Empirical Validation of Object Oriented Data Warehouse Design Quality Metrics

Jaya Gupta1, Anjana Gosain2, and Sushama Nagpal3

1,2 University School of Information Technology, GGS Indraprastha University, Delhi
jaya_gupta8@yahoo.com, anjana_gosain@hotmail.com
3 Computer Engineering Dept, Netaji Subhash Institute of Technology, Delhi
sushmapriyadarshi@yahoo.com

Abstract. Data warehouses have been developed that stores information enabling the knowledge worker to make better and faster decisions. As a decision support information system, a data warehouse must provide high level quality of data and quality of service. Various metrics have been defined and theoretical validated to measure the quality of the data warehouse in a consistent and objective manner and if quality measured, it can be managed and improved. Now, in this paper we will use these design quality metrics and empirically validated these metrics by conducting an experiment using regression analysis and deriving the conclusions according to the analysis so that they can be used by researchers and users.

Keywords: Object Oriented Conceptual Modeling, Metrics, Regression Analysis, Empirical Validation.

1 Introduction

Data warehouses have been developed to answer the increasing demands of information required by the top managers and economic analysts of organizations. Quality is the key issue in the building the data warehouse and gives the confidence that particular information meets some context specific quality requirements [8]. A lack of quality in the data warehouse can have disastrous consequences for the organizations. One of the main issues that influence the quality is designing the data warehouse using various data models i.e. conceptual, logical and physical [14]. Conceptual modeling forms the basis of the data warehouse and is concerned with the real world view and understanding of data. Various conceptual model metrics are defined to measure a quality factor in a consistent and objective manner [9] and gives the best ways to help professionals and researchers. Here our goal is to empirically validate the object oriented conceptual model metrics using the correlation and regression analysis technique.

2 Related Work

In this the related work is defined for the two main topics covered in this paper: (i) Object Oriented Conceptual Multidimensional modeling (ii) Quality metrics proposed in the data warehouse.
2.1 Object Oriented Conceptual Multidimensional Modeling

An approach in [5] has been proposed as an object-oriented (OO) conceptual MD modeling approach. This proposal is a profile of the Unified Modeling Language (UML) which use the standard extension mechanisms (stereotypes, tagged values and constraints) provided by the UML. The extension used the Object Constraint Language OCL for expressing well-formedness rules of new defined elements. Another approach is given in [12] i.e. YAM2 allows the representation of several semantically related star schemas, as well as summarizability and identification constraints. In [6] the authors propose an approach that provides a theoretical foundation for the use of object-oriented databases and object-relational databases in data warehouse, multidimensional database, and online analytical processing applications.

2.2 Quality Metrics for the Data Warehouses

Metrics are proposed for OO conceptual modeling and theoretical validation and empirical validation is done in [1] but the models taken were few in number. Various OO conceptual model quality metrics are proposed in [4] but are not empirically validated. Si-Said and Prat [3] have proposed some metrics for measuring multidimensional schemas analyzability and simplicity. But these metrics proposed so far has not been empirically validated. Various metrics have been proposed [2] to assure the quality of data warehouse logical models validated both formally and empirically [8]. In [4] authors present a framework to design metrics in which each metric is part of a quality indicator we wish to measure. In [15] there is a review of research in conceptual model quality and identifies the major theoretical and practical issues which need to be addressed.

3 Conceptual Modeling

Multidimensional modeling has been widely accepted as the foundation of data modeling for data warehouses [10]. The first design steps accomplished in data warehouses involve producing a conceptual schema by using a conceptual model that conveniently represents the multidimensional modeling properties. Conceptual modeling describes entity classes, and characteristics i.e. attributes and associations between pairs of those things of significance i.e. relationships. We now need objective metrics for this purpose and we should empirically validate those metrics.

3.1 Object Oriented Conceptual Modeling

In conceptual modeling, we have used the extension of the UML (Unified Modeling Language). This is an object-oriented conceptual approach for data warehouses that easily represents main data warehouse properties at the conceptual level[1]. Tables 2 and 3 summarize the defined stereotypes along with a brief description and the corresponding icon in order to facilitate their use and interpretation. These stereotypes are classified into class stereotypes (Table 2) and attribute stereotypes (Table 3).