Chapter IV

A Lattice-Based Framework for Interactively and Incrementally Mining Web Traversal Patterns

Yue-Shi Lee, Ming Chuan University, Taiwan, R.O.C.
Show-Jane Yen, Ming Chuan University, Taiwan, R.O.C.

Abstract

Web mining is one of the mining technologies, which applies data mining techniques in large amounts of Web data to improve the Web services. Web traversal pattern mining discovers most of the users' access patterns from Web logs. This information can provide the navigation suggestions for Web users such that appropriate actions can be adopted. However, the Web data will grow rapidly in the short time, and some of the Web data may be antiquated. The user behaviors may be changed when the new Web data is inserted into and the old Web data is deleted from Web logs. Besides, it is considerably difficult to select a perfect minimum support threshold during the mining process to find the interesting rules. Even the experienced experts also cannot determine the appropriate minimum support. Thus, we must constantly

adjust the minimum support until the satisfactory mining results can be found. The essences of incremental or interactive data mining are that we can use the previous mining results to reduce the unnecessary processes when the minimum support is changed or Web logs are updated. In this chapter, we propose efficient incremental and interactive data mining algorithms to discover Web traversal patterns and make the mining results to satisfy the users' requirements. The experimental results show that our algorithms are more efficient than the other approaches.

Introduction

With the trend of the information technology, huge amounts of data would be easily produced and collected from the *electronic commerce* environment every day. It causes the Web data in the database to grow up at amazing speed. Hence, how should we obtain the useful information and knowledge efficiently based on the huge amounts of Web data has already been the important issue at present.

Web mining (Chen, Park, & Yu, 1998; Chen, Huang, & Lin, 1999; Cooley, Mobasher, & Srivastava, 1997; EL-Sayed, Ruiz, & Rundensteiner, 2004; Lee, Yen, Tu, & Hsieh, 2003, 2004; Pei, Han, Mortazavi-Asl, & Zhu, 2000; Yen, 2003; Yen & Lee, 2006) refers to extracting useful information and knowledge from Web data, which applies data mining techniques (Chen, 2005; Ngan, 2005; Xiao, 2005) in large amount of Web data to improve the Web services. Mining *Web traversal patterns* (Lee et al., 2003, 2004; Yen, 2003) is to discover most of users' access patterns from Web logs. These patterns can not only be used to improve the Web site design (e.g., provide efficient access between highly correlated objects, and better authoring design for Web pages, etc.), but also be able to lead to better marketing decisions (e.g., putting advertisements in proper places, better customer classification, and behavior analysis, etc.)

In the following, we describe the definitions about Web traversal patterns: Let $I = \{x_1, x_2, ..., x_n\}$ be a set of all Web pages in a Web site. A *traversal sequence S* = $<w_1, w_2, ..., w_m>$ $(w_i \in I, 1 \leq i \leq m)$ is a list of Web pages, which is ordered by traversal time, and each Web page can repeatedly appear in a traversal sequence, that is, backward references are also included in a traversal sequence. For example, if there is a path which visits Web page <B>, and then go to Web page <G> and <A> sequentially, and come back to Web page <B>, and then visit Web page <C>. The sequence <BGABC> is a traversal sequence. The *length* of a traversal sequence *S* is the total number of Web pages in *S*. A traversal sequence with length *l* is called an *l-traversal sequence*. For example, if there is a traversal sequence α = <AGDFAB>, the length of α is 6 and we call α a 6-traversal sequence. Suppose that there are two traversal sequences α = $<a_1, a_2, ..., a_m>$ and β = $<b_1, b_2, ..., b_n>$ $(m \leq n)$, if there

Related Content

Distributed Privacy Preserving Clustering via Homomorphic Secret Sharing and Its Application to (Vertically) Partitioned Spatio-Temporal Data
Can Brochmann Yildizli, Thomas Pedersen, Yucel Saygin, Erkay Savas and Albert Levi (2011). *International Journal of Data Warehousing and Mining (pp. 46-66).*
www.igi-global.com/article/distributed-privacy-preserving-clustering-via/49640?camid=4v1a

A Survey on Fuzzy Association Rule Mining
Harihar Kalia, Satchidananda Dehuri and Ashish Ghosh (2013). *International Journal of Data Warehousing and Mining (pp. 1-27).*
www.igi-global.com/article/survey-fuzzy-association-rule-mining/75613?camid=4v1a

Towards Big Linked Data: A Large-Scale, Distributed Semantic Data Storage
Bo Hu, Nuno Carvalho and Takahide Matsutsuka (2013). *International Journal of Data Warehousing and Mining (pp. 19-43).*
www.igi-global.com/article/towards-big-linked-data/105118?camid=4v1a

Organic Knowledge Management for Web-Based Customer Service

www.igi-global.com/chapter/organic-knowledge-management-web-based/27910?camid=4v1a