Model Warehouse Design: New Ideology And Challenges

Pooja R. Gupta, Dr. Sanjay Pachauri, Dr. Pradeep Chouksey

Abstract: This Model management is an important issue in many different contexts and domains. Due to the specific characteristics of models, specific systems are required for model management, in order to model, store, retrieve, and manipulate models in an efficient and effective way. Thus, there need to create model repositories where the knowledge artifacts known as models will reside, as primitive data reside in data base management systems (DBMSs). This paper introduces model characteristics and its importance in present scenario. The problems and challenges with model management are also presented here. The research direction towards a model base management system (PBMSs) by abstracting common features of models and management requirements is discussed.

Keywords: Model bases; Model-Base Management Systems; , Data models; Knowledge.

I. INTRODUCTION

Now a day's many applications huge volumes of data [8]. Many techniques have been developed on how to extract knowledge, especially in the context of data mining. The Results of such operations are abstract and compact representations of the original data, which called models [1]. The data processing methods such as model recognition, data mining and knowledge extraction result in knowledge artifacts like clusters, association rules, decision trees and others [12]. The volume of extracted models from various knowledge discovery applications is increasing rapidly and so does the need for effective and efficient model management. There are no tools or systems that are designed to deal exclusively with this problem [5]. All current technologies just try to extend existing system to avail models but this is not effective because of special characteristic of models. Thus, an ordinary DBMS is not adequate for the management of this kind of information not only because of its special structure and properties, but furthermore because of its diversity. Models can generated through clustering schemes, association rules, classification rules, probabilistic rules and many others. Moreover models can be found in images, signals, text and music and of course in the World Wide Web. Figure 1 depicts the presentation of model for huge amount of data. Here 'P' represents the individual model. The first objective of this paper is to discuss the necessity of model Management. The second objective is to discuss problems and challenges with model management. The third objective is to present to categories of the models according their structure.

II. LITERATURE SURVEY

[4] Author considers the larger problem of modeling, storing, and querying models, in a database-like setting and use a Model-Base Management System (PBMS) for this purpose. Specifically, (a) they formally define the logical foundations for the global setting of model management through a model that covers data, models, and their intermediate mappings; (b) They present formalism for model specification along with safety restrictions and (c) They introduce predicates for comparing models and query operators. The main Contributions of this paper are the organization of models and data in a Model Warehouse environment that (a) is based on the formal foundations of a well specified logical model and (b) provides the necessary mechanisms for the querying of models and data.

[11] Presents new way of conceptual design of model-bases. They discus UML could be used and extended to this end. They address the main issues in static modeling, including the representation of relationships between models, and briefly present some issues related to functional and dynamic modeling. Author just shown how it would be possible to conceptually model a model-base from the static, functional, and dynamic points of view by properly extending UML through the stereotyping mechanism. This is better way to express the semantics of models and their components, and believe that adopting UML is still preferable since it is a standard de facto for most software engineering applications. [8]Author review the concept of models and their applicability in several research. They examine the different types of models that are extracted from a data set, in order to gather the necessary requirements for the definition of a model model. This model constitutes the heart of the Model Base Management System. Authors try to integration of existing approaches towards a novel logical integration of models into a data model, language and base management system support. They reviewed different types of models from many areas, the current efforts on modeling data mining operations were addressed along with the corresponding results.

III. CHALLENGES AND PROBLEMS WITH PBMS

1. Require a specific system to store and reuse these models in order to fulfill requirements of the users for later decision making.

2. Define a valid mapping between the model base and the raw data in order to be able to switch between raw data and their model base.

3. Require a specific data structure or schema to store models because their structure can be much more complex as well as diverse.

4. Model semantics are much richer than the raw dataset so system needs to preserve it.

5. Models' behavior/functionality is significantly more complex [7]. This involves also the similarity issue that is becoming much more complex and involves multiples dimensions of similarity, such as

i. intra-model vs. inter-model similarity
ii. Structural vs. value based similarity etc.

6. An intelligent model retrieval language needs to be incorporate.
7. System must be able to compare models with specified operations.
8. The models need to be updated time to time. There should be a clear policy for updating models without creating inconsistencies.

9. Since raw data may be very heterogeneous, several kinds of knowledge artifacts exist that can represent hidden knowledge. Clusters, association rules are common examples of such knowledge artifacts, generated by data mining applications. So the Model Base Management System must be able to handle this heterogeneity.

10. Models are special kind of data. So we need to put them a very specialized storage system that is called in this paper 'Model Warehouse'. This system must be able to handle all kind of models.

IV. MODEL BASE MANAGEMENT SYSTEM

Models share some characteristics that make traditional DBMSs unable to represent and manage them. Models [2], thus, can be regarded as knowledge units that effectively describe entire subsets of data (in this sense, they are compact). Models can be generated from different application contexts resulting in very heterogeneous structures [13]. Moreover, often, heterogeneous models have to be managed together [9]. Understanding the structure of models is essential for the design of a PBMS since a general model should be developed for the representation of all these artifacts in both logical and physical level of a PBMS [4].

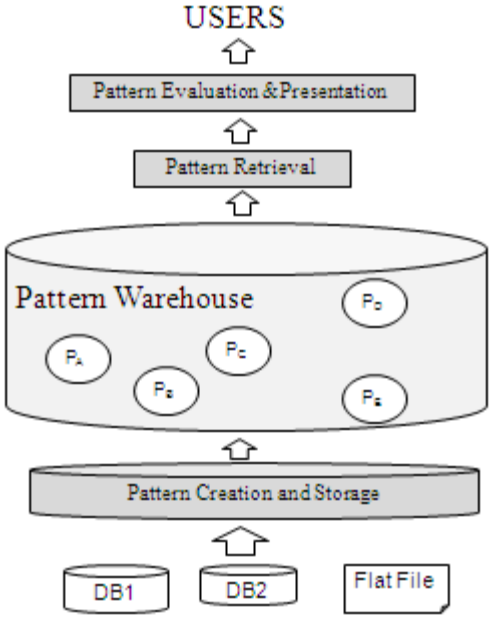

Figure 2: *Architecture of PBMS*

Model management is an important issue in many different contexts and domains. The most important context in which model management is required business intelligence and data mining. Business intelligence concerns a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprises in business decisions. Data mining is one of the fundamental activities involved in business intelligence applications besides querying and reporting, OLAP processing, statistical analysis, and forecasting [10]. The knowledge units resulting from the data mining tasks can be quite different. Figure 2 depicts the working architecture of model base management system. A Model Base Management System (PBMS) is a system for handling (storing/processing/retrieving) models defined over raw data in order to efficiently support model matching and to exploit model-related operations generating information [6]. The set of models managed by a PBMS is called here model warehouse. Some required functionalities of model warehouse are:

(a) User-defined model types support.
(b) Relation between raw data and models
(c) Quality Measures
(d) Temporal features
(e) Hierarchies over types

The model warehouse can be classified according to the following aspects:
(i) The chosen architecture to manage models together with data;
(ii) Architecture to manage models and data separately.

Like any DBMS, PBMS also required at least two different types of languages must be provided: the Model Manipulation Language (PML), providing the basic operations by which models can be manipulated (e.g. extracted, synchronized, deleted), and the Model Query Language (PQL), supporting model retrieval [3]. PQL queries take as input models and data sets and returns models.

V. CONCLUSIONS

Models refer to knowledge artifacts used to represent in a concise and rich in semantics way huge quantity of heterogeneous raw data or some of their characteristics. Models are relevant in any knowledge intensive application, such as data mining, information retrieval, or image processing. We introduce new approach aiming at the definition of a system architecture that efficiently represents, maintains and manages models. The cornerstone of this new approach will be the model concept, aiming at representing huge volumes of information in an effective way. We have discussed some challenged and problems for designing model warehouse.

ACKNOWLEDGMENT

This work is supported by research grant from MANIT, Bhopal, India under Grants in Aid Scheme 2010-11, No. Dean(R&C)/2010/63 dated 31/08/2010.

REFERENCES

[1] Catania B. "A Framework for Data Mining Model Management" , In Proc. of ECML-PKDD'04, LNAI 3202, 87-98, 2004.
[2] Evangelos Kotsifakos and Irene Ntoutsi "Database support for data mining models", ACM-PCI'05

Proceedings of the 10th Panhellenic conference on Advances in Informatics, Pages 14-24,2005

[3] Imielinski, T. & Virmani, A. " MSQL: A Query Language for Database Mining", Data Mining and Knowledge Discovery, 2(4), 373–408,1999.

[4] Manolis Terrovitis , Panos Vassiliadis and Spiros Skiadopoulos," Modeling and language support for the management of model-bases", Data & Knowledge Engineering-ScienceDirect-2007.

[5] Ralph Kimball , Margy Ross, "The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling", John Wiley & Sons, Inc., New York, NY, 2002

[6] Rizzi S., Bertino E., Catania B., Golfarelli M., Halkidi M., Terrovitis M., Vassiliadis P., Vazirgiannis M., Vrahnos E.. Towards a logical model for models. Proc. ER Conference, Chicago, IL, USA, 2003.

[7] Stephan Schneider &Dirk Frosch-Wilke," Analysis models in dimensional data modeling", Proceeding ICDEM'10 Proceedings of the Second international conference on Data Engineering and Management , Pages 109-116 Springer-Verlag Berlin, Heidelberg ©2012.

[8] Vazirgiannis M., Halkidi M., Tsatsaronis G., Vrachnos E., et al. A Survey on Model Application Domains and Model Management Approaches. PANDA Technical Report TR-2003-01, 2003. Available at http://dke.cti.gr/panda.

[9] Jaesoon Park, Youngwok Kim and , Youngmin Cha, "The Concept of Model Warehouse and Contemplate an Application in Integrated Network Data Ware" Telecommunication Network Lab., Korea Telecom, 2002

[10] Michael Eldridge, "Enterprise Data Warehouse: A Models Approach to Data Integration", Microsoft IT showcase, © 2010 Microsoft Corporation.

[11] Stefano Rizzi "UML-Based Conceptual Modeling of Model-Bases". Proceedings of the International Workshop on Model Representation and Management, Heraklion, Hellas, March 2004.

[12] Vivek Tiwari & Vipin Tiwari , " Association Rule Mining-A Graph based approach for mining Frequent Itemsets". IEEE International Conference on Networking and Information Technology (ICNIT 2010) Manila, Philippines,IEEE 2010, ISBN:978-4244-7577-3.

[13] [13]Vivek Tiwari & Dr. R.S.Thakur, "A level wise Tree Based Approach for Ontology-Driven Association Rules Mining", CiiT International Journal of Data Mining and Knowledge Engineering, 2011.