International Journal of Innovative
Computing, Information and Control
Volume **5**, Number **8**, August **2009**

 ISSN 1349-4198
pp. **2351–2367**

MINING HYBRID SEQUENTIAL PATTERNS BY HIERARCHICAL MINING TECHNIQUE

Kuen-Fang Jea, Ke-Chung Lin and I-En Liao

Department of Computer Science
National Chung-Hsing University
250, Kuo Kuang Rd., Taichung 402, Taiwan
{ kfjea; phd9212 }@cs.nchu.edu.tw; ieliao@nchu.edu.tw

Received January 2008; revised May 2008

Abstract. *Unlike sequential patterns, hybrid sequential patterns display not only the path but also the relationship among transaction items. The information provided by the collection of hybrid sequential patterns is useful in improving the analysis of marketing strategies, such as, browsing web pages, discovering customers' behaviors and so on. The process of mining hybrid sequential patterns in a database, however, becomes complicated by the huge number of candidate patterns. In this paper, we propose a hierarchical mining technique to deal with this complexity. The unique features of this new technique include: counting hybrid sequential patterns by class, and examining database transactions in a top-down manner. This results in scanning a database, at most, twice. Using the technique, we develop an efficient mining algorithm and conduct a simulation to study its performance. There are three major contributions in this paper. First, our proposed pattern-class concept provides a new way to count a group of patterns simultaneously. Second, we propose a novel decomposition model to lower the I/O cost in counting patterns from a large database. And third, we prove the correctness of counting patterns in the pattern decomposition model in this paper.*
Keywords: Hybrid sequential pattern, Mining, Hierarchical mining technique

1. **Introduction.** The sequential pattern [3,4] is an ordered list of frequent itemsets in a customer's sequence. It reveals the customer's purchasing behaviors in transactional databases, from which the policy-maker may make an informed business decision. It is useful in various applications for improving the quality of analysis, including the analysis of web surfing [12], transactional customers' behaviors [3,4], network alarm patterns [13], DNA sequences [11] and so on.

Mining sequential patterns from a database is known as the sequential mining problem. Agrawal [3] introduced the sequential mining problem. Several algorithms for solving the problem have since been proposed [3,4,7-9,13]. AprioriAll [3] and GSP [4] use a candidate generation method, such that the sequential patterns at a level can be used to construct candidate patterns at the next level. FreeSpan [7] and PrefixSpan [11] project the sequence database into several smaller ones by the frequent 1-patterns and recursively grow up the sequential pattern in each projected database for mining all sequential patterns. MEMISP [8] scans the database once to read all sequences into memory. Then, it mines sequential patterns by recursively finding the items that form a sequential pattern and generating an index set for further mining. SPADE [13] generates sequential patterns in the lattice by an intersection of id-lists in the vertical database. Three database scans complete the mining process. With an efficient lattice search technique and simple join operations, SPADE discovers all sequential patterns.