Missing Values Estimation for Skylines in Incomplete Database

By: Alwan, A (Alwan, Ali)[1; Ibrahim, H (Ibrahim, Hamidah)[2]; Udzir, N (Udzir, Nurizura)[2]; Sidi, F (Sidi, Fatimah)[2]

INTERNATIONAL ARAB JOURNAL OF INFORMATION TECHNOLOGY
Volume: 15 Issue: 1 Pages: 66-75
Published: JAN 2018
Document Type: Article

Abstract
Incompleteness of data is a common problem in many databases including web heterogeneous databases, multi-relational databases, spatial and temporal databases, and data integration. The incompleteness of data introduces challenges in processing queries as providing accurate results that best meet the query conditions over incomplete database is not a trivial task. Several techniques have been proposed to process queries in incomplete database. Some of these techniques retrieve the query results based on the existing values rather than estimating the missing values. Such techniques are undesirable in many cases as the dimensions with missing values might be the important dimensions of the user’s query. Besides, the output is incomplete and might not satisfy the user preferences. In this paper we propose an approach that estimates missing values in skylines to guide users in selecting the most appropriate skylines from the several candidate skylines. The approach utilizes the concept of mining attribute correlations to generate an Approximate Functional Dependencies (AFDs) that captured the relationships between the dimensions. Besides, identify the strength of probability correlations to estimate the values. Then, the skylines with estimated values are ranked. By doing so, we ensure that the retrieved skylines are in the order of their estimated precision.

Keywords
Author Keywords: Skyline Queries; Preference Queries; Incomplete Database; Query Processing; Estimating Missing Values

Author Information
Reprint Address: Alwan, A (reprint author)

Organization-Enhanced Name(s)
International Islamic University Malaysia

Addresses:
[1] Int Islamic Univ Malaysia, Kulliyyah Informat & Commun Technol, Kuala Lumpur, Selangor, Malaysia


Publisher
ZARKA PRIVATE UNIV, COLL COMPUTING & INFORMATION SOC, PO BOX 132222, ZARQA, 13132, JORDAN

Journal Information
Impact Factor: Journal Citation Reports

Categories/Classification
Research Areas: Computer Science, Engineering
Cited References: 37

Showing 30 of 37  View All in Cited References page

1. From complete to incomplete information and back
   By: Antova, L.; Koch, C.; Olteanu, D.
   SIGMOD Pages: 713-724 Published: 2007
   Times Cited: 9

2. 10(106)) worlds and beyond: efficient representation and processing of incomplete information
   By: Antova, Iyublena; Koch, Christop; Olteanu, Dan
   VLDB JOURNAL Volume: 18 Issue: 5 Pages: 1021-1040 Published: OCT 2009
   Times Cited: 16

3. Salsa: computing the Skyline without scanning the whole sky
   By: Bartolini, Ilaria; Ciaccia, Paolo; Patella, Marco.
   CIKM Pages: 405-414 Published: 2006
   Times Cited: 15

4. An analysis of four missing data treatment methods for supervised learning
   By: Batista, GEAPA; Monard, MC
   APPLIED ARTIFICIAL INTELLIGENCE Volume: 17 Issue: 5-6 Pages: 519-533 Published: MAY-JUN 2003
   Times Cited: 179

5. Fast high-dimensional data search in incomplete databases
   By: Beng Chin Ooi; Cheng Hian Goh; Kian-Lee Tan
   Proceedings of the Twenty-Fourth International Conference on Very-Large Databases Pages: 357-67 Published: 1998
   Times Cited: 2

6. Finding Skylines for Incomplete Data
   By: Bharuka, R.; Kumar, P.
   P 24 AUSTR DAT C Pages: 109-117 Published: 2013
   Times Cited: 7

7. Finding Superior Skyline Points from Incomplete Data
   By: Bharuka, R.; Kumar, P.
   P INT C MAN DAT Pages: 35-44 Published: 2013
   Times Cited: 2

8. The Skyline operator
   By: Borzsnyi, S; Kessmann, D; Stocker, K
   Book Group Author(s): IEEE COMPUTER SOCIETY; IEEE COMPUTER SOCIETY; IEEE COMPUTER SOCIETY
   17TH INTERNATIONAL CONFERENCE ON DATA ENGINEERING, PROCEEDINGS Book Series: IEEE International Conference on Data Engineering Pages: 421-430 Published: 2001
   Times Cited: 624

9. On First-Order Query Rewriting for Incomplete Database Histories
   By: Bruyere, Veronique; Decan, Alexandre; Wijten, Jef
   TIME 2009: 16TH INTERNATIONAL SYMPOSIUM ON TEMPORAL REPRESENTATION AND REASONING, PROCEEDINGS Pages: 54-61 Published: 2003
   Times Cited: 2

10. Indexing incomplete databases
    By: Canahuate, G; Gibas, M; Ferhatosmanoglu, H
    ADVANCES IN DATABASE TECHNOLOGY - EDBT 2006 Book Series: LECTURE NOTES IN COMPUTER SCIENCE Volume: 3896 Pages: 884-901 Published: 2006
    Times Cited: 4

11. On high dimensional skylines
    By: Chan, Chee-Yong; Jagadish, H. V; Tan, Kian-Lee; et al.
    Times Cited: 45
12. **Finding k-dominant skylines in high dimensional space**  
   BY: Chan, C.-Y.; Jagadish, H.; Tan, K.-L.; et al.  
   ACM SIGMOD INT C M Pages: 503-514 Published: 2006

13. **Searching Dimension Incomplete Databases**  
   BY: Cheng, Wei; Jin, Xiaoming; Sun, Jian-Ta.; et al.  
   IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING Volume: 26 Issue: 3 Pages: 725-738 Published: MAR 2014

14. **Skyline with presorting**  
   BY: Chomicki, J.; Godfrey, P.; Gryz, J.; et al.  
   19TH INTERNATIONAL CONFERENCE ON DATA ENGINEERING, PROCEEDINGS Book Series: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON DATA ENGINEERING (SERIES) Pages: 717-719 Published: 2003

15. **Skyline with presorting: Theory and optimizations**  
   BY: Chomicki, J.; Godfrey, P.; Gryz, J.; et al.  
   Intelligent Information Processing and Web Mining, Proceedings Book Series: ADVANCES IN SOFT COMPUTING Pages: 595-604 Published: 2005

16. **Efficient High Dimension Data Clustering using Constraint-Partitioning K-Means Algorithm**  
   BY: George, Aloysius  
   INTERNATIONAL ARAB JOURNAL OF INFORMATION TECHNOLOGY Volume: 10 Issue: 5 Pages: 467-476 Published: SEP 2013

17. **Maximal vector computation in large data sets**  
   BY: Godfrey, Parke; Shipley, Ryan; Gryz, Jarek.  
   P 31 INT C VER LARG Pages: 229-240 Published: 2005

18. **Models for Incomplete and Probabilistic Information**  
   BY: Green, T. J.; Tannen, V.  

19. **Local and Global Approximations for Incomplete Data**  
   BY: Gryzmal-Butse, Jerzy W.; Rzasa, Wojciech  
   TRANSACTIONS ON ROUGH SETS VIII Book Series: Lecture Notes in Computer Science Volume: 5084 Pages: 21-34 Published: 2008

20. **Rough set approach to incomplete data**  
   BY: Gryzmal-Butse, JW  
   ARTIFICIAL INTELLIGENCE AND SOFT COMPUTING - ICAISC 2004 Book Series: Lecture Notes in Artificial Intelligence Volume: 3070 Pages: 50-55 Published: 2004

21. **Title: [not available]**  
   BY: GRZYMALABUSSE JW  
   LECT NOTES COMPUT SC Volume: 3100 Pages: 78 Published: 2004

22. **A comparison of several approaches to missing attribute values in data mining**  
   BY: Gryzmal-Butse, J. W.; Hu, M.  
   Rough sets and current trends in computing Pages: 379-385 Published: 2000  
   Publisher: Springer Science & Business Media Heidelberg, Berlin

23. **Evaluating top-k queries over incomplete data streams**  
   BY: Haghani, P.; Michel, S.; Aberer, K.  
   P 38 ACM C IN KNOWL Pages: 877-886 Published: 2009

24. **An evaluation of k-nearest neighbour imputation using Likert data**  
   Times Cited: 43
25. Skyline query processing for incomplete data
   By: Khalefa, Mohamed E.; Mokbel, Mohamed F.; Levandoski, Justin J.
   Book Group Author(s): IEEE
   2008 IEEE 24TH INTERNATIONAL CONFERENCE ON DATA ENGINEERING, VOLS 1-3 Book Series: IEEE International Conference on Data Engineering Pages: 556-565 Published: 2008

26. Efficient progressive skyline computation
   By: Kian-Lee Tan; Pin-Kwang Eng; Beng Chin Ooi
   Proceedings of the 27th International Conference on Very Large Data Bases Pages: 301-10 Published: 2001

27. Shooting stars in the sky: an online algorithm for Skyline queries
   By: Kossmann, D.; Ramsak, R.; Rost, S.
   Proceedings of the Twenty-eighth International Conference on Very Large Data Bases Pages: 275-86 Published: 2002

28. An incomplete database approach to global query processing
   By: Otsuka, S; Miyazaki, N
   Book Group Author(s): IEEE
   TWELFTH INTERNATIONAL CONFERENCE ON INFORMATION NETWORKING (ICOIN-12), PROCEEDINGS Pages: 337-342 Published: 1998

29. An optimal and progressive algorithm for skyline queries
   By: Papadias, D.; Tao, Y.; Fu, G.; et al.
   SIGMOD Pages: 467-478 Published: 2003
   [Show additional data]

30. Completeness of Queries over Incomplete Databases
   By: Razniewski, Simon; Nutt, Werner
   PROCEEDINGS OF THE VLDB ENDOWMENT Volume: 4 Issue: 11 Pages: 749-760 Published: AUG 2011