ISSN (Print): 0974-6846 ISSN (Online): 0974-5645

Optimum Session Interval based on Particle Swarm Optimization for Generating Personalized Ontology

S. Chitra^{1*} and B. Kalpana²

¹Department of Computer Science, Government Arts College, Coimbatore, Tamil Nadu, India; schitra789@gmail.com

²Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India

Abstract

A semantic web usage mining method is suggested to identify the association between consumer emotions and buying behaviors by utilizing the web log data. Fuzzy logic is used to signify the temporal conception and resource attributes for the requested URLS of web access activities. From this, a Personal Web usage Ontology is created which facilitates semantic web applications. But the limitation is less efficient in terms of accuracy and user satisfaction. Thus an innovative technique which is called Optimum Session Interval based Particle Swarm Optimization (OSIPSO) is introduced. This technique is used to find the optimum session interval. The Particle swarm optimization has no overlapping and mutation computation and it is proficient in global search. Additionally, an associative classification is used to enhance the accuracy. Associative classification is a combination of associative rule mining and classification rule mining. An experimental result shows that the proposed work has a high accuracy and high efficient.

Keywords: Associative Classification, Emotion and Behavior Profiling, Ontology Generation, Semantic Web, Particle Swarm Optimization

1. Introduction

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. Human motions are a significant factor of human behaviors in web mining analysis¹. The relations between customer sensations and their buying behaviors have been well recognized^{2,3}. In the web applications, the consumer emotions and behaviors are significant to enhance the performance.

The customers self-report is used to analyze the emotions of the particular resources. After the user visit the website there exists an opportunity for the users to record their suggestions in the emotional state. By utilizing the records provided by the users it is possible to analyze the emotional influence of the users. Web usage mining is an essential approach which is used to detain the consumer access pattern and to realize recurrent user access patterns⁴. A semantic web usage mining is a method whereas it is used to relate each demanded webpage with one or more ontologies for better understanding of web navigation.

A semantic web usage mining⁵ is one of the techniques which is used to automatic creation of periodic web access pattern. Earlier, web usage mining methods⁶ are focused on mining frequent access patterns which have occurred recurrently within the entire duration of all the user access sessions. So, this method analyzes the frequently used resources at a particular time period. Furthermore, ontology is generated to collect web access behaviors and emotional influence of the users for the specific resources. To improve the classification accuracy in the proposed work, Optimum Session Interval based on Particle Swarm Optimization (OSIPSO) is introduced to identify the

optimum session interval. Particle swarm optimization is used in the search space of a given problem to determine the settings or parameters necessary to maximize a particular objective.

2. Previous Research

In the previous studies, some of the web usage mining approaches are recommended for extracting statistical information and also user access patterns. Many of the ontology generation techniques have been also investigated. These techniques predominantly focus on generating a concept hierarchy for creating ontologies.

Pei J et al.⁶ suggested mining access pattern from the weblogs. The web access pattern is nothing but a sequence of accesses followed by the users in a frequent manner. In this work, to concentrate on the mining access patterns from web logs effectually⁷. A web access pattern tree is a new data structure suggested for mining access patterns from weblogs. This tree accumulates compressed critical information for access pattern mining and also assists the development of new algorithms for mining access patterns in a large set of log pieces.

Stumme et al. proposed semantic web mining for analyzing the records of web usage. The main intent of semantic web mining is to merge two research areas. One is called semantic web and web mining. The main objective of the web mining is to discover the meaning of web resources and their usage. The main motivation of the semantic web is to develop the current web by machine-processing information in order to facilitate for semantic-based tools supporting the human user.

The user query is a set of keywords which is written in a natural language. When the user sends a query to the web search engines use index files for recovering the documents. Indexes may be keywords, terms, syntactic or semantic structures. Amalia Todirascu et.al designed a prototype of a system for querying the web in a natural language⁹. The semantic resources are used to filter the search and a data-driven methodology is adopted for resource acquisition. The description logic is used to signify the domain hierarchy and to provide efficiency and fault tolerance for unfinished data. The description logic provides Logic inference mechanisms which are used to enlarge dynamically the domain model, and to complete missing information discovered from the user query.

Taxonomies or conceptual hierarchies are critical for any knowledge-based system. Cimiano et al. proposed a new method for repeated achievement of taxonomies from texts based on formal concept analysis¹⁰. This method is based on the postulation that verbs pose strong selection limitations on their arguments. The conceptual hierarchy is then built on the basis of the inclusion relations between the extensions of the selection limitations of all the verbs, whereas the verbs themselves present intentional descriptions for each concept. After that, to formalize the design in terms and this method is used to attain a concept hierarchy for the tourism domain out of texts.

Quan et al. suggested a framework which is called Fuzzy Ontology Generation. This framework consists of the succeeding steps: Analysis of fuzzy normal method, creation of fuzzy ontology and Semantic Representation adaptation. Furthermore, an approximating calculation technique is proposed which assists the generated fuzzy ontology to be incrementally provided with new instances¹¹. By using the uncertainty data, this framework is used to create ontology as it can symbolize the uncertainty in the emotion and make a concept hierarchy from the vaguest information in automatically. This framework has also been used to create Machine Service Ontology for Semantic Ontology in this research.

Dai and Mobasher¹² used domain ontology to improve web usage mining for conventional web usage logs, but the mapping from requesting URLs to ontological entities lacks reliability, particularly for dynamic websites. By mapping every demanded URL to one or more conceptions, a method of semantic improvement of web usage logs has accomplished¹³. So, the particular user interest is determined and uses the association rule mining to the semantically improved weblogs.

The ontology can be generated manually by utilizing the ontology editor. In ontology formation process, the collection of knowledge attainment is used with machine learning techniques. Many of the methods are suggested for creating ontologies. These methods include Natural Language Processing, association rule mining¹⁴ and hierarchical clustering. Furthermore, these techniques mainly focus on generating concept hierarchies from text documents or relational databases.

3. Generation of Personal Web usage Ontologies

A Personal Web Usage Ontologies is created by individual users by using the semantically enriched web usage logs.

This method follows the four steps. Create personal Web Usage Lattice, Creation of Global Web Usage Lattice, Generating Web usage ontology and Generation of Personal Web Usage Ontologies.

3.1 Web usage Lattice Creation

The preprocessing process consists of data cleaning, user recognition and session detection. The main idea of the preprocessing is to eradicate unsuccessful requests, redundant data and to identify all personal access sessions for every individual user. The web access session of the user is defined $S = \langle (URL_1, t_1), (URL_2, t_2), \dots (URL_n, t_n) \rangle$ is a succession of *URL*, with timestamp t_i. The user interest level of a particular URL is nothing but time spent for a specific URL. The estimation of duration of a specific URL is $d_i = (t_{i+1} - t_i)$. Every URL, in the user access session is related to setting of resource attributes $M_{ri} \subseteq M_r$ for denoting the semantics of the content in a specific URL. The user access session is treated as a sequence of sets of resource attributes M_{ij} instead of a sequence of individual URL, and it is denoted as $S = \langle (M_{r1}, t_1, d_i), (M_{r2}, t_2, d_2), \dots (M_{rn}, t_n, d_n) \rangle$. The estimation of the level of interest depends on the total duration for each resource attribute $m_k \in M_r$ during the user access session. The duration of the resource attribute is evaluated by to take summation of the sequence of requests URLs and the duration of the URL.

 $K = (G, M_p, M_r, I)$ is a fuzzy periodic Web Usage Context, where G denotes set of user access sessions, $M_{\rm g}$ is a set of periodic attributes, *M* is a set of resource attributes, I is a fuzzy set of the domain to denote the associations between user access sessions and attributes. Each fuzzy relation $R(g, m) \in I$ is represented by a membership value $\mu(g, m) \in [0, 1]$. Each user access session $g \in G$ can also be represented as a fuzzy set on the domain. For a periodic attribute $m_p \in M_p$, the membership value $\mu_p(g, m_p)$ in a user access session $g \in G$ can be computed using the period of g. Also, for a resource attribute, the membership value $\mu_r(g, m_r)$ in a user access session $g \in G$ can be computed using the total duration of M. If Z(g, m) is less than the Z(m) the user the membership of the resource attribute is zero. If $Z(g, m_r)$ is greater than $Z(m_r)$ membership of the resource attribute is one. Z(m) is nothing but the proportion of the total duration of accessing the resource in all web access sessions of the user, which denotes the user's global interest of the resource. Z(g, m)is the proportion of the duration of accessing the resource within the user access session g, weighted by an emotional influence factor e_r derived from the consequent ΔE using one of the following rules:

Rule 1: -E: This is the baseline situation where $e_r = 1$; ignoring emotional influence.

Rule 2: $+E_1$. If $\Delta E < 3$ then $e_r = 0.5$; otherwise $e_r = 1$; we only repress resources with negative emotional influence. Rule 3: $+E_2$, e_r is derived by using the formula, $e_r = 0.1$ ΔE + 0.7, thus the larger value is assigned to e_z with increasing ΔE , i.e., $e_r = 0.8, 0.96, 1, 1.1, 1.2$

Z(g, m) denotes the user local interest and emotional influence of the resource. For a given web usage context $K = (G, M_s, M_s, I)$ the set of common to user access sessions are defined as $A \subseteq G$ and the set of user access sessions are defined as $B \subseteq M_p$. The fuzzy support of a set of attributes is defined as how frequently the set of attributes common to user access sessions and set of user access sessions which have all the same attributes. The fuzzy confidence is defined as the number of times the user access the similar resource attributes in the set of user access sessions and the set of resource attribute frequent to user access sessions.

For creating the global web usage lattice, the set of selected periodic attributes M_p and resource attributes M_r for all users are taken into consideration. $W_G = \{B_k\}$ represents the set of entire web access activities. $|W_G|$ denotes the total number of global web access activities. There

should be a total of $\left(\frac{a}{i}\right) \times \left(\frac{b}{j}\right)$ global web access activities with i periodic attributes and j resource attributes, where $\frac{a}{i}$ and $\frac{b}{j}$ symbolize the number of combinations. Every web access activity has direct sub-activities. Each web access activity has direct sub activities. The Global Web Usage Lattice is created by using the sub activity associations.

3.2. Generation of Ontology

Ontology includes taxonomy with a set of inference regulations. The appearance of taxonomy is defined as the set of domain concepts and the associations among them. By utilizing a class and hierarchy mapping, and property mapping, the Global Web Usage Ontology can be created. There are some benefits of web ontology that does not modify the original ontology and preserves the incremental ontology creation. The instance mapping is used to integrate the personal web usage lattice of a user with the global web usage ontology.

Finally, the user satisfaction level is determined. The satisfaction for the overall web personalization is defined as,

$$Satisfaction = \frac{\sum\limits_{PR_i \in PR_a} satisfaction(PR_i)}{|PR_a|}$$

where, PR_a is a subset of personalized resources. PR_i is set of personalized resources. The evaluation of satisfaction is used to measure how well the user is interested in the personalized resources.

Algorithm 1: Generation of personalized Ontology

Input: Set of weblogs

Output: Satisfaction level of user

- 1. Take a set of Web Usage Logs
- 2. // Creation of Personal Web Usage Lattice
- 3. Pre-processing in weblogs
- 4. Sequence of web access session

$$S = <(URL_1, t_1), (URL_2, t_2), ..., (URL_n, t_n)>$$

- 5. Estimation of duration $d_i = (t_{i+1} t_i)$.
- 6. Every URL_i has set resource attributes $M_{ri} \subseteq M_r$

7.
$$d(S, m_k) = \sum_{i=1}^{n} \alpha_{ki} d_i$$
, where

8. // Estimation of level of interest

$$\alpha_{ki} = \begin{cases} 1, & \text{if } m_k \in M_{ri} \\ 0, & \text{otherwise} \end{cases}$$

9.
$$\mu(g,m) = \begin{cases} \mu_p(g,m), & \text{if } m \in M_p \\ \mu_r(g,m), & \text{if } m \in M_r \end{cases}$$

10. // Membership function // Where G = set of user access sessions, M_p = set of periodic attributes, M_r = A set of resource attributes, I =fuzzy set of the domain

- 11. $\mu_p(g, m_p) = \max_{t \in p(g)} {\{\mu_p(t, m_p)\}}, where$
- 12. // The membership function of periodic attributes

$$\mu_r(g, m_r) = \begin{cases} 0, & \text{if } z(g, m_r) < \frac{1}{2}Z(m_r) \\ \frac{2z(g, m_r)}{z(m_r)} - 1, & \text{if } \frac{1}{2}Z(m_r) \le Z(g, m_r) \le Z(m_r) \\ 1, & \text{if } z(g, m_r) > Z(m_r), \end{cases}$$

13. // The membership function of resource attributes

$$Z(m_r) = \frac{\sum\limits_{gk \in G} d(g_k, m_r)}{\sum\limits_{gk \in G} (te(g_k) - ts(g_k))}$$

- 14. // Total duration of accessing the resource
- 15. Generate Periodic association pattern

$$Sup(B) = \frac{\sum_{g \in B^*} (\mu_p(g) \times \mu_r(g))}{|G|}$$

16. // Fuzzy support value

$$Conf(v(B)) = prob((B \cap M_r) | (B \cap M_p))$$

$$= \frac{S \operatorname{up}(B)}{S \operatorname{up}(B \cap M_p)}$$

- 17. // Fuzzy confidence value
 - // Creation of global web usage lattice
- 18. Personal and global web usage Ontology generation

$$\sum_{R_i \in PR}$$
 satisfaction(PR_i)

19. $Satisfaction = \frac{\sum\limits_{PR_i \in PR_a} satisfaction(PR_i)}{|PR_a|}$

// Computation of satisfaction level

20. where PR_a is a subset of personalized resources. PR is set of personalized resources

4. Optimum Session Interval based on Particle Swarm **Optimization (OSIPSO)**

In the previous method the accuracy is less in terms of user satisfaction level. So, in this article Optimum Session Interval based on Particle Swarm Optimization (OSIPSO) is introduced. The session interval chosen in this method is very significant for improving the accuracy. The optimum session interval is identified by using particle swarm optimization algorithm.

The particle swarm optimization is a computational method which optimizes a problem by incessantly trying to improve an applicant solution with observes to a given determination of quality. In every iteration process, each candidate solution is calculated by the objective function being optimized, deciding the fitness of that solution. In the PSO algorithm, every particle preserves the position, gathered the candidate solution, fitness value and velocity. In this algorithm, during the execution of the process the best fitness value is taken as individual best fitness value. Among the all particles in the swarm, the PSO algorithm evaluates the best fitness value which is called global fitness value. The candidate solution attained this fitness is called global best candidate solution.

There are three major steps in PSO algorithm:

- 1. Evaluation of fitness value for every particle
- 2. Update the individual and global best fitness and positions
- 3. Update the velocity and position for every particle

Algorithm 2: Optimum Session Interval based Particle Swarm Optimization (OSIPSO) algorithm

- 1. Initialize N number of particles in the swarm, a position of a particle is denoted as X_i and velocity is denoted as V_i . Let *pbest* represents the best wellknown position of particle i and gbest signifies the best position of the entire swarm
- 2. Initialize the particle's position X_i
- 3. For each particle i = 1, 2 ... N
- 4. Calculate fitness value for every particle
- 5. //For fitness calculation satisfaction is taken from the previous algorithm
- 6. Fitness

satisfaction

Number of user sessions × Number of particles

// Computation of fitness

- 7. If fitness value is superior than the best fitness value (pBest)
- 8. Set present value as the new pBest
- 9. Until a termination criterion is met
- 10. Select the particle with best fitness value of all particles as the gbest
- 11. For every particle
- 12. // Calculation of particle velocity
- 13. $V_i(t+1) = wv_i(t) + c_1r_1[\hat{x}_i(t) x_i(t)]$ $+ c_2 r_2 [g(t) - x_i(t)]$

// where, the index of the particle is represented by i, $v_i(t)$ is the velocity of particle i at time t, $x_i(t)$ is the position of particle i at time t, parameters w, c1, and c2 are coefficients

- 14. Update particle position
- 15. $x_i(t+1) = x_i(t) + v_i(t+1)$
- 16. Until some stopping condition is met

5. Associative Classification **Technique**

The associative classification method is introduced to assimilate the classification rule mining and association rule mining. The main intent of classification rule mining is to discover a small set of regulations in the database to form a precise classifier. The main intention of association rule mining to discover all regulations in the database that satisfy some minimum support and minimum confidence constraints. The target of mining is not decided in association rule mining. In contrast to that, the target is pre-determined in the classification rule mining. In this work, the amalgamation of both associative and classification rule mining, which is pertinent for realistic applications? By assimilating these two techniques a new method is proposed which is called associative classification to improve the accuracy.

In classification there are two phases: (1) training data (2) testing data

The training phase contains a periodic association access pattern of the users with support and confidence values. In the training data, already we have the set of emotional behavior of particular resources for the users in every session. The fuzzy support and confidence values are indicated the quality of such periodic association access patterns. From this we obtain the number of users interested in particular resources in a particular session. In the testing phase, a test data is given for identifying the accuracy level. In this phase, the obtained periodic association access pattern is matched with training data and evaluate the accuracy.

Algorithm: Associative classification algorithm

Input: Input data set

Output: Classification accuracy

1. Generate periodic web access pattern by using Algorithm 1

2.
$$Sup(B) = \frac{\sum\limits_{g \in B^*} (\mu_p(g) \times \mu_r(g))}{|G|}$$

// Fuzzy support value

3. $Conf(v(B)) = prob((B \cap M_r) | (B \cap M_p))$

$$=\frac{Sup(B)}{Sup(B\cap M_p)}$$

// Fuzzy confidence value

- 4. Generate the association rule in the training phase
- 5. Compute the periodic web access pattern in the testing phase
- 6. Apply the association rule which is computed in the training phase
- 7. Match the rule and predict the classification accuracy

6. Performance Validation

The experimental results are evaluated for the existing Web Usage Ontology Generation (WUOG) approach and the proposed Optimum Session Interval based Particle Swarm Optimization (OSIPSO) including associative classification. The raw web server log data were acquired from a web forum at Nanyang Technological University, Singapore. The web forum includes seven main topics and 57 subtopics. Access data of the top 50 users were used in the experiments.

6.1 Satisfaction Level

Satisfaction level evaluates how probable it is that a user is involved in one of the personalized resources in the period-supported sessions (Figure 1).

The satisfaction level is measured for the existing Web usage Ontology Generation (WOG) approach and the proposed Optimum Session Interval based Particle Swarm Optimization (OSIPSO) including associative classification. Compared to the existing WOG method, the proposed method achieves a high satisfaction level.

6.2 Accuracy

The accuracy is measured for the existing Web usage Ontology Generation (WOG) approach and the proposed Optimum Session Interval based Particle Swarm Optimization (OSIPSO) including associative classification. Furthermore an associative classification is used to evaluate the classifying accuracy. Compared to the existing WOG method the proposed method achieves high accuracy (Figure 2).

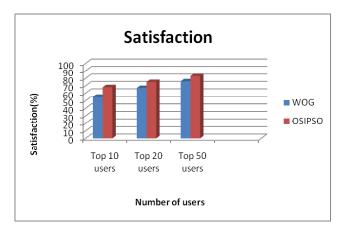


Figure 1. Satisfaction level.

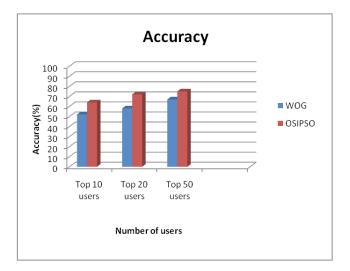


Figure 2. Accuracy.

7. Conclusion

For automatic creation of the Personal Web Usage Ontology from web usage logs, a semantic web usage mining technique is suggested. The Personal Web Usage Ontology is used to analyze the user web access behavior and emotional influence about the particular resources. This Personal Web Usage Ontology is utilized by software agents to give Semantic Web services. Furthermore, the user queries are processed by search engines and re-rank the search results based on the user behavior which is obtained from Personal Web Usage Ontology. This method finds the consumer emotions in every session interval. If the session interval changes, the accuracy level also changes. So, to find optimum session interval, Optimum Session Interval based on Particle Swarm Optimization (OSIPSO) is proposed. Using Particle Swarm Optimization algorithm the best session interval is to be chosen. Furthermore, the user queries are processed by search engines and re-rank the search results based on the user behavior which is obtained from Personal Web Usage Ontology.

For future work, there is a need to reduce the gap between the methodical demonstration of preferences and customers' actual preferences. Mapping between customer necessities and customer preference ontologies is a significant concept. A new method of preference selection from customers is anticipated for mapping the customer necessities and customer preference ontologies.

8. References

- 1. Dolan RJ. Emotion, cognition, and behavior. Science. 2002; 298(5596):1191-94.
- 2. Weinberg P, Gottwald W. Impulsive consumer buying as a result of emotion. J Business Research. 1982; 10(1):43-57.
- 3. Piron F. A comparison of emotional reactions experienced by planned, unplanned and impulse purchasers. Advances in Consumer Research. 1993; 20(341-44):199-204.
- 4. Srivastava J, Cooley R, Deshpande M, Tan P-N. Web usage mining: discovery and applications of usage patterns from web data. ACM SIGKDD Explorations. 2000; 1(2):12-23.
- 5. Fong ACM, Zhou B, Hui SC, Tang J, Hong GY. Generation of personalized ontology based on consumer emotion and behavior analysis. IEEE Transcations on Active Computing. 2012; 3(2):152-64.
- 6. Pei J, Han J, Mortazavi-Asl B, Zhu H. Mining access patterns efficiently from web logs. Proc Fourth Pacific-Asia Conf. Knowledge Discovery and Data Mining. Current Issues and New Applications. 2000; 396-407.
- 7. Mobasher B, Dai H, Luo T, Nakagawa M. Effective personalization based on association rule discovery from web usage data. Proc Third Int'l Workshop Web Information and Data Management. 2001; 9-15.

- 8. Stumme G, Berendt B, Hotho A. Usage mining for and on the semantic web. Proc NSF workshop next generation data mining. 2002 Nov; 77-86.
- 9. Todirascu A, De Beuvron F, Ga^{lea} D, Rousselot F. Using description logics for ontology extraction. Proc First Workshop Ontology Learning at the 14th European Conf. Artificial Intelligence; 2000 Aug.
- 10. Cimiano P, Staab S, Tane J. Deriving concept hierarchies from text by smooth formal concept analysis. Proc GI Workshop Lehren-Lernen-Wissen-Adaptivit. 2003; 72-9.
- 11. Quan TT, Hui SC, Fong ACM, Cao TH. Automatic fuzzy ontology generation for semantic web. IEEE Trans Knowledge and Data Eng. 2006 Jun; 18(6):842-56.
- 12. Dai H, Mobasher B. Using ontologies to discover domainlevel web usage profiles. Proc Second Semantic Web Mining Workshop at ECML/PKDD; 2002.
- 13. Oberle D, Berendt B, Hotho A, Gonzalez J. Conceptual user tracking. Proc First Int'l Atlantic Web Intelligence Conf. 2003 May; 155-64.
- 14. Maedche A, Staab S. Discovering conceptual relations from text. Proc 14th European Conf Artificial Intelligence. 2000 Aug; 321-25.