter process could immediately filtration the data that the system could not analyze and represent, and has the capability to suggest up-to time information. Collaborative Filtering as a Device Learning problem utilizes a administered understanding algorithm on especially developed knowledge, whose illustration dimension has been decreased following the applying of dimensionality reduction techniques[4].

1) Cosine-based likeness: Otherwise named vector-based likeness, this definition considers a couple of items and their evaluations as vectors, and characterizes the comparability between them as the path between these vectors :

$$\text{sim(k,j)} = \cos(\vec{k}.\vec{j}) = \frac{\vec{k}.\vec{j}}{||\vec{k}||_2 * ||\vec{j}||_2} \quad (1)$$

2) Pearson (relationship)- based likeness: That likeness measure depends on what considerably the evaluations by standard clients for many goods digress from normal appraisals for anyone items:

$$\text{Sim(l,m)} = \frac{\sum_{s\epsilon S}(Z_{s,l}-\bar{Z}_l)(Z_{s,m}-\bar{Z}_m)}{\sqrt{\sum_{s\epsilon S}(Z_{s,l}-\bar{Z}_{s,l})^2}\sqrt{\sum_{s\epsilon S}(Z_{s,m}-\bar{Z}_m)^2)}} \quad (2)$$

3) Adjusted cosine similarity: That similarity estimation is a revised type of vector-based similarity where we take into the way that diverse clients have unique appraisals applications; therefore, several clients may charge items exceptionally when all is claimed in done, and the others can offer points decrease evaluations being an preference [5]. To get rid of this drawback from vector-based similitude, we take normal appraisals for every single simple customer out of every client's evaluating for the pair of points in question:

$$\text{Sim(l,m)} = \frac{\sum_{s\epsilon S}(Z_{s,l}-\bar{Z}_l)(Z_{s,m}-\bar{Z}_m)}{\sqrt{\sum_{s\epsilon S}(Z_{s,l}-\bar{Z}_{s,l})^2}\sqrt{\sum_{s\epsilon S}(Z_{s,m}-\bar{Z}_m)^2)}} \quad (3)$$

C. *Demographics-based:* the RS charges sources based on similarity between age (age, sexuality, state of residence) of people who rated the research high and these of an individual querying the RS. The key advantage of a demographics-based approach is that it generally does not involve a total history of customer appraisals of the sort expected by combined systems.

- Obtaining data difficulty: It is difficult to acquire personal infonnation from the users. Web users norn1ally prevent participating in a relationship with the Web sites. This is certainly caused by because of lack of belief in the solitude plan of today's Web sites. Therefore, users both withhold private data or offer false data.
- Grey sheep issue: It identifies users with opinions that are uncommon, and therefore they do not recognize or argue consistently with any number of users.

D.*Social Network-based:* the RS charges sources based on choices suggested by people discussing a social relationship with an personal querying the RS. This approach is generally used along with collaborative filtration techniques [6].

E.*Hybrid RS:* the unit charges those items to be advised based on a variety of the techniques identified above. An entire classification of cross techniques, record numerous hybridization techniques to mix couples of recommender algorithms [7].

IV. GENETIC ALGORITHM

Genetic algorithm (GA), first proposed by John Holland [1975], are nature-inspired optimization strategy that can be advantageously used for several optimization problems. Genetic algorithm begins with an original population of individuals generated at random. Each individual in the population presents a possible solution to the issue under concern and every individual is considered for fitness with regards to the optimization task to be solved. The persons evolve through successive iterations, called generations. Throughout each era, every individual in the population is considered using fitness function. Range within the people is completed in a fitness-proportionate way: the more fit a person, the significantly much more likely it will be selected for reproduction in tothese time [Back et aI., 2000], ergo increasing successive generations. The population of the following era is done through genetic operators, specifically crossover and mutation. Crossover wants two people and yields two new persons while mutation alters one specific to produce a simple new offspring. The procedure remains before termination situation is satisfied. For recommender systems, an individual designs are changing over amount of time in gentle of the most up-to-date ratings [8]..

V. AMAZON RECOMMENDER SYSTEM

Recommendation methods are most readily useful observed because of their use on e-commerce Those sites, wherever they pick feedback in relation to a customer's passions to make a listing of recommended items. Several applications use only those things that clients purchase and obviously cost to signify their passions, nevertheless they have the ability to also use different characteristics, including services and products regarded, demographic information, topic passions, and favorite artists. At Amazon.com, the utilization of advice methods to modify the online hold for each simple customer. The hold radically changes predicated on customer passions, showing coding games to a computer software manufacture and baby toys to a new mother. The click-through and transformation charges — two crucial measures of Web-based and e-mail marketing achievement — considerably surpass those of untargeted content such as for instance marketing ads and top-seller lists. E-commerce advice methods frequently function in a hard environment [9].

A. *Decision Tree*

A decision tree is a graphical representation of possible solutions to a decision centered on specific conditions. It's named a decision tree since it begins with an individual field (or root), which then branches down into numerous options, just like a tree. Decision trees are valuable, not merely because they're graphics that allow you to see everything you are thinking, but also since making a decision tree needs an organized, reported thought process. Usually, the biggest restriction of our choice making is that people can only just select from the known alternatives. Decision trees support formalize the brainstorming process therefore we can recognize more potential solutions [10].
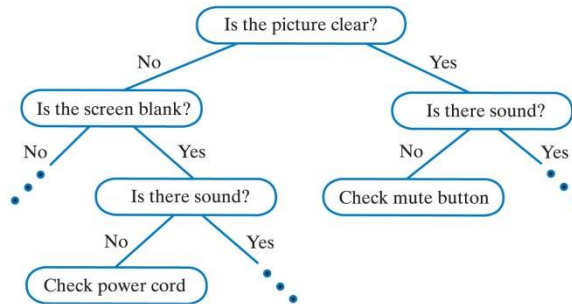
Figure 1: Decision Tree

B. *Random forest algorithm*

Work with synthetic neural frameworks, generally referred to as neural frameworks, has been motivated from the solace of its introduction by the affirmation that the cerebrum figures in an absolutely unmistakable way from the ordinary mechanized computer. Consider, for event, human vision, which will be a data handling undertaking. It is the limit of the visual framework to give a layout of the surroundings about us and, more huge, to give the information we need to talk with nature [11].
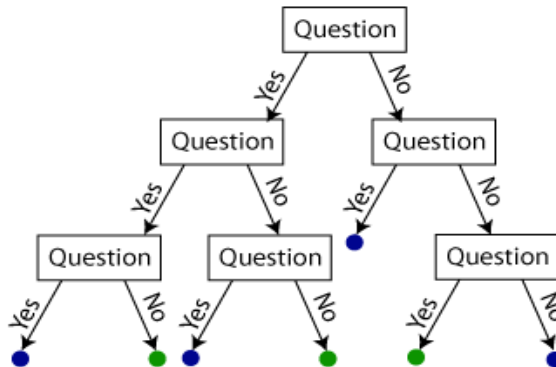
Figure 2: Random Forest

C. *Linear model*

The restorative methodology is often utilized as a light emanating diode method for learning. Direct scientific techniques are normal a part of that learning procedure. In the common, real, and sociologies, alongside in operation and style, direct structures are of phenomenal use inside similarly arranging periods of look at and assessment of the following information. In different spots it offer a quick prologue to straightforward and various direct relapse structures, and investigation of-change (ANOVA) models [12].
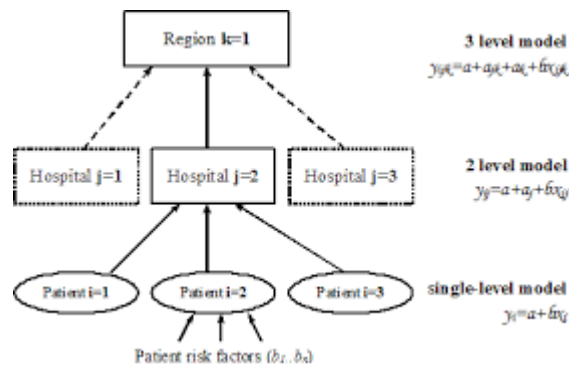
Figure 3: Linear model

1) *Multiple linear regression model:*

The impact is for the most part energized by a few indicator variable. Case in point, the produce of a seed may be set up by the aggregate sum of nitrogen, potash, and phosphate composts utilized. These elements are overseen by the experimenter, by and by the give are frequently settled by wild components, for example, for case these connected to climate [13].

VI. LITERATURE SURVEY

Guy Shani et al. [1] proposed that typical Recommender systems adopt a fixed view of the recommendation method and address it as a forecast problem. To succeed in exercise, an MDP-based Recommender process should employ a powerful original design; and the majority of this paper is concerned with the technology of this kind of model. Te suggested use of an n-gram predictive design for generating the original MDP. Our n-gram design causes a Markov-chain model of person conduct whose predictive precision is higher than that of existing predictive models.

Panagiotis Symeonidis et al. [2] proposed that providing reason to a suggestion allows credibility to a recommender system. Some recommender techniques (Amazon.com etc.) take to to explain their tips, in an effort to restore client popularity and trust. But their explanations are bad, as they are centered entirely on rating knowledge, ignoring this content data. Our model program MoviExplain is really a film recommender program that gives both accurate and justifiable recommendations.

G. Adomavicius et al. [3] proposed that context-aware recommender systems (CARS) make more applicable tips by establishing them to the specific contextual situation of the user. This informative article examines how contextual information may be used to produce more intelligent and useful recommender systems. It offers an summary of the multifaceted concept of context, examines a few approaches for incorporating contextual information in endorsement method, and demonstrates the utilization of such approaches in several request parts where several types of contexts are exploited.

Nguyen Thai-Nghe et al. [4] proposed a book method which uses recommender system techniques for academic information mining, particularly for predicting scholar performance. To validate this method, we assess recommender system techniques with conventional regression techniques such as for example logistic/linear regression by utilizing academic information for intelligent tutoring systems. Experimental results show that the proposed method can increase prediction results.

Sanjeevan Sivapalan et al. [5] proposed that recommender techniques are adjusting from novelties used by several E-commerce sites, to serious business methods that are re-shaping the world of E-commerce. They presented an explanation of how recommender techniques help E- commerce sites increase sales, and analyze six sites that use recommender techniques including many sites that use multiple recommender system.

Mustansar Ali Ghazanfar [6] proposed that Recommender methods apply equipment learning techniques for filtering invisible information and can estimate whether a user would really like confirmed resource. The Demographic recommender methods classify people or goods centered on the personal attribute and produce suggestion based on demographic categorizations. These methods suffer from scalability, data sparsity, and cold-start issues leading to poor quality guidelines and decreased coverage.

Paolo Massa et al. [7] proposed that recommender Systems based on Collaborative Selection recommend to people goods they could like. They propose to displace this with the use of a trust full, an algorithm in a position to propagate trust within the trust system and to calculate a trust fat that may be used instead of the likeness weight. An scientific evaluation on Epinions.com dataset suggests that Recommender Systems that take advantage of trust data are the utmost effective in expression of precision while preserving a good coverage.

J. Bobadilla [8] proposed that recommender techniques allow us in similar with the web. These were initially predicated on demographic, content-based and collaborative filtering. This short article has an breakdown of recommender techniques along with collaborative selection strategies and algorithms; in addition, it explains their development, offers an original classification for these techniques, identifies aspects of future implementation and advances particular places picked for past, present or future value.

International Journal of Innovative Research in Computer and Communication Engineering

VII.   SIMULATION RESULTS

In this we compare the correlation parameter r of linear model, decision tree, random forest, neural network of existing system with proposed system which uses genetic algorithm that is it will gives best values in comparison to others.

A.*correlation coefficient(r):*The correlation coefficient is just a evaluate that decides the amount to which two variables movements are associated. The range of values for the correlation coefficient is -1.0 to 1.0. If a calculated correlation is greater than 1.0 or less than -1.0, an error has been made. A correlation of -1.0 suggests a perfect negative correlation, while a correlation of 1.0 suggests an ideal good correlation.

$$N_{yx} = \frac{cov(p_y, p_x)}{\delta_y \delta_x} (4)$$

B. *Coefficient of determination(R):*The coefficient of determination of a linear regression design could be the quotient of the variations of the repaired values and seen values of the dependent variable. When we denote z_i considering that the seen values of the dependent variable, as its suggest, and considering that the repaired cost, then the coefficient of perseverance is:

$$r^2 = \frac{\sum(\hat{z}_i - \bar{z})^2}{\sum(z_i - \bar{z})^2} \ (5)$$

C. *RMSE (Root Mean Square Error):* Root-mean-square error is a measure of the differences between values predicted by a model or an estimator and the values actually observed. It can be explained as:

$$\text{RMSE} = \sqrt{\frac{1}{SN}\sum_{j=1}^{S}\sum_{i=1}^{N}(f(j,i) - f'(j,i))^2} \ (6)$$

D. *MSE(Mean Square Error):*The MSE assesses the caliber of an estimator (i.e., a mathematical purpose mapping a style of information to a parameter of the population from that the info is sampled) or a predictor (i.e., a purpose mapping arbitrary inputs to a sample of prices of some arbitrary variable). Classification of an MSE varies centered on whether one is describing an estimator or even a predictor. The MSE of the predictor might be estimated by:

$$\text{MSE} = (\frac{1}{n}\sum_{j=1}^{n}(\hat{Z}_J - Z_j)^2) \ (7)$$

i.e., the MSE is the mean $(\frac{1}{n}\sum_{j=1}^{n}.)$of *the* square of the problems $(\hat{Z}_J - Z_j{}^2).$).This is an simply computable amount for a specific test (and thus is sample-dependent).

Table1 showing the analysis Mean square error (MSE) and Root mean square error (RMSE) respectively as it is known that needs to be minimized,The values of Mean square error parameter also needs to be minimized.

Table1: MSE & RMSE COMPARISON TABLE

Parameters	Decision Tree	Linear model	Neural network	Random forest	Genetic algorithm
MSE	0.28	0.06	0.06	0.06	0.83
RMSE	1.14	1.43	1.44	1.42	0.37

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

It is clearly shown in the table that the MSE of the proposed technique is lower among other algorithms, The values of RMSE of proposed genetic based algorithm is also lesser than the earlier techniques. Therefore proposed technique is effective than Decision tree, Linear model, neural network, random forest and genetic algorithmas shown in fig 4.
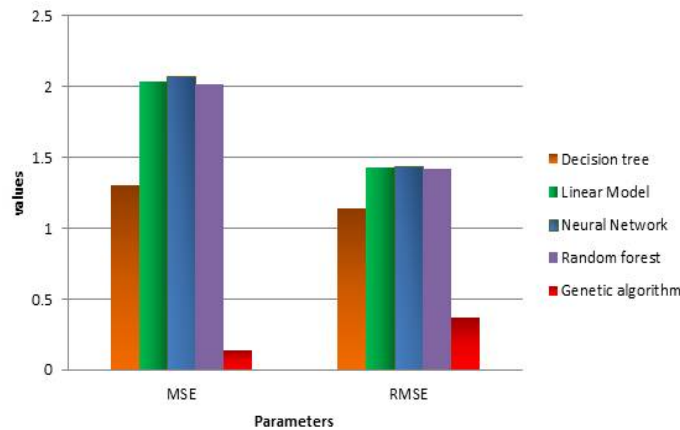

Figure 4: MSE & RMSE GRAPH

Table2 showing the analysis Co-relation coefficient (r) and Coefficient of determination (R) respectively as it is known that needs to be maximized, Co-relation coefficient also needs to be maximized.

Table2: r &R COMPARISON TABLE

Parameters	Decision Tree	Linear model	Neural network	Random forest	Genetic algorithm
R	0.53	0.25	0.24	0.25	0.93
R	0.28	0.06	0.06	0.06	0.83

It is clearly shown in the graph that the values of Co-relation coefficient parameter of the proposed technique is higher among other algorithms, the values of Coefficient of determination parameter of proposed genetic based algorithm is also higher than the earlier techniques. Therefore proposed technique is effective than Decision tree, Linear model, neural network, random forest and genetic algorithm as shown in fig 5.
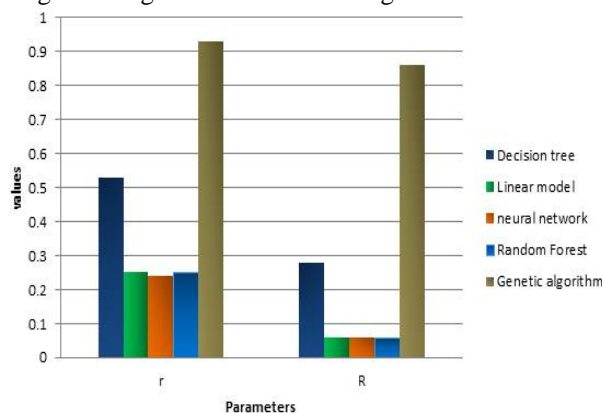

Figure5: r & R GRAPH

VIII. CONCLUSION

In this paper we have evaluated the performance of decison tree, Neural network and linear regression based machine learning algorithms for prediction of demand in social networks. For improving the performance of neural

network algorithm we have used genetic algorithm. The comparison is shown between the proposed algorithm as well as the existing algorithm. The proposed work shows better results than the existing algorithm.

REFERENCES

1. An MDP-based Recommender System by Guy Shani, Ronen I. Brafman, David Heckermanin J. Mach. Learn. Res. (2012)
2. MoviExplain: A Recommender System with Explanations byPanagiotis Symeonidis, Alexandros Nanopoulos, Yannis Manolopoulos
3. Context-aware recommender systems by G. Adomavicius, A. Tuzhilin
4. Recommender system for predicting student performance by Nguyen Thai-Nghe, Lucas Drumond, Artus Krohn-Grimberghe, Lars Schmidt-Thieme
5. Recommender systems in e-commerce by Sanjeevan Sivapalan, Alireza Sadeghian, Hossein Rahnama, Asad M. Madni
6. A scalable, accurate hybrid recommender system by Mustansar Ali Ghazanfar, Adam Prugel-Bennett
7. Trust-aware recommender systems by Paolo Massa, Paolo Avesani
8. Recommender systems survey by J. Bobadilla, F. Ortega, A. Hernando, A. Guti
9. Comparative recommender system evaluation by Alan Said, Alejandro Bellogín
10. Hybrid web recommender systems by Robin Burke
11. Application of Dimensionality Reduction in Recommender System -- A Case Study by Badrul M Sarwar, George Karypis, Joseph a Konstan, John T Riedl
12. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions by Gediminas Adomavicius, Alexander Tuzhilin
13. Hybrid recommender systems: Survey and experiments by Vipul Vekariya, G. R. Kulkarni
14. A Social Network-Based Recommender System (SNRS) by J. He, W.W. Chu
15. Recommender Systems by Linyuan Lü, Matus Medo, Chi Ho Yeung, Yi-Cheng Zhang, Zi-Ke Zhang, Tao Zhou