IJCAT - International Journal of Computing and Technology, Volume 3, Issue 9, September 2016
ISSN : 2348 - 6090
www.IJCAT.org

Discovering of Data Dependencies in Relational Data Base: A Recent Overview

[1] Pakhale Sham Laxman, [2] Vijay Kumar Verma

[1] Computer Science & Engineering Department
M. Tech (CSE) IV Sem. Lord Krishna College of technology
Indore M.P. India 453331

[2] Computer Science & Engineering Department
Asst. Prof Lord Krishna College of technology
Indore M.P. India - 453331

Abstract - Data Dependencies plays an important role in the design of relational databases. The discovery of Data dependency from databases has recently become a significant research problem. Data dependencies represent domain knowledge and can be used to verify database design and assess data quality. Data normalization is a common mechanism employed to support database designers to ensure the correctness of their design. Functional dependencies are relationships between attribute of a database relation, a functional dependency state that the value of an attribute is uniquely determined by the values of some other attributes. Functional dependency plays a key role in database normalization. Discovering FDs can also help a database designer to decompose a relational schema into several relations through the normalization process to get rid or eliminate some of the problems of unsatisfactory database design. In This paper we proposed an overview of various methods developed recently for Discovering Data Dependencies.

Keywords **- Normalization, Functional Dependency, Database, Schema, Attribute.**

1. Introduction

A functional dependency is a statement X →Y requiring that X functionally determines Y where X; Y⊆ R. The dependency is satisfied by a database instance r if for any two tuples t_1; $t_2 \in r$, if $t_1[X] = t_2[X]$ then $t_1[Y] = t_2[Y]$. X is called the left-hand side (lhs) or the determinant and Y is called the right-hand side (rhs) or the dependent. Types Data Dependency discovery has attracted a lot of research interests from the communities of database design; machine learning and knowledge discovery. Three typical types of dependencies are often involved in the discovery, functional dependencies (FDs), conditional functional dependence and inclusion dependencies (INDs).

The aim of dependency discovery is to find important dependencies holding on the data of the database. These discovered dependencies represent domain knowledge and can be used to verify database design and assess data quality. In recent years, the demand for improved data quality in databases has been increasing and a lot of research effort in this area has been given to dependency discovery.

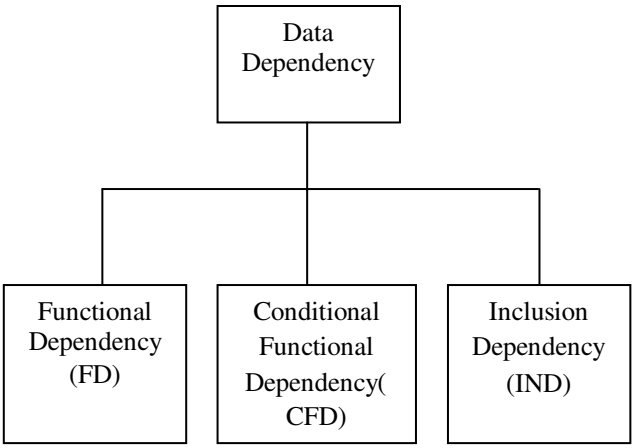

Figure 1 Types of data dependency

2. Methods for Discovery Data Dependencies

There are two basic approach are used for discovery data dependencies.

1. **Top-Down**
2. **Bottom-Up.**

The top-own methods start with generating candidate FDs level-by-level, from short lhs to long lhs, and then check

the satisfaction of the candidate FDs for satisfaction against the relation or its partitions.

The bottom-up methods, on the other hand, start with comparing tuples to get agree-sets or difference-sets, then generate candidate FDs and check them against the agree-sets ordifference-sets for satisfaction.

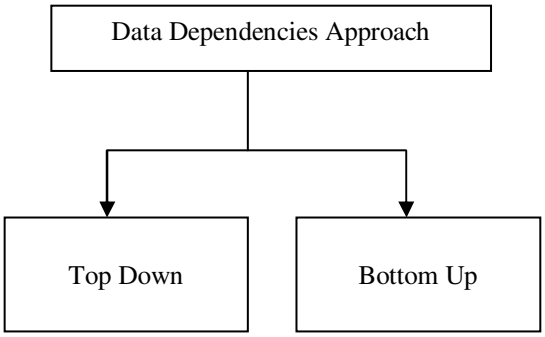

Figure 2 Basic Approach for data dependency

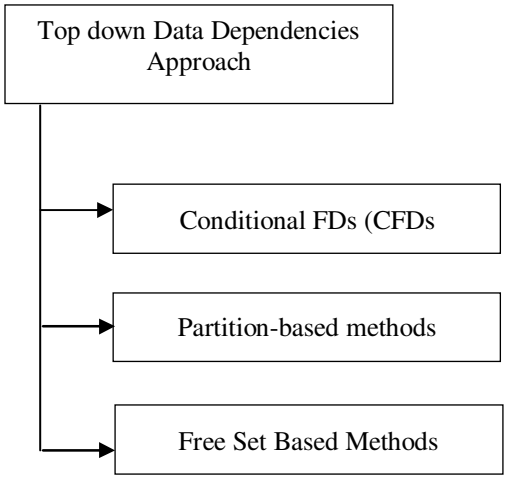

Figure 3 Top down approach based methods

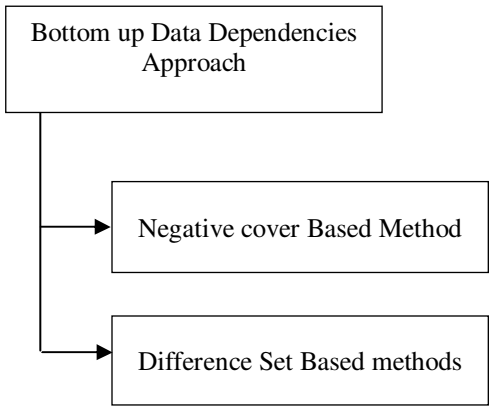

Figure 4 Bottom Up approach based methods

3. Literature Review

In 2008 KatalinTundeJanosiRancz&VioricaVarga proposed "A Method for Mining Functional Dependencies in Relational Database Design Using Formal Concept Analysis".They presented a new methodto optimize and extend a previous research on FCA and databases, byanalyzing the functional dependencies in order to correctly build databaseschemata. They intended to mine functional dependencies in a relational database table. The novelty of proposed method is that it builds invertedindex files in order to optimize the construction of the formal context offunctional dependencies. Proposed method provides FCA visualization which makes easier to managedatabase schema and normal forms[1].

In 2008 Jalal Atoum,Dojanah Bader and Arafat Awajanproposed "Mining Functional Dependency from Relational DatabasesUsing Equivalent Classes and Minimal Cover". They proposed a new algorithm called FD_Discover for discoveringFunctional Dependencies (FDs) from databases. Proposed algorithm employs some concepts from relationaldatabases design theory specifically the concepts of equivalences and the minimal cover. The suggested new algorithm(FD_Discover) to discover FDs utilizes the concepts of equivalent properties and minimal(Canonical) cover of FDs[2].

In 2009 Fabien De Marchi proposed "CLIM:CLosed Inclusion dependency mining in databases". They proposed INclusion Dependency (IND) miningwhich is a classical data mining problem, with many applicationsin databases and data analysis. They optimize aclosure operator, as it is done for support-based pattern mining. They show that IND mining problem can besolved by closed set mining,which is a new concept and has benefices over existing methods[3].

In 2010 Jixue Liu, Jiuyong Li Chengfei&Liu Yongfeng Chen proposed "Discover Dependencies from Data - A Review".They proposed a study over functional and inclusion dependency and show that functional dependency discovery is important to knowledgediscovery, database semantics analysis, database design, and dataquality assessment. They gives details about functional dependency, conditionalfunctional dependency, approximate functional dependencyand inclusion dependency discovery and XML functional dependencies[4].

In 2011 Wenfei Fan , Floris Geerts , Jianzhong Li , Ming Xiong "Discovering Conditional Functional Dependencies" They investigates the discovery of conditional functional dependencies (CFDs) and show that CFDs are a recent extension of functional dependencies

(FDs) by supporting patterns of semantically related constants, and can be used as rules for cleaning relational data. They provide three methods for CFD discovery, CFD Miner, CTANE, and Fast CFD[5].

In2012 ThiernoDiallo&JeanMarcPetitproposed "Discovering Editing Rules For Data Cleaning"They proposed a pattern mining techniques for discovering eRs from existing source relations (possibly dirty) with respect to masterrelations (supposed to be clean and accurate).They proposed a new semantics of eRs taking advantage ofboth source and master data.Theproposed techniques address the discovery problem of eRs and heuristics to clean data[6]. In 2013 Sujoy Dutta proposed "Mining Full Functional Dependency toAnswer Null Queries and Reduce Imprecise Information Based on Fuzzy Object Oriented Databases". They proposed new concepts of fuzzy functional dependency and extended to fullfunctional dependency on similarity based fuzzy object oriented data model. In addition they also proposed a data miningalgorithm discover all functional dependencies among attributes [7].

In 2014 P.Andrew, J.Anishkumar& S.Balamurugan, proposed "Investigations on Methods Developed forEffective Discovery of FunctionalDependencies"They gives details about various methods to discover functional dependencies from data and effectivepruning for the discovery of conditional functional dependencies.They also showed that Functional dependencies and Fast FDs a heuristic-driven, Depth-first algorithm for mining FD from relation instances areelaborated[8]. In 2015 Thorsten Papenbrock and Jens Ehrlich proposed "Functional Dependency Discovery:An Experimental Evaluation of Seven Algorithms".They describe, evaluate, and compare theseven most cited and most important algorithms on this same problem. They classify the algorithms into three different categories,explaining their commonalities. They also describe allalgorithms with their main ideas. The descriptions provideadditional details [9].

4. Comparative Study

We proposed a comparative study based on technique used in the algorithm.

Table 1 Comparision based on techniques

Algorithm Name	Techniques used
TANE	Used data Based partition concepts and cardinality
FD_Mine	Used data Based partition and Equivalent class
Dep-Miner	Used Difference Set based concepts
FD_Discover	Used concepts of Equivalent Classes and Minimal Cove

5. Conclusion

Normalization play important role in Data base design and provides several advantage to remove redundancy in the database. Data dependency helps in nomination. We present a study over recently proposed data dependency discovery algorithms. Each and every algorithms use a different concepts for efficiently discover data dependency. Without Identify correct data dependency we cannot design a good database and improve anomalies. Some New Data dependency is discovered like conditional functional dependency, Inclusion Dependency for XML data.

References

[1] KatalinTundeJanosiRancz And VioricaVarga "A Method For Mining Functional Dependencies In Relational Database Design Using FCA"Studia Univ. Babes Volume LIII, Number 1, 2008

[2] Jalal Atoum, Dojanah Bader and Arafat Awajan"Mining Functional Dependency from Relational DatabasesUsing Equivalent Classes and Minimal Cover" Journal of Computer Science 4 (6): 421-426, 2008

[3] Fabien De Marchi"CLIM:CLosed Inclusion dependency Mining in databases" work has been partially Funded by the French National ResearchAgency DEFIS 2009

[4] Jixue Liu Jiuyong Li&Chengfei Liu "Discover Dependencies from Data - A Review" School of Management, Xian University of Architecture November 8, 2010

[5] Wenfei Fan , Floris Geerts , Jianzhong Li , Ming Xiong "Discovering Conditional Functional Dependencies" IEEE Transactions on knowledge and data Engineering, Vol.23, No. 5, May 2011

[6] Thierno Diallo &Jean Marc Petit "Discovering Editing Rules For Data Cleaning" article was presented at the 9th InternationalWorkshop on Quality in Databases (QDB) 2012.

[7] Sujoy Dutta "Mining Full Functional Dependency to Answer Null Queries and Reduce Imprecise Information Based on Fuzzy ObjectOriented Databases" (IJCSET) Vol. 4 No. 03 Mar 2013

[8] P.Andrew, J.Anishkumar& S. Balamurugan "Investigations on Methods Developed for Effective Discovery of FunctionalDependencies" IJIRCCE (An ISO 3297: 2007 Certified Organization)Vol. 3, Issue 2, February 2015

[9] Thorsten Papenbrock Jens& Ehrlich Jannik Marten Functional Dependency Discovery:An Experimental Evaluation of Seven Algorithms Proceedings of the VLDB Endowment, Vol. 8, No. 10 2015