International Journal of Innovative
Computing, Information and Control
Volume **7**, Number **9**, September **2011**

ICIC International ©2011 ISSN 1349-4198
pp. 5515–5528

# MINING EMERGING PATTERNS FROM TIME SERIES DATA WITH TIME GAP CONSTRAINT

Hsieh-Hui Yu[1], Chun-Hao Chen[2] and Vincent S. Tseng[1,*]

[1]Department of Computer Science and Information Engineering
National Cheng Kung University
No. 1, University Road, Tainan 701, Taiwan
digital@idb.csie.ncku.edu.tw; *Corresponding author: tsengsm@mail.ncku.edu.tw

[2]Department of Computer Science and Information Engineering
Tamkang University
No. 151, Ying-Chuan Road, Taipei 251, Taiwan
chchen6814@gmail.com

Received April 2010; revised August 2010

Abstract. *Discovery of powerful contrasts between datasets is an important issue in data mining. To address this, the concept of emerging patterns (EPs) has thus been introduced by Dong and Li. EPs are a set of itemsets whose support changes significantly from one dataset to another. Although an increasing number of works focus on this topic with regard to relational databases, few have considered mining EPs in time series. In this paper, we thus propose a framework named PIPs-SAX for mining EPs from time series data. The framework contains two phases: the first phase is data transformation and the second is the EPs mining. The first phase transforms the time series data into a symbolic representation based on the SAX and PIPs algorithms. In the second phase, we propose an algorithm, called TSEPsMiner, to mine time series EPs with a time gap constraint. Experiments on financial data collected from the Taiwanese stock exchange were also made in order to evaluate the effectiveness of the proposed framework.*
**Keywords:** Emerging patterns, Contrast sets, Time series data analysis, Symbolic aggregative approximation (SAX), Perceptually important points (PIPs)

1. **Introduction.** The discovery of powerful contrasts between datasets is an important issue in data mining. To address this, the concept of emerging patterns (EPs) was first introduced by Dong and Li [11]. EPs are a set of itemsets whose support changes significantly from one dataset to another. The discovery of EPs can essentially be regarded as a variant of association rule mining [2-4], and the itemsets found using EPs are considered as meaningful features that can be used to distinguish or describe the differences among a collection of datasets, while association rule mining discovers rules that describe the current situation.

Recently, an increasing amount of attention has focused on this topic with regard to relational databases [6,11,14,15,18,20,21], but few studies have considered sequential or time series databases. For both of these latter two databases, the occurrence order of each item is meaningful, and so, general methods for mining EPs in relational databases cannot be directly applied to them. Chan et al. thus introduced the concept of emerging substrings (ESs) to address the problem of mining EPs in sequential databases [9], utilizing a suffix tree-based framework.

Lin et al. [23] also proposed an approach to mining contrast sets. To the best of our knowledge, it was the first work for mining contrast sets from time series data. It tries to discover pairs of significant subsequences that can differentiate two sets of data. For