International Journal of Multidisciplinary Research and Development
Online ISSN: 2349-4182, Print ISSN: 2349-5979, Impact Factor: RJIF 5.72
www.allsubjectjournal.com
Volume 3; Issue 5; May 2016; Page No. 52-54

Analysis of theoretical validation (Distance framework approach) of conceptual metrics

[1] Sushil Kumar, [2] Pradeep Sangwan
[1] Ph. D Scholar, OPJS University, Churu, Rajasthan, India.
[2] Ph.D. Scholar, India.

Abstract
The paper discuss the importance of theoretical validation of quality metrics for quality evaluation of data warehouse conceptual model metrics. The quality metrics play significant role in quality evaluation of data warehouse models at conceptual and logical level.

Keywords: Theoretical Validation, Quality metrics, Data warehouse, Conceptual models.

1. Introduction
Data warehouse store large amounts of information about organizations. This information can be made use of for predicting futuristic trends. To enable fast and accurate extraction of useful information the quality aspect of data warehouses is to be focused upon. One of the major factors affecting the quality of data warehouses is quality metrics. The focus of the paper is on the study of various quality metrics proposed by researchers and discussing their importance in quality evaluation using theoretical validation techniques.

2. Literature Review
The design of a data warehouse system follows an incremental process starting with conceptual phase, followed by logical phase and finally physical phase. The conceptual design phase is the most important phase of the data warehouse design process as it lays the foundation for effective software development. Several quality metrics have been proposed by researchers to evaluate the quality of conceptual and logical models of data warehouse development. The related literature is as follows:
Serrano *et al.* 2007 [1]. has proposed quality metrics for quality evaluation of conceptual models. The proposed metrics are theoretically and empirically validated to prove their utility. The metrics proposed by him for a schema S are as follows:
- NDC(S) Number of dimension classes.
- NBC(S) Number of base classes.
- NC(S) Total number of classes, NC(S) = NDC(S) + NBC(S) + 1
- RBC(S) Ratio of base classes. Number of base classes per dimension class.
- NAFC(S) Number of FA attributes of the fact class.
- NADC(S) Number of D and DA attributes of the dimension classes.
- NABC(S) Number of D and DA attributes of the base classes.
- NA(S) Total number of FA, D and DA attributes, NA(S) = NAFC(S) + NADC(S) + NABC(S)
- NH(S) Number of hierarchy relationships DHP(S) Maximum depth of the hierarchy relationships.
- RSA(S) Ratio of attributes. Number of attributes FA divided by the number of D and DA attributes.

Serrano *et al.* 2008 [2]. Proposed structural metrics for quality evaluation of logical models and carried out an empirical study to investigate their significance in quality evaluation of logical models. The proposed metrics are as follows:
- NFT (Sc). Number of fact tables in the schema.
- NDT (Sc). Number of dimension tables in the schema.
- NFK (Sc). Number of foreign keys in all the fact tables of the schema.
- NMFT (Sc). Number of facts in the fact tables.
- Genero *et al.* 2007 [3]. proposed 3 size metrics and 8 structural metrics for conceptual models as follows:

Size metrics
- Number of Classes (NC) The total number of classes in a class diagram.
- Number of Attributes (NA) The number of attributes defined across all classes in a class diagram (not including inherited attributes or attributes defined within methods).
- Number of Methods (NM) The total number of methods defined across all classes in a class diagram, not including inherited methods.

Structural metrics
- Number of Associations (NAssoc) The total number of association relationships in a class diagram
- Number of Aggregations (NAgg) The total number of aggregation relationships (each "wholepart" pair in an aggregation relationship).
- Number of Dependencies (NDep) The total number of dependency relationships.
- Number of Generalizations (NGen) The total number of generalization relationships (each "parent-child" pair in a generalization relationship).
- Number of Generalization Hierarchies (NGenH) The total number of generalization hierarchies, i.e. it counts the total number of structures with generalization relationships.
- Number on Generalization Hierarchies (NAggH) The total number of aggregation hierarchies, i.e. it counts the total numbers of "whole-part" structures within a class diagram.
- Maximum DIT (MaxDIT). The maximum DIT value obtained for each class of the class diagram. The DIT

value for a class within a generalization hierarchy is the longest path from the class to the root of the hierarchy

- Maximum HAgg (MaxHAgg) The maximum HAgg value obtained for each class of the class diagram. The HAgg value for a class within an aggregation hierarchy is the longest path from the class to the leaves.
- Calero *et al.* 2001 [4]. Proposed various metrics for different configurations of data warehouse schemas.
- Table metrics: NA, NFK
- Star metrics: NDT, NT, NADT, NAFT, NA, NFK, RSA, RFK
- Schema metrics: NFT, NDT, NSDT, NT, NAFT, NADT, NASDT, NA, NFK, RSDT, RT, RFK, RSDTA

Also these metrics were theoretically validated to prove their relevance.

Shull *et al.* 2008 [5] focused on the role of replications in empirical study. Replications were categorized in two types:

- Exact replication
- Conceptual replication

Exact replication was further classified as dependent and independent replication. Goals, benefits, limitations of each were discussed with due emphasis on documentation.

Moody, 2005 [6]. Discussed various theoretical and practical issues in evaluating the quality of conceptual models with special emphasis on experimental techniques.

Lucia *et al.* 2010 [7]. Conducted three sets of controlled experiments aimed at analyzing whether UML class diagrams are more comprehensible than ER diagrams during data models maintenance. The results indicated that UML class diagram subjects achieved better comprehension levels.

kpodjedo *et al.* 2011 [8]. Performed an investigation to find the usefulness of elementary design evolution metrics to identify defective classes. It was shown that design evolution metrics make significantly better predictions of defect density than other metrics and thus help in reducing testing effort by focusing test activity on reduced volume of code.

3. Theoretical Validation and its Importance

DISTANCE (Serrano *et al.* 2007) [1]. Framework guarantees that the metrics defined and validated using the framework are in a ratio scale. The theoretical validation of metric using Distance framework aims to provide answer to following questions:

- Does the proposed metric provide a measure (numeric value) to evaluate specific attribute of data warehouse conceptual model? The specific attribute measured in terms of proposed metric is understand ability based on the structural complexity of model.
- Is the metric capable enough to transform one configuration of data warehouse conceptual model to other on application of finite sequence of elementary transformations?
- This distance based measure construction process as discussed by Serrano *et al.* (2007) [1]. consists of five steps:
- Step 1. Find a measurement abstraction: The step aims to map data warehouse conceptual model onto its set of relationships between facts and dimensions. The output of this step is a set of measurement abstractions M containing existing relationships between facts and dimensions.
- Step 2. Model distances between measurement abstractions: The step outputs a set of elementary transformations that can transform relationship sets of one model to relationship sets of other model. The input is taken in the form of measurement abstractions obtained in previous step.
- Step 3. Quantify distances between measurement abstractions: This step aims to give a count of shortest possible elementary transformations to transform relationship set of one model to relationship set of other model.
- Step 4. Find a reference abstraction: To generalize the approach that can be applicable to any number of conceptual data warehouse models a base case having lowest possible value of proposed metric is identified and is output of this step.
- Step 5. Define the software measure: This step gives the numerical count of the proposed metric measured with respect to base case as identified in previous step for any data warehouse conceptual model.

4. Research Implication

The literature review conducted in the paper gives a detailed overview of the various quality metrics proposed by various researchers and their importance in predicting the quality of conceptual models. Theoretical validation gives the measure to evaluate the capability of a metric to measure specific attribute of conceptual models. The current research work carried out in this paper will help the future researchers to propose new metrics and to evaluate their significance using more effective theoretical validation techniques.

5. Conclusion

The paper discuss in detail theoretical validation technique based on DISTANCE framework to measure the significance of quality metrics in predicting the quality of conceptual data warehouse models.

The present work can be extended by proposal on new quality metrics and to check their significance using theoretical validation techniques. Also new, more effective techniques of theoretical validation can be explored by researchers.

References

1. M Serrano, J Trujillo, C Calero, M Piattini, Metrics for data warehouse conceptual models understandability, Information and Software Technology, 2007; 49(8): 851-870.
2. Serrano MA, Calero C, Sahraoui HA, Piattini M. Empirical studies to assess the understandability of data warehouse schemas using structural metrics, Software Quality Journal, 2008; 16(1): 79-106.
3. Genero M, Poels G, Piattini M. Defining and validating metrics for assessing the understandability of entity–relationship diagrams, Data & Knowledge Engineering, 2008; 64(03): 534-557.
4. Calero C, Piattini M, Pascual C, Serrano M. Towards Data warehouse Quality Metrics, International Workshop on Design and Management of Data Warehouses (DMDW'01), 2001.
5. Shull F, Jeffrey C, Carver, Sira Vegas, Natalia Juristo. The Role of Replications in Empirical Software Engineering. Journal of Empirical Software Engineering, Springer, 2008. 13: 211-218.

6. Moody D. Theoretical and practical issues in evaluating the quality of conceptual models current state and future directions' Data & Knowledge Engineering, 2005; 55(3):243-276.
7. Lucia A, Carmine, Gravino, Oliveto R, Tortora G. An experimental comparison of ER and UML class diagrams for data modeling' Journal of empirical software engg, Springer, 2010. 15: 455-492.
8. Kpodjedo S, Ricca F, Galinier F, Gueheneuc Y, Antoniol G. Design evolution metrics for defect prediction in object oriented systems, Empirical Software Engineering, 2011; 16(01): 141-175.