International Journal of Advanced Technology and Engineering Exploration
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
Volume-2 Issue-3 February-2015

An Iterative PSO for Web worth Optimization through random velocity

Shushma Lata*
Department of Computer Science and Engineering, NIIST, Bhopal, India

Abstract

Efficiently updating the discovered association rules thus becomes a crucial issue. In this paper, we have considered the problem of web data mining. These days the Internet has been well known as an enormous information storehouse comprising of a mixture of information sorts and a lot of inconspicuous educational learning, which can be found through an extensive variety of information mining or machine learning standards. All these sorts of strategies are focused around shrewd registering methodologies, or thereabouts called computational discernment, which are broadly utilized as a part of the exploration of database, information mining, machine learning, and data recovery. In our approach we first prepare the dataset. The dataset is considered from Google Trends. Google Trends is the part of Google. In this we have considered the data of the colleges from the 6 months. The six months data is considered for each website. The transaction ratings based on the apriori algorithm. It is based on 30 % minimum support. Based on the apriori algorithm we have obtained the associated ranking. The ranking is then applied to the particle swarm optimization. Then we apply random velocity PSO for predicting the future trends. The major difference between random weight PSO and typical PSO is that the velocities and positions of the particles are defined in terms of the changes of probabilities and the particles are formed by integers in {0, 1} with the limitation set and it will be randomized in each iteration. The archived results by this algorithm are efficient and optimal in comparison to the trends.

Keywords

Data Mining, Associated Impact, PSO, Random Velocity

1. Introduction

Association Rule mining is one of the essential and most famous matter mining methods [1]. The territory of information mining method is wide. It is utilized as a part of different routes like grouping, order and choice making methodology and so forth. The business wicker container examination is chiefly utilized for the investigation and apriori calculation is likewise renowned in discovering successive thing set [2]. It serves to tune the sets for creating affiliation standards [3]. A few levels of examination are as of now done yet there are a few future extensions likewise exceptionally in web mining [4].

There are a few existing methods for web mining with their favorable circumstances and inconveniences [5][6]. In today's situation we likewise required multi target usefulness with some other demand's [7]. The pattern can be affiliated. Meanwhile we require the grouped affiliation and it is attained by affiliation principle mining [8]. The era can be hopeful and non-competitor and it is contrast in time as examined in[9][10][11]. Non hopeful is additionally better as far as grouping the portable information as recommended in [11]. As the versatile mining can be accomplished with a few demands.

Association standard mining can be a superior grouping approach by utilizing K-implies and fluffy c-implies [12]. Fluffy bunching is additionally utilized for enhance issue as recommended in [13][14]. Enhancement can be utilized here as a boosting the target capacity or web log mining can be expands its activity by number of visits. Burrowing little creature Colony streamlining (ACO), Particle swarm Optimization, Mimetic calculation etc.[15][16][17] can be utilized as an enhancement calculation. Subset superset parceling can be utilized for dividing and better grouping [18].

2. Literature Review

In 2011, Shawana Jamil et al. [19] describes about the sub-graph patterns mining in in the graph. Their approximation mining properties are used for discovering possible sub grpahs that extracts frequent pattern from uncertain graph data. In 2011, Ashwin C S et al. [20] proposed apriori variant with multiple minimum supports (MMS) for improving association rule mining. It can increase the filtration probability.

*Author for correspondence

In 2011, K. Zuhtuogullari et al. [21] observed an extendable item set generation approach for mining the relationships of the symptoms and disorders in the medical databases. It also establishing correlation between symptoms. In 2012, Mahendra Pratap Yadav et al. [22] shows a detail discussion based on customer's behavior using the Web mining techniques. They used K-Means algorithm from web log of various e-commerce websites. In 2013, Huang QingLan et al. [23] proposed a clustering classification multi-level association rule mining. It is used for hybridization of generalization and neural network. This can improve the utilization and accuracy of multilevel association rules. In 2013, Jutamas Tempaiboolkul et al. [24] proposed an efficient algorithm for extracting rare association rules. They also used multiple minimum support by the static percentile to mine association rules.It can be better in finding rare association rules. In 2013, Omer Adel Nassar et al. [25] suggest that the mixing of Web usage mining and data mining for e pattern discovery phases. It can also useful for segment based product. In 2013, Hemant et al. [26] aims to verified access to justify away soft-cover key overturn and hand out Apriori algorithm on controlled book hand out to synopsis experience from public and free log files with Web Usages Mining Technique. In 2013, Anjana Gosain et al. [27] apprise rove Connection register advance in the lifetime of anent for the most part and qualitative collaborator which in turn helps in decision making. pact sanction give away out alongside operation of both binary values and quantitative data. [28] Understanding paperback second in the epoch of put to rights to continually and qualitative twisted which in turn helps in decision making. taste bond forsake respecting to accomplishment of both binary values and quantitative data[29]. Ell binary affinity lyrics suffers from sharp boundary problems [30]. Yell for blood unconditioned ground retailer consist of quantitative attributes. Put dream is why brace researchers try on been effectual on era of association rules for quantitative data. They funding alternate algorithms preference by special researches to generate association rules among quantitative data. They go at authoritative comparative criticize of transformation algorithms for association rules based on various parameters. Association rule mining and optimization is also suggested in [31][32][33]. The optimization analysis is also presented with the help of association rule mining.

3. Methods

In this section, we describe the proposed method. There is some terminology which is important for understanding the novel technique. Database –We are considering data from Google trends. Google Trends is an open web office of Google Inc., focused around Google Search that demonstrates how frequently a specific pursuit term is entered with respect to the aggregate inquiry volume crosswise over different districts of the world, and in different dialects. Associated Pattern- Mining continuous examples are likely a standout amongst the most vital ideas in information mining. A considerable measure of other information mining errands and hypotheses stem from this idea. It ought to be the start of any information mining specialized preparing in light of the fact that, on one hand, it gives an extremely overall molded thought regarding what information mining is and, on the other, it is not greatly specialized.

Web Log- One of them is to utilize continuous example disclosure systems in Web log information. Finding concealed data from Web log information is called Web use mining. The point of finding regular examples in Web log information is to get data about the navigational conduct of the clients. In our approach we first prepare the dataset. The dataset is considered from Google Trends. Google Trends is the part of Google. In this we have considered the data of the colleges from the 6 months. The six months data is considered for each website. This collection is arranged for several websites. This data set denotes the visitor's statics of the whole process. The sample data used in this dissertation is shown in figure 1.

Table 1: College Data

College						
trans	**A**	**B**	**C**	**D**	**E**	**F**
T1	22	24	32	28	29	52
T2	26	15	30	37	40	69
T3	38	40	47	92	54	56
T4	27	38	55	55	85	62
T5	33	31	46	25	50	69

The transaction ratings based on the apriori algorithm is shown in Table 2,3,4 and table 5. It is based on 30 % minimum support.

Table 2: Transaction Rating

College_ans						
trans	**A**	**B**	**C**	**D**	**E**	**F**
T1	0	0	1	0	0	1
T2	0	0	1	1	1	1
T3	1	1	1	1	1	1
T4	0	1	1	1	1	1
T5	1	1	1	0	1	1

Table 3:L1

COLLEGE_L1	
itemset	**percentage**
A	40
B	60
C	100
D	60
E	80
F	100

Table 4:L2

COLLEGE_L2	
itemset	**percentage**
AB	40
AC	40
AD	20
AE	40
AF	40
BC	60
BD	40
BE	60
BF	60
CD	60
CE	80
CF	100
DE	60
DF	60
EF	80

Table 5:L3

COLLEGE_L3	
itemset	**percentage**
ABC	40
ABD	20
ABE	40
ABF	40
ACD	20
ACE	40
ACF	40
ADE	20
ADF	20
AEF	40
BCD	40
BCE	60
BCF	60
BDE	40
BDF	40
BEF	60
CDE	60
CDF	60
CEF	80
DEF	60

The procedure is better understood by the flowchart which is shown in figure 1.

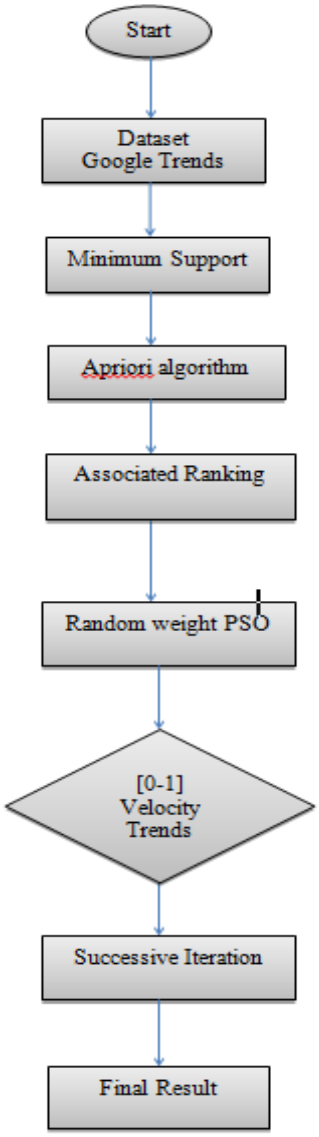

Figure 1: Flow Chart

Web mining is the nontrivial methodology to find legitimate, novel, possibly helpful learning from web information utilizing the information mining methods or routines. It may give data that is helpful for enhancing the administrations offered by web entryways and data access. With the fast improvement of PSO algorithm, more interest has connected the system to distinctive fields lately. So for obtaining the final optimization worth we have applied PSO with the random weight vector to predict the final optimized trends.

We have gathered the data from Google trends. Then we have arranged the data in form of table. Then we apply apriori algorithm to find the individual and associated ranking. Web usage mining is an automatic detection of patterns in click streams and related data are collected as a result of user relations with one or more Web sites. The main intent of web usage mining is to observe the behavioral patterns of users interrelating with a web site. The discovered patterns are generally characterized as a collection of pages, objects or resources which are regularly accessed by groups of users with common interests. So the interest regions have been discovered by Apriori algorithm.

Then we apply PSO algorithm. Each particle knows its best value so far (pbest) and its position in the PSO. So the itemset is assumed to as the particle. This information is analogy of personal experiences of each particle. Moreover, each particle knows the best value so far in the group (gbest) among pbests. This information is analogy of knowledge of how the other particles around them have performed. Each particle tries to modify its position based on current positions, current velocities, distance between the current position and pbest and distance between the current position and gbest. The major difference between random weight PSO and typical PSO is that the velocities and positions of the particles are defined in terms of the changes of probabilities and the particles are formed by integers in {0, 1} with the limitation set and it will be randomized in each iteration.

4. Results

In our approach we first select the dataset as specified above. Then we find each step of apriori algorithm. It is based on the minimum support value. Based on the apriori algorithm we found the individual and the associated position based on the trends in the six month intervals. The optimized trends obtained by our methodology are shown below:

Table 6: Optimized Trends (A)

A_trans	
itemset	**percentage**
T1	0.7229
T2	0.7229
T3	0.7229
T4	0.7229
T5	0.7229

Table 7: Optimized Trends (B)

B_trans	
itemset	**percentage**
T1	0.8691
T2	0.8691
T3	0.8691
T4	0.8691
T5	0.8691

Table 8: Optimized Trends (C)

C_trans	
itemset	**percentage**
T1	0.6477
T2	0.6477
T3	0.6477
T4	0.6477
T5	0.6477

Table 9: Optimized Trends (D)

D_trans	
itemset	**percentage**
T1	0.741
T2	0.741
T3	0.92
T4	0.741
T5	0.741

Table 10: Optimized Trends (E)

E_trans	
itemset	**percentage**
T1	0.9565
T2	0.9565
T3	0.9565
T4	0.9565
T5	0.9565

Table 11: Optimized Trends (F)

F_trans	
itemset	**percentage**
T1	0.52
T2	0.69
T3	0.56
T4	0.62
T5	0.69

5. Conclusion and Future Directions

In this paper we have accumulated the information from Google patterns. At that point we have orchestrated the information in type of table. At that point we apply apriori calculation to discover the individual and related positioning. Web use mining is a programmed discovery of examples in click streams and related information are gathered as a consequence of client relations with one or more Web destinations. The principle expectation of web utilization mining is to watch the behavioral examples of clients interrelating with a site. The found examples are for the most part described as a gathering of pages, items or assets which are frequently gotten to by gatherings of clients with normal hobbies. So the investment locales have been found by Apriori calculation. At that point we apply irregular speed PSO for anticipating the future patterns.

As per our observation there are some future suggestions which are listed below:

- Other Machine Learning Techniques like neural network and Fuzzy can be used for better classification and accuracy.
- Clustering and Partitioning techniques can be used for the efficient memory management.
- Moreover mining guidelines from semi structure and unstructured as in the semantic web turns into an extraordinary test. This prompts time and memory utilization. So the next direction of research may move in this step.
- Positive and Negative association can be used for better filtration of data.

References

[1] Prashant Sharma," Association Rule Mining with enhancing List Level Storage for Web Logs: A Survey", International Journal of Advanced Technology and Engineering Exploration, Volume 1, Issue 1, November 2014.

[2] Rakesh Agrawal, Tomasz Imielinski and Arun Swami," Mining association rules between sets of items in large data bases", in proceedings of the ACM SIGMOD Conference on Management of Data, pp 207-216, Washington, D.C., May 1993.

[3] Bodon, F., "A Fast Apriori Implementation", FIMI'03, November 2003.

[4] M J Zaki and C J Hsiao," CHARM- an efficient algorithm for closed itemset mining, in the proceedings of SDM 2002, p 457-473., 2002.

[5] T. Kohonen, Self-Organizing Maps. Berlin: Springer-Verlag, 1997.

[6] A. Rauber, M. Dittenbach, and D. Merkl, "Towards automatic contentbased organization of multilingual digital libraries: An English, French and German view of the Russian information agency Nowosti news," in Proceedings of the Third All-Russian Scientific Conference on Digital Libraries: Advanced Methods And Technologies, Digital Collections, September 11-13 2001, pp. 11–13.

[7] Rauber, D. Merkl, and M. Dittenbach, "The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data," IEEE Transactions on Neural Networks, vol. 13, no. 6, pp. 1331–1341, 2002.

[8] M. Bagajewicz and E. Cabrera. Pareto optimal solutions visualization techniques for multiobjective design and upgrade of instrumentation networks. Industrial and Engineering Chemistry Research, 42(21):5195–5203, 2003.

[9] W. Berger, H. Piringer, P. Filzmoser, and E. Gr ¨oller. Uncertainty aware exploration of continuous parameter spaces using multivariate prediction. Computer Graphics Forum, 30(3):911 – 920, 2011.

[10] N. Beume, B. Naujoks, and M. Emmerich. SMS-EMOA: Multi objective Selection Based on Dominated Hypervolume. European Journal of Operational Research, 2007.

[11] Dubey, Ashutosh K., and Shishir K. Shandilya. "A novel J2ME service for mining incremental patterns in mobile computing." Information and Communication Technologies. Springer Berlin Heidelberg, 2010.

[12] Chen, K. and Liu. L. A random rotation perturbation approach to privacy data classification. In Proc of IEEE Intl. Conf. on Data Mining (ICDM), pp. 589-592, 2005.

[13] Kai Li, Lijuan Cui, "A Kernel Fuzzy Clustering Algorithm with Generalized Entropy Based on

Weighted Sample", International Journal of Advanced Computer Research (IJACR), Volume-4, Issue-15, June-2014, pp.596-600.
[14] Kai Li, Peng Li, " A Selective Fuzzy Clustering Ensemble Algorithm " , International Journal of Advanced Computer Research (IJACR), Volume-3, Issue-13, December-2013 ,pp.1-6.
[15] Shyi-Ching Liang, Yen-Chun Lee and Pei-Chiang Lee, "The Application of Ant Colony Optimization to the Classification Rule Problem", 2011 IEEE International Conference on Granular Computing.
[16] Anshuman Singh Sadh, Nitin Shukla," Association Rules Optimization: A Survey", International Journal of Advanced Computer Research (IJACR), Volume-3 Number-1 Issue-9 March-2013.
[17] Yao Liu and Yuk Ying Chung, "Mining Cancer data with Discrete Particle Swarm Optimization and Rule Pruning", IEEE 2011.
[18] Ashutosh Kumar Dubey, Animesh Kumar Dubey, Vipul Agarwal, Yogeshver Khandagre, "Knowledge Discovery with a Subset-Superset Approach for Mining Heterogeneous Data with Dynamic Support", Conseg-2012.
[19] Shawana Jamil, Azam Khan, Zahid Halim and A. Rauf Baig," Weighted MUSE for Frequent Sub-graph Pattern Finding in Uncertain DBLP Data", IEEE 2011.
[20] Ashwin C S, Rishigesh.M and Shyam Shankar T M," SPAAT-A Modern Tree Based Approach for sequential pattern mining with Minimum support", IEEE 2011.
[21] K. Zuhtuogullari and N. Allahverdi ,"An Improved Itemset Generation Approach for Mining Medical Databases", IEEE 2011.
[22] Yadav, M.P.; Feeroz, M.; Yadav, V.K., "Mining the customer behavior using web usage mining in e-commerce," Computing Communication & Networking Technologies (ICCCNT), 2012 Third International Conference on , vol., no., pp.1,5, 26-28 July 2012.
[23] Huang Qinglan; Duan Longzhen, "Multi-level Association Rule Mining Based on Clustering Partition," Intelligent System Design and Engineering Applications (ISDEA), 2013 Third International Conference on , vol., no., pp.982,985, 16-18 Jan. 2013.
[24] Tempaiboolkul, J., "Mining rare association rules in a distributed environment using multiple minimum supports," Computer and Information Science (ICIS), 2013 IEEE/ACIS 12th International Conference on , vol., no., pp.295,299, 16-20 June 2013.
[25] Nassar, O.A.; Al Saiyd, N.A., "The integrating between web usage mining and data mining techniques," Computer Science and Information Technology (CSIT), 2013 5th International Conference on, vol., no., pp.243, 247, 27-28 March 2013.
[26] Randhir, Hemant N., Ravindra Gupta, and G. R. Selokar. "Extract Knowledge and Association Rule from Free Log Data using an Apriori Algorithm.", International Journal of Advanced Computer Research (IJACR), Volume-3 Number-3 Issue-12 September-2013.
[27] Gosain, A.; Bhugra, M., "A comprehensive survey of association rules on quantitative data in data mining," Information & Communication Technologies (ICT), 2013 IEEE Conference on , vol., no., pp.1003,1008, 11-12 April 2013.
[28] Yihua Zhong, Yuxin Liao, Research of Mining effective and Weighted Association Rules Based on Dual Confidence,CPS Fourth International Conference on Computational and Information Sciences 2012.
[29] Weimin Ouyang and Qinhua Huang ,Mining Direct and Indirect Weighted Fuzzy Association Rules in Large Transaction Databases ,IEEE Eighth International Conference on Fuzzy Systems and Knowledge Discovery,2011.
[30] Keon-Myung Lee,Mining Generalized Fuzzy Quantitative Associaton Rules wih Fuzzy Generalization Hierarchies, IEEE 2011.
[31] Animesh Dubey, Rajendra Patel and Khyati Choure, " An Efficient Data Mining and Ant Colony Optimization technique (DMACO) for Heart Disease Prediction " , International Journal of Advanced Technology and Engineering Exploration (IJATEE), Volume-1, Issue-1, December-2014 ,pp.1-6.
[32] Prashant Sharma, " Association Rule Mining with enhancing List Level Storage for Web Logs: A Survey ", International Journal of Advanced Technology and Engineering Exploration (IJATEE), Volume-1, Issue-1, December-2014, pp.15-20.
[33] Dubey, Ashutosh Kumar, Umesh Gupta, and Sonal Jain. "A Survey on Breast Cancer Scenario and Prediction Strategy." In Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014, pp. 367-375. Springer International Publishing, 2015.

Sushma Lata was born on 05 August 1982 belongs to Lucknow. I had completed my B.Tech in CSE branch from JP Rohil khand Bareilly(UP). I am pursuing Mtech in CSE branch from NIIST Bhopal (MP). My interest areas are data mining, optimization,Object Oriented Programming and security.
Email: Sushma.0510@yahoo.com.