

# A Review: Study of various Ant Colony Optimization Techniques

Chandana A. Nighut<sup>1</sup>, Samira F. Nigrel<sup>2</sup>

Lecturer, Computer Engineering Department, Atharva College of Engineering, Mumbai, Maharashtra<sup>1</sup>

Lecturer, Information Technology Department, Atharva College of Engineering, Mumbai, Maharashtra<sup>2</sup>

**Abstract:** The hasty e-commerce development has made equally trade community and consumers face a novel situation. Due to severe competition on one hand and the customer's option to choose from numerous substitutes business community has realized the essential of clever marketing strategies and relationship management. Web usage mining attempts to find out useful facts from the secondary data obtained from the communications of the users with the Web. The paper concerns some Ant Colony Clustering and optimization technique for Web server log file to analyze user's interest, which is useful for giving suggestion about specific user's interest.

**Keywords:** Web Usage Mining, Ant Colony Optimization, ACLUSTER, ACCANTCLUST

## I. INTRODUCTION

This Web usage mining is the application of data mining techniques to web click-stream data in order to dig out usage patterns. As web sites prolong to raise in size and difficulty, the outcome of web usage mining have become significant for a number of application such as web site design, trade and advertising decision support, personalization, usability studies and network traffic analysis. [3] Data clustering is the process of finding similarities in data and placing alike data into groups and dissimilar data items into another group. Clustering is helpful whenever there is enormous amount of data. Several clustering techniques are used on web log data [1]. The study of ant colonies behaviour and their self-organizing capabilities is of interest to knowledge retrieval/ management and decision support systems sciences, because it provides models of distributed adaptive organization, which are useful to solve difficult optimization, classification, and distributed control problems, among others [2].

Real ants are capable of finding the shortest path from a food source to their nest without using visual cues by exploiting pheromone information. While walking, ants deposit pheromone on the ground, and follow, in probability, pheromone previously deposited by other ants. [2]

In Ant colony optimization (ACO) pheromone trails are used. Artificial ants may carry one or more objects and may drop them according to given probabilities. These agents do not communicate directly with each other, but they may influence themselves through the configuration of objects on the floor. Thus after a while, these artificial ants are able to construct groups of similar objects, a problem which is known as data clustering. [4]

Section II gives the motivation of this paper. Section III is devoted to Literature Survey. Conclusion is presented in Section IV.

## II. MOTIVATION

Millions of users access web sites in all over the world. When they access a websites, a bulky amount of records generated in log files which is extremely significant

because many times user repeatedly access the same type of web pages and the evidence is maintained in log files. These series can be considered as a web access pattern which is helpful to find out the user behaviour.

In recent years, there has been an increasing number of research works done with regard to web usage mining. The main motivation of this survey is to know what research has been done on Web usage mining to analyse the hourly and daily web traffic volume. [5]

## III. LITERATURE REVIEW

Web mining - is the application of data mining techniques to discover patterns from the Web. According to analysis targets, web mining can be divided into three different types, which are Web usage mining, Web content mining and Web structure mining.

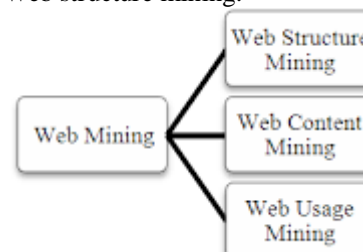


Fig.1. Web Mining

Web Content Mining (Analyse the content of web pages as well as results of web Searching) Web content mining is a process of extracting up information from texts, images and other contents. The technologies that are mainly used in web content mining are NLP (Natural language processing) and IR (Information retrieval). Web Structure Mining (Hyperlink Structure) Web structure mining is a process of extracting up information from linkages of web pages. Web structure mining is the process of using graph theory to analyse the node and connection structure of a web site. This graph structure can provide information about ranking and enhance search results of a page through filtering. Web Usage mining (analysing user web navigation) Web usage mining is a process of extracting

information from user how to navigate web sites. Web usage mining also known as web log mining, aims to discover interesting and frequent user access patterns from web browsing data that are stored in web server logs, proxy server logs or browser logs.[6]

#### A. Web usage mining

A lot work has been performed on removing various pattern information from Web logs and the application of the discovered data range from improving propose and structure of a Web site to enabling business organizations to function more efficiently. Jespersen et al [7] proposed an hybrid approach for analysing the visitor click sequences. A combination of hypertext probabilistic grammar and click fact table approach is used to mine Web logs which could be also used for general sequence mining tasks. Mobasher et al [8] proposed the Web personalization system which consists of offline tasks related to the mining of usage data and online process of automatic Web page customization based on the knowledge discovered. LOGSOM proposed by Smith et al [9], utilizes self-organizing map to organize web pages into a two-dimensional map based solely on the users' navigation behaviour, rather than the content of the web pages. LumberJack proposed by Chi et al [10] builds up user profiles by combining both user session clustering and traditional statistical traffic analysis using K-means algorithm. Joshi et al [11] used relational online analytical processing approach for creating a Web log warehouse using access logs and mined logs (association rules and clusters). A comprehensive overview of Web usage mining research is found in [12] [3].

#### B. ACLUSTER Algorithm

Title Ramos et al proposed ACLUSTER algorithm to follow real antlike behaviours as much as possible. Moreover, the strategy not only allows guiding ants to find clusters of objects in an adaptive way (if, by any reason, one cluster disappears, pheromone tends to evaporate on that location), as the use of embodied short-term memories is avoided (since this transition probabilities tends also to increase pheromone in specific locations, where more objects are present). In order to model the behaviour of ants associated to different tasks, as dropping and picking up objects, we suggest the use of combinations of different response thresholds. ACLUSTER uses combinations of two independent response threshold functions, each associated with a different environmental factor (or, stimuli intensity), that is, the number of objects in the area, and their similarity.[10] Abraham et al proposed web Usage Mining using artificial ant colony clustering and genetic programming. He proposed an ant clustering algorithm to discover Web usage patterns (data clusters) and a linear genetic programming approach to analyse the visitor trends. Empirical results clearly shows that ant colony clustering performs well when compared to a self-organizing map (for clustering Web usage patterns) even though the performance accuracy is not that efficient when compared to evolutionary-fuzzy clustering (iminer) approach.[2]

#### C. ACCANTCLUST Algorithm

Propose an accelerated ant based clustering algorithm (ACCANTCLUST) which is based on chemical recognition system of ants and this algorithm finds the number of clusters automatically. In ANTCLUST algorithm, when meeting between two ants is simulated, if the meeting is between two ants with no nest, and if they accept each other, these two ants are placed in a new nest. If they do not accept each other, no nest is created. In the proposed algorithm, if the ants do not accept each other, two new nests are created and the ants are placed in two different nests. The advantage of the proposed algorithm generates optimum clusters even if the last two steps of ANTCLUST are removed and thereby reducing the time taken to form clusters [13].

#### D. Ants NESTMATE

Figures When the size of the cluster goes on increasing due to increase in users or growth of interest of users it becomes inevitable need to optimize the clusters. This paper proposes a cluster optimizing methodology based on ants nestmate recognition ability and is used for eliminating the data redundancies that may occur after the clustering done by the web usage mining methods. This paper [14] proposes an AntClusterTrack algorithm; a cluster optimization algorithm which takes its input from a neural network based training process, ART1. The clusters obtained by L2 layer of ART1 is feed into an ant based clustering approach that checks for the similarity of the pheromone values of the artificial ants. This is done on the fact that ants belonging to the same nest will have similar odor. In this algorithm clusters are considered as the ants nest and the url combinations in each cluster is considered as the artificial ants.

#### E. Session Simulation Based on ants

For certain period of web sessions we need to mine the secondary data (web logs) derived from the users' interaction with the web pages. So for this,[15] at first Ant-based clustering algorithm is applied to pre-processed log files to extract frequent patterns, then it is displayed in an interpretable format and secondly decision tree method is used to find and predict user's navigation behaviour. Decision trees are used in classification and prediction. It is simple yet a powerful way of knowledge representation. Two type of approaches are used were the offline phase is based on Ant based clustering and the online phase is based on decision trees. The experimental results represent that the approach can improve the quality of clustering for user navigation pattern in web usage mining systems. These results can be used for predicting user's next request in the huge web sites.

#### F. Lumer-Faieta Algorithm

Ant-based clustering is applied to pre-processed logs to extract frequent patterns for pattern discovery and then it is displayed in an interpretable format in [16]. In this paper, a new method is proposed for extracting patterns from web logs based on ant clustering algorithm. We apply ant-based clustering for pattern discovery, other

similar methods applied ant colony clustering to segregate visitors. Some methods applied Markov models for modelling user web navigation behaviour. But the proposed method has the similarity and speed of ant-based clustering algorithm rather than other clustering algorithms.

#### IV. CONCLUSION

In this paper we survey the area of web usage mining. Hierarchical clustering's complexity increases when it is iterated for more than 100 times. Ant based algorithm can be used for usage data optimization. From the survey we can say that ACCANTCLUST is better in all terms.

#### REFERENCES

- [1] Sonia P Chaudhary, Hitesh Gupta, S.J. Patil, "Web Log Clustering using FCM and Swarm Intelligence Based Algorithms", IJRSET, Vol. 2, Issue 1, January 2013.
- [2] A. Abraham, Vitorino Ramos, "Web Usage Mining Using Artificial Ant Colony Clustering and Genetic Programming", Evolutionary Computation 2003, CEC'03.
- [3] J. Srivastava, R. Cooley, "web usage mining discovery and application of interesting patterns from web data", SIGKDD Explorations 2000.
- [4] A. Colomi, M. Dorigo, and V. Maniezzo, "Distributed optimization by ant colonies", in Proceedings of the First European Conference on Artificial Life, eds., F. Varela and P. Bourguine, pp. 134-142. MIT Press, Cambridge, Massachusetts, 1991.
- [5] Ujwala Patil, Sachin Pardeshi, "A Survey on User Future Request Prediction: Web Usage Mining", IJEATE, ISSN 2250-2459, Volume 2, Issue 3, March 2012.
- [6] Aparna N. Gupta, Prof. Arti Karndikar, "A Review: Study of Various Clustering Techniques in Web Usage Mining", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 3, March 2014.
- [7] Jespersen S.E., Thorhauge J., and Bach T., A Hybrid Approach to Web Usage Mining, Data Warehousing and Knowledge Discovery, LNCS 2454, Y. Kambayashi, W. Winiwarter, M. Arikawa (Eds.), pp. 73-82, 2002.
- [8] Mobasher B., Cooley R. and Srivastava J., Creating Adaptive Web Sites through Usage-based Clustering of URLs, In Proceedings of 1999 Workshop on Knowledge and Data Engineering Exchange, USA, pp.19-25, 1999.
- [9] Smith K.A. and Ng A., Web page clustering using a self-organizing map of user navigation patterns, Decision Support Systems, Volume 35, Issue 2, pp. 245-256, 2003.
- [10] Jespersen S.E., Thorhauge J., and Bach T., A Hybrid Approach to Web Usage Mining, Data Warehousing and Knowledge Discovery, LNCS 2454, Y. Kambayashi, W. Winiwarter, M. Arikawa (Eds.), pp. 73-82, 2002.
- [11] Joshi, K.P., Joshi, A., Yesha, Y., Krishnapuram, R., Warehousing and Mining Web Logs. Proceedings of the 2nd ACM CIKM Workshop on Web Information and Data Management, pp. 63-68, 1999.
- [12] Cooley R., Web Usage Mining: Discovery and Application of Interesting patterns from Web Data, Ph. D. Thesis, Department of Computer Science, University of Minnesota, 2000.
- [13] H. hannahinbarani, k. thangavel "clickstream intelligent clustering using accelerated ant colony algorithm" ©2006 IEEE.
- [14] Annaalphy l, s. prabakaran "cluster optimization for improved web usage mining using ant nestmate approach" @IEEE 2011.
- [15] V. sujatha, Dr. punithaval li "an approach to user navigation pattern based on ant based clustering and classification using decision trees" @ 2010
- [16] Kobra Etminani, Mohammad-R. Akbarzadeh-T., Noorali Raeeji Yanehsari, "Web Usage Mining: users' navigational patterns extraction from web logs using Ant-based Clustering Method", IFSA-EUSFLAT 2009