International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 22 (2017) pp. 11818-11823

User Interesting Navigation Pattern Discovery Using Fuzzy Correlation Based Rule Mining

[1]D. Uma Maheswari
Research Scholar,Department of Computer Science,
Karpagam Academy of Higher Education, Karpagam University, Coimbatore
Tamil Nadu, India.
Orcid Id: 0000-0002-3944-5918

[2]Dr. R.Gunasundari
Associate Professor, Department of Information Technology,
Karpagam Academy of Higher Education, Karpagam University, Coimbatore,
Tamil Nadu, India.

Abstract

The explosive growth of Internet has given rise exclusively on the web server log. The prediction of user navigation pattern can be determined using web log file which contains the massive hidden valuable information pertaining to the visitors. The extraction of useful information from these data has proved to be very useful for optimizing Web Sites and Promotional Campaigns for marketing. But discovering the frequent web user navigation pattern is a very toughest challenge in web log file. The paper concentrates the web usage mining topic; it is one of the intensive research quarters as its latent for the tailored services and adaptive web sites. This proposed work focuses on adapting a fuzzy based inference system to generate a set of rules from the clustered pattern more efficiently and effectively for identifying the significant user navigation pattern. This paper starts with the preprocessing of the given weblog, followed by clustering them and finding correlation based Fuzzy rule mining. These rules provide knowledge that helps to improve the website designing, in advertising, web personalization while comparing with two existing approaches fuzzy association rule mining and association rule mining.

Keywords: web log file, fuzzy, association rule, navigation, correlation, patterns.

INTRODUCTION

At the present time, Web becomes the vertebrae of information. The key problem for the internet users is being unable to retrieve functional and applicable in sequence. By analyzing the browsing patterns of the users will assist the organizations to suggest the further relevant trap leaf according to the current gain of the user. Analyzing and modelling web navigation behavior is helpful in understanding the demands of Web users.

Application of mining of web log files of a web server consists of three main tasks:

Preprocessing stage eliminates all the unrelated user's request from web server log files resulted in log reduction chased by users and sessions identification. In pattern discovery, similar navigation pattern and form clusters from web log files are extracted.

In pattern analysis, the set of generated rules from the clustered dataset are analyses.Rule discovery using Association rule is the major technique frequently used technique in is one of the data mining tasks, it will determine and reveal the relationship among data. A set of transactions are used for analysis and using it with significant association among the dataset is identified in a given database. While, the size of data is really bigger, the association desires the huge investigation space depending on the support count.

Rules discovery finds general rules in the format X=>Y, meaning that, when page X is visited in a transaction t1 then the page Y will also be visited in the same transaction t1. Different confidence and support values may be used for these kinds of rules [15]. Confidence is defined as the proportion between the number of transactions holding both items of rule and number of transactions enclosing just the antecedent. Support is the gain of transactions in the rule is true.

Literature Survey

This section discusses about the various existing approaches on web user navigation pattern analysis from the web log data.

In paper [1] the authors have predicted the users' navigation pattern in the web log file with high precision, a fuzzy clustering based hybrid algorithm with weighted association rules are presented. This algorithm contains two phases, specifically offline and online phases.

Anagha et al [2] have proposed the Constraint based web mining approach used to decrease the size of association rules obtained from Web log. The method provides evidence which

is effective in dropping the repetition of information and also advances the effectiveness of mining tasks.

AlMurtadha et al. [3] have focused on enhancing the forecast of the next visited web pages and recommended it to the present mysterious user by conveying them to the finest navigation profiles acquired by preceding navigations of analogous interested users. Castellano et al [4] has exploited a usage based web recommendation namely NEWER which determines the prospective of Computational Intelligence techniques to recommend dynamically interesting pages to users according to their preferences.

Almurtadha et al [5] have focused on Profile Aggregation based on clustering of Transactions which is an enhanced recommendation system for website development and understanding user navigation pattern in web. In [6], Fuzhi et al have presented a recommendation algorithm in two stages using K-means clustering in Mobile E-commerce. K.Thangadurai et al [7] have proposed Rough set based Clustering and this technique guarantees the rough K-means clustering as a most promising pattern discovery technique and evaluated with the conventional K-means and weighted K-Means clustering methods for various data sets.

Crowley [8] has pioneered a model based on temporal pattern called as Temporal Tree Associative Rule a data mining approach. This can be used both for inaccuracy and temporal uncertainty of temporal events to be stated as Symbolic Time Sequences. Maragatham. G and Lakshmi. M [9] has proposed an algorithm that capable to mine the temporal association rules based on utilities by adjusting the support with pertinent to the time periods and utility.

Mobasher et al. [10] have projected an effectual and highly optimal technique for Web personalization based on the determining rule using association from usage data. It is used for developing a recommender system. In this work, the proposed system anticipated a scalable framework for recommender systems using association rule mining from click stream data.

Suneetha and Krishnamoorti [11] have recommended an enhanced version of Apriori algorithm to extracts appealing correlations, frequent patterns, and associations along with web pages visited by users in their browsing sessions.

This paper proposes a fuzzy correlation rule mining in which from the clustered data interesting navigation pattern of web users are determined efficiently by eliminating the misleading rules generated by traditional association rule mining.

Fuzzy Association Rule Mining

Fuzzy association rule mining [12, 13] is a method to locate out the fuzzy itemsets or fuzzy attributes which frequently occur together from a fuzzy dataset

For a given database D with transactions T = { t1,t2, t3,..., tn} with pages visited items I = {i1, i2, i3, .. in} and converted fuzzy transactions T′ ={ t1′ t2′ t3′,.. tn'}

A fuzzy association rule, say F_X-> F_Y , holds in fuzzy dataset T with fuzzy support f_{supp} $(\{F_X,F_Y\})$ and with fuzzy confidence $f_{conf}(\{F_{x,->}F_Y\})$ defined as follows

$$f_{supp}(\{F_X, F_Y\}) = \frac{\sum_{i=1}^{n} \min(f_j(t_i) | f_j \in \{F_X, F_Y\})}{N}$$

$$f_{conf}(F_X \rightarrow F_Y) = \frac{\text{fsupp}(\{F_X, F_Y\})}{\text{fsupp}(F_X)}$$

If f_{supp} $(\{F_{x->},F_Y\})$ is greater than or equal to the user-predefined minimal fuzzy support (fs) and f_{conf} $(F_{x->},F_Y)$ is also greater than or equal to the user-predefined minimal fuzzy confidence ($f\,c$), then, fuzzy association rule $F_{x->},F_Y$ is considered as an interesting fuzzy association rule, and it means that F_x and F_y frequently occur together in session.

PROPOSED SYSTEM

This section discusses about the proposal of a system that discover user navigation based interesting patterns in these weblogs. The web log file is in the form of sequence of events where each record represents the session with the concern page navigation. The access sequence events are clustered based on the fuzzy roughest theory based Jaccard Distance Similarity measure. Here the similarity between the sessions is identified and framed different clusters. The sessions with higher similarity are grouped in a same cluster where as the dissimilar one are presented in the other cluster. This process is performed to discover the user navigation pattern. This proposed work extends the earlier work to identify the interesting navigation pattern of the web log file by proposing a correlation based fuzzy association rule mining. The Rules with higher support, confidence and correlation produces more promising result to determine the user navigation which will be produced as recommendation to optimize website design.

Our proposed system would involve the following steps:

- The input is a set of Weblogs file from msnbc.com site
- These records are stored into the database.
- Using a Fuzzy Rough Set Jaccard Similarity based User Session Clustering these entries are divided into clusters.
- Now, Correlation based fuzzy association rule mining is applied on these clusters, to obtain interesting pattern of user navigation on web pages having fuzzy minimum support, Fuzzy minimum confidence and Fuzzy minimum correlation.

- As a result of Fuzzy association rule mining, interesting patterns can be discovered and client's web usage can be evaluated.

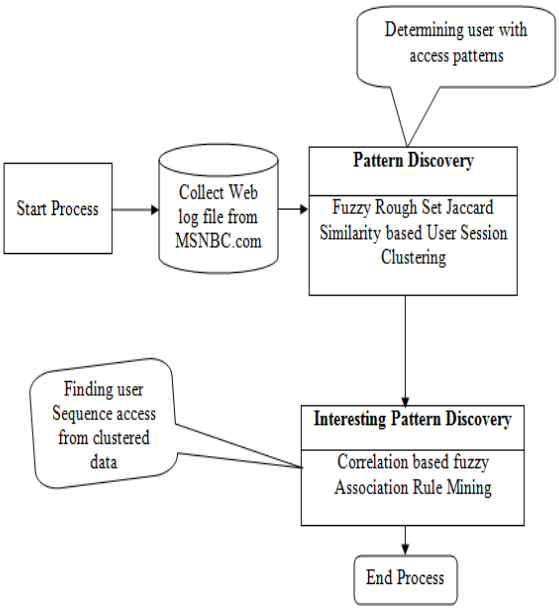

Figure 1: Proposed work

About the Web Log File

In this proposed work, the bits of Web logs as the sequences of events to find the associations among web pages on the origin of sequential patterns over a period of time. Let EV be a set of events. A Web log piece or (Web) access sequence seq = ev1,ev2,ev3,…evn (evi ∈ EV) for (1<= i <= n) is a chain of procedures, while n is called the length of the access sequence. An access sequence with length n is also called n-sequence. In an access sequence seq, repetition is allowed. Duplicate references to a page in a web access chain imply the reverse traversals, Refreshes or reloads. For example, 2, 2, 3 and 2, 3 are two unusual access sequences, in which 2 and 3 are two events. The Data used for the testing comes from Internet Information Server (IIS) logs for msbc.com and news related portions of msn.com for one whole day. Each chain in the data set corresponds to page views of a user during that day. There are 1 million records and we selected 60,000 samples. Each event in the sequence corresponds to a request for a page. Requests are record barely at the stage of page sort. There are 17 page categories represented using their corresponding integers assigned, namely front page (1) , news (2), tech(3), local(4), opinion (5), on-air(6), misc(7), weather (8), health (9), living (10), business(11), sports (12), summary (13), bulletin board service (14), travel (15), msn-news (16), and msn-sports (17).

Fuzzy Rough Set Jaccard Similarity based User Session Clustering

In order to determine the number of clusters in this algorithm, the Fuzzy Roughest Theory Based Jaccard Distance Similarity Clustering [FRSTJSC} is used to cluster the obtained sessions and the method which is suggested in our previous work [14] is used. The users' sessions are clustered based on the similarities among the visited web pages. After applying the FRSTJSC method to the obtained sessions, several cluster centers are resulted as based on the Fuzzy rough set JACCARD similarity to get the following clusters.

C1={S1,S2,S3,S5,S7,S8}

C2={S4,S6}

C3={S2,S7,S9}

C4={S1,S3,S8,S10}

Each resultant cluster indicates a subset of the users' sessions.

Fuzzy Correlation Rule Mining

This section discusses about the proposed work of correlation based fuzzy association rule mining in web log interesting pattern discovery. From the resultant cluster set obtained using FRSTJSC for pattern discovery rules are generated.

Now let us consider that FC = {fc1, fc2... fcn} be a set of fuzzy cluster itemset. Let S= {s1, s2, s3...sm} be a set of sessions and each session represented with access sequence of the web user. The set {fc1(s1), fc2(s1),..fcn(s1)}

where fci(s1) represents the membership degree that s1 belong the fuzzy cluster fci. Here in terms of membership representation fci = $\mu_{fci}(s1)$, fci(s1) ∈ [0 1]. The user defined minimal fuzzy support is termed as $Support_{fuzzy}$. The user defined minimal fuzzy confidence is denoted as $Confidence_{fuzzy.}$ User defined minimal fuzzy correlated is coined as ρ. The following procedure shows step by step process of proposed correlation based fuzzy rule mining.

Steps

1. For each fuzzy cluster Item fci ∈ FC, fsupport(fci) is determined.
2. Assume $Level_1$ ={ FCi| FCi ∈ FC, fsupport(FCi)>=$Support_{fuzzy}$} is the first level frequent fuzzy cluster item whose size is equal to 1.
3. Let CC_2 = {FC_A, FC_B} is the set of candidate combination set generated by the join of $Level_1$ itself. FC_A, FC_B∈ $Level_1$ and FC_A≠ $FC_{B=.}$ The number of the fuzzy items of each element of CC_2= 2.
4. For each element of cc_2 say ({FC_A, FC_B}) the fuzzy support (*fsupport* ({FC_A, FC_B}) and the fuzzy

correlation coefficient ($\rho_{A,B}$) between FC_A and FC_B are computed.

$$\text{fsupport}(\{FC_A,FC_B\}) = \frac{\sum_{i=1}^{n} \min(fcj(Si) \mid fcj \in \{FC_A, FC_B\})}{n}$$

If *f support* ($\{FC_A, FC_B\}$) is greater than

$$\rho_{A,B} = \frac{P_{A,B}}{\sqrt{P_A^2 . P_B^2}}$$

$$P_{A,B} = \frac{\sum_{i=1}^{n} (\mu_A(x_i) - \overline{\mu_A}).(\mu_B(x_i) - \overline{\mu_B})}{n-1}$$

$$\overline{\mu_A} = \frac{\sum_{i=1}^{n} (\mu_A(x_i))}{n}$$

$$\overline{\mu_B} = \frac{\sum_{i=1}^{n} (\mu_B(x_i))}{n}$$

$$P_A^2 = \frac{\sum_{i=1}^{n} (\mu_B(x_i) - \overline{\mu_B})^2}{n-1}$$

$$P_B^2 = \frac{\sum_{i=1}^{n} (\mu_A(x_i) - \overline{\mu_A})^2}{n-1}$$

5. If *fsupport* ($\{FC_A, FC_B\}$) is greater than or equal to $support_{fuzzy}$, and $P_{A,B}$, is greater than or equal to ρ then the combination (FC_A, FC_B,)is an element of $Level_2$. Therefore the *Level*2 is the set of large (or frequent) combinations of two fuzzy itemsets of Level1.

6. Next, each CCk , k >= 3 , can be generated by $Level_{k-1}$ joint with itself. Suppose (FC_A, FC_B,) and (FC_C, FC_D) are two elements of Level k −1 , and one of the fuzzy cluster itemsets of (FC_A, FC_B,) say FC_A , is equal to one of the fuzzy cluster itemsets of ((FC_C, FC_D,) say FC_C , and the total number of the fuzzy items of the combination (FC_A { FC_B , FC_D }) is equal to k , and it will also be a element of CC k .

7. Next, for each element of CCk, the fuzzy support and the fuzzy correlation coefficient are still used to select the elements of $Level_k$.

8. When each $Level_k$, k >= 2 , is obtained, for each element of $Level_k$, say (F_X ,F_Y) , two fuzzy rules, F_x->F_y and F_y->F_x , can be generated. If the fuzzy confidence of rule is greater than or equal to $confidence_{fuzzy}$, then it is considered as an interesting fuzzy correlation rule.

9. The algorithm continues until no next CC_{k+1} can be generated.

EXPERIMENTAL RESULT

The Experimental Result was conducted using msnbc.com and news related portion of msn.com. The proposed work enhanced the traditional fuzzy association based rule mining by adapting correlation based mining and was implemented using MATLAB. 60,000 samples were selected. Each event in the sequence corresponded to a request for a page. This work extended our previous approach of fuzzy rough set based clustering technique. With the input of clustered dataset result was conducted based on the fuzzy support vs frequent itemset genearation, fuzzy confidence vs interesting rule discovery and fuzzy correlation vs interesting rule discovery. The proposed work performance was compared with two existing approaches namely association rule and fuzzy association rule which have higher chance of generating misleading rules.

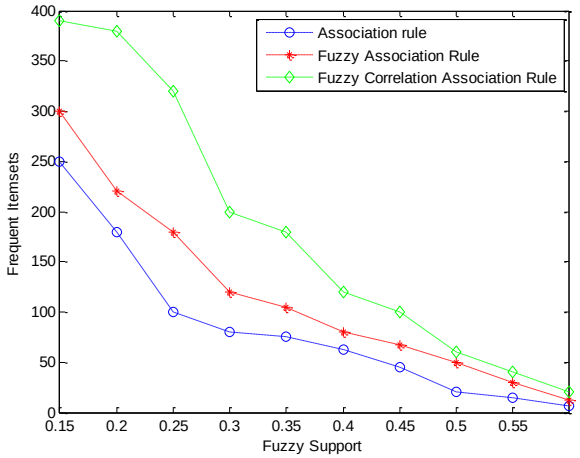

Figure 2: Performance comparison of association rule, fuzzy association rule and proposed fuzzy correlation association rule based on Fuzzy support vs Frequent Itemsets.

Figure2 depicts the variance between the number of large itemsets generated from the association rule, fuzzy association rule and the new correlation based fuzzy rule mining approach using different fuzzy support values. The number of large itemsets increases as the minimum support decreases.

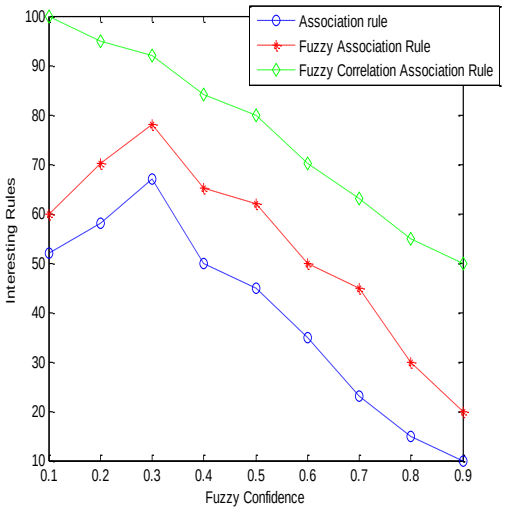

Figure3: Performance comparison of association rule, fuzzy association rule and proposed fuzzy correlation association rule based on Fuzzy Confidence vs Frequent Itemsets

Figure 3 shows the number of interesting rules generated using user specified fuzzy confidence with different values. The comparison shows that the more interesting rules are generated by the proposed work due to the generation for higher number of frequent itemsets using fuzzy support while comparing the other two existing approaches.

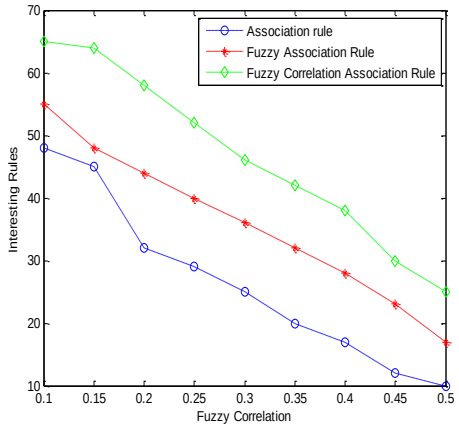

Figure 4*:* Performance comparison of association rule, fuzzy association rule and proposed fuzzy correlation association rule based on Fuzzy Correlation vs Frequent Itemsets

In Figure 4, the interesting rules for user navigation pattern is generated based on the fuzzy correlation mining in which while comparing the fuzzy confidence it generates less number of rules. By using fuzzy confidence alone fuzzy rules can mine centers for discovering fuzzy clustered itemset which frequently occur.

These exposed rules are identified as fuzzy association rules. Conversely, two fuzzy clustered itemsets which frequently occur together can not imply which is forever an interesting relationship among them. To handle with this circumstance, fuzzy correlation analysis is used to aid in discovering the fuzzy correlation rules. The proposed work outperforms the remaining existing approaches using its property of fuzzy correlation analysis.

CONCLUSION

Mining of interesting and useful knowledge from the user navigation behavior over the internet help to develop a optimized web design for marketing companies or online sales. Designing and offering users wish to look for in websites is the vital aspire of web usage mining. In this approach, this aspire is satisfied by using fuzzy correlation based rule mining technique on clustered data obtained by fuzzy rough set theory and this sessions are clustered based on the similarity of web page access sequence and then using the clustered dataset the fuzzy rules are mined using correlation analysis based on the membership degree of each sessions on different clusters. To avoid the misleading rule generation as done by the traditional techniques this proposed work substitute's structure to avoid the generation of misleading fuzzy rules. The result of the proposed work also shows the most promising result on the usage of less search space by considering the cost and time factor for mining interesting relationship between fuzzy clusters itemsets generated based on fuzzy correlation analysis and thus, the discovered fuzzy rules are called fuzzy association correlation rules.

REFERENCES

[1] Zeynab Liraki, v, Javad Mirabedini Predicting theusers' Navigation Patterns in Web, using Weighted Association Rules and Users' Navigation Information, International Journal of Computer Applications (0975 – 8887) Volume 110 – No. 12, January 2015.

[2] Anagha Shastri , Dipti Patil, V.M.Wadha Constraint-based Web Log Mining for Analyzing Customers' Behaviour, International Journal of Computer Applications (0975 – 8887) Volume 11 No.10, December 2010.

[3] AlMurtadha, Y.M., M.N.B. Sulaiman, N. Mustapha and N.I. Udzir,2010. Mining web navigation profiles for recommendation system. Inform.Technol. J., 9:790-796. DOI:10.3923/itj.2010.790.796

[4] Castellano, G., Fanelli, A.M., & Torsello, M.A. (2011). NEWER: A system for NEuro-fuzzy Web recommendation. Applied soft Computing,11(1), 793-806.

[5] AlMurtadha, Y.,Sulaiman, M..N.B., N. Mustapha and N.I. Udzir,(2011). IPACT: Improved web page recommendation System Using Profile Aggregation Based on Clustering of Transactions, American Journal of Applied Sciences, 8(3),277-283.

[6] Fuzhi ZHANG, Huilin LIU, jinbo CHAO, A Two-stage Recommendation Algorithm based on K- means Clustering in Mobile E-commerce, Journal of Computational Information Systems 6:10 (2010) 3327-3334.

[7] Dr.K.Thangadurai, M.Uma, Dr.M.Punithavalli, A Study On Rough Clustering Global Journal of Computer Science and Technology Vol. 10 Issue 5Ver. 1.0 July 2010 P a g e | 55.

[8] Mathieu Guillame-Bert, James L. Crowley, "NewApproach on Temporal Data Mining for Symbolic Time Sequences: Temporal Tree Associate Rules," ictai, pp.748-752, 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence, 2011.

[9] Maragatham. G and Lakshmi. M ―A Strategy for Mining Utility based Temporal Association Rules‖ ,IEEE 2010, pp 38-41.

[10] B. Mobasher, R. Cooley, and J. Srivastava, "Automatic personalization based on web usage mining," Communications of the ACM, vol. 43, no. 8, pp. 142–151, 2000.

[11] K. R. Suneetha and R. Krishnamoorti, "Web log mining using improved version of apriori algorithm," International Journal of Computer Applications, vol. 29, no. 6, pp. 23–27, 2011.

[12] Dubois, D., Hüllermeier, E. and Prade, H., "A Systematic Approach to the Assessment of Fuzzy Association Rules," Data Mining and Knowledge Discovery Journal, Vol. 13, No. 2, (2006), 167 – 192.

[13] Xie, D. W., "Fuzzy Association Rules discovered on Effective Reduced Database Algorithm," In Proc. IEEE Conf.on Fuzzy Systems, 2005.

[14] Uma Maheswari , Dr. A. Marimuthu, An Ensemble Fuzzy Rough Set Jaccard Similarity measure based Approach on User Session Clustering, International Journal of Computer Systems (IJCS), ISSN: 2394-1065 Series: Volume 03, Number 04, April 2016.

[15] M. Henri Briand, M. Fabrice Guillet, M. Patrick Gallinari, M. Osmar Zaaiane, "Web Usage Mining: Contributions to Intersites Logs Preprocessing and Sequential Pattern Extraction with Low Support", World Academy of Science, Engineering and Technology 48 2008.