Mining Web Transaction Patterns in an Electronic Commerce Environment

Ching-Huang Yun and Ming-Syan Chen
Department of Electrical Engineering
National Taiwan University
Taipei, Taiwan, ROC
E-mail:{ chyun@arbor.ee.ntu.edu.tw, mschen@cc.ee.ntu.edu.tw }

Abstract

In this paper, we explore a new data mining capability which involves mining Web transaction patterns for an electronic commerce (EC) environment. We propose an innovative mining model that takes both the traveling patterns and purchasing patterns of customers into consideration. First, we develop algorithm WR to extract meaningful Web transaction records from Web transactions so as to filter out the effect of irrelevant traversal sequences. Second, we devise algorithm WTM for determining the large transaction patterns from the Web transaction records obtained.

1 Introduction

Some existing electronic commerce environments [1][2], Web pages are usually designed as shop-windows. Customers can visit these Web pages and make Web transactions through the Web interface. It is known that mining information from such an EC system can provide very valuable information on consumer buying behavior and the quality of business strategies can then be improved [4]. Consequently, we shall explore in this paper a new data mining capability which involves mining *Web transaction patterns* for an EC environment.

First, for each Web transaction, we develop algorithm *Web-transaction-Record (WR)* algorithm, to extract meaningful Web transaction records from a given Web transaction. Each Web transaction record is represented by the form: <*path: a set of purchases*>, where a purchase, denoted by N(i), means that item i was purchased in node N along the path. After all the Web transaction records are derived from Web transactions, algorithm *Web Transaction Mining (WTM)* is developed for determining the large transaction patterns from the Web transaction records. Similarly to DHP [6], algorithm WTM utilizes the purchasing patterns for the candidate transaction pattern generation in the pattern discovering procedure. An illustrative example is given for the algorithm proposed.

This paper is organized as follows. Preliminaries are given in Section 2. Algorithms for determined Web transaction patterns is described in Section 3. This paper concludes with Section 4.

2 Preliminaries

Let N = $\{n_1, n_2, ..., n_g\}$ be a set of nodes in the EC environment and I = $\{i_1, i_2, ..., i_h\}$ be a set of items sold in the system. We then have the following definitions.

Definition 1 *Let* $<s_1 s_2 ... s_y : n_1\{i_1\}, n_2\{i_2\}, ..., n_x\{i_x\}>$ *be a transaction pattern, where* $i_m \subseteq I$ *for* $1 \leq m \leq x$, *and* $\{n_1, n_2, ..., n_x\} \subseteq \{s_1, s_2, ..., s_y\} \subseteq N$. *Then,* $<s_1 s_2 ... s_y : n_1\{i_1\}, n_2\{i_2\}, ..., n_x\{i_x\}>$ *is said to* **pattern-contain** *a transaction pattern* $<w_1 w_2 ... w_q : r_1\{t_1\}, r_2\{t_2\}, ..., r_p\{t_p\}>$ *if and only if* $\{s_1 s_2 ... s_y\}$ *contains* $\{w_1 w_2 ... w_q\}$ *and* $\{n_1\{i_1\}, n_2\{i_2\}, ..., n_x\{i_x\}\}$ *contains* $\{r_1\{t_1\}, r_2\{t_2\}, ..., r_p\{t_p\}\}$.

Definition 2 *A Web transaction is said to* **pattern-contain** $<w_1 w_2 ... w_q : r_1\{t_1\}, r_2\{t_2\}, ..., r_p\{t_p\}>$ *if one of its Web transaction records pattern-contains* $<w_1 w_2 ... w_q : r_1\{t_1\}, r_2\{t_2\}, ..., r_p\{t_p\}>$.

A Web transaction consists of a set of purchases along the corresponding nodes in its traversal path. A transaction pattern is a large transaction pattern if there is a sufficient number of Web transactions pattern-containing it. It is worth mentioning that by taking both the traveling patterns and purchasing patterns into consideration, the problem of mining Web transaction patterns is in nature different from those addressed in prior works [3][5].

3 Algorithms for Web Transaction Patterns

In general, a Web transaction, generated from electronic commerce services, consists of a traversal path and a list of items purchased along the path. Given a Web transaction of a customer, algorithm WR is devised to derive Web transaction records to capture the customer traveling and purchasing behaviors in an EC environment.

Algorithm WR

Step 1. For each Web transaction, constructing a customer transaction tree by mainly incorporating customer transaction records, which correspond to nodes

with purchases, as branches. Each customer transaction record includes the traversal path and the items purchased in the last node of this path.
Step 2. Determining all the Web transaction records by traversing customer transaction tree obtained in Step 1 in a depth-first manner.
Step 3. Storing the Web transaction, including the Web transaction records and the corresponding WT_ID, into the database.

Algorithm WTM

A transaction pattern with k-purchase <$s_1s_2...s_p$: $n_1\{i_1\}$, $n_2\{i_2\}$,..., $n_k\{i_k\}$> is called a large k-transaction pattern, if there are a sufficient number of Web transactions pattern-containing it. Let C_k be a candidate set of large k-transaction patterns and T_k represent the set of the large k-transaction patterns. Similarly to DHP [6], WTM utilizes large transaction patterns for generating candidate transaction patterns. Furthermore, WTM employs a sophisticated hash tree, called Web transaction tree, to store candidate transaction patterns. WTM hashes not only each node but also each purchase in the path. According to each Web transaction record of a Web transaction, the support of a candidate transaction pattern is determined by the number of Web transactions that pattern-contain this candidate transaction pattern. WTM then obtains the large transaction patterns by destructing the Web transaction tree. Consider the example scenario in Figure 1. In the first pass, where WTM constructs the Web transaction tree by hashing each Web transaction record to construct the Web transaction tree and counts the support of individual purchases. Then, WTM destructs the Web transaction tree for deriving T_1, the set of large 1-transaction patterns. In each subsequent pass, WTM starts with the large transaction patterns found in the previous pass for generating new candidate transaction patterns to be stored in a Web transaction tree. Then, WTM proceeds to the counting of supports and finally reaches the generation of large transaction patterns.

After all large transaction patterns are obtained, one can derive the Web-transaction association rules from the large transaction patterns. In this example, <ABCE : B$\{i_1\}$, C$\{i_2\}$, E$\{i_4\}$> is one large 3-transaction pattern with support = 2 and <AB : B$\{i_1\}$> is one large 1-transaction pattern with support = 3. As a result, we can derive one Web-transaction association rule <ABCE : B$\{i_1\}$ $\Longrightarrow$ C$\{i_2\}$, E$\{i_4\}$> with the support equal to support(<ABCE : B$\{i_1\}$, C$\{i_2\}$, E$\{i_4\}$>) = 2 and the confidence equal to $\frac{support(<ABCE:B\{i_1\},C\{i_2\},E\{i_4\}>)}{support(<AB:B\{i_1\}>)}$ = 67%.

4 Conclusion

In this paper, we explored a new data mining capability which involves mining Web transaction patterns. First, we developed algorithm WR to extract

Database

Web Transactions		
WT_ID	Path	Purchases
100	ABCE	B{i_1}, C{i_2}, E{i_4}
	ABFGH	B{i_1}, H{i_6}
	ASJL	S{i_7}, L{i_9}
200	ABCE	B{i_1}, C{i_2}, E{i_4}
	ASJLQ	S{i_7}, Q{i_{10}}
300	ABCE	B{i_1}, E{i_4}
	ABFG	B{i_1}, G{i_5}
	ASJL	S{i_7}, J{i_8}, L{i_9}
400	ABD	D{i_3}
	ABFG	G{i_5}
	ASJLQ	S{i_7}, J{i_8}, Q{i_{10}}

C1

Path	Purchase	Sup.
AB	B{i_1}	3
ABC	C{i_2}	2
ABD	D{i_3}	1
ABCE	E{i_4}	3
ABFG	G{i_5}	2
ABFGH	H{i_6}	1
AS	S{i_7}	4
ASJ	J{i_8}	2
ASJL	L{i_9}	2
ASJLQ	Q{i_{10}}	2

T1

Path	Purchase	Sup.
AB	B{i_1}	3
ABC	C{i_2}	2
ABCE	E{i_4}	3
ABFG	G{i_5}	2
AS	S{i_7}	4
ASJ	J{i_8}	2
ASJL	L{i_9}	2
ASJLQ	Q{i_{10}}	2

C2 ($|C2|=28$)

Path	Purchases	Sup.
ABC	B{i_1}, C{i_2}	2
ABCE	B{i_1}, E{i_4}	3
⋮	⋮	⋮
AS	B{i_1}, S{i_7}	0
ASJ	B{i_1}, J{i_8}	0
⋮	⋮	⋮
ASJLQ	J{i_8}, Q{i_{10}}	1
ASJLQ	L{i_9}, Q{i_{10}}	0

T2

Path	Purchases	Sup.
ABC	B{i_1}, C{i_2}	2
ABCE	B{i_1}, E{i_4}	3
ABCE	C{i_2}, E{i_4}	2
ASJ	S{i_7}, J{i_8}	2
ASJL	S{i_7}, L{i_9}	2
ASJLQ	S{i_7}, Q{i_{10}}	2

C3

Path	Purchases	Sup.
ABCE	B{i_1}, C{i_2}, E{i_4}	2

T3

Path	Purchases	Sup.
ABCE	B{i_1}, C{i_2}, E{i_4}	2

Figure 1: An illustrative example for mining Web transaction patterns

meaningful Web transaction records from Web transactions so as to filter out the effect of irrelevant traversal sequences. Second, we devised algorithm WTM for determining the large transaction patterns from the Web transaction records obtained.

Acknowledgments

The authors are supported in part by the National Science Council, Project No. NSC 89-2219-E-002-007 and NSC 89-2213-E-002-032, Taiwan, Republic of China.

References

[1] http://www.amazon.com/.

[2] http://www.aol.com/.

[3] R. Agrawal and R. Srikant. Fast Algorithms for Mining Association Rules in Large Databases. *Proceedings of the 20th International Conference on Very Large Data Bases*, pages 478–499, September 1994.

[4] A. G. Buchner and M. Mulvenna. Discovery Internet Marketing Intelligence through Online Analytical Web Usage Mining. *ACM SIGMOD Record*, 27(4):54–61, Dec. 1998.

[5] M.-S. Chen, J.-S. Park, and P. S. Yu. Efficient Data Mining for Path Traversal Patterns. *IEEE Transactions on Knowledge and Data Engineering*, 10(2):209–221, April 1998.

[6] J.-S. Park, M.-S. Chen, and P. S. Yu. Using a Hash-Based Method with Transaction Trimming for Mining Association Rules. *IEEE Transactions on Knowledge and Data Engineering*, 9(5):813–825, October 1997.