

New OLAP Operators for Missing Data

Maha Ben Kraiem*,** Kais Khrouf*, Jamel Feki***, Franck Ravat**, Olivier Teste**

*MIR@CL Laboratory, University of Sfax
Airport Road Km 4, P.O. Box. 1088, 3018 Sfax, Tunisia
Maha.BenKraiem@yahoo.com, Khrouf.Kais@isecs.rnu.tn

** IRIT, University of Toulouse
118, route de Narbonne, 31069 Toulouse Cedex 9, France
{Ravat, Teste}@irit.fr

*** University of Jeddah, FCIT, IS dept
Saudi Arabia
jfeiki@uj.edu.sa

Abstract. Data analysis of social networks is often impeded by the problem of missing data. Recent studies highlight the negative effects of this problem mainly regarding querying process. The analysis of data social networks would be severely distorted when limited to filled fields (i.e., not null valued fields) whereas missing data are ignored. To overcome the missing data problem, we provide in this paper an extension of classical *Drilldown* and *Rollup* operators in order to support analyses on multidimensional datasets containing missing values of dimension members.

1 Introduction

In the last decade, many social networks such as Facebook, LinkedIn and Twitter have been developed, and they made users perceive the Web as a place where they exchange feelings and opinions as well as contents. However, despite these tools ease the sharing and collaboration between users; they may cause new challenges concerning the relevant exploitation of these User-Generated Contents (UGC) for decision making systems. Thus, new multidimensional models have been proposed for OLAP purposes. The multidimensional modeling comes with a set of specifics such as missing data. Missing data in social networks is a long standing but relatively poorly understood problem. The analysis of social networks is even thwarted by missing values. There are several ways in which researchers can cope with missing values, which are frequently found in data collected in empirical research. The easiest way is to simply ignore the missing data. However, restricting analyses to the observed responses (i.e., not null fields) results in serious loss of information and then decreases the power of statistical results. Some other missing data treatments include weighting procedures, model-based procedures, and imputation. Facing to great amount of missing data in large volumes of data sets, we set a twofold purpose, first increase the efficiency of analysis and, secondly, help the analysts. For this reason, we extend the two classical *Drilldown* and *Rollup* operators; this extension enables the analyst to handle missing data on dimension members. In this context, our previous work