International Journal of Innovative Computing, Information and Control
Volume **6**, Number **8**, August **2010**

ICIC International ©2010 ISSN 1349-4198
pp. 3425–3436

ASSOCIATION RULE MINING FOR EVALUATION OF REGIONAL ENVIRONMENTS: CASE STUDY OF DAPENG BAY, TAIWAN

Chen-Yuan Chen[1], Shiahn-Wern Shyue[2,*] and Chin-Jui Chang[3]

[1]Department of Computer Science
National Pingtung University of Education
No. 4-18, Ming Shen Rd., Pingtung 90003, Taiwan
cyc@mail.npue.edu.tw

[2]Department of Marine Environment and Engineering
National Sun Yat-sen University
No. 70, Lianhai Road, Gushan District, Kaohsiung City 80424, Taiwan
*Corresponding author: swshyue@mail.nsysu.edu.tw

[3]Department of Information Management
Transworld Institute of Technology
No. 1221, Jen-Nang Road, Chia-Tong Li, Douliou, Yunlin 64063, Taiwan

Received February 2009; revised August 2009

Abstract. *Data mining is commonly used in a wide range of different fields. It is an important tool for analyzing information in research areas where there is a huge collection of data. The aim of this study is to determine the various patterns that characterize multi-dimensional marine environments. In particular, we wish to determine the association patterns, the characterize Taiwan's Dapeng Bay. The processes of association rule (AR) mining and decision tree (DT) analysis are the main methodologies used in this study. We also utilize Weka, a comprehensive suite of Java class libraries for implementing many machine learning algorithms and the Clementine 10.1 software package for data mining analysis. Applications constructed from the AR model and decision tree model can be applied in other scientific domains. This paper describes the data mining process for Dapeng Bay, which is located in the south of Taiwan.*
Keywords: Data mining, Association rule (AR), Decision tree (DT), Cluster analysis (CA)

1. **Introduction.** Artificial intelligence is currently widely used in real-time applications, such as robot, image identification and data mining. Data mining is the process of extracting hidden patterns from large amounts of data. The technique is becoming an increasingly important tool for transforming data into information [1,2,14,15]. Recent oceanographic and environmental projects generally involve the use of large interdisciplinary multi-variable data sets [3,16-18]. These data sets have to be quality checked, reliably stored, and then analyzed. It is necessary to manage large quantities of interdisciplinary data on marine and coastal environments. These data sets can contain numerous variables from different sources and with different structures (oceanographic, meteorological, biological and so on). It is a very complicated task to combine these multi-disciplinary data sets, verify them, and analyze them jointly [4,19,20].

The association model is one tool that has been successfully employed for predictive models. It is possible to construct models with a high degree of predictive accuracy with regard to data preparation and experimental techniques. If the model can be visualized, it is possible to gain valuable insights into the data, such as the relative importance and relevance of the attributes [5]. Cluster analysis (CA) is a way of grouping objects (cases)