

PTCR-MINER: AN EFFECTIVE RULE-BASED CLASSIFIER ON MULTIVARIATE TEMPORAL DATA CLASSIFICATION

CHAO-HUI LEE AND VINCENT S. TSENG*

Department of Computer Science and Information Engineering
National Cheng Kung University
No. 1, University Road, Tainan 701, Taiwan

*Corresponding author: tsengsm@mail.ncku.edu.tw

Received May 2010; revised September 2010

ABSTRACT. *Multivariate temporal data are hybrid data. Numeric and categorical data type could be consisted of. Most past researches cannot be operated directly on the multivariate temporal data with both types. Additionally, no useful and readable rules are provided in their methods for advanced classification analysis. We proposed Progressive Temporal Class Rule Miner (PTCR-Miner) algorithm to achieve the classification on multivariate temporal data with a rule-based designed. Through our algorithm, all really useful classification rules are discovered. The rules follow the purification concept we defined, which makes rules comprehensible and intuitive for general users on data classification. We did several experiments to evaluate our method with a multivariate temporal data simulator. Experimental results show PTCR-Miner performs effectively and efficiently on the different simulated multivariate temporal datasets. That means the discovered rules are really helpful and comprehensible for data classification. Furthermore, the rule-based and flexible architecture enables PTCR-Miner more applicable to different areas of multivariate temporal data classification.*

Keywords: Rule-based, Progressive, Multivariate temporal data, Classification, Data mining

1. Introduction. Useful information discovery is a main purpose of data mining. With the complexity of recorded data, different kinds of data mining methods are proposed continuously. Multivariate temporal dataset is a popular recorded format recently, which describes the states of an event using different variables with the time. The completeness of these data type is preferred in many study areas, such as weather data and medical data. For classification of multivariate temporal data, there are relatively rare suitable methods and it is due to the complexity of the data type. Many temporal datasets are hybrid data, which contain categorical and numeric values. Most data mining methods focus only on numerical data or only on categorical data and their modules or methods cannot apply to this kind of dataset appropriately. Additionally, the datasets recorded in multivariate temporal mode are generally significant for advanced analyses or diagnoses. Therefore, the supply of the information about classification is important for the studies on multivariate temporal data. Many temporal related researches focused on the time series data which consisted of only numeric values and performed high accurate classification results. Several different studies applied similar concepts to multivariate time series data which consisted of multiple numeric time series data. The datasets are regarded as a matrix and transformed into lower dimension format for easier similar measurements [23,24]. However, many temporal variables are recorded in categorical type and even in different sample rate for a multivariate temporal dataset. Thus, a dataset with many