Genetic algorithm based efficient social network Recommender System

Sandeep Kaur*1, Mini Ahuja*2
*Student & Computer Science and Engineering & Guru Nanak Dev University
Amritsar, India

Abstract —The rapid development of the net has relatively increased the quantity of social web site users. At the same time frame there is a remarkable increase in demand of customers for unique web. A recently available statistical record suggests that the 70% of the internet traffic suffers from obstruction problem. Therefore evaluating the particular demand of provided site may possibly improve the efficiency of social sites. That paper shows that the use of hybridization of knowledge mining techniques can be done to enhance the precision rate more for social network advised system. Most of the present techniques are restricted to some significant top features of social networks. The employment or genetic algorithm has been dismissed to enhance the precision rate more for national networks. The precision rate of today’s techniques can be acquired to be poor therefore development is necessary to cause them to become more regular. Therefore in order to reduce these constraints a fresh approach have been planned which will use genetic algorithm to boost the prediction rate further.

Keywords — Social Networks, Recommender systems, Types of Recommendation system, Genetic algorithm, Random forest algorithm.

I. INTRODUCTION

The ‘new research of networks’, or complicated networks research, came to be in the late 1990s when large electronic information models and successful computational resources allowed investigating large degree complicated systems. In the network approach to complicated programs the communicating components and their communications are mapped to network nodes and hyperlinks, see fig 1.1. This is completed in the trust that the primary qualities of the device aren’t in the details of the weather but rather in their relationship patterns. Even though this simplification might appear primitive, the simplicity and flexibility of the network approach have caused it to be one of the major tools in the modern complicated program research. Recently, the target of network research has been shifting towards performance and dynamics of networks. Performance is frequently linked to densely linked models of nodes, which are called areas, adventures, or clusters. The basic problem is always to exactly define what takes its community and how such structures could be successfully determined in large true networks. Another concurrent focal topic of network research is dynamics – improvements of the network framework and/or dynamics of techniques happening on networks. Over the last decade, network versions have turned out to become a useful software in the research of network dynamics, either by giving possible artificial network structures for simulations of dynamical techniques, or by seeking to explain the growth and modify of the framework of true networks. Presently, there’s a large amount of versions which effort to recapture the traits of various types of real-world networks. Social networks that road the connectivity patterns of an individual are an important type of networks and numerous versions have now been planned to explain their features using recommender system [1].

Figure 1: (a) A pulling featuring the Königsberg links on the 18th century, and (b) the related system addressing the landmasses and the linking bridges.

II. SOCIAL NETWORKS

The social network can be an exceptional way to get clients and retain current ones. The true value is the way marketers can interact making use of their market at an individual level. In place of simply ‘Offer! Offer! Offer!, cultural network is all about the two-way communication which helps to construct an extended 9 expression relationship with the customers (Bharti, 2011). From one-way
communication to multi-user interaction threads, the social media is growing the level of communication for both individuals and institutions equally (Coleman, 2013). Social media marketing has taken the web from one-way communication to strong involvement in real time interactions (Soils, 2010). Firms have the ability to transform the methods they choose to develop the products and markets. Social media marketing has changed their means of communication with the customers, means of getting potential customers, means of preserving them and finally the means of promotion. It's rightly said, “a great audio should be a great listener” and the social media is genuine the firms to ‘hear more’ to the customers before performing in accordance. The firms are adopting new strategies to incorporate social media to their regular interaction with the customers (Jones and Monk, 2009). By increasing the publicity of a brand related context to the prospective clients, the cultural networking website marketing leads to viral spreading of the messages about a brand, which provides as a very powerful promotional strategy. Viral marketing and brand growth are regarded as being co-related. They're the outcomes of an effective social networking marketing strategy. Customers heavily depend on the buy suggestions of others, believing in others' activities and later discussing their very own activities to supply suggestions to others. Word-of-mouth promotion is recognized as while the strongest, many influential and most powerful tool of promotion. Marketers used to pay an incredible number of their budget on public relations and different connected actions to facilitate the word-of-mouth promotion. Social media marketing marketing is just a relatively low priced, rapid and efficient solution to spread word-of-mouth amongst the online clients. Viral marketing stimulates the word-of-mouth interaction by using the social network internet sites and the clients' purposes to generally share the communications with their friends (peers) on these networks. The viral campaign strategy is more prone to succeed when the communications are transported through the friends who share common interest, thus appealing to the mark clients about the brand. The amount of referrals developed by the clients leads to cost-effective promotional techniques.

A. Applications of Social Networks

Generally, centered on kind of their application, social network environments are split into following teams:

1. Friendly surroundings wherever persons mostly conversation or reveal ideas. These teams are formed to be able to produce relationships and reveal photos and movies among friends for free.

2. Company surroundings wherever consumers perform business responsibilities as sellers or costumers. For sellers, these teams are formed due to monetary advantages and for costumers for providing needs.

3. Common surroundings which are formed to be able to examine and reveal information and create understanding as well.

4. Compound setting wherever there's a combination of friendly setting and business or medical environment users.

III. RECOMMENDER SYSTEMS

The Recommender techniques were introduced as a computer-based intelligent approach to deal with the issue of information and solution overload. They could be applied to effectively give individualized firms in lots of e-business domains, benefiting equally the client and the merchant. Recommender Practices may gain the client by producing to him a few ideas on goods he is assumably planning to like. At the same time frame figure, the business enterprise is likely to be gained by the increase of income that'll usually happen when the buyer is found having an raise of goods he might probably discover appealing. The 2 basic entities that can come in just about any Recommender System are an individual (sometimes also called customer) and them. A client is someone who employs the Recommender Process giving his see about numerous things and gets a few ideas about new things from the machine. 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. Web information advice fall into three categories:

1. Recommendation on the basis of the similarity of website pages
2. Recommendation on the basis of the choices and behaviors of individual groups
3. Recommendation on the basis of the preference and behaviour of personal individual.

IV. GENETIC ALGORITHM

The Genetic Algorithm Method of Recommender Programs The decision of similarity purpose is an important situation for the device and depends seriously on the issue itself.
The main difficulty in profile matching computation applying Formulae is due to the inclusion of all of the ranked films for a certain user in the profile directly. In line with the lightweight user product that difficulty gets eliminated by using just a fixed number of functions for the users regardless of the number of films they have rated. Now a suitable likeness function is required to match users. Euclidean range function [Han & Kamber, 2006] is utilized in our function that is given by:

\[
dis(v_x, v_y) = \sqrt{\sum_{i=1}^{18} (JIM_{v_x}(i) - JIM_{v_y}(i))^2}
\] (1)

RS that uses for similarity computation because the Euclidean-based RS (EBRS). The planned person design contains 18 characteristics contributing similarly throughout likeness computation in the EBRS. That doesn’t catch the true to life situation where each individual areas different weights to different features. These weights are susceptible to transform with time and adjusting tastes of each and every user. A highly effective learning unit to capture these weights is required.

Ujjin & Bentley [2004] applied the major search to get person things for variety wavelengths tailoring the suggestion method to the choices of every person individual but with a extensive profile. Every individual areas an alternative solution concern on each feature, which may be referred to as feature weight. To be able to apply actually individualized RS, these weights must be got and fine-tuned to reveal each user’s preference. By imposing characteristics weights to Formula, genetic algorithm can be properly used to obtain these weights producing a genetic-algorithm

-based RS (GBRS). For this process the weighted Euclidean range purpose [Han & Kamber, 2006] is used for likeness computation:

\[
dis(v_x, v_y) = \sqrt{\sum_{i=1}^{18} N_i \times (JIM_{v_x}(i) - JIM_{v_y}(i))^2}
\] (2)

GBRS using the small consumer design retains memory-based RS accuracy, and model-based RS scalability. The user design increases the internet procedure for generating some like-minded persons within which a memory-based collaborative RS is moved out. These subsections offer a short introduction to GA and the conditioning function. Within our test, GA adapts the function weights to recapture an individual things for different characteristics [Ujjin & Bentley, 2004]. The event loads of client V a are exhibited as several loads, weight(va) := (NI ‘... D 1 8), wherever d is how many features. The genotype of Ni is just a real-valued numbers. When the fat for almost any function is zero, that function is ignored. That permits function variety to be flexible to each user’s preference.

V. PROPOSED ALGORITHM

A genetic algorithm (GA) procedures a population of competitive prospect alternatives represented by some chromosomes, 0. Each chromosome in the people represents a probable means to fix an optimization issue. An over-all platform of GA

A. Procedure: Genetic Algorithm

begin
    s = 0;
    initialize M(s);
    evaluate M(s);
    while (not termination condition) do
        begin
            s = s + 1;
            select M(s) from M(s - 1);
            alter M(s);
            evaluate M(s);
        end
    end

Figure 2: Flow chart of proposed methodology

Subsequent Houck et al. [1995], a chromosome representation describes every individual in the populace of interest that establishes how the issue is
structured in the GA and also establishes the genetic operators which is often used. Each chromosome comprises a schedule of genes from a specific alphabet. An alphabet can contain binary figures, traveling period numbers, integers, patterns, etc. In Holland’s unique design [Holland, 1975], the alphabet was limited by binary digits. Nevertheless, it’s been revealed that more natural representations are more efficient and generate greater solutions. One of use representation of chromosome is just a real-valued representation which involves determines how the issue is organized in the GA and also determines the genetic operators which can be usually used. Each chromosome comprises a schedule of genes from a certain alphabet. An alphabet may include binary results, traveling time numbers, integers, patterns, etc. In Holland’s special design [Holland, 1975], the alphabet was limited by binary digits. Nonetheless, it’s been unveiled that more normal representations. The chromosomes search place, fie, is:

\[ \Omega_{2} = \{ \theta_{t} \in \mathbb{R} | \theta_{t}^{\min} \leq \theta_{t} \leq \theta_{t}^{\max}, \ldots, \theta_{r}^{\min} \leq \theta_{r} \leq \theta_{r}^{\max} \} \quad (3) \]

All the genes \( i \) in the chromosome will undoubtedly be changed in the confined place fie during the genetic operations. Top of the and lower bounds of \( i \) must certainly be written by designer. After a developed chromosome by genetic procedures moves beyond fie, then a original chromosome will undoubtedly be retained. Initially, a citizenry of people IS developed and then every individual is evaluated for fitness with regards to the optimization task to be solved. Collection within the population is performed in a fitness-proportionate way: the more fit an individual, the more likely it is usually to be opted for copy in to the next generation. The fantastic recently produced people (if exist) change today’s poor people to produce the newest citizenry for the next generation. GA terminates if the mandatory price of conditioning is reached or probably the most volume of decades is elapsed.

**B. Fitness Function**

Locating a suitable Function purpose is often a complex problem for GA applications [Goldberg, 1989] and for the application being mentioned for the reason that chapter it’s not a small job since every set of loads in the GA populace is used for similarity computation and therefore the RS needs to be re-run on the whole repository for each and every simple new pair of loads to be able to computes their exercise [Ujjin & Bentley, 2004]. An unwanted (good) pair of loads might (should) develop a bad (good) town pair of guys and women for the powerful client, and ergo bad (good) recommendations. One process to obtain the exercise purpose is by reformulating the matter as a administered understanding task. For this type of purpose this scores of the powerful client are arbitrarily split in to two disjoint versions, always check always scores collection (66%) and understanding scores collection (34%). To get the exercise record for the developed pair of loads, the RS must certanly be purpose and the believed scores for each and every simple film in training scores collection must certanly be computed. The average of the differences between this and believed scores of all shows in training scores collection is used whilst the exercise record for that pair of weights.

\[ \text{Fitness}(v_{t}) = \frac{1}{|A_{t}|} \sum_{j=1}^{t} |s_{t,j} - s_{j}^{*}| \quad (4) \]

wherever \( |A_{t}| \) could be the cardinality of working out ratings set equivalent to a given productive user.

**VI. LITERATURE SURVEY**

Ernesto Damiani et al. [1] proposed that team-based organizational structures are actually commonly followed for activities such as for instance product growth, customer service and process-improvement initiatives due to their increased likelihood of earning better decisions and fixing problems. However, group effort often encounters issues like information overload or misconceptions as a result of aim misalignment. The setting ahead of the indisputable fact that computer-supported effort situations may have an optimistic affect group effort by raising group customers attention, focusing interest on job delivery, and fostering the volume of conversation among group members.

Marco de Gemmis et al. [2] proposed that Recommender methods are filters which recommend items or data that could be fascinating to users. These methods analyze the past conduct of a person, construct her page that stores information about her passions, and exploit that page to locate possibly fascinating items. The main issue of this process is that it may offer exact but probably clear recommendations, since suggested items are just like those an individual currently knows. The planned technique enriches a graph-based advice algorithm with history knowledge which allows the machine to profoundly realize the things it deals with.

Martinez-Cruz et al. [3] proposed that recommender programs consider and filter the large amount of data on the Internet, so they really may be used to help users along the way of opening to applicable information. Here building an ontology to characterize the confidence between users utilizing the fuzzy linguistic modeling, so that in the endorsement technology method we do not take into consideration users with related scores record but users by which each user can trust.
Cataldo Musto et al. [4] report proposes a platform for suggestion of advantage allocation methods which combines case-based thinking with a book diversification strategy to guide financial advisors in the job of proposing varied and customized expense portfolios. The performance of the framework has been considered in the form of an fresh period done against 1172 actual customers, and results reveal that the produce received by recommended portfolios overcomes that of portfolios planned by human advisors generally in most fresh adjustments while conference the preferred risk profile. Furthermore, our diversification strategy shows promising results with regards to equally selection and average yield.

Andre Luiz Vizine Pereira et al. [5] proposed that Recommender Programs (RSs) are powerful and common tools for e-commerce. To construct their tips, RSs take advantage of different knowledge resources, which record the faculties of items, customers, and their transactions. Despite recent innovations in RS, the cool begin issue is still a relevant concern that deserves further attention, and arises as a result of lack of prior information about new customers and new items. To decrease system destruction, a cross approach is presented that mixes collaborative filter tips with demographic information. The approach is based on a preexisting algorithm, SCOAL (Simultaneous Co-Clustering and Learning), and provides a hybrid endorsement approach that can handle the (pure) cool begin issue, wherever no collaborative information (ratings) is available for new users.

VII. PERFORMANCE EVALUATION

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.

1. PSNR (Peak Signal to Noise Ratio): Maximum Top sq sound percentage could be the percentage involving the most probable price of the signal and the ability of the corrupting noise. It is calculated in decibels (db). It could be discussed as:

$$\text{PSNR} = 10 \log_{10} \left( \frac{M^2}{MSE} \right)$$  \hspace{1cm} (5)

2. Accuracy: The accuracy is calculated as percentage change of predicted target with actual target with acceptable error. Detail is an explanation of arbitrary mistakes, a calculate of mathematical variability. Accuracy has two meanings:

1. more typically, it is an explanation of systematic mistakes, a calculate of mathematical error;

2. instead, the ISO describes accuracy as describing equally types of observational problem above (preferring the term trueness for the most popular meaning of accuracy).

$$\text{Accuracy} = \frac{100}{M} \sum_{j=1}^{M} g_j$$ \hspace{1cm} (6)

$$g_j = \begin{cases} 1 & \text{if } \text{abs}(\hat{T}_j - T_j) \leq \text{Err} \\ 0 & \text{otherwise} \end{cases}$$ \hspace{1cm} (7)

R-predicted target, T-actual target, Err-acceptable error and M- is total number of instances.

3. Error rate: Error rate of a conversation channel. The frequency with which mistakes or sound are introduced to the channel. Problem charge may be tested with regards to erroneous bits acquired per bits transmitted. For instance, one or two mistakes per 100 000 bits may be considered a common charge for a slender group point-to-point line. The circulation of mistakes is usually non standard: mistakes tend ahead in bursts. Therefore the mistake charge of a place may be given with regards to percentage of error-free seconds. one charge is stated as an adverse energy of five: one charge of just one single bit per 100 000 will be stated being an mistake charge of 10–5.

$$\text{Bit-Error-rate(BER)} = \frac{1}{2} \left( 1 - \text{efficiency} \left( \sqrt{\frac{E_b}{N_o}} \right) \right)$$ \hspace{1cm} (8)

Another method of presenting error rate is to think about the errors as caused by adding the info signal to an underlying error signal.

Table 1 showing the analysis PSNR, Accuracy and error rate respectively as it is known that PSNR needs to be minimized, accuracy also needs to be maximized and error rate needs to be minimized.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Decision Tree</th>
<th>Linear model</th>
<th>Neural network</th>
<th>Random forest</th>
<th>Genetic algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>35.24</td>
<td>32.31</td>
<td>32.17</td>
<td>32.80</td>
<td>50.12</td>
</tr>
<tr>
<td>Accuracy</td>
<td>50.21</td>
<td>36.1</td>
<td>35.28</td>
<td>36.24</td>
<td>60.93</td>
</tr>
<tr>
<td>Error rate</td>
<td>49.79</td>
<td>63.9</td>
<td>64.72</td>
<td>63.76</td>
<td>39.07</td>
</tr>
</tbody>
</table>

Performance comparison of machine learning methods in the prediction of demand in social network by PSNR, Error rate and Accuracy. Accuracy score on testing data set.
Figure 3: Comparison Graph
It is clearly shown in the fig that the PSNR of the proposed technique is higher among other algorithms, accuracy is also higher than other algorithms. Also the error rate 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 algorithm. Therefore proposed technique is effective than Decision tree, Linear model, neural network, random forest and genetic algorithm.

VIII. CONCLUSIONS
In this paper we have evaluated the performance of decision tree, Neural network and linear regression based machine learning algorithms for prediction of demand of customer 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 performance result shows that Genetic algorithm method outperforms over other machine learning methods in the prediction. The proposed work shows better results than the existing algorithm. Thus in near future we shall propose hybrid genetic algorithm for social system. Also contrast of the proposed strategy will be done with several other evolutionary formulas for social system advised system.

REFERENCES