Chapter XV

Online Analytical Mining for Web Access Patterns

Joseph Fong, City University of Hong Kong, Hong Kong

Hing Kwok Wong, City University of Hong Kong, Hong Kong

Anthony Fong, City University of Hong Kong, Hong Kong

ABSTRACT

The WWW and its associated distributed information services provide rich world-wide on-line information services, where objects are linked together to facilitate interactive access. Users seeking information from the Internet traverse from one object via links to another. It is important to analyze user access patterns, which helps improve web page design by providing an efficient access between highly correlated objects, and also assists in better marketing decisions by placing advertisements in frequently visited documents. We need to study the user surfing behavior through examining the web access log, browsing frequency of web pages and computing the average duration of visitors. This chapter offers an architecture to store the derived web user access paths in a data warehouse, and facilitates its view maintainability by use of metadata. The system will update the user access paths pattern with the data warehouse by the data operation functions in the metadata. Whenever a new user access path occurs, the view maintainability is triggered by a constraint class in the metadata. The data warehouse can be analyzed on the frequent pattern tree of user access paths on the web site within a period and duration. The result is an online analytical mining path traversal pattern. Performance studies have been done to demonstrate the effectiveness and efficiency of the system with the following contributions: an architecture of online analytical mining using frame model metadata, a methodology of implementing the online analytical mining, and the resultant cluster of web pages frequently visited by users for marketing use.

Figure 1: The Knowledge Discovery and Data Mining process
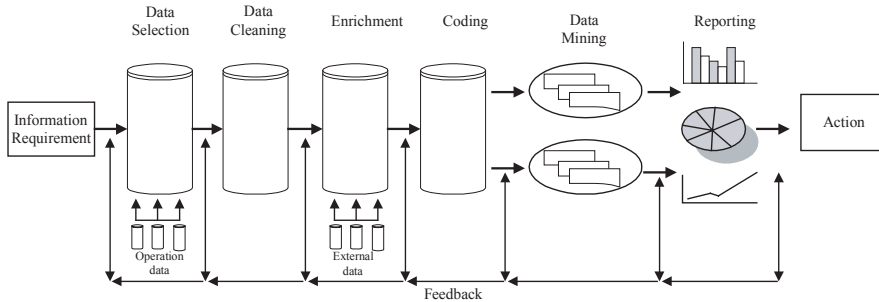
INTRODUCTION

Today, with the advent of the web and electronic commerce, nearly every organization has a web site where tremendous amounts of customer data have been generated and collected. These customer data contain a wealth of potentially accessible information. However, the explosive growth of data will inevitably lead to a situation such that it is increasingly difficult to access the desired information. As a result, there are great demands for analyzing data and transforming them into useful information and knowledge. Therefore, Knowledge Discovery and Data Mining (KDD) has become an important field in recent years to address the need for analyzing data in very large data repositories.

KDD is the process of automatic extraction of implicit, novel, useful, and understandable patterns in large databases. There are many steps in the KDD process, which include data selection, data cleaning, enrichment, coding, data-mining task, algorithm selection, and interpretation of discovered knowledge (Adriaans & Zantinge, 1996). This process tends to be interactive, incremental and iterative. Figure 1 illustrates the steps of the knowledge discovery process.

There is a relationship between the activities of data mining and data warehouse – the architecture foundation of decision support systems. The data warehouse sets the stage for effective data mining. The data mining can be done without data warehouse, but the data warehouse can improve the chances of success in data mining (Inmon, 1996).

Background

As the usage of the World Wide Web explodes, a massive amount of data is generated by web servers in the form of web access logs. It is a rich source of information for understanding web user surfing behavior. Web usage mining is one type of web mining activity that involves the automatic discovery of user access patterns on one or more web servers. Also, it applies data mining algorithms to web access logs to locate the regularities in web users' access patterns.

Analysis of these access data provides useful information for server performance enhancements, restructuring web sites, and direct marketing in electronic commerce. As a result, web usage mining has been used in improving web site design, business and marketing decision support, user profiling, and web server system performance, etc.

Among methods of discovering various knowledge in large databases, the association rule has attracted great attention in database research communities in recent years (Agrawal,

Related Content

ONTOMETRIC: A Method to Choose the Appropriate Ontology
Adolfo Lozano-Telloand Asunción Gomez-Perez (2004). *Journal of Database Management (pp. 1-18).*
www.irma-international.org/article/ontometric-method-choose-appropriate-ontology/3308

Designing Information Systems Capabilities to Create Business Value: A Theoretical Conceptualization of the Role of Flexibility and Integration
Christoph Schlueter Langdon (2006). *Journal of Database Management (pp. 1-18).*
www.irma-international.org/article/designing-information-systems-capabilities-create/3355

Information Mediation Using Metamodels: An Approach Using XML and Common Warehouse Metamodel
Luyin Zhaoand Keng Siau (2007). *Journal of Database Management (pp. 69-82).*
www.irma-international.org/article/information-mediation-using-metamodels/3375

A Comparison of the FOOM and OPM Methodologies for User Comprehension of Analysis Specifications
Judith Kabeliand Peretz Shoval (2005). *Information Modeling Methods and Methodologies: Advanced Topics in Database Research (pp. 175-194).*
www.irma-international.org/chapter/comparison-foom-opm-methodologies-user/23014

Context-Aware Query Processing in Ad-Hoc Environments of Peers
Nikolaos Folinas, Panos Vassiliadis, Evaggelia Pitoura, Evangelos Papapetrouand Apostolos Zarras (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications (pp. 1844-1866).*
www.irma-international.org/chapter/context-aware-query-processing-hoc/8008