

Efficient Algorithm and Framework for Interesting Patterns Generation for optimization of Association rule mining using genetic Algorithm for Social Networking Sites

Karuna Nidhi Pandagre

Department of Computer Science, Aisect University,
Bhopal, India

S. Veenadhari

Department of Computer Science, Aisect
University,
Bhopal, India

ABSTRACT

Social Networking (SN) plays an important part in real life because people is already connected in any types of social network like facebook, twitter, linkedin, instagram etc. SN gives a platform to people to share their ideas to each other. Moreover this type of online community, people may share their common interest. SN is a medium of making link with people having unique or common interest and goals. But there are various challenges in SN services. The major challenge is privacy it means keeping your information private isn't just about your own choices. It's about your friends' choices, too. In this article a novel Algorithm proposed for Interesting Patterns Generation (AIPG) and framework for interesting patterns generation from weblog. This approach is also helpful in SN

General Terms

Patterns Generation Algorithms and Framework.

Keywords

Social Network, Private Data, Weblog, Visitors

1. INTRODUCTION

Now days SN services are rapidly growing, lots of social networks are available such as facebook, twitter, linkedin, instagram etc [1]. In social networking services any person can connect another person have shown the person using dot in the give figure and maintained its relationships through lines. In this network any number of visitors (v1, v2, v3 ...) can communicate each other. Web usage mining extracts useful information by using the web server log. By web log data can collect the behavioral patterns of the users. These patterns are used to de ne pattern discovery and associations between documents [2].

1.1 Web Server Log

Server log is a text file, which records every activity of visitors. The log file used or maintained on server log should

be in a proper format which is called as file format [3]. There are various server file formats are available and contain different types of information

such as Date- The date on which the activity occurred., Time- The time of activity occurred, Client IP addresses-The IP address is address of machine, User name-The name of that particular visitor; and anonymous user indicated by a "Hyphen" (-), Service Name- It tells the internet service name, Server IP address-The IP address of machine.

1.2 Web Usage Mining

Web usage mining is one of the applications of data mining technique.[4][19]. There are various tools are available for analyzing weblog such as Rapid Miner, Weblog Expert, Deep Log Analyzer etc.[14][18], and every tool offered some or the other feature which was better than the rest. Web usage mining is one of the prominent research area due to these following reasons [5][10]. Thus personalization for a user can be achieved through web usage mining. The second is to identify frequent access patterns for the users. Those patterns are to improve the overall performance of future accesses. Common access patterns improve the actual design of web pages and for making other medications to a website [6] [11] [22]. Usage patterns can be used for business intelligence in order to improve sales and advertisement by providing product recommendations. Web usage mining approach is very efficient for identifying user's behavior in a SNS. Through this approach advertisement recommendation is possible in SNS. There are various ways to predict user behavior such as a particular timestamp, which kind of web services mostly accessed by a user. Identify commonly used browsers by user [9] [20]. Suggest friend/group in social networking with various constraints.

To predict new advertisements dynamically for users based on accessed services. To provide next product/items or service information according to current service accessed. Provide promotional users and discount information to users.

```

199.72.81.55 - - [01/Jul/1995:00:00:01 -0400] "GET /history/apollo/ HTTP/1.0" 200 6245
unicomp6.unicomp.net - - [01/Jul/1995:00:00:06 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
199.120.110.21 - - [01/Jul/1995:00:00:09 -0400] "GET /shuttle/missions/sts-73/mission-sts-73.html HTTP/1.0" 200 4085
burger.letters.com - - [01/Jul/1995:00:00:11 -0400] "GET /shuttle/countdown/liftoff.html HTTP/1.0" 304 0
199.120.110.21 - - [01/Jul/1995:00:00:11 -0400] "GET /shuttle/missions/sts-73/sts-73-patch-small.gif HTTP/1.0" 200 4179
burger.letters.com - - [01/Jul/1995:00:00:12 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 304 0
burger.letters.com - - [01/Jul/1995:00:00:12 -0400] "GET /shuttle/countdown/video/livevideo.gif HTTP/1.0" 200 0
205.212.115.106 - - [01/Jul/1995:00:00:12 -0400] "GET /shuttle/countdown/countdown.html HTTP/1.0" 200 3985
d104.aa.net - - [01/Jul/1995:00:00:13 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
129.94.144.152 - - [01/Jul/1995:00:00:13 -0400] "GET / HTTP/1.0" 200 7074
unicomp6.unicomp.net - - [01/Jul/1995:00:00:14 -0400] "GET /shuttle/countdown/count.gif HTTP/1.0" 200 40310
unicomp6.unicomp.net - - [01/Jul/1995:00:00:14 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 200 786
unicomp6.unicomp.net - - [01/Jul/1995:00:00:14 -0400] "GET /images/KSC-logosmall.gif HTTP/1.0" 200 1204
d104.aa.net - - [01/Jul/1995:00:00:15 -0400] "GET /shuttle/countdown/count.gif HTTP/1.0" 200 40310
d104.aa.net - - [01/Jul/1995:00:00:15 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 200 786
d104.aa.net - - [01/Jul/1995:00:00:15 -0400] "GET /images/KSC-logosmall.gif HTTP/1.0" 200 1204
129.94.144.152 - - [01/Jul/1995:00:00:17 -0400] "GET /images/ksclogo-medium.gif HTTP/1.0" 304 0
199.120.110.21 - - [01/Jul/1995:00:00:17 -0400] "GET /images/launch-logo.gif HTTP/1.0" 200 1713
ppptky391.asahi-net.or.jp - - [01/Jul/1995:00:00:18 -0400] "GET /facts/about_ksc.html HTTP/1.0" 200 3977
net-1-141.eden.com - - [01/Jul/1995:00:00:19 -0400] "GET /shuttle/missions/sts-71/images/KSC-95EC-0916.jpg HTTP/1.0" 200 34029
ppptky391.asahi-net.or.jp - - [01/Jul/1995:00:00:19 -0400] "GET /images/launchpalms-small.gif HTTP/1.0" 200 11473
205.189.154.54 - - [01/Jul/1995:00:00:24 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
waters-gw.starway.net.au - - [01/Jul/1995:00:00:25 -0400] "GET /shuttle/missions/51-1/mission-51-1.html HTTP/1.0" 200 6723
ppp-mia-30.shadow.net - - [01/Jul/1995:00:00:27 -0400] "GET / HTTP/1.0" 200 7074
205.189.154.54 - - [01/Jul/1995:00:00:29 -0400] "GET /shuttle/countdown/count.gif HTTP/1.0" 200 40310
alyssa.prodigy.com - - [01/Jul/1995:00:00:33 -0400] "GET /shuttle/missions/sts-71/sts-71-patch-small.gif HTTP/1.0" 200 12054
ppp-mia-30.shadow.net - - [01/Jul/1995:00:00:35 -0400] "GET /images/ksclogo-medium.gif HTTP/1.0" 200 5866
dial22.lloyd.com - - [01/Jul/1995:00:00:37 -0400] "GET /shuttle/missions/sts-71/images/KSC-95EC-0613.jpg HTTP/1.0" 200 61716
smyth-pc.moorecap.com - - [01/Jul/1995:00:00:38 -0400] "GET /history/apollo/apollo-13/images/70HC314.GIF HTTP/1.0" 200 101267
205.189.154.54 - - [01/Jul/1995:00:00:40 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 200 786
ix-or12-01.ix.netcom.com - - [01/Jul/1995:00:00:41 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
ppp-mia-30.shadow.net - - [01/Jul/1995:00:00:41 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 200 786
ppp-mia-30.shadow.net - - [01/Jul/1995:00:00:41 -0400] "GET /images/MOSAIC-logosmall.gif HTTP/1.0" 200 363
    
```

Fig 1: Sample Log

1.3 Association Rule Mining (ARM)

In ARM and Frequent Pattern (FP) mining technique [16][16], the user sets a minimum threshold value and the pattern values are greater than those itemset are frequent itemsets otherwise it's called as a rare itemsets. Moreover the size of the mining outcomes depends on the user specified threshold. However, it is a very challenging task to be determined user specified threshold in real world application. If user specified threshold is too low, then it will generate more number of patterns. Some patterns may not be interesting.

1.4 Genetic Algorithms

The pioneering work of J. H. Holland in the 1970's proved to be a significant contribution for scientific and engineering applications. Since then, the output of research work in this field has grown exponentially although the contributions have been, and are largely initiated, from academic institutions world-wide [6][13]. It is only very recently that it has been able to acquire some material that comes from industry. The concept of this is somehow not clearly understood. However, the obvious obstacle that may drive engineers away from using GA is the difficulty of speeding up the computational process, as well as the intrinsic nature of randomness that leads to a problem of performance assurance [21][23]. There are various terms in genetics Chromosome, Initial Population, Crossover, Offspring, Fitness and Mutation. There are various representation and selection techniques of chromosome [12]

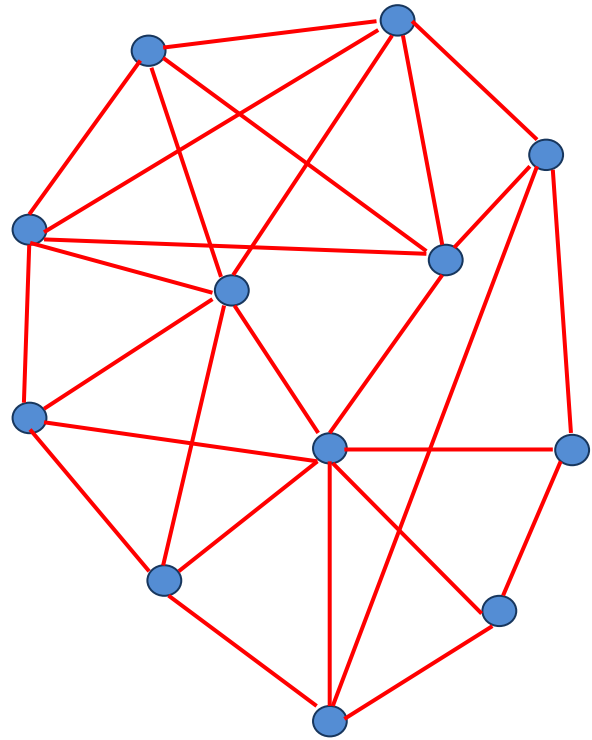


Fig 2: Basic structure of Social Network.

Table 1. Chromosome Representation through encoding

Encoding name	Representation chromosome
Binary encoding	Chromosome A=101100101100101011100101 Chromosome B=111111100000110000011111
Permutation Encoding	Chromosome A= 1 5 3 2 6 4 7 9 8 Chromosome B= 8 5 6 7 2 3 1 4 9
Value Encoding	Chromosome A=1.2324 5.3243 0.4556 2.3293 2.4545 Chromosome B= ABDJEIFJDHDIERJFDLDFLEGT

	Chromosome C=(back), (back), (right), (forward), (left)
Tree Encoding	Every chromosome is a tree of some objects, such as functions or commands in programming language

Genetic Algorithm

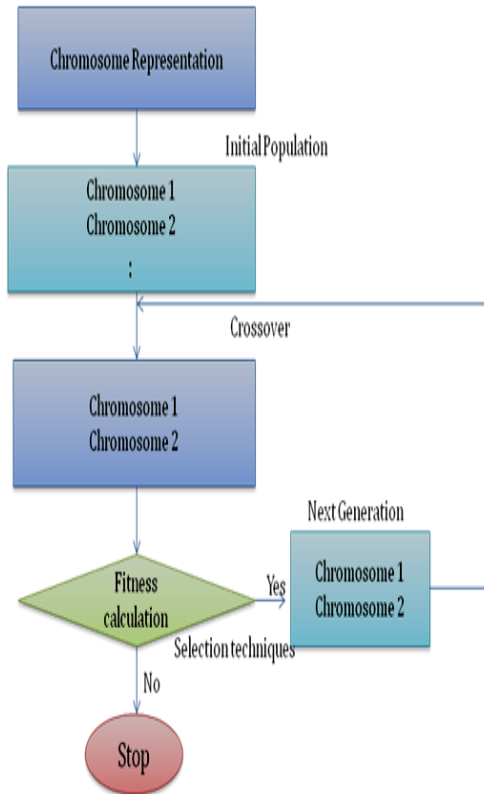


Fig 3: Genetic Algorithm Process

Crossover:

Mix and match two chromosome and produce two new chromosomes i.e. offspring's. There are two points for crossover such as Initial Strings, Offspring.

Single point

11000101 01011000 01101010 → 11000101 01011001 01111000

00100100 10111001 01111000 → 00100100 10111000 01101010

Two Point

11000101 01011000 01101010 → 11000101 01111001 01101010

00100100 10111001 01111000 → 00100100 10011000 01111000

Uniform

11000101 01011000 01101010 → 01000101 01111000 01111010

00100100 10111001 01111000 → 10100100 10011001 01101000

Arithmetic

11001011 + 11011111 = 11001001 (AND)

11001011 + 11011111 = 11001011 (OR)

Heuristic

Offspring1 = BestParent + r*(BestParent – WorstParent)

Offspring2 = BestParent

Tree crossover: In both parent one crossover point is selected, parents are divided in that point and exchange part below crossover point to produce new offspring.

Fitness Function: Fitness function is a mathematic function to calculate the survivability of chromosome mate. The fit chromosome will be the part of mating pool for the next generation.

Mutation:

Table 2. Mutation Table

In Numeric values:	In Real values:
Flipping of bits with some fixed probability 1100010101 -> 1100110101	Any arithmetic calculation at gene position which contain the fixed probability. Eg: If δ in the range[0,1] and v is gene $v + -\delta * v$

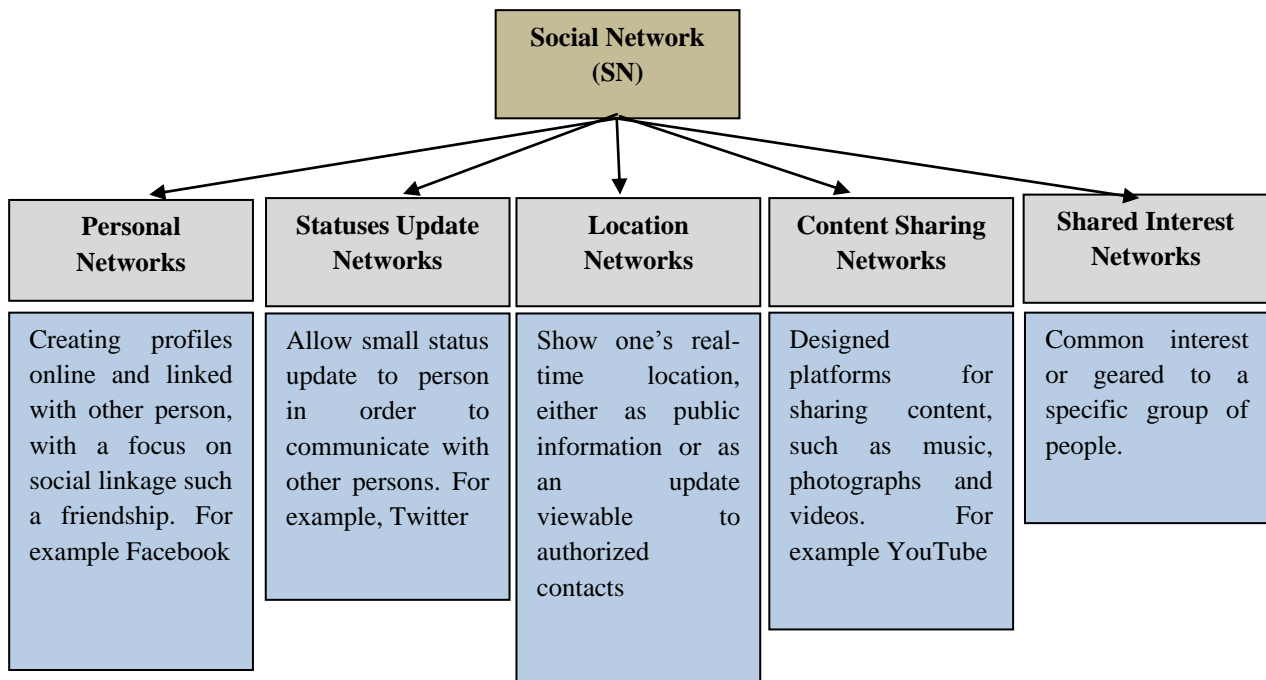


Fig 4: Genetic Algorithm Process Classification of Social Network

2. RELATED WORK

K. Zhang et al. [6] propose a genetic algorithm to solve the influence maximization problem. S. Chang et al. [7] developed refined research model a quantitative empirical study was designed and conducted to investigate the relationships between important constructs including effort expectancy, social influence, privacy concern, perceived risk, trust, and continuance intention. Users' trust on SNS is mainly influenced by effort expectancy, social influence, and perceived risk. The impacts of influential factors are different between Facebook and LinkedIn user groups. Empirical results imply that when the positive influences, such as effort expectancy of the online community, users from different SNS have different needs in social reputation, or high security compliance, in consideration of the negative influence of perceived risk to users' trust when continuing to use the SNS. This study presents a conceptual framework using extant theories and models to form a foundation for SNS incorporating the important factors of trust and privacy concerns. B. Jang et al. [8] extract the characteristics of online news and data from social network services, including the differences and similarities between them. It found the following differences: First, the news responds to of social events but content on social network services is related to personal interests. Second, the news is continually related to a specific issue or set of issues, whereas topics of conversation change daily in social network services. Third, items from the news can be identified with a single keyword in searches, whereas more keywords are needed to extract the desired information from social network services at the same time, It found that the words mentioned in both the news and on social network services. were similar, and both were used for commercial purposes. Analysis revealed that the news is related to the keyword generally, uses same words repeatedly, and its range of topics is narrow and public. On the contrary, social network services are not related to the keyword often, and their range of topics is wide and personal. Furthermore the analysis showed that the ranking algorithm improves the topic detection rate and catches the topic quickly. This paper provides useful information to better understand the

characteristics of online news and data from social network services. S. Rathore et al. [10] constructed a labeled dataset of the two most prominent SNSs – Twitter and Facebook. L. Bahri et al. [11] discuss the issues related to privacy preservation between centralization and decentralization, and provide a review of available research work on decentralized privacy preserving services for social networks. K. F. Man et al. [12] introduces GA and developed a framework of a design tool for industrial engineers.

3. PROPOSED WORK

3.1 Algorithm:

Algorithm for Interesting Patterns Generation (AIPG)

Terms using AIPG:

Weblog W Binary Encoding BE, Transaction ID TID, Chromosome-I C-I,

Input:

File with TID

Output:

Strong Patterns

Read W and select Webpage P parameter.

Generate Binary File from P

Construct BE file.

Read BE of TID and Items in form of chromosome to create population.

Select initial population (all C-I) on basis minimum threshold fitness satisfied.

Select two C-I as parent from initial population performs arithmetic crossover (logical AND) to produce offspring. Because of arithmetic crossing over only common alleles transfer in offspring, Inspection of mutation. Inverting Bit of offspring 1 is replaced with 0 or vice versa.

Calculate threshold fitness for offspring. If offspring satisfy minimum threshold then select both parent chromosome-I as frequent and all associated ancestor chromosomes are frequent and select offspring in population for reproduction.

Repeats steps 5, 6, 7, until find all possible frequent chromosome-I or Generated all population satisfied minimum threshold fitness.

Read BE of TID and Items in form of chromosome to create population.

Select initial population (all C-I) on basis minimum threshold fitness satisfied.

3.2 Genetic Operation to find survival population

(Frequent Pattern) of chromosome-I

Selection of Initial population or Reproduction:

In transaction dataset there have item which represent in a particular TID (or how many object are access). Here items are treated as chromosome in form of binary encoded.

Binary encoding of each item is represents its sale in corresponding TID's.

And Define fitness function $f(x)$ for the chromosome-I.

3.3 Fitness Function:

Fitness function is a mathematic function to calculate the survivability of chromosome. Calculate percent support probability ratio of each chromosome-I or individual.

Fitness Calculation:

1= indicate item sold in particular TID.

0= indicate item not sold in particular TID.

So Fitness function support

$$F(s) = \frac{\text{Total Number of true bit in chromosome-I}}{\text{Total length of chromosome-I}}$$

$$F(s) = \frac{\text{TDAc}}{\text{TLC}}$$

Where :

TDAc= Total dominants alleles in chromosome-I

TLC= Total length of chromosome-I

Percent support $f(Sp) = f(s) * 100 = \frac{\text{TDAc}}{\text{TLC}} * 100$

Assume the Transaction dataset which is shown in Table 3

Table 3: Transactional Dataset

TID	Item dataset
T10	{P1,P2,P5,P6}
T20	{P2,P3,P6}
T30	{P1,P2,P4}
T40	{P2,P3,P7}
T50	{P1,P2,P4,P6}
T60	{P1,P3}
T70	{P2,P3}
T80	{P1,P3,P6}
T90	{P1,P2,P3,P6}
T100	{P1,P2,P3,P5}

Scan the uncertain transactional dataset and perform value encoding scheme on transactions and items in the form of chromosome to create population.

For finding initial population select chromosome from set of all chromosomes on the basis of minimum threshold value which is calculated by defines fitness function.

Select any two chromosomes as parent from initial population and perform arithmetic crossing over (minimum ()) to produce offspring. Only minimum value alleles are transferred in offspring.

Perform inspection of mutation. For this value of inverted bit in offspring is replace by mean of corresponding bits in its both parents.

Use fitness function to calculate fitness of offspring. If offspring satisfies minimum threshold then consider as population for reproduction.

If selected chromosomes, crossover with their parent to produce offspring and also satisfy minimum threshold fitness then select all subsets of parent chromosome consider as frequent. Repeat steps 3, 4, and 5, until all possible offspring are found as frequent generated or all population (maximal itemsets) which satisfies minimum threshold fitness.

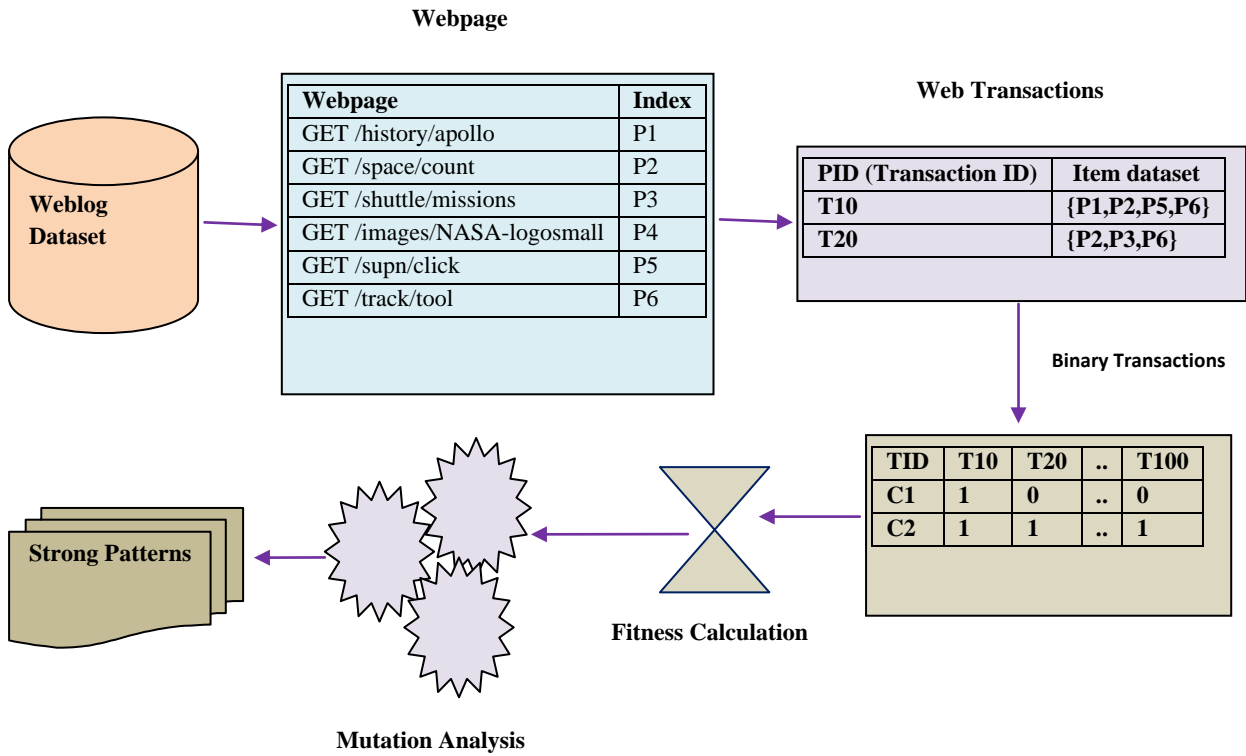


Fig 5: Proposed Framework

Table 4: Initial Population

There are ten chromosomes (items) in dataset D, their binary encoding.

TID	T 1 0	T 20	T 30	T 40	T 50	T 60	T 70	T 80	T 90	T1 00
CI1	1	0	1	0	1	1	0	1	1	1
CI2	1	1	1	1	1	0	1	0	1	1
CI3	0	1	0	1	0	1	1	1	1	1
CI4	0	0	1	0	1	0	0	0	0	0
CI5	1	0	0	0	0	0	0	0	0	1
CI6	1	1	0	0	1	0	0	1	1	0
CI7	0	0	0	1	0	0	0	0	0	0

3.4 Fitness Calculation

Evaluate fitness support percents of each Individual.

$$F(Sp) = (TDAc/TLC) * 100$$

$$\text{Chromosome-I1} = (7/10) * 100 = 70\%$$

$$\text{Chromosome-I2} = (8/10) * 100 = 80\%$$

$$\text{Chromosome-I3} = (7/10) * 100 = 70\%$$

$$\text{Chromosome-I4} = (2/10) * 100 = 20\%$$

$$\text{Chromosome-I5} = (2/10) * 100 = 20\%$$

$$\text{Chromosome-I6} = (5/10) * 100 = 50\%$$

$$\text{Chromosome-I7} = (1/10) * 100 = 10\%$$

It assume minimum support fitness= 20%

Then Initial population of individuals are:

Table 5: Second Initial Population

TI D	T 10	T 20	T 30	T 40	T 50	T 60	T 70	T 80	T 90	T1 00
CI 1	1	0	1	0	1	1	0	1	1	1
CI 2	1	1	1	1	1	0	1	0	1	1
CI 3	0	1	0	1	0	1	1	1	1	1
CI 4	0	0	1	0	1	0	0	0	0	0
CI 5	1	0	0	0	0	0	0	0	0	1
CI 6	1	1	0	0	1	0	0	1	1	0

one set Frequent Items are = {P1, P2, P3, P4, P5, P6}

Crossing over (Recombination):

Choose two chromosome from initial population on basis of fitness value

Step A:

Chromosome- I 1 1 0 1 0 1 1 0 1 1 1

Chromosome- I 2 1 1 1 1 1 0 1 0 1 1

AND arithmetic crossing over

Offspring (child) {I1, I2}c= 1 0 1 0 1 0 0 1 1

Fitness value of offspring= (5/10)*100=50%

Offspring satisfy minimum threshold fitness so both parents selected as frequently associated.

4. RESULT ANALYSIS

Result A: Two set frequent item = {P1,P2}

Step B: Suppose it is performing crossover between {P1, P2} C chromosome and Chromosome-I3.

Offspring (child) {P1, P2} c= 1 0 1 0 1 0 0 1 1

Chromosome-I3 = 0 1 0 1 0 1 1 1 1

Arithmetic AND crossing over

Offspring (child) {P1, P2, P3}c= 0 0 0 0 0 0 0 1 1

Fitness value of Offspring (child) {P1, P2, P3}c = (2/10)*100=20%

Offspring {P1, P2, P3}c is survivable so

Result: B Three frequent item set = {P1, P2, P3}, as per property its all two item sub sets are frequent= { {P1, P2}, {P1,P3}, {P2,P3}}.

StepC:

Suppose it select two individuals {P1, P2, P3}c×Chromosome-P4

Offspring (child) {P1,P2,P3}c= 0 0 0 0 0 0 0 1 1

Chromosome-P4 = 0 0 1 0 1 0 0 0 0

Arithmetic AND Crossing over

Offspring (child) {P1, P2, P3, P4}c= 0 0 0 0 0 0 0 0 0

Fitness of Offspring (child) {P1, P2, P3, P4} c = 0% .

Result C: Thus offspring is not survival or four set itemset {P1, P2, P3, P4} is not frequent.

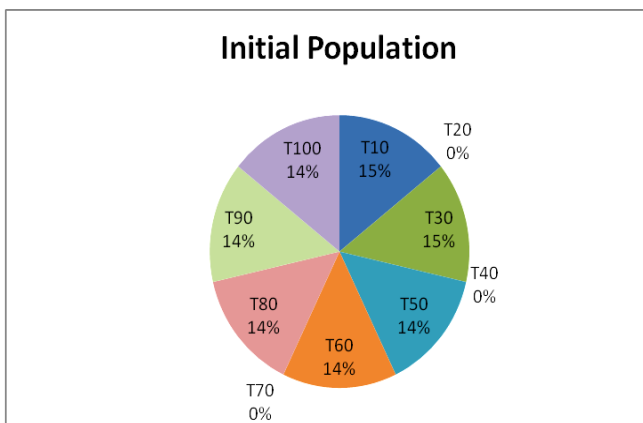


Fig 6: Initial Population Analysis

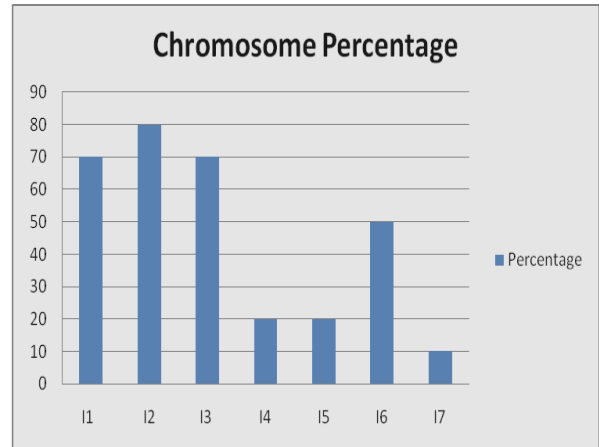


Fig 7: Chromosome Percentage Analysis

4.1 Mutation Analysis:

Once crossing over was performed, Mutation may take place. Mutation use to maintain genetics diversity form one generation to next.

It can analyze mutation in binary encode chromosome for inverting a bit '1' replace with '0' and reverse is also applicable.

Chromosome-I= 1100010000= 1000100000

Using mutation can increase or decrease fitness of individual. It is depend on mutant. But in the case of solving real problem when analyze mutation operation it depends on what are the goals of searching solution in problem space. Arithmetic mutation transfer only common alleles (character) from parents to Children. Survivability depends on the how many common alleles in both parent chromosomes. In the arithmetic crossing over possible of mutation is Zero percent.

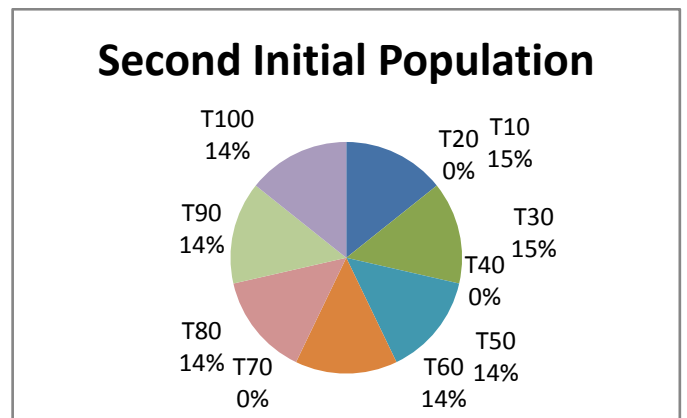


Fig 8: Second Initial Population Analysis

One set data item = {P1, P2, P3, P4, P5, and P6}.

Two set data item = {{P1,P2},{P2,P3},{P1,P3},{P2,P3}, {P1,P4},{P2,P4},{P1,P5},{P2,P5},{P1,P6},{P2,P6}and {P3,P6}}.

Three set data item= {{P1, P2, P3}, {P1, P2, P4}, {P1, P2, P5} and {P1, P2,P6}}.

This genetic algorithm based approach needs only one datasets scan. It generates less number of candidates or offspring and not requires level by level candidate generation as traditional approaches. This is good approach to get maximal itemsets. It is easy to implement as parallel process

to get frequent itemsets. In this case each pair of chromosome can run on individual processor and perform logical AND operations. In next pass find only survival offspring and repeat same process to get higher size offspring. Further Generate the association rules from the frequent item set. Optimization of these rules using genetic algorithms can find best association rules.

5. CONCLUSION

Now a days, large number of users are using social networking services and lots of social networking service providers are available to provide services at free of cost. When click on the social network that means all record maintain by the social network server. Social network server analyzes all record in the form of weblog. In this paper proposed an efficient algorithm and framework for finding strong visitor patterns from weblog. It also applied a genetic approach for optimizing those patterns. The proposed algorithm is useful in real life for different perspectives. It predicts the diseases based on the previous medical transactions. It predicts the trends for future based of sales transactions also it may be used by the business peoples for improving their business in e-commerce. It is also helpful for multidimensional association rules mining from relational databases. The frequent itemset are useful for advertisement management, tourist guidance, and product value analysis. This is also useful for gene prediction based on genetic optimization approach.

6. ACKNOWLEDGMENTS

Our thanks to the anonymous reviewers and the authors of references.

7. REFERENCES

- [1] Chu H., Deng D. and Park J. 2011. Live Data Mining Concerning Social Networking Forensics Based on a Facebook Session Through Aggregation of Social Data, *IEEE Journal On Selected Areas In Communications*, Vol. 29, No. 7.
- [2] Bakariya B., Thakur G. 2014. User Identification Framework in Social Network Services Environment, *Informatica Economică* vol. 18, no. 2, pp 15-23.
- [3] Yuan M., Chen L., Yu P. and Yu T. 2013. Protecting Sensitive Labels in Social Network Data Anonymization, *IEEE Transactions on Knowledge and Data Engineering*, VOL. 25, NO. 3.
- [4] Hajian B. and White T. 2011. Modelling In-fluence in a Social Network: Metrics and Evaluation, *IEEE International Confer-ence on Privacy, Security, Risk, and Trust*, *IEEE International Conference on Social Computing*, 2011.
- [5] G. Thakur, Bakariya B. and Mohbey K. 2012. An Inclusive Survey on Data Preproc-essing Method Used in Web Usage Min-ing, presented in Seventh International Conference on Bio-Inspired Computing: Theories and Application, (BIC-TA 2012), ABV-Indian Institute of Informa-tion Technology and Management Gwalior, December 14 - 16.
- [6] Zhang K., Du H., Feldman M. 2017. Maximizing influence in a social network: Improved results using a genetic algorithm, *Physica A*, pp. 20–30.
- [7] Chang S., Liu A. and Shen W. 2017. User trust in social networking services: A comparison of Facebook and LinkedIn, *Computers in Human Behavior*.
- [8] Jang B. and Yoon J. 2018. Characteristics Analysis Of Data From News And Social Network Services, *IEEE Access*.
- [9] Ha T. and Hoang D. 2017. An assistive healthcare platform for both social and service networking for engaging elderly people, *23rd Asia-Pacific Conference on Communications (APCC)*.
- [10] Rathorea S., Sangaiahb A. and Parka J. 2017. A novel framework for internet of knowledge protection in social networking services, *Journal of Computational Science*, pp.55–65.
- [11] Bahri L., Carminati B. and Ferrari E. 2018. Decentralized privacy preserving services for Online Social Networks, *Online Social Networks and Media*, Vol.6, pp. 18–25.
- [12] Man K., Tang K., and Kwong S. 1996. Genetic Algorithms: Concepts and Applications, *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS*, VOL. 43, NO. 5.
- [13] Bakariya B., Thakur G. 2017. Pattern Mining Approach for Social Network Service, *National Academy Science Letters*, Springer, Volume 40, Issue 3, pp 183–187, June 2017.
- [14] Bakariya B., Thakur G. 2015. An Efficient Algorithm for Extracting High Utility Itemsets from Web Log Data, *The Institution of Electronics and Telecommunication Engineers (IETE) Technical Review*, Volume-32, Issue-02, March- 2015, Pages 151-160.
- [15] Bakariya B., Chaturvedi K., Singh KP and Thakur GS. 2016. Efficient Approach for Mining Top-k Strong Patterns in Social Network Service, *5th International Conference on Eco-friendly Computing and Communication Systems (ICECCS-2016)* Dec. 8-9, IEEE, MANIT Bhopal.
- [16] Chaturvedi K., Patel R. and DK. 2015. Fuzzy C-Means based Inference Mechanism for Association Rule Mining: A Clinical Data Mining Approach, *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS*, Volume 6, Issue 6, Pages 103-110, 2015.
- [17] Bakariya B. and Thakur G.S. 2016. Mining Rare Itemsets from Weblog Data, *National Academy Science Letters*, Springer, Volume-39, Issue-05, pp 359–363.
- [18] Hassani M., TöwsAlfredo D. and Seidl C. 2017. BFSPMiner: an effective and efficient batch-free algorithm for mining sequential patterns over data streams, *International Journal of Data Science and Analytics*, pp 1–17.
- [19] Birla B., Patel S. and Sunhare H. 2013. Comprehensive Framework for Pattern Analysis through Web Logs Using Web Mining: A Review , *International Journal of Computer Science and Mobile Computing*, *IJCSMC*, Vol. 2, Issue. 4, , pg.32–37.
- [20] Thilagu M. and Nadarajan R. 2012. Efficiently Mining of Effective Web Traversal Patterns with Average Utility, *Procedia Technology*, Volume 6, Pages 444-451.
- [21] Kuramochi M. and Karypis G. 2004. An efficient algorithm for discovering frequent subgraphs, *IEEE*

Transactions on Knowledge and Data Engineering,
Volume: 16 Issue: 9.

- [22] Bin W. and Zhijing L. 2003. Web mining research, Computational Intelligence and Multimedia Applications, 2003. ICCIMA.
- [23] Singh A., Sharma D. and Pathak A. 2013, Web Usage Mining: A Concise Survey on Tools and Applications.

International Journal of Computer Applications, Vol. 74,
Issue 1, pp.1-7.

- [24] Pandagre, Ms Karuna Nidhi, and S. Veenadhari. Efficient Approach for Finding Strong Patterns from Weblog using Web Usage Mining. International Journal of Engineering Science Invention ,Vol 7, Issue 1, pp.06-14