May 2013 IJMIE Volume 3, Issue 5 ISSN: 2249-0558

PARALLEL ALGORITHM BASED CONSUMER BEHAVIOR ANALYSIS FOR GENERATING PERSONALIZED ONTOLOGY SYSTEM

S.Lakshmi Priya*

Mr. S.Varatharajan**

Abstract

Consumer face a dramatic problem related to web search. They expect the most relevant and efficient results. While the result often disappoint the consumer and also their precious time. This paper propose a semantic web usage mining approach for discovering periodic internet access patterns from elucidated net usage logs which incorporates data on client emotions and behaviors through self-reporting and behavior tracking. We use fuzzy logic to represent real-life temporal ideas (e.g., morning) and requested resource attributes (ontological domain ideas for the requested URLs) of periodic pattern-primarily based web access activities. These fuzzy temporal and resource representations, that contain both behavioral and emotional cues, are incorporated into a private Web Usage Lattice that models the user's web access activities. From this, we tend to generate a private Net Usage Ontology, which enables semantic web applications such as customized web resources recommendation. Finally, we have a tendency to demonstrate the effectiveness of our approach by presenting experimental leads to the context of personalized net resources recommendation with varying degrees of emotional influence.

* P.G Stubent, Department of Information Technology, PSN Collage of Engineering and Technology, Tirunelveli

** Associate Professor, Department of Information Technology, PSN Collage of Engineering And Technology, Tirunelveli

Keywords— Emotion and behavior profiling, Behavioral tracking, semantic web, Adaptation in mid to long-term interaction, consumer habits, personalization, recommender system, weblog mining, knowledge discovery, ontology generation.

I. INTRODUCTION

Emotion is central to the quality and range of everyday human experience [1]. Emotion provides the principal currency in human relationships as well as the motivational force for what is best and worst in human behavior. Emotion exerts a powerful influence on reason and, in ways neither understood nor systematically researched, contributes to the fixation of belief. In a narrower context, the relationships between consumer emotions and their buying behaviors also seem well documented. In addition, with more and more people connected to the Internet, today's technology savvy consumers are likely to use the web to find information pertinent to products and services before they commit to a purchase. Emotions have been found to influence a person's web surfing behaviors [2]. Discovering and modeling consumers' emotions and surfing habits and behaviors are important for many web applications
such as personalized web search and recommendation for business applications [4].

Here, we use self-report to incorporate emotions into a personalized consumer profile and web access patterns to model the consumer's mid to long-term web surfing behaviors. Specifically, users are asked to record any change in their emotional state at the end of each web access request. The information is used to gauge the emotional influence of the accessed resources on the user. To capture consumers' access patterns, many web usage mining techniques, have been developed for mining statistical information and user access patterns in terms of association and sequence of requested resources [5].

Recent research has focused on mining web usage data for the Semantic Web is known as semantic web usage mining, The idea is to associate each requested webpage with one or more ontological entities to better understand the pattern of web navigation. The discovered knowledge can potentially be used for semantic web applications, such as personalized web content recommendation. Web mining techniques can be applied to help create the Semantic Web. A backbone of the Semantic Web is ontologies, which at present are often hand-crafted. The content of the Semantic Web as being represented by ontologies and meta-data [6].

In this paper, we propose a semantic web usage mining approach for automatic generation of periodic pattern based web usage ontology for the Semantic Web. Our proposed approach mines periodic access patterns, which occur frequently in a particular period, e.g., every morning, directly from web usage logs that have been semantically enriched with information on topics and emotional influence. Such periodic access patterns are very useful for mid to long-term behavioral tracking.

The ontology will accumulate personal information on web access behaviors and habits, as well as the emotional influence of the accessed resources. So, those who have concerns about privacy may choose not to use our system. Many ontology generation techniques have been investigated [7]. These techniques focus mainly on generating concept hierarchy for building ontology from free text documents or relational databases. Our proposed approach aims at extracting semantics from semantically enriched web usage logs automatically and generates personalized web usage ontology for the Semantic Web.

The rest of this paper contained as follows: In Section 2, we review the related work. Section 3 presents our approach for automatic generation of personalized web usage ontology based on periodic pattern web access using semantically enriched weblogs.Finally, Section 4 concludes the paper.

II. RELATED WORK

A. *Semantic Web Uasage Mining*

The Web Usage Mining is the application of data mining methods to the analysis of recordings of Web usage, most often in the form of Web server logs. Since traditional web usage logs only record requested URLs but not the semantics of contents requested by the users, it is difficult to use such logs for tracking the users' actual web access behaviors, emotions, and interests.

In response, a number of semantic web usage mining techniques [8] have been proposed. For semantic enrichment of web usage logs by mapping each requested URL to one or more concepts from the ontology of the underlying website. It clusters groups of sessions with specific user interests from the semantically enhanced weblogs, and applies association rule mining to the semantically enhanced weblogs [9].

Concept-logs (C-logs) by enriching each web server log record with keywords from a taxonomy representing the semantics of the requested URLs. C-logs were analyzed in [10] with MINE RULE (a query language for association rule mining) for discovering access patterns.

Most semantic web usage mining techniques focus only on discovering simple usage statistics and common access patterns of user groups. Further, the discovered knowledge should be represented as ontology to enable Semantic Web applications.

B. *Ontology Generation*

The term ontology can be defined in many different ways. Ontology as an explicit specification of a set of objects, con-cepts, and other entities that are presumed to exist in some area of interest and the relationships that hold them. As implied by the general definition, an ontology is domain de-pendent and it is designed to be shared and reusable. Usually, ontologies are defined to consist of abstract concepts and rela-tionships (or properties) only. In some rare cases, ontologies are defined to also include instances of concepts and relation-ships. For this purpose, it is defined as an ontology to be a set of concepts *C* and relationships *R*. The relationships in *R* can be either taxonomic or non-taxonomic.

Ontology languages are semantic markup languages for defining ontologies. We use Web Ontology Language (OWL) [11], which was proposed as W3C Recommendation, for ontology specification. OWL facilitates greater machine interpretability of web content than XML [12], RDF, and RDF Schema [13] by providing additional vocabularies along with a formal semantics.

Many approaches have been investigated for generating ontology [14]. These include Natural Language Processing (NLP) techniques , association rule mining , hierarchical clustering , translation from relational databases , and Formal Concept Analysis (FCA) [15]. However, these techniques focus mainly on constructing concept hierarchies from text documents or relational databases.

C. Fuzzy Association Rule Mining

Researchers used different association rule algorithms to find the patterns and relationships between access sequences and user sessions. Then introduced the concept of n-gram sequence models on top of the traditional association rule algorithms. One of the problems of association rule techniques is that each item is distinct from the others, therefore any two items are either considered to be the same or are totally different. Another problem of association rule techniques is that, when applied in a large data set, the size of the resultant rule set will be extremely large. Therefore, in a real-time application system, to match a rule with input query becomes a critical issue.

The fuzzy association rule technique is used to improve the prediction accuracy and fuzzy index tree for fast matching of rules [16]. The general fuzzy rule is mined from this case base by the fuzzy association rule technique. The adaptive fuzzy rule will be mined from the web log and using the fuzzy association rule technique as well. The general fuzzy rule will do the prediction based on the matching of case similarity, while the adaptive rules are used to fine-tune and adopt some unusual patterns. An adaptation engine is defined to control the shifting between similarity and adaptation. The adaptation engine is generated by learning the criteria and conditions of adaptation based on the prediction accuracy and recall rate of the rules.

D. Periodic Pattern Mining

Discovering periodic patterns from time series databases is an important data mining task for many applications, such as behavioral tracking. According to the type of patterns, periodic patterns can be divided into periodic association rules and periodic sequential patterns. Periodic association rules are rules that associate with a set of events that occur periodically; such association rules hold only during certain time intervals but not others. Periodic sequential pattern mining can be viewed as an extension of sequential pattern mining [17] by taking into account the periodic characteristics in the time series data.

III. Web Usage Ontology Generation

Our goal is to generate the automatic personalized web usage ontology for individual users from semantically enriched web usage logs. Fig. 1 shows the proposed approach which consist of four major steps: Personal Web Usage Lattice (PWUL) Construction, Global Web Usage Lattice (GWUL) Construction, Global Web Usage Ontology (GWUO) Generation, and, finally, PWUO Generation.

The semantic web usage logs is given as inputs, it first identifies a set of periodic attributes and a set of resource attributes enhanced with user-reported emotional influence to represent periodic web access activities. It constructs a PWUL from the access sessions of the user. Then constructs a GWUL to represent all periodic based global web access activities and hierarchical relationships between these activities. GWUL contains a large number of global web access activities PWUL is just a small sublattice of GWUL.

Generates GWUO from GWUL by mapping global web access activities and their hierarchical relationships into activity classes and their properties. Then PWUO for a user by mapping the personal web access activities in PWUL into activity instances of the corresponding activity classes in GWUO. Written in OWL, PWUO is the knowledge base that can subsequently provide personalization facility to the users.

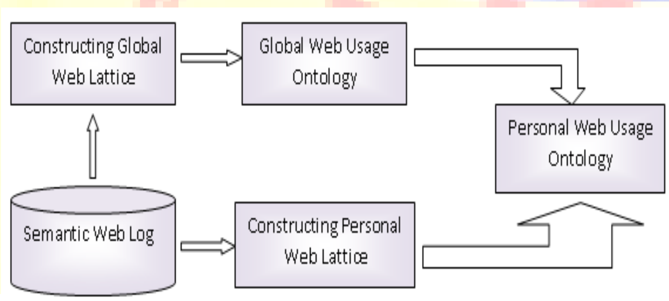

Fig 1 Architecture of Personalized Web Usage Ontology

A. *Semantic Web Logs*

Semantically enriched web log contain requested URL with one or more ontological entities such as concepts, attributes, and relations, to better describe the patterns of web navigation. We adapt a frame work in conceptual user tracking [18] which provide semi-automatically generated web log. This web log contain pre-defined data and an emotional influence score ΔE reported by user. Table 1 contain semantically enriched weblog.

UserID	TimeStamp	URL	ΔE	Topic
User1	12/Jan/2011 08:20.01	URL1	3	#Topic1, #Topic3,......
User1	12/Jan/2011 08:21:10	URL2	2	#Topic7, #Topic5,.......
User2	12/Jan/2011 08:21:32	URL7	4	#Topic1, #Topic8,.......
User1	12/Jan/2011 08:22:50	URL3	2	#Topic3, #Topic1,.......
User3	12/Jan/2011 08:33:10	URL3	1	#Topic6, #Topic2,.......
User1	12/Jan/2011 09:10:02	URL5	4	#Topic7, #Topic3,.......
User3	12/Jan/2011 09:17:32	URL6	3	#Topic2, #Topic1,.......
User2	12/Jan/2011 09:26:17	URL4	5	#Topic3, #Topic7,.......

Table 1:Semantically Enriched Web Log

Each entry in the table can have N user accessed resource at a specific time and emotional influences. The web access activity can be represented by a set of periodic attributes M_p and resource attributes M_r We also define eight real-life temporal concepts: Early Morning, Late Morning, Noon, Early Afternoon, Late Afternoon, Evening, Night, and Late Night, as periodic attributes. More generally, we could also use days of the week (e.g., Monday, Tuesday, etc.) or other real-life temporal concepts (e.g., weekdays, weekend, etc.) as periodic attributes.

B. *Personalized Web Usage Lattice*

The PWUL constructed by preprocessing performed on the weblog data to extract the most relevant information for further processing. We perform preprocessing tasks on the semantically enriched weblogs similarly to those for traditional web server logs as discussed in [19], i.e., data cleaning, user identification, and session identification. The purposes are to discard unsuccessful requests, unnecessary data and to identify all personal access sessions for each individual user. A user's web access session is a sequence of URL requested in a time stamp. Parallel algorithm [20] can be used to find the time stamp value.

Construct the Web Usage Context for a user from this preprocessed user access sessions. Fuzzy periodic Web Usage Context based on the preprocessed access sessions. From the definition of fuzzy periodic web usage context, we can further establish the following: the set of attributes common to user access sessions, the set of user access sessions having the same attributes, fuzzy support of a set of attributes, and the notion of web access activity.

Given a Web Usage Context K =(G, Mp, Mr, I),

the fuzzy support of a set of attributes B ≤ Mp U Mr and

B ≠φ is defined as

$$Sup(B) = \frac{\sum_{g \in b'} (\mu(g) * \mu_r(g))}{|G|}$$

Where,

$$\mu_p(g) = \begin{cases} min_{m_p \in (B \cap M_p)} \{m_p(g, m_p)\}, & if\ B \cap M_p = \phi \\ 1, & otherwise, \end{cases}$$

$$\mu_r(g) = \begin{cases} min_{m_r \in (B \cap M_r)} \{m_r(g, m_r)\}, & if\ B \cap M_r = \phi \\ 1, & otherwis \end{cases}$$

Then identify the activity relationships in order to construct PWUL. The fuzzy support of a web access activity v(B) is defined as Sup(v(B))= Sup(B) and the fuzzy confidence

Of v(B) is defined as

$$Conf\left(v(B)\right) = prob\left((B \cap M_r \mid (B \cap M_p)\right) = \frac{Sup\ (B)}{Sup\ (B \cap M_p)}$$

Where prob (·) is a condition probability. For virtual web access activity v(ø), We define Sup(v(ø)) = 1.0 and conf(v(ø)) = 1.0.

Some efficient approaches [21], [22], [23] have been proposed in traditional FCA for computing concept lattice. Among them, the TITANIC algorithm [24] is one of the most efficient lattice construction algorithms, especially for weakly correlated data in very large data sets. We have modified the TITANIC algorithm with fuzzy support and fuzzy confidence for constructing PWUL from the Web Usage Context. From PWUL, inference rules can be extracted to deduce the user's periodic association access patterns. In particular, fuzzy logic can be applied to infer association access patterns of a user.

Given a minimum support MinSup ε[0,1] and a minimum confidence MinConf ε[0,1], we call a periodic association access pattern interesting if its fuzzy support value is not less than MinSup and its fuzzy confidence value is not less than MinConf. In the Table 2, if we set MinSup = 0.1 and MinConf =0.15, then the last two entries will be discarded, leaving only six interesting periodic association access patterns.

Table 2

Periodic Access Patterns

Activity	Periodic Association Access Pattern	Sup	Conf
1	$\{Evening(0.8)\} \Rightarrow \{Chat(0.9)\}$	0.21	0.62
2	$\{Night(1.0)\} \Rightarrow \{Chat(0.9)\}$	0.30	0.45
3	$\{Night(1.0)\} \Rightarrow \{Games(0.9)\}$	0.39	0.58
4	$\{Evening(0.8)\} \Rightarrow \{Sports(0.8), Chat(0.6)\}$	0.14	0.41
5	$\{Evening(0.8), Night(0.8)\} \Rightarrow \{Chat(0.9)\}$	0.13	0.66
6	$\{Night(1.0)\} \Rightarrow \{Games(0.7), Chat(0.5)\}$	0.16	0.24
7	$\{Late\_Afternoon(0.6), Evening(0.5)\} \Rightarrow \{Sports(0.8), Chat(0.6)\}$	0.06	0.60
8	$\{Evening(0.8), Night(0.6)\} \Rightarrow \{Sports(0.8), Games(0.6), Chat(0.5)\}$	0.06	0.30

C. Global Web Usage Lattice

Construct a GWUL of all users. We represent a global web access activity using the set of selected periodic attributes Mp and resource attributes Mr.

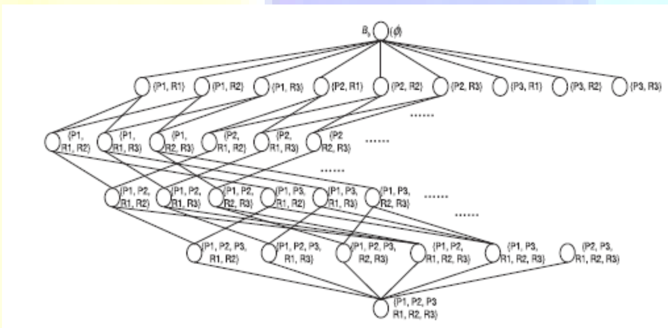

Fig 2 Global Web Usage Lattice

Fig. 2 shows a GWUL that contains three periodic attributes “P1 (Late Afternoon),” “P2 (Evening),” and “P3 (Night),” and three resource attributes “R1 (Sports concept),” “R2 (Games concept),” and “R3 (Chat concept).”

D. Global Web Usage Ontology Construction

Ontology consists of a taxonomy with a set of inference rules. The taxonomy can be expressed as a set of domain concepts (i.e., classes of objects) and the relationships among them (i.e., class hierarchy). Based on the formal definition of ontology [28], we define GWUO. It can be generated through class and hierarchy mapping and property mapping. The property mapping coompries attribute property mapping, taxonomy mapping and quality mapping.

OWL allows extension of imported definitions without the need to modify the original ontology and supports incremental ontology construction, making it easy to incrementally update the activity class hierarchy. When new periodic attributes or resource attributes are introduced, we just need to create new activity classes with new attributes and insert new subclass relationships between the new and existing activity classes to extend the activity class hierarchy.

E. *Personal Web Usage Ontology Construction*

Generate PWUO by combining the PWUL of a user with the GWUO using instance mapping. Instance mapping generates a set of activity instances IP from the corresponding activity classes in the GWUO and an activity instance hierarchy $<I_P$. As expected, the number of activity instances in PWUO of a user is much smaller than the number of activity classes in GWUO.

IV. Conclusion

Presented an approach for automatic generation of Personal Web Usage Ontology of periodic access patterns from web usage logs that have been semantically enriched with information on emotional influence and resource topics. Over time, the knowledge base can capture both consumer web access behavior and emotional influence of the web resources on the user. By varying the degrees of emotional influence, we found that emotional influence contributed positively to the results. In future we consider, To reduce the gap between the systematic representation of preferences and customers' actual preferences. Mapping between customer needs and customer preference ontologies.

References

[1] R.J. Dolan, "Emotion, Cognition, and Behavior," Science, vol. 298 no. 5596, pp. 1191-1194, 2002.

[2] J. Kalbach, Designing Web Navigation. O'Reilly, 2007.

[3] D.H. Choi and B.S. Ahn, "Eliciting Customer Preferences for Products from Navigation Behavior on the Web: A Multicriteria Decision Approach with Implicit Feedback," IEEE Trans. Systems, Man, and Cybernetics, Part A, vol. 39, no. 4, pp. 880-889, July 2009.

[4] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," ACM SIGKDD Explorations, vol. 1, no. 2, pp. 12-23, 2000.

[5] R. Missaoui, P. Valtchev, C. Djeraba, and M. Adda, "Toward Recommendation Based on Ontology-Powered Web-Usage Mining," IEEE Internet Computing, vol. 11, no. 4, pp. 45-52, July/Aug. 2007.

[6] A. Maedche and S. Staab, "Ontology Learning for the Semantic Web," IEEE Intelligent Systems, vol. 16, no. 2, pp. 72-79, Mar./ Apr. 2001.

[7] P. Cimiano, S. Staab, and J. Tane, "Deriving Concept Hierarchies from Text by Smooth Formal Concept Analysis," Proc. GI Workshop Lehren-Lernen-Wissen-Adaptivit, pp. 72-79, 2003.

[8] G. Stumme, B. Berendt, and A. Hotho, "Usage Mining for and on the Semantic Web," Proc. NSF Workshop Next Generation DataMining, pp. 77-86, Nov. 2002.

[9] R. Meo, P.L. Lanzi, M. Matera, and R. Esposito, "Integrating Web Conceptual Modeling and Web Usage Mining," Proc. WebKDDWorkshop Web Mining and Web Usage Analysis, pp. 105-115, 2004.

[10] P. Fraternali, M. Matera, and A. Maurino, "Conceptual-Level Log Analysis for the Evaluation of Web Application Quality," Proc. First Conf. Latin Am. Web Congress, pp. 46-57, Nov. 2003.

[11] D.L. McGuinness and F. Van Harmelen, "OWL Web Ontology Language Overview," W3C Recommendation, http://www. w3.org/TR/owl-features/, Feb. 2004.

[12] F. Yergeau, T. Bray, J. Paoli, C.M. Sperberg-McQueen, and E. Maler, "Extensible Markup Language (XML) 1.0 (Third Ed.)," W3CRecommendation,http://www.w3.org/TR/2004/REC-xml-20040204/, Feb. 2004.

[13] G. Klyne and J.J. Carroll, "Resource Description Framework (RDF): Concepts and Abstract Syntax," W3C Recommendation, http://www.w3.org/TR/rdf-concepts/, Feb. 2004.

[14] A. Maedche and S. Staab, "Ontology Learning for the Semantic Web," IEEE Intelligent Systems, vol. 16, no. 2, pp. 72-79, Mar./ Apr. 2001.

[15] C. De Maio, G. Fenza, V. Loia, and S. Senatore, "A Multi Facet Representation of a Fuzzy Ontology Population," Proc. IEEE/WIC/ ACM Int'l Joint Conf. Web Intelligence and Intelligent AgentTechnology, pp. 401-404, 2009.

[16] C. Wong, S. Shiu, and S. Pal, "Mining Fuzzy Association Rules for Web Access Case Adaptation," Proc. Fourth Int'l Conf. Case-Based Reasoning, Case-Based Reasoning Research and Development, pp. 213-220, 2001.

[17] R. Agrawal and R. Srikant, "Mining Sequential Patterns," Proc. 11th Int'l Conf. Data Eng., pp. 3-14, Mar. 1995.

[18] D. Oberle, B. Berendt, A. Hotho, and J. Gonzalez, "Conceptual User Tracking," Proc. First Int'l Atlantic Web Intelligence Conf., pp. 155-164, May 2003.

[19] R. Cooley, B. Mobasher, and J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns," Knowledge andInformation Systems, vol. 1, no. 1, pp. 5-32, 1999.

[20] P. Njiwoua and E.M. Nguifo, "A Parallel Algorithm to build Concept Lattice," Proc Groningen Intrl. Information Tech. conf. for students , pp103-107, 1993.

[21] G. Stumme, R. Taouil, Y. Bastide, N. Pasquier, and L. Lakhal, "Computing Iceberg Concept Lattices with TITANIC," DataKnowledge Eng., vol. 42, no. 2, pp. 189-222, 2002

[22] L. Nourine and O. Raynaud, "A Fast Algorithm for Building Lattices," Information Processing Letters, vol. 71, nos. 5/6, pp. 199-204, 1999.

[23] R. Godin, R. Missaoui, and H. Alaoui, "Incremental Concept Formation Algorithms Based on Galois (Concept) Lattices," Computational Intelligence, vol. 11, pp. 246-267, 1995.

[24] E. Bozsak, M. Ehrig, S. Handschuh, A. Hotho, A. Maedche, B.Motik, D. Oberle, C. Schmitz, S. Staab, L. Stojanovic, N. Stojanovic, R. Studer, G. Stumme, Y. Sure, J. Tane, R. Volz, and V. Zacharias, "KAON—Towards A Large Scale Semantic Web," Proc. Third Int'lConf. E-Commerce and Web Technologies, pp. 304-313, Sept. 2002.

Autour's Profile :

S.Lakshmi Priya doing her M.Tech in Information Technology at PSN College of Engineering And Technology, Tirunelveli. She received her B.Tech in Information Technology from VPMM Engineering College, Krishnankovil in 2007. Her research interests include datamining, artificial Intelligent.