International Journal of
Information
and
Management
Sciences

International Journal of Information and Management Sciences
27 (2016), 129-145. DOI:10.6186/IJIMS.2016.27.2.4

Discovering Time-Interval Sequential Patterns by a Pattern Growth Approach with Confidence Constraints

Huan-Jyh Shyur, Chichang Jou, Chi-Bin Cheng and Chih-Yu Yen

Tamkang University

Abstract

Sequential pattern mining is to discover frequent sequential patterns in a sequence database. The technique is applied to fields such as web click-stream mining, failure forecast, and traffic analysis. Conventional sequential pattern-mining approaches generally focus only the orders of items; however, the time interval between two consecutive events can be a valuable information when the time of the occurrence of an event is concerned. This study extends the concept of the well-known pattern growth approach, PrefixSpan algorithm, to propose a novel sequential pattern mining approach for sequential patterns with time intervals. Unlike the other time-interval sequential pattern-mining algorithms, the approach concerns the time for the next event to occur more than the timing information with its precedent events. To obtain a more reliable sequential pattern, a new measure of the confidence of a sequential pattern is defined. Experiments are conducted to evaluate the performance of the proposed approach.

Keywords: Sequential pattern mining, time-interval sequential patterns, pattern growth, confidence.

1. Introduction

With the fast development of internet and information technologies, the amount of accumulated electronic data has been growing exponentially, which also fertilizes the chance of extracting useful information to assist in market analysis, flaw detection, decision making, etc. Data mining is a technology to find implicit, previously unknown, and potentially useful information from such a large amount of data. Many types of data mining techniques have been proposed to serve different purposes of information extraction, among which sequential pattern mining is one of the important endeavors [5, 9, 25]. Sequential pattern mining was first introduced by Agrawal and Srikant [2] in the mid 1990s to discover patterns that occur frequently in a sequence database. A typical example of a sequential pattern is that a customer who, having bought a camera, returns to buy an auxiliary lenses.