Chapter 90
Association Rule Mining in Developmental Psychology

D. A. Nembhard
Pennsylvania State University, USA

K. K. Yip
Pennsylvania State University, USA

C. A. Stifter
Pennsylvania State University, USA

ABSTRACT

Developmental psychology is the scientific study of progressive psychological changes that occur in human beings as they age. Some of the current methodologies used in this field to study developmental processes include Yule's Q, state space grids, time series analysis, and lag analysis. The data collected in this field are often time-series-type data. Applying association rule mining in developmental psychology is a new concept that may have a number of potential benefits. In this paper, two sets of infant-mother interaction data sets are examined using association rule mining. Previous analyses of these data used conventional statistical techniques. However, they failed to capture the dynamic interactions between the infant-mother pair as well as other issues relating to the temporal characteristic of the data. Three approaches are proposed in this paper as candidate means of addressing some of the questions that remain from previous studies. The approaches used can be applied to association rule mining to extend its application to data sets in related fields.

1. INTRODUCTION

Development psychology is the scientific study of progressive psychological changes that occur in human beings as they age. This field examines change across a broad range of topics including motor skills and other psycho-physiological processes, problem solving abilities, conceptual understanding, acquisition of language, moral understanding, and identity formation.

Generally, multivariate time series data are collected to answer important questions in developmental psychology research. These include questions related to emotional response patterns

DOI: 10.4018/978-1-4666-2455-9.ch090

(Rovine, Molenaar, & Corneal, 1999), marital interactions (Laurenceau, Feldman-Barrett, & Rovine, 2005), parent-child interactions (Stifter & Braungart, 1995), cognitive processing (Gilmore & Johnson, 1998), therapeutic interventions (Molenaar, 1987), the development of motor coordination (Fitzpatrick, 1998), and postural models (Newell, 1998), among others. Often data are collected in the form of daily reports, diaries, relatively dense behavioral observations, electronic responses (beepers, online studies), or physiological responses (ECG, brainwaves). In these areas, common methodologies include statistical analysis such as ANOVA, MANOVA, regression model, state space grids, Yule's Q, and time series analyses (Bukowski, Adam, & Santo, 2006). See Bishop, Feinberg, and Holland (1975), Everitt (1996), Kiess (1996), Welkowitz, Ewen, and Cohen (1986), and Wilcox (2002) for implementations of these methods in developmental psychology and related fields.

ANOVA and MANOVA are statistical models based on the analysis of variance, which test for significant differences between means of different groups of observations. Regression models are used to predict one dependent variable from one or more independent variables. Both of these methods require the analyst to clearly define the dependent and independent variables. Yule's Q is based on the odds ratio and a symmetric measure taking on values between -1 and +1, where the extremes imply perfect negative or positive association, respectively, with 0 representing no association. We remark that these methodologies are useful and straightforward when the researchers have a well defined set of questions to address. However, there may often exist underlying information and relations among the variables for which no questions have been formulated. In these cases, potentially valuable knowledge may go undetected. Additionally, temporal characteristics of these data while potentially of interest are not addressed by the existing set of statistical tools commonly used for these data.

To address some of these limitations, researchers have developed the state space grid, a methodology based on dynamic system principles (Hollenstein, 2005), as a graphical approach that utilizes observational data and quantifies these data according to two ordinal variables that define the state space for the system. Observations across time are plotted as a trajectory in the state space. Another method, time series analysis is a methodology that investigates the underlying structures in a time series (i.e., sequence of data points measured at successive times), in order to make forecasts. Common models used in time series analysis include Autoregressive Moving Average models (ARMA) and Autoregressive Integrated Moving Average models (ARIMA). These methods analyze the dynamics of data across time. However, the resulting models are only applied to an individual. If for example, an experiment has collected times series data for 10 different subjects, using these two methodologies can efficiently describe the 10 subjects individually, but the dynamic of the all the subjects as a group are not addressed.

As a means of overcoming some of these limitations, we will examine association rule mining. We remark that association rule mining has been shown to be a useful method in extracting information and uncovering hidden correlations in data. However, the approach is new to the field of developmental psychology. Further, conventional association rule mining does not consider the temporal nature that exists in some data. Each data point taken across time is usually treated as an individual record in the data set, where each record in the data set is assumed to be independent. A goal of association rule mining is to find associations between different attributes in the data set among these records. Hence, the correlation of the data with respect to time is collapsed. In order to deal with temporal data, researchers in the data mining field have developed advanced algorithms which adapted the ideas of temporal association rules and sequential pattern mining (e.g., Das,