Research Article Impact Factor: 0.621 ISSN: 2319-507X
Sandesh Dhawale, IJPRET, 2014; Volume 2 (8): 661-671 IJPRET

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

IMPLEMENTING AND APPLYING DOMINANT GRAPH INDEX IN DOMINANT RELATIONSHIP ANALYSIS

MR. SANDESH SUDHAKAR DHAWALE[1], MR. A B CHOUGULE[2]

PG Student, MCERC, Nasik, MH.

Accepted Date: 15/02/2014 ; Published Date: 01/04/2014

Abstract: Dominant relationship concept has recently used for answering Top-k queries. Here the concept of dominant relationship has extended to do the business analysis. More specifically here we have presented new type of analysis, called as dominant relationship analysis. This dominant relationship analysis can help product manufacturers to design their product effectively while remaining profitable. We have used Dominant Graph(DG) index for doing dominant relationship analysis. Dominant Graph captures dominant relationship between product and customers.

Keywords: Dominant Graph, Dominant Relationship Analysis, Generalized Dominant Relationship Queries.

PAPER-QR CODE

Corresponding Author: MR. SANDESH SUDHAKAR DHAWALE

Access Online On:

www.ijpret.com

How to Cite This Article:

Sandesh Dhawale, IJPRET, 2014; Volume 2 (8): 661-671

INTRODUCTION

Dominant relationship concept has recently used for answering top-k queries. Given a dataset S of n-dimensions. Let D={D1, D2,...,Dn} be the set of dimensions. Let x and y be two products, we can say that product x dominates product y if x is better than y in at least one dimension and equal to or better than it in remaining dimensions. Dominant relationship between products and customers is captured using Dominant Graph(DG) index. We use dominant graph for answering following five types of queries 1) Generalized Linear Optimization Queries (GLOQ) 2) Generalized Subspace Analysis Queries (GSAQ) 3) Generalized Comparative Dominant Queries (GCDQ) 4) Generalized Skyline Product Queries (GSPQ) 5) Generalized Subspace Skyline Query (GSSQ). These queries are called as Generalized Dominant Relationship Queries(GDRQ). Analysis done using GDRQs can help product manufacturer to position their product in market more efficiently.

The paper is organized as follows. Section II discuses related work. Section III presents problem statement. Section IV discusses five generalized dominant relationship queries. Section V discusses computation of dominant graph index.

Section VI presents strategies to answer GDRQs. Section VII presents conclusion.

Related Work

Many researchers have shown interest and contributed to development of customer-product Relationship Analysis methods. Ling Zhu [1] proposed a method called DADA to answer various forms of dominance relationship queries which can help manufacturers in positioning the product in market effectively. Lei Zou and Lei Chen [2] have investigated connection between top-k queries and dominant relationship in records and based on which proposed efficient layer based indexing structure Dominant Graph to answer top-k queries. Tom Brijs [3] have proposed and implemented methods for profit oriented association rule discovery. It shows that with the use of frequent itemset it is possible to identify cross-sale potential of a product item and use this information for better product selection. Konrad Stocker [4] proposed and implemented method for determinig skyline points in large dataset i.e points which are not dominated by any other point. Yao J [5] proposed data mining technique for sensitivity analysis. It helps manufacturers to identify sensible factors that play an important role in improving total profit. The research reviewed here mainly focuses on how to find patterns that maximize the profit or can take actions into effect in a given transaction data set. Manufacturers are also interested in creating a set of best possible products such that the newly created products are not dominated by the products in the existing market and in finding all the subspaces where the given product is in skyline. This is an area which has so far been left untouched by the researchers in data-mining field and hence there is need of work to be done.

Problem Statement

To propose a Dominant Relationship Analysis Tool for product manufacturers. This analysis tool extracts important information from a given large dataset. The extracted information will be very useful for manufacturers to do the business analysis and to determine the position of their product in the market more effectively.

Generalized Dominant Relationship Queries

In this section we first introduce five GDRQs and DG.

A. Generalized Linear Optimization Query(GLOQ)

The motivation for GLOQ is the observation that manufacturers do not have infinite resources and hence they have to take into consideration many conditions when they want to position their product in the market, e.g. making laptop lighter requires better components which causes to increase in price of laptop. We model such conditions using plane L.

Given a palne L, a set of objects C in space D, We wish to find some products which intersect L and dominate most points in C.

B. Generalized Subspace Analysis Query (GSAQ)

The motivation to GSAQ is the observation that manufacturers may be interested in analysing dominant relationship in the subspaces of product dimension D.

Given a set of points C and a point p in n-dimensions of D the most basic GSAQ is to compute for each subspace D' the number of points dominating or dominated by p.

C. Generalized Comparative Dominant Query (GCDQ)

The motivation behind GCDQ is to compare the set of dominated objects between competitive products.

Given set of objects A, B, C GCDQ determines how many objects in C are dominated by some objects in A and not by any objects in B. This is useful for manufacturers to determine the number of customers who are solely dominated by his products.

Given a set of objects A, B, C GCDQ also determines how many objects in C are dominated by both objects in A and objects in B. This is useful to manufacturers to determine the number of customer preferences which are dominated by his products and by competitor's product in common

D. Generalized Skyline Product Query(GSPQ)

The motivation behind GSPQ is the observation that manufacturers may be interested to produce products which are not dominated by any existing products in the market.

Given a set Te of existing products in market and source and source tables T1, T2,....Tn of subproducts. GSPQ produces new products such that newly created products are not dominated by any existing products in the market.

E. Generalized Subspace Skyline Query(GSSQ)

The motivation to GSSQ is the observation that manufacturers could be interested in analyzing that in which subspace product is undominated by customer preferences.

Given a set C of customer preferences and a point p in n-dimensional space, GSSQ determines all subspaces where point p is not dominated by any customer preference.

Defining and Computing Dominant Graph

In this section we define DG and present algorithms for inserting and deleting records from DG.

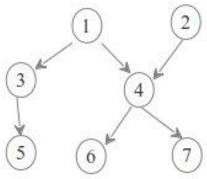
A. Defining DG

Dominant Graph(DG) captures dominant realtionship between customer preferences and products. We use DG to answer DRQs.

Definition Given a set S of records in a multidimension space, S has k nonempty maximal layers Li, i=1.....k. The records r in ith maximal layer and records r' in (i+1)th layer form a bipartite graph gi , i= 1....(m-1). There is a directed edge from r to r' in gi if and only if record r dominates r'.

TID	X	Y
1	350	100
2	300	300
3	650	200
4	500	400
5	750	250
6	550	600
7	600	500

Database



Dominant Graph for Database

All attributes are minimal attributes meas minimal values are preferred for them.

B. Computing DG

Insert Algorithm

Input: i) Dominant graph of the database.

ii) r = record to be inserted.

1. If r is not dominated by any record in first layer of DG then

2. set n=0.

3. else

4. All record Pi at first layer of DG that dominates r are collected to form the set P.

5. Do DFS search from each Pi to find the longest path L, and each record in dominates r.

6. Set n = |L|.

7. Insert r into the (n+1)th layer of DG.

8. If r dominates some records Ci in the (n+1)th layer of DG then

9. All the descendants records of Ci (including Ci) are collected to form the set S.

10. For each record O in S do

11. O is degraded into it's next layer.

12.Build parrent children relationship between O and records in current next layer.

13. If O has some other parrent A that is not in set S then

14. Delete the directed edge from A to O.

15. Build parrent children relationship between records in nth layer and r.

16. Build parrent children relationship between records in r and n+2 th layer

17. Report the updated DG.

Delete Algorithm

Input: i) Dominant Graph of the database.

ii) r = record to be deleted.

1. Find the position of r.

2. Delete the r from nth layer of DG.

3. for each child Ci of r do

4. if Ci has no another parent in nth layer then

5. Insert Ci into nth layer.

6.Build parent children relationship records at (n-1) th layer and Ci.

7. Do Del(Ci, n+1).

Del(O,m)

1. Delete O from mth layer of DG.

2. for each child Ci of O do

3. if Ci has no another parrent in mth layer then

4. Insert Ci into mth layer.

5. Build the parrent children relationship between records in (m-1)th layer and Ci.

6. Del(Ci, m+1)

Answering Gdrqs Using DG

A. Generalized Linear Optimization Query (GLOQ)

Given a plane, L, and a set of objects, C, in an N-dimensional space of D, Linear Optimization Query can be defined as determining any point p on the plane L which dominates most objects from C. Any point that lies on the plane L is profitable to manufacturer

GLOQ Algorithm

Input: A plane L, set of objects C .

Output: Points on plane L which dominate maximum

number of objects from C.

1. Start from the point <1,1,...1> in N-dimensional space.

2. At any stage if the cell is at the bottom left of plane L, iterate continually on it's children until we find the cell on the plane L.

3. If the cell is on the plane L add it to the result cell set.

4. Obtain the cells from result cell set which dominate maximum number of objects from set C using GLOQ_Max_Dominating function.

Function GLOQ_Max_Dominating (result cell set, DG for set C)

1. Find the position of the cell in DG for set C.

2. Determine number of objects in set C are dominated by the cell.

3. Repeat step 1 and 2 for each cell in result cell set.

4. Return the cells which dominate more objects in set C than any other cell in result cell set.

B. Generalized Subspace Analysis Query(GSAQ)

Given a set C of customer preferences and a point p in the N -dimensional space D. (1) Determine number of customer's preferences dominating point p in dimension D'. (2) Determine number of customer's preferences dominated by point p in dimension D'. Where D'⊆ D.

GSAQ Algorithm

Input: i) C = A set of objects in N-dimensional space D.

ii) D' = A subspace of dimension D. iii) A point p.

Output: i) Number of objects in C dominated by point p in

subspace D'.

ii) Number of objects in C dominating point p in

subspace D'.

1. Let I1,I2...Ik be dimensions in subspace D'

2. Build dominant graph for each dimension in subspace D' for dataset C.

3. Determine objects in C dominated by point p for dimension I1,I2...Ik of subspace D'.

4. Let S1, S2...Sk be the sets of objects dominated by point p for the dimension I1, I2...Ik respectively.

5. Determine objects in C which are incomparable with point p for the dimension I1,I2...Ik of subspace D'.

6. Let S1', S2',...Sk' be the sets of objects which are incomparable with point p for dimension I1, I2,...Ik respectively.

7. Objects in C which are dominated by p in subspace D' are obtained by following equation. {S1US1'} ∩ {S2US2'} ∩ ....∩ {SkUSk'}.

8. Objects in C which are dominating point p in subspace D' are obtained by following equation C - ({S1US1'} ∩ {S2US2'} ∩ ....∩ {SkUSk'}).

C. Generalized Comparative Dominant Query (GCDQ)

Comparative Dominant Queries are used to compare the set of dominated objects between competitive products. First the concept of group dominant is introduced. Group Dominant, gdominating(A,C,D) :- Given two sets of objects A and C in an N dimensional space of D, we define gdominating(A, C, D) as the set of objects in C which are dominated by some object from A.

GCDQ− (A, B, C, D): Given three sets of objects in the N -dimensional space of we define CDQ− (A, B, C, D) as: |gdominating(A, C, D) − gdominating(B, C, D)|.

GCDQ∩ (A, B, C, D): Given three sets of objects in the N -dimensional space of D, we define CDQ∩(A, B, C, D) as: |gdominating(A, C, D) ∩ gdominating(B, C, D)|.

GCDQ Algorithm

Input: Dominant graph for three sets of objects A

Output: i) Set of objects in C which are dominated by some bjects in A and not by any objects in B.

ii) Set of objects in C which are dominated by some objects in A and some objects in B

1. Take an object x from set A.

2. Find the position of that object x in dominant graph.

3. Determine objects in C which are dominated by the object x.

4. Repeat steps 1, 2, 3 for each object in set A.

5. Let S1 be the set of objects in C which are dominated by objects in A.

6. Take an object y from set B.

7. Find the position of the object y in dominant graph.

8. Determine objects in C which are dominated by the object y.

9. Repeat steps 6, 7, 8 for each object in set B.

10. Let S2 be the set of objects in C which are dominated

by objects in B.

11. CDQ− (A, B, C, D) = S1 – S2.

12. CDQ∩ (A, B, C, D) = S1 ∩ S2.

D. Generalized Skyline Product Query

Given a set Te of existing products in market, a set of best possible products is to be created from source tables T1, T2,...Tn of sub products such that the newly created products are not dominated by any existing products in Te.

GSPQ Algorithm

Input: i) Existing product table Te.

ii) Source tables of sub products T1, T2, ...Tn.

Output: Table of newly created skyline products.

1. Build dominant graph for existing product dataset Te.

2. Obtain tables T1' , T2' , ..., Tn' which contain skyline points from source tables T1,T2,...Tn respectively.

3. Generate new product table Tp by using sub products in tables T1', T2'...Tk'.

4. Find a set Tp' of all products from Tp which dominate all the products in skyline of DG of Te.

5. Return Tp'.

E. Generalized Subspace Skyline Query(GSSQ)

Given a set C of customer preferences and a point p in the N -dimensional space D. Subspace Skyline Query can be defined as determining all possible subspaces where point p is not dominated by any customer preference.

GSSQ Algorithm

Input: i) C: An N dimensions set of customer preference.

ii) A point p .

Output: All the subspaces where p is in skyline.

1. Create DG for each dimension in D for dataset C.

2. Take point p from dataset S.

3. Obtain the sets S1,S2....Sn such that each point in Si dominates p in the dimension Di.

4. if S1∩S2....∩Sk == NULL then Point p is a skyline point in subspace {D1,D2,...Dk}.

5. For all combinations of dimensions in D check whether p is in skyline or not.

CONCLUSIONS

In this paper, we have introduced Dominant Relationship Analysis and proposed five types of queries to illustrate the various aspects of this new form of analysis. To support the queries, we have further used an efficient indexing structure Dominant Graph. In our future work, we will look at how these three types of queries can be integrated efficiently and elegantly.

REFERENCES

1. Ling Zhu, Cuiping Li, Anthony K. H. Tung, Shan Wang, "Microeconomic analysis using dominant relationship analysis", Springer, Knowledge and Information Systems, 2011.

2. Lei Zou, Member, IEEE, and Lei Chen, Member, "IEEE , Pareto-Based Dominant Graph: An Efficient Indexing Structure to Answer Top-K Queries ",IEEE transactions on Knowledgeand Data, Engineering, Vol. 23, no. 5, May 2011.

3. Brijs T, Swinnen G, Vanhoof K, Wets G, "Using association rules for product assortment decisions: a case study", Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pp 254–260, 1999.

4. Börzsönyi S, Kossmann D, Stocker K , "The skyline operator", Proceedings of the 17th international conference on data engineering , pp 421–430, 2001.

5. Yao J, "Sensitivity analysis for data Mining", Proceedings of the 22nd international conference of the North American fuzzy information processing society, pp 272–277, 2003.