



# **Investigations on Methods Developed for Effective Discovery of 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 to discover functional dependencies from data. Effective pruning for the discovery of conditional functional dependencies is discussed in detail. Di conditional Functional Dependencies and Fast FDs a heuristic-driven, Depth-first algorithm for mining FD from relation instances are elaborated. Privacy preserving publishing micro data with Full Functional Dependencies and Conditional functional dependencies for capturing data inconsistencies are examined. The approximation measures for functional dependencies and the complexity of inferring functional dependencies are also observed. Compression - Based Evaluation of partial determinations is portrayed. This survey would promote a lot of research in the area of mining functional dependencies from data.

**KEYWORDS:** Effective Pruning, Conditional Functional Dependency (DFD), Mining, Data Anonymization, Similarity Constraints.

## **I. INTRODUCTION**

Now-a-days we can note many spreading usage of location –aware devices such as many GSM mobile phones, GPS enabled PDA's, location sensors, and active RFID tags. Due to this device usage scenario, the device generate a large collection of moving data objects with the help of trajectory data, all these data are used for various data identification and analysis process. For instance consider traffic control, one can hack the control unit of traffic control management. Therefore it is way clear that a hacker may collects many temporal data to cover sensational messages of an organization and especially he/she can discover many personal information of third party/check points of many premises. Typically personal data (data privacy) are been fetched. Due to user's identity replacement which is actual like terminal i.e. QID is a moving data are linked to external information to re-identify individual existence, thus the attacker can be able to track and trace the anonymous moving objects back into individuals. Even though the location privacy has already been accepted as an important problem and effective privacy-preserving solutions to publish the trajectories data. These trajectories data might be defined by user itself and by data mining the databases. In this world's technology for positioning systems, the location of the trajectories data can be predicted very accurately. The location data can be obtaining through the score pairs i.e. longitude and latitude. The location can also be finding out by QIDs by identifying the frequent mining pattern technique. The QID mining looks for the frequently mined pattern and correlated with the threshold defined by the user. Even though privacy has been protected there are few open problems the two fundamental that are taken as objectives of our project:

1. Identifying secured moving data objects with high probability (Granularities of QID Location)
2. Quick & Efficient discovery of QID of moving data objects

## **II. EFFECTIVE PRUNING FOR THE DISCOVERY OF CONDITIONAL FUNCTIONAL DEPENDENCIES**

In this paper, the author describes about the extension of traditional dependency called CFD which is to detect and repair the inconsistent data. According to review, CFDs render 100% confidence association rules. Authors



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

proposed an algorithm for pruning criteria and to prune these search space. The author also proposed an algorithm based on number of medium to large datasets which is faster than consistent CFD and exhibits a linear time performance in the size of dataset.

The pruning algorithm is to prune or remove the search space, unnecessary closures and generators. This pruning a criteria was evaluated on read data sets and the proposed algorithm is faster than the constant CFD. The author also showed how Chi Square can be used to measure the interestingness of CFD's.

### III. DI CONDITIONAL FUNCTIONAL DEPENDENCIES

In this paper, authors demonstrated about the CFD which is an extension of functional dependencies. The finding of the quality of CFD is a complex task which involves manual effort. For identifying the object, the constant CFD plays an important role. For this discovery problem CFD had introduced new challenges. The author provides these three methods for discovery.

First method, CFD miner for mining closed item set and is utilized to discover constant CFD. It is also used for identification of objects that is essential to data cleaning and integration.

Secondly method, CTANE is a level wise algorithm which is popular in mining FDs.

Third method, Fast CFD is based on the depth-first approach using in Fast FD.

It chooses the item set which is nearer to the attribute in order to reduce the search space.

By comparing the above mentioned algorithm, CFD miner is faster than the fast CFDs and CTANE for constant CFD discovery CTANE does not scale the duty of relation but Fast CFD can do this. This algorithm provides a cleaning tool for the users to choose various different applications.

### IV. FAST FDS: A HEURISTIC-DRIVEN, DEPTH-FIRST ALGORITHM FOR MINING FD FROM RELATION INSTANCES

In this paper, the author presented computing minimal FDs from different sets with the help of heuristic-driven, depth-first search. The author indicated that Fast FDs is ambitious for each of the following classes of standard related relation instances,

- (i) Random integer-valued instances of varying correlation factors.
- (ii) Random Bernoulli instances,
- (iii) Real-life ML repository relation instances.
- (iv) The experiments in this paper that for a huge relation Fast FDs are convincingly better for are classes in the case of inherent space efficiency of the depth-first search method.

The heuristic-based, depth-first search methods are solution to Artificial intelligence (AI) problems. The author also suggested us to do the future work is to consider the incremental dependency inference problem.

### V. PRIVACY RESERVING PUBLISHING MICRO DATA WITH FFD

In this paper, the author explained about the data publishing which has becomes a problem towards individual privacy. Recent research informs that the different background knowledge has made the threats to the privacy of published data. In this paper, the author brings out a study of privacy threat from FFD which is utilized as a part of adversary knowledge. There are several existing anonymizations principle is k-anonymity, l-diversity etc... to prevent against an FFD-based to prevent against an FFD-based privacy attacks but none of them does it. So, the author formalize the FFD-based privacy attack and also explained the privacy model, (d,l) influence to contest the FD-based attack and it is also exhibited by experimental study.

### VI. CONDITIONAL FUNCTIONAL DEPENDENCIES FOR CAPTURING DATA INCONSISTENCIES

In this paper, the author proposed that in variation to the traditional functional dependencies (FDs), the conditional functional dependencies are mainly utilized or used to design schema and the conditional functional



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

dependencies is to control the release of information by some relates data's . In this paper, the consistency problem for CFDs is NP-complete and the implication problem for CFDs is COND-complete.

## VII. ON APPROXIMATION MEASURES FOR FUNCTIONAL DEPENDENCIES

In this paper, the author examined the issue of how to measure the degree to which a functional dependency is relative.

The initial motivation lies in the fact that relative functional dependencies perform potentially interesting patterns existent in a table. This kind of identification is a valuable data mining problem.

Firstly, the author developed an approximation measure by axiomatizing the degree to which  $X \rightarrow Y$  is relative. It is also proven that a unique unnormalized measure satisfies these axioms up to a multiplicative constant.

## VIII. ON THE COMPLEXITY OF INFERRING FUNCTIONAL DEPENDENCIES

In this paper, the author stated that the dependency inference problem is to identify a cover for the set of FDs which hold in a given relation. The author said that the problem has applications in relational database design and in query optimization. The author showed that this problem is resolved using a Brute-force algorithm in time for a relations with row  $S()$  attributes(n).

## IX. COMPRESSION - BASED EVALUATION OF PARTIAL DETERMINATIONS

In this paper , the author determining the problem of partial determination and the compression based method are used to evaluate the above problem . This is viewed as generalizations of both FD and association rules . It will extending the measures used for evaluating the support and confidence .

Partial determinations are generalizations of functional dependencies . It can be expressed as  $X \rightarrow dY$  . Where  $d$  is the number . The set of  $x$  will be referred as LHS , and  $Y$  will be referred as RHS . The partial determination is used for both the  $X \rightarrow dY$  and  $pdx \rightarrow dY$  .

The future work of this plan is to extending with other strategies like genetic algorithms and combinations of search algorithm .The new compression – based measures are used to evaluate the partial determination and this is used for the search . This partial determination is a useful form of KDD since it is more expressive .The other measures of the partial determination is MDC based functions . This will avoiding the over fitting the data .

## X. CONCLUSION AND FUTURE WORK

This paper detailed about various methods to discover functional dependencies from data. Effective pruning for the discovery of conditional functional dependencies is discussed in detail. Di conditional Functional Dependencies and Fast FDs a heuristic-driven, Depth-first algorithm for mining FD from relation instances are elaborated. Privacy preserving publishing micro data with Full Functional Dependencies and Conditional functional dependencies for capturing data inconsistencies are examined. The approximation measures for functional dependencies and the complexity of inferring functional dependencies are also observed. Compression - Based Evaluation of partial determinations is portrayed. This survey would promote a lot of research in the area of mining functional dependencies from data.

## 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 Spyrtos, Partition semantics for relations, PODS, 1985, pp. 261–275.
4. Jeremy T. Engle, Edward L. Robertson, Depth first algorithms and inferencing 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.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

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 ICDE '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 Lakhal, 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>).
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.Chranyaa, 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 for Protecting 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.



ISSN(Online): 2320-9801  
ISSN(Print): 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

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, December 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, December 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 KalaingarKarunanidhi Institute of Technology, Coimbatore, Tamil Nadu, India. Their areas of research interests include Network Security, Cloud Computing and Database Security.



**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, Kalaingar 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.**