Binary data flow visualization on factorial axes

Alfonso Iodice D'Eза, Francesco Palmumbo

1 Dipartimento di Matematica e Statistica
Università di Napoli Federico II
Complesso Universitario di Monte S. Angelo, via Cintia
I-80126, Napoli, Italy
iodiced@unina.it

2 Dipartimento di Istituzioni Economiche e Finanziarie
Università di Macerata
via Criscenti, 20
I-62100, Macerata, Italy
palmumbo@unime.it

Abstract. Data streams are one of the most relevant new data sources, they refer to flows of
data that come at a very high rate. Let us consider a stock exchange market, where n different
stocks with p considered attributes (e.g. price, quantity, seller/buyer id, . . . ) are negotiated all
day long. The distinguishing feature in data streams analysis is that the focus is on transient
relations. The present paper proposes a visualization tool exploiting Multidimensional Data
Analysis (MDA) techniques to represent the evolving association structures among attributes
over different time-frames. The general aim is to detect the stability of the deviation from
independence in the occurrence of an observed set of attributes stored as binary stream.

1 Introduction

In recent years, enhancements in monitoring activities and collecting data determined the need
for a different approach in knowledge extraction: new data are produced at a faster rate than the
capability of analyzing them.

Information mining through traditional data mining systems becomes often inadequate. Data
streams are one of the most relevant new data sources, they refer to flows of data that come
at a very high rate. Let us consider a stock exchange market, where n different stocks are negociated;
for each interval time-unit (seconds or minutes) a n × p array is added to the database, with
p indicating the number of considered attributes (e.g. price, quantity, seller/buyer id, . . . ). These
features make data streams leading to data structure unusual in the data analysis and statistical
data mining framework. New and more appropriate techniques should be taken into account to
usefully extract knowledge without storing the data for a long term [12].

The most relevant changing feature in data mining systems dealing with high-speed data
streams is the necessity of analyzing data in a single pass: iterative procedures will lead to unfeasible
solutions. Further ideal features of a data stream mining system are described in the proposal
by [4].

Data streams mining, in a wide sense, can be considered an evolution of data mining. The
development of data mining techniques has been fed in the last decade by statisticians and
computer scientists. Similarly, the data stream analysis should be enhanced with the contribution
of researchers from different areas.

Data stream analysis techniques can be roughly divided into: i) data-based, indicating techni-
niques aiming to summarise or reduce the amount of streams to be analyzed; ii) task-based tech-
niques, facing the crucial problem of adapting existing algorithms to the new computational costs;