

# A Survey on Strategies Developed for Mining Functional Dependencies

P.Andrew<sup>1</sup>, J.Anishkumar<sup>2</sup>, Prof.S.Balamurugan<sup>3</sup>, S.Charanyaa<sup>4</sup>

Department of IT, Kalaignar Karunanidhi Institute of Technology, Coimbatore, TamilNadu, India<sup>1,2,3</sup>

Senior Software Engineer Mainframe Technologies Former, Larsen & Tubro (L&T) Infotech, Chennai TamilNadu, India<sup>4</sup>

**ABSTRACT:** This paper details about various methods prevailing in literature on strategies developed for mining functional dependencies. Efficient discovery of functional and approximate dependencies of partition is discussed in detail. An efficient algorithm for discovery functional and approximate dependencies called TANE is examined. Mining functional dependencies from data and Depth first algorithms and influencing for AFD mining are examined in detail. Polynomial time queries over inconsistent databases with functional dependency and foreign keys are reviewed. This survey would promote a lot of research in the area of mining functional dependencies.

**KEYWORDS:** Functional Dependency, Depth-First Algorithm, Foreign Key, Data Anonymization, Information Mining.

## I. INTRODUCTION

Several authors have undergone many survey and created their algorithms in securing moving object data. The main objective is that is to hide the spatio-temporal data to the anonymous user. The major reason is that to protect the isolated data. Since data privacy preserving is been our major goal we can do various research in this field. Ghinita et al. [44][2009] considered two conceal mechanisms in which an adversary background knowledge of maximum speed to infer more specific location information. Based on the velocity of the movement of the object the attacker could assume that the object is under a particular movement as an example if Alice is walking on the road based on her velocity of speed the attacker could find that Alice is waling if she is driving a car the attacker could find that she is driving. So based on the velocity based adversary knowledge the trajectory data could be predicted. So Ghinita considered two types of attacks: (1) the initiative without back ground information more on sensitive location map. (2) The initiative with such background information. In the first case, the privacy requirement is not allow to an attacker to diagnose the user location in the sub-region in the pretext region. In the second case, the privacy requirement direct that the probability of association between the user and the location. She consider two types of transformation on trajectory databases, temporal and spatial cloaking. The author planned two alternatives in achieving temporal cloaking: request deferral and postdating. The space and time error are generated in the temporal cloaking and spatial cloaking. In this for the low velocity the request are safe but in the case that the velocity increase the request are deferred/postdated. As the velocity increases the request for temporal data should be processed while moving. Correlation-based adversary knowledge: In this data publishing most of the attacker uses this correlation-based adversary knowledge because the attacker attain by correlating the timestamps of the user. The attacker finds the highest probability of the user location by forward motion and the backward motion model during a period of time the attacker correlates the user location by their movement. Jin et al [45]. implemented two protocols for publishing the data. In the first they cluster the location of the users and then the data is published if the forward breach probability is below the user-defined threshold. Then they recluster the data at a period of time and find out the backward breach probability, by correlating forward breach probability and backward breach probability the data is published if the data does not reach the user-defined threshold. But these approaches are more effective to defend against the attacks but not to stop it.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2015

## II. EFFICIENT DISCOVERY OF FUNCTIONAL AND APPROXIMATE DEPENDENCIES G PARTITION

In this paper, the author gave a new approach to find FDs, with partitioning the sets of rows with respect to their attribute values. These partitioning will inturn make us easy and very efficient to work with it and the rows also can be identified easily. The efficient in practice is a new algorithm which is been used in these experiments. By this algorithm, the author shown that the running time is been increased in the several orders of magnitude over the previously published result and also the author said that these algorithm will be applicable also for a larger databases.

## III. TANE: AN EFFICIENT ALGORITHM FOR DISCOVERY FUNCTIONAL AND APPROXIMATE DEPENDENCIES

In this paper, the author defines about the discovery of FDs, which uses an important database analysis technique called TANE, an efficient algorithm for finding FD's from large database. TANE is an algorithm which is based on partitioning or splitting the rows that will inturn decide the validity of the FD's. This will make easier and an efficient way while the same thing dealing with FDs. In this paper, the experimental result shows that the algorithm TANE is fast interms of practice. For a bench mark DB the running times are improved by several magnitudes among the previous works. The algorithm which was defined is also useful in the case of larger datasets.

## IV. MINING FUNCTIONAL DEPENDENCIES FROM DATA

In this paper, the author proposes an efficient rule discovery algorithm called FD-Mine which is used for mining functional dependencies from data. The author identified the equivalence among the attributes by using Armstrong's axioms for FDs that is used to decrease the size of the dataset and the number of FDs to be checked. The author initially described four effective pruning rules which inturn reduces the size of space allocated to search i.e., search space. The number of FDs that is to be verified is decreased by skipping the search for FDs that are implied logically by FDs which is been discovered already. The author also presented the FD-Mine algorithm that incorporates the four pruning rules to the mining process. The FD-Mine algorithm described in this paper, shows that the pruning does not inturn will loss the useful information. In this paper, the author put forth the future work can also done to extend the FD-Mine algorithm in order to discover other data dependencies like multi-valued dependencies and conditional dependencies . And the FD-Mine approach may be applied to the data cleaning phase of datamining, inturn to reduce the size of a relation without losing any useful information.

## V. DEPTH FIRST ALGORITHMS AND INFLUENCING FOR AFD MINING

In this paper, the author mentioned that the task of searching a powerset lattice is part of many problem domains. The algorithm Approximate Functional Dependencies (AFDs) is used for finding lattices which uses bottom-up breadth first search (BU-BFS). In this paper, the author introduces a part of the MOLS framework which is an method used for managing and utilizing gathered information about the status of rules in the lattice. These gather information has some limited agility with BU-BFS. The ability of depth first search (DFS) algorithm is used in order to infere the information is also been explored in this paper. The paper focus on algorithms that is motivated by the idea which improves performance from an algorithm is more likely to persist as users make different approximation measured. In this paper, the author presented a machine independent evaluation where difference in the performance can be attributed to the algorithm directly.

## VI. POLYNOMIAL TIME QUERIES OVER INCONSISTENT DATABASES WITH FUNCTIONAL DEPENDENCY AND FOREIGN KEYS

In this paper, the author addressed the problem of efficient computing consistent answers to queries over relational databases that may be inconsistent to FDs and foreign key constraints which is obtained from repaired databases repair strategy. The author considered a particular set of FD known as canonical and a repair strategy that has only tuple updates and insertions are also allowed to restore the consistency. The tuples can be inserted when the foreign key are validated. The author proposed an approach that allows us to obtain a unique repaired database that may be computed

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2015

in polynomial time. The author also identified the classes of constraints that has a consistent answers to particular classes of conjunctive queries that can be computed in polynomial time.

SQL> select \* from employ;

TID	ID	NAME	JYEAR	WAGES	SUPUSR
1	E1	AAAAA	2001	5000	Carol
2	E2	BBBBB	2005	4000	Alex
3	E3	CCCCC	2009	2500	Blent
4	E4	DDDDD	2013	1000	King

SQL> select \* from tb\_edetails;

TID	PH_NUMBER	ADDRESS	GENDER
1	7788994455	Alexion Road	Male
2	8877994455	Carol Road	Female
3	9988774455	Red wood road	Male
4	9977884455	Rosewood Road	Female

SQL> |

In these tables, TID is assigned as primary key in **EMPLOY** table and TID is used as foreign key in **TB\_EDETAILS** table. Using this tables and primary key constraints we can identify the FD's in the table through the java program.

## VII. CONCLUSION AND FUTURE WORK

This paper detailed about various methods prevailing in literature on strategies developed for mining functional dependencies. Efficient discovery of functional and approximate dependencies of partition is discussed in detail. An efficient algorithm for discovery functional and approximate dependencies called TANE is examined. Mining functional dependencies from data and Depth first algorithms and influencing for AFD mining are examined in detail. Polynomial time queries over inconsistent databases with functional dependency and foreign keys are reviewed. This survey would promote a lot of research in the area of mining functional dependencies.

## REFERENCES

1. Shaoxu Song, Lei Chen, "Efficient discovery of similarity constraints for matching dependencies", Data & Knowledge Engineering, Elsevier, 2013.
2. Pado Atzgeni, Valeria D. Antonellis, Relational Database Theory, The Benjamin/Cummings Publishing Company, Inc., 1993
3. Stavros S. Cosmadakis, Paris C. Kanellakis, Nicolas Spyrtatos, Partition semantics for relations, PODS, 1985, pp. 261–275.
4. Jeremy T. Engle, Edward L. Robertson, Depth first algorithms and inferring for AFD mining, IDEAS, 2009, pp. 54–65.
5. Wenfei Fan, Floris Geerts, Xibei Jia, Anastasios Kementsietsidis, Conditional functional dependencies for capturing data inconsistencies, TODS 33(2) (6) (2008) 1–48.
6. Wenfei Fan, Floris Geerts, Jianzhong Li, Ming Xiong, Discovering conditional functional dependencies, TKDE, 2010.
7. Sergio Flesca, Filippo Furfaro, Sergio Greco, Ester Zumpano, Repairing inconsistent xml data with functional dependencies, Encyclopedia of Database Technologies and Applications (2005) 542–547.
8. Chris Giannella, Edward Robertson, On approximation measures for functional dependencies, Information Systems 29 (6) (2004) 483–507.
9. Yka Huhtala, Juha Karkkainen, Pasi Porkka, Hannu Toivonen, Efficient discovery of functional and approximate dependencies using partitions, ICDE, 1998.
10. Yka Huhtala, Juha Karkkainen, Pasi Porkka, Hannu Toivonen, Tane: an efficient algorithm for discovering functional and approximate dependencies, The Computer Journal 42 (2) (1999) 100–111.
11. Ronald S. King, James Oil, Discovery of functional and approximate functional dependencies in relational databases, Journal of Applied Mathematics and Decision Sciences 7 (1) (2003) 49–59.
12. Jyrki Kivinen, Heikki Mannila, Approximate dependency inference from relations, LNCS 646 — Database Theory ICDT '92, 1992, pp. 86–98.
13. Tony T. Lee, Tong Ye, A relational approach to functional decomposition of logic circuits, TODS 36(2) (13) (2011) 1–13, (30).
14. Jiuyong Li, Jixue Liu, Hannu Toivonen, Jianming Yong, Effective pruning for the discovery of conditional functional dependencies, The Computer Journal (2012), (<http://comjnl.oxfordjournals.org/content/early/2012/06/23/comjnl.bxs082.short>).
15. Jixue Liu, Chengfei Liu, Jiuyong Li, Yongfeng Chen, Discover dependencies from data — a review, TKDE 24 (2) (2012) 251–264, (<http://www.computer.org/portal/web/csdl/doi/10.1109/TKDE.2010.197>).
16. Stephane Lopes, Jean-Marc Petit, Lotfi Lakhil, Efficient discovery of functional dependencies and Armstrong relations, LNCS 1777 — 7th International Conference on Extending Database Technology (EDBT): Advances in Database Technology, 1777, 2000, pp. 350–364.
17. David Maier, The Theory of Relational Databases, Computer Science Press, 1983. (<http://web.cecs.pdx.edu/~maier/TheoryBook/TRD.html>).

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2015

18. Heikki Mannila, Kari-Jouko Rih, On the complexity of inferring functional dependencies, *Discrete Applied Mathematics* 40 (1992) 237–243.
19. Heikki Mannila, Kari-Jouko Rih, Algorithms for inferring functional dependencies from relations, *Data and Knowledge Engineering* 12 (1) (1994) 83–99.
20. Victor Matos, Becky Grasser, Sql-based discovery of exact and approximate functional dependencies, *Proceeding ITiCSE-WGR '04 Working Group Reports from ITiCSE on Innovation and Technology in Computer Science Education, 2004*, pp. 58–63.
21. Cristian Molinaro, Sergio Greco, Polynomial time queries over inconsistent databases with functional dependencies and foreign keys, *DKE* 69 (7) (2010) 709–722.
22. Noel Novelli, Rosine Cicchetti, Fun: an efficient algorithm for mining functional and embedded dependencies, *ICDT, 2001*, pp. 189–203.
23. Noel Novelli, Rosine Cicchetti, Functional and embedded dependency inference: a data mining point of view, *Information Systems* 26 (7) (2001) 477–506.
24. Iztok Savnik, Peter A. Flach, Bottom-up induction of functional dependencies from relations, *AAAI Workshop of KDD, 1993*, pp. 174–185.
25. Hui Wang, Ruilin Liu, Privacy-preserving publishing microdata with full functional dependencies, *DKE* 70 (3) (2011).
26. Catharine Wyss, Chris Giannella, Edward Robertson, Fastfids: a heuristic-driven, depth-first algorithm for mining functional dependencies from relation instances — extended abstract, *DaWaK, 2001*, pp. 101–110.
27. Hong Yao, Howard J. Hamilton, Mining functional dependencies from data, *Journal of Data Mining and Knowledge Discovery* 16 (2) (2008) 197–219.
28. B.Powmeya , Nikita Mary Ablett ,V.Mohanapriya,S.Balamurugan,"An Object Oriented approach to Model the secure Health care Database systems,"In proceedings of International conference on computer , communication & signal processing(IC<sup>3</sup>SP)in association with IETE students forum and the society of digital information and wireless communication,SDIWC,2011,pp.2-3
29. Balamurugan Shanmugam, Visalakshi Palaniswami, "Modified Partitioning Algorithm for Privacy Preservation in Microdata Publishing with Full Functional Dependencies", *Australian Journal of Basic and Applied Sciences*, 7(8): pp.316-323, July 2013
30. Balamurugan Shanmugam, Visalakshi Palaniswami, R.Santhya, R.S.Venkatesh "Strategies for Privacy Preserving Publishing of Functionally Dependent Sensitive Data: A State-of-the-Art-Survey", *Australian Journal of Basic and Applied Sciences*, 8(15) September 2014.
31. S.Balamurugan, P.Visalakshi, V.M.Prabhakaran, S.Chraryaa, S.Sankaranarayanan, "Strategies for Solving the NP-Hard Workflow Scheduling Problems in Cloud Computing Environments", *Australian Journal of Basic and Applied Sciences*, 8(15) October 2014.
32. Charanyaa, S., et. al., , A Survey on Attack Prevention and Handling Strategies in Graph Based Data Anonymization. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(10): 5722-5728, 2013.
33. Charanyaa, S., et. al., Certain Investigations on Approaches forProtecting Graph Privacy in Data Anonymization. *International Journal of Advanced Research in Computer and Communication Engineering*, 1(8): 5722-5728, 2013.
34. Charanyaa, S., et. al., Proposing a Novel Synergized K-Degree L-Diversity T-Closeness Model for Graph Based Data Anonymization. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(3): 3554-3561, 2014.
35. Charanyaa, S., et. al., , Strategies for Knowledge Based Attack Detection in Graphical Data Anonymization. *International Journal of Advanced Research in Computer and Communication Engineering*, 3(2): 5722-5728, 2014.
36. Charanyaa, S., et. al., Term Frequency Based Sequence Generation Algorithm for Graph Based Data Anonymization *International Journal of Innovative Research in Computer and Communication Engineering*, 2(2): 3033-3040, 2014.
37. V.M.Prabhakaran, Prof.S.Balamurugan, S.Charanyaa," Certain Investigations on Strategies for Protecting Medical Data in Cloud", *International Journal of Innovative Research in Computer and Communication Engineering Vol 2, Issue 10, October 2014*
38. V.M.Prabhakaran, Prof.S.Balamurugan, S.Charanyaa," Investigations on Remote Virtual Machine to Secure Lifetime PHR in Cloud ", *International Journal of Innovative Research in Computer and Communication Engineering Vol 2, Issue 10, October 2014*
39. V.M.Prabhakaran, Prof.S.Balamurugan, S.Charanyaa," Privacy Preserving Personal Health Care Data in Cloud" , *International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 2, October 2014*
40. P.Andrew, J.Anish Kumar, R.Santhya, Prof.S.Balamurugan, S.Charanyaa, "Investigations on Evolution of Strategies to Preserve Privacy of Moving Data Objects" *International Journal of Innovative Research in Computer and Communication Engineering*, 2(2): 3033-3040, 2014.
41. P.Andrew, J.Anish Kumar, R.Santhya, Prof.S.Balamurugan, S.Charanyaa, " Certain Investigations on Securing Moving Data Objects" *International Journal of Innovative Research in Computer and Communication Engineering*, 2(2): 3033-3040, 2014.
42. P.Andrew, J.Anish Kumar, R.Santhya, Prof.S.Balamurugan, S.Charanyaa, " Survey on Approaches Developed for Preserving Privacy of Data Objects" *International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 2, October 2014*
43. S.Jeevitha, R.Santhya, Prof.S.Balamurugan, S.Charanyaa, " Privacy Preserving Personal Health Care Data in Cloud" *International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 2, October 2014*.
44. K.Deepika, P.Andrew, R.Santhya, S.Balamurugan, S.Charanyaa, "Investigations on Methods Evolved for Protecting Sensitive Data", *International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 4, Decermber 2014*.
45. K.Deepika, P.Andrew, R.Santhya, S.Balamurugan, S.Charanyaa, "A Survey on Approaches Developed for Data Anonymization", *International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 4, Decermber 2014*.
46. S.Balamurugan, S.Charanyaa, "Principles of Social Network Data Security" LAP Verlag, Germany, ISBN: 978-3-659-61207-7, 2014
47. S.Balamurugan, S.Charanyaa, "Principles of Scheduling in Cloud Computing" Scholars' Press, Germany,, ISBN: 978-3-639-66950-3, 2014
48. S.Balamurugan, S.Charanyaa, "Principles of Database Security" Scholars' Press, Germany, ISBN: 978-3-639-76030-9, 2014

## BIOGRAPHY

**P.Andrew and J.Anishkumar** are currently pursuing their B.Tech. degree in Information Technology at Kalaigarnarunaidhi Institute of Technology, Coimbatore, Tamil Nadu, India. Their areas of research interests include Network Security, Cloud Computing and Database Security.

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2015



**Prof.S.Balamurugan** obtained his B.Tech degree in Information Technology from P.S.G. College of Technology, Coimbatore, Tamil Nadu, India and M.Tech degree in Information Technology from Anna University, Tamil Nadu, India respectively. He is currently working towards his PhD degree in Information Technology at P.S.G. College of Technology, Tamil Nadu, India. At present he holds to his credit **65 papers International Journals and IEEE/ Elsevier International Conferences**. He is currently working as Assistant Professor in the Department of Information Technology, Kalaignar Karunanidhi Institute of Technology, Coimbatore, Tamil Nadu, India affiliated to Anna University TamilNadu, India. He is **State Rank holder** in schooling. He was **University First Rank holder** M.Tech. Semester Examinations at Anna University, Tamilnadu, India. He served as a Joint Secretary of IT Association, Department of Information Technology,

PSG College of Technology, Coimbatore, Tamilnadu, India. He is the **recipient of gold medal and certificate of merit** for best journal publication by his host institution **consecutively for 3 years**. Some of his professional activities include invited Session Chair Person for two Conferences. He has guided 16 B.Tech projects and 2 M.Tech. projects. He has won a best paper award in International Conference. His areas of research interest accumulate in the areas of Data Privacy, Database Security, Object Modeling Techniques, and Cloud Computing. He is a life member of ISTE,CSI. **He has authored a chapter in an International Book "Information Processing" published by I.K. International Publishing House Pvt. Ltd, New Delhi, India, 978-81-906942-4-7. He is the author of 3 books titled "Principles of Social Network Data Security", ISBN: 978-3-659-61207-7, "Principles of Scheduling in Cloud Computing" ISBN: 978-3-639-66950-3, and "Principles of Database Security", ISBN: 978-3-639-76030-9.**



**S.Charanyaa** obtained her **B.Tech** degree in Information Technology and her **M.Tech** degree in Information Technology from Anna University Chennai, Tamil Nadu, India. She was **gold medalist** in her B.Tech. degree program. She has to her credit **27 publications in various International Journals and Conferences**. Some of her outstanding achievements at school level include **School First Rank holder in 10<sup>th</sup> and 12<sup>th</sup> grade**. She was working as Software Engineer at Larsen & Turbo Infotech, Chennai for 3 years where she got promoted as Senior Software Engineer and worked for another 2 years. She worked at different verticals and worked at many places including Denmark, Amsderdam handling versatile clients. She is also the recipient of **best team player award for the year 2012 by L&T**. Her areas of research interest accumulate in the areas of

Database Security, Privacy Preserving Database, Object Modeling Techniques, and Cloud Computing. **She is the author of 3 books titled "Principles of Social Network Data Security", ISBN: 978-3-659-61207-7, "Principles of Scheduling in Cloud Computing" ISBN: 978-3-639-66950-3, and "Principles of Database Security", ISBN: 978-3-639-76030-9.**