Extraction of User Navigation Pattern Based on Particle Swarm Optimization

Prof. Mehul Barot
Research Scholar, C.U Shah University, Wadhwan City, 363030, Gujarat, India
m.p.barot@gmail.com

Dr. Kalpesh H. Wandra
2Dean, C.U Shah University, Wadhwan City, 363030, Gujarat, India
Khwandra@rediffmail.com

Dr. Samir B. Patel
3Assistant Professor, PDPU - Raisan,Gandhinagar
samir.patel@sot.pdpu.ac.in

Abstract: With current projections regarding the growth of Internet sales, online retailing raises many questions about how to market on the Net. A Recommender System (RS) is a composition of software tools that provides valuable piece of advice for items or services chosen by a user. Recommender systems are currently useful in both ~~the~~ research and in the commercial areas. Recommender systems are a means of personalizing a site and a solution to the customer's information overload problem. Recommender Systems (RS) are software tools and techniques providing suggestions for items and/or services to be of use to a user. These systems are achieving widespread success in e-commerce applications nowadays, with the advent of internet. This paper presents a categorical review of the field of recommender systems and describes the state-of-the-art of the recommendation methods that are usually classified into four categories: Content based Collaborative, Demographic and Hybrid systems. To build our recommender system we will use fuzzy logic and Markov chain algorithm.

Keywords:*Recommender System; Information Filtering; Prediction; Classification; User based; Item base; Fuzzy Logic;Particle SwarmOptimization; SwarmIntelligence; Agglomerative; Pattern Analysis.*

*****

1. INTRODUCTION TO PARTICLE SWARM OPTIMIZATION

Particle Swarm Optimization is a population based algorithm proposed by Kennedy J, Eberhart RC that is based upon the cognitive and social behavior of swarm. Initially PSO was designed to simulate how birds are seeking food, defined as cornfield vector. PSO has been successfully applied in several areas such as clustering problem [6][9], image processing [7], function optimization [1] etc. The main advantage of PSO is that it has less parameter to adjust. PSO does not have any complicated evolutionary operators such as crossover, mutation like that in genetic algorithm [9].

PSO follows a stochastic optimization method based on Swarm Intelligence (SI). The fundamental idea is that each particle represents a potential solution which it updates according to its own experience and that of neighbors [8]. The PSO algorithm searches in parallel using a group of individuals. Individuals or particles in a swarm, approach to the optimum through its present velocity, previous experience and the experience of its neighbors [8]. PSO searches the problem domain by adjusting the trajectories of moving points in a multidimensional space. The motion of individual particles for the optimal solution is governed through the interactions of the position and velocity of each individual, their own previous best performance and the best performance of their neighbors [2]. For a swarm of n particles the ith particle is represented by a position denoted as xi = (xi1, xi2, . . . , xin) where n is the number of particles. Except the position, each particle of a swarm is represented in D dimensional space with a velocity vi = (vi1, vi2. . . vid)[2]. The Standard PSO method updates the velocity and position of each particle according to the equations 1 & 2.

$$v_i^{t=1} = w * v_i^t + r_1 * \alpha\left(\vec{b} - \overrightarrow{x_i^t}\right) + r_2 * \beta\left(\vec{n} - \overrightarrow{x_i^t}\right) \text{------ (1)}$$

$$x_i^{\vec{t}+1} = x_i^{\vec{t}} + v_i^{\vec{t}+1} \text{-------- (2)}$$

Where is vt current velocity and xt is current position of particle, both α and β are positive constants and r1 and r2 are uniform random number ranging from 0 to 1, b and n are personal best positions found by ith particle and all the particles respectively, w is inertia weight. PSO algorithm is very fast, simple and easy to understand and implement. It also has a very few parameters to adjust [2] and requires little memory for computation.

2. SWARM OPTIMIZATION RELATED WORK

The applications of PSO as defined in [10] are that the PSO is used to calculate he cluster center vector and compare with k-means and fuzzy c-means. PSO offers better performance than the traditional clustering analysis algorithms.

For web session clustering, [5] used the "Euclidean Distance" to calculate the similarity and applied the Particle Swarm Optimization (PSO) for web session clustering. By applying the Euclidean Distance (ED) measure he performed the session

clustering. But they did not compare the results with other PSO based web session clustering techniques, they compared the results with K-mean but PSO and K-Mean are completely different in nature and produce different results.

[4] Proposed the PSO based on Sequence clustering technique which is also adopted in our model. Similarity measure for sequences clustering was defined as ratio of common items and unique items in two sequences, and then set the similarity in the order of occurrence of items in two sequences. In this paper, only similarity measure was changed from Euclidean Distance to S3M and the authors calculated the similarity based on the longest common sequence (LCS). For large amount of data, the time and space complexity can be big issue. The authors compared the results with K-Mean only.

In [11] authors combine both hierarchical clustering and partitioned clustering techniques and added Swarm intelligence to the process to give the novel PSO based Hierarchical Agglomerative data clustering technique. HSPO clustering works in an agglomerative manner starting from a relatively large number of particles and combining down to only one final particle.

[3] Proposed technique in which they enhance the idea and technique of [5]. They modified the session vector and experimented3 different similarity measures such as Angular Separation, Canberra Distance and Spearman Distance instead of Euclidean Distance Alone.

3. The Proposed Approach for PSO

In our proposed approach, we have enhanced the idea of [11] [5]. We introduced new Similarity Measure in order to calculate similarity between two paths instead of Euclidean Distance and Angular Separation. Our main goal was extracting most frequent paths which are accessed by different users and apply the results in web site modification in order to attract more users. In our approach we define complete Web Usage mining Methodology based on swarm intelligence. Our approach follows mainly 3 steps:

1) Pre Processing,
2) Pattern discovery and
3) Pattern analysis.

In preprocessing phase there are five steps such as Log cleaning, User identification, Session Identification, Path Completion which we follow in order to get preprocessed web log file for further processing. After getting complete path of user navigation

Hierarchical Clustering algorithm is followed for generating pattern and for better visualization, we have used Dendogram.

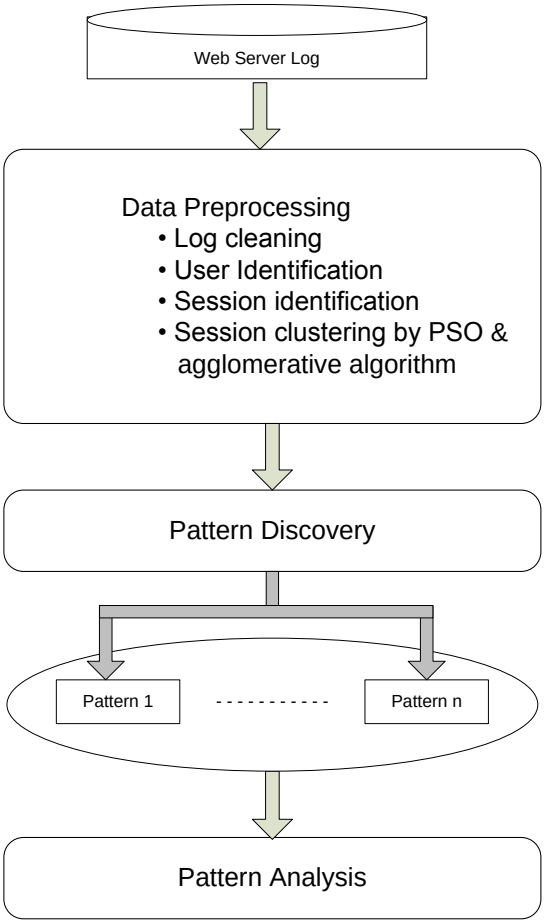

Figure1: ProposedArchitecture

Log cleaning is a Process of removing irrelevant entries from log file which are not required for further processing. Our web log file contains 5 Fields, only necessary fields and rows are selected rest of them are dropped. Entries which are discarded

– All JPEG, GIF file, Java Scripts, other audio/video files
– HTTP status code having failed status
– entries occurred from the crawlers or spiders

Next is User identification, it differentiates the log records according to users for the analysis. We differentiated each user of the site by its host name or IP Address.

Next is User identification, it differentiates the log records according to users for the analysis. We differentiated each user of the site by its host name or IP Address.

After identifying different users we did Session Identification based on time oriented heuristic. We have taken 30 minutes threshold as a default time-out.

 ISSN: 2321-8169
1320 – 1323

$$Sim\ (p_1, p_2) = \sum_{i=1}^{n} \delta_{ij} K^{-i} \text{------- (3)}$$

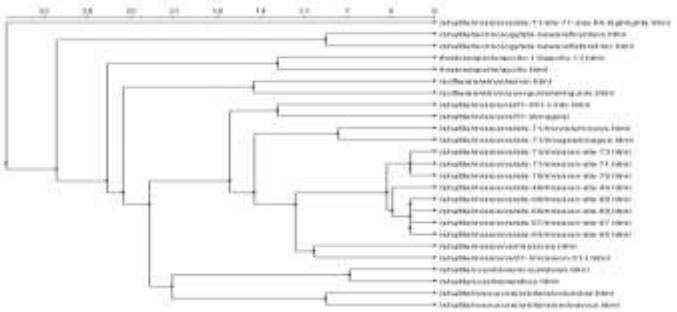

Figure2: Dendogram offinaloutputofour approach

For Next Phase of our approach we did web session clustering by applying hybridization of Particle Swarm Optimization with Hierarchical agglomerative algorithm. In web session clustering output of each iteration of PSO was directly fed into Hierarchical Agglomerative algorithm as an input. So for that we initialize each particle by assigning them a random velocity and positions in two dimensional input spaces. Next, we iterate the similarity value and nominate Wining Particle based on personal best position (pBest) then after we update new velocity and position. Position and Velocity of particles are updated in each iteration in order to find most suitable position. In last step, we applied Hierarchical agglomerative clustering to the winning Particles. These winning Particles are singlet clusters used as input. In last step we generated a dendogram from output of Hierarchical Agglomerative algorithm for better visualization and structured information as shown in Figure 2. We have applied equation 3 in order to final similarity between two navigational paths.

4. Experimental Results And Analysis

To measure performance of our approach we tested our algorithm on NASA web log file1 and analyzed the log entries having http requests of one day 1st July of 1995. We have taken existing system of alam's work for comparative study and analysis of our proposed system. First of all we compared two graphs of most frequently accessed path against frequency of paths of existing approach and our approach. Figure 3 shows top N frequently accessed patterns fond from existing system. Figure 4 shows top N frequently accessed patterns from our system. Here we got long navigational paths which are good Address.

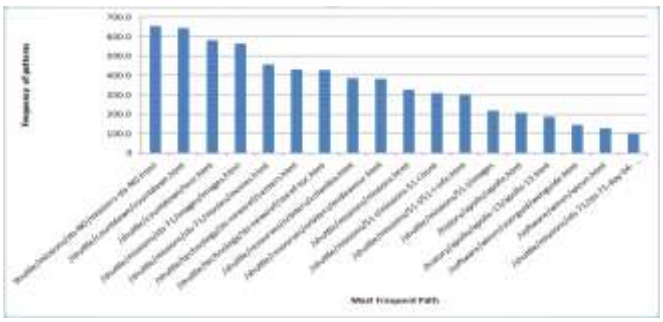

Figure3:Top N frequently accessed paths from existing system

Lastly we compared our approach with existing in terms of intra cluster distance. Main objective of our approach was to minimize intra cluster distance as fitness function. Figure 5 shows analysis of both systems in terms of fitness values. Here we gain better results over existing system as values of fitness function were lower than existing.

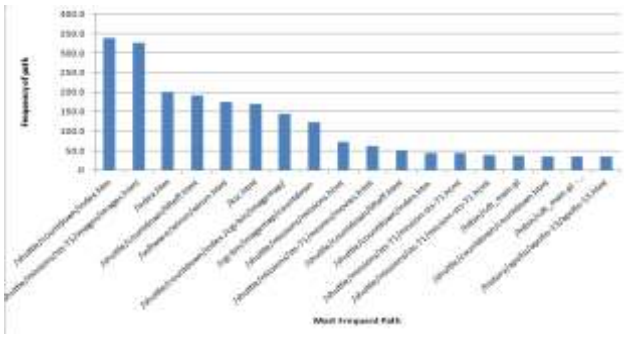

Figure 4 TopN frequentlyaccessedpaths from ourapproach

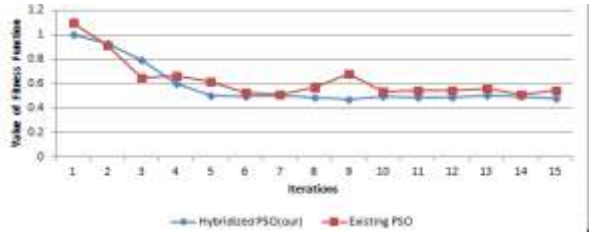

Figure5:Intra-cluster distancein bothsystem

5. Conclusion

In this paper, we presented our proposal to use PSO and its Hybridization in order to get better clustering of User Navigation pattern. These navigation Patterns can be used in website modification which help to attract more users towards website. Combination of Particle Swarm Optimization with Hierarchical agglomerative algorithm takes more time for large dataset but it reduces intra cluster distance which fulfill overall objective of our research. Hence, when time taken by system is not an issue then hybridized PSO is good option to extract user navigation patterns.

We consider some directions for future work. Firstly, we will consider effect of time measure and efficiency on hybridized PSO and it can be adapted to other visualization techniques in order to get more information from output such as top N frequent Path.

References

[1] PaterliniS, KrinkT (2006) *Differentialevolutionand particleswarmoptimizationinpartitionalclustering*

[2] NiuY, ShenL (2006) *An adaptive multi-objective particleswarmoptimization forcolorimagefusion.* Lecturenotesincomputerscience,LNCS

[3] http://setosa.io/ev/markov-chains/

[4] https://en.wikipedia.org/wiki/Fuzzy_logic

[5] http://whatis.techtarget.com/definition/fuzzy-logic

[6] http://in.mathworks.com/help/fuzzy/what-is-fuzzy-logic.html?requestedDomain=in.mathworks.com

[7] ShafiqAlam,GillianDobbie,PatriciaRiddle,M.Asif Naeem. *ParticleSwarmOptimizationBasedHierarchical Agglomerative Clustering*.2010IEEE/WIC/ACM International ConferenceonWebIntelligenceand IntelligentAgentTechnology.

[8] Chen CY, Ye F (2004) *Particle swarm optimization algorithmanditsapplication toclusteringanalysis*.In: Proceedingsofthe2004IEEEinternationalconferenceon networking,sensingandcontrol.Taipei,Taiwan.

[9] Lu.H and T.T.S Nguyen(2009). Experimental Investigation of PSO based Web User Session Clustering 2009. International journal of Soft computing and Pattern Recognition 978-0-7695-3879-2/09 IEEE DOI 10.1109./SoCPaR.2009.127.

[10] Alam S. , G. Dobbie, et al. Particle Swarm based Clustering of Web usage data 2008. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent technology.

[11] http://setosa.io/ev/markov-chains/