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 12, December 2015

A Review on Exploring the record's Standard In a Relation for better understanding

Shilpa Devasthali

II Year M.E Student, Dept. Computer Engineering, BSIOTR, Savitribai Phule Pune University, Wagholi, Pune, India

ABSTRACT: For the given database table of ranked records, there is a problem in finding selection condition.we have to find the qualified record that shows its ranking among the qualified tuples. In this paper we study the Standing Maximization Problem. this will give the approximate solution for the given problem. It will show the object promotion and characterization. we also show the hardness of problem and for that solution proposed the greedy methods for high accuracy. Our solution on real database will confirm the effectiveness and efficiency .

KEYWORDS: Standing maximization Problem,NP-Hardness, relational databases

I. INTRODUCTION

To rank the record as per user preferences there are certain types of operations are used. theae operations include top-k and skyline queries.the top-k operation and the skyline query is used to calculate the highest grade. with the help of these queries superior object can be defined.
for eg: consider the following

Table 1.
A Ralation with CS PhD Graduates

Name	Age	Location	Expertise	Publications
Brown	30	N.America	System	14
Smith	27	N.America	Database	8
Suzuki	32	Asia	Theory	9
Muller	28	Europe	Theory	15
Dubolis	26	Europe	System	12
Martin	31	Europe	Database	17
Kim	28	Asia	Database	10
Chen	26	Asia	Theory	12
Gupta	26	Asia	System	13

A relation with CS Phd Graduates. This table contains the attributes as name,age,location,expertise and publication. For measuring the quality of graduates, consider publications as the measuring attribute. if we go accordingly then kin does not have a good ranking. But if we restrict the relation with (age<30) and (expertise ='databases'),then kin's ranking is 1st

II. PROBLEM STATEMENT

The input for this is the relation R(D,M), query tuple(tq) such that tq ϵ R,and support threshold sup, $0<sup\leq1$. R.D is the set of predicate attribute and R.M is the set of measuring attribute for ranking tuple in R.
If we ssume that, t.M>t'M then t is considered better than t'.

III. OBJECTIVE

Find the conjunction of selection predicates C on R.D such that:
i)tq is included in σ_c R
ii)there are at least sup.|R|tuples in σ_c R and
iii)percentile rank pr(t_q, σ_c R) is maximized

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 12, December 2015

IV. BACKGROUND

It includes the following:
i)Rank based analysis and query by output
ii)Promosion analysis
iii)Query refinement
iv)Recerse top-k queries

i)Rank based analysis and query by output: Rank based analysis is the study which chooses the most useful attribute that are most influenced in ranking of item from the given query. And the query by ouput worked on queries. it finds the query such that it matches at least on semantic that it inherits from the given database. for our problem this is useful for finding the retrieval hit of target tuple.

ii)Promotion analysis: promotion analysis is based on the region based promotion queries but this required large space and also it is very expensive. for efficient processing the Region based promotion cube framework is designed. the generalisation of SMP(symmetric Multiprocessing)also solve our problem.For the high cost of exploring all possible conditions they turn to use Marerialize algorithm .It found that only materialization will not give the most effective output hence adaptive approach is used as it gives the optimal structure.

iii)Query refinement: Query refinement is the technique which isolates the outliers from the input .
Our paper , in contrast is to work on predicate that affects the whole input. mishra and koudas invented the iffective and interesting way for query refinement.they proposed the system that works on range and equality predicates on numerical and catagorial attributes.here only constrain is the query output size .

iv)Reverse Top-K queries: in this technique there is some preferences given to the attributes of a relation .for example: with the help of top –k queries will give the set of ranked object (product) that are customer interected in . but with the help of Reverse top-k queries it will show the set of customers that find a product appealing. this also shows how to automatically provide a meaningful interpretation for average rating of product. using this type of query we can formalise andquantify the compititiveness between product based relationship between the set of their potential customers.
All the work above study cannot solve the SMP.

V. METHODOLOGY

Problem's input:
i) Ordinal
ii) Hierarchical
iii) Binary or categorical
i)**Ordinal**: if attribute is type of ordinal then we can define equality as (e.g.,age=28)
if range is predicate then (e.g.,26<age<28)
ii) **Hierarchical**: assumption is that there is hierarchy of values .
Hence we can generalise the lowest granularity value of tuple (e.g.,location=Boston,location=USA)
iii)**Binary and categorical attribute:** only possible predicate is equality on value of query tuple.

VI. BASE METHOD

To solve the SMP the base method is Naïve Algorithm. This algorithm search in depth first manner all selection predicates on all attribute that contain the value of query object t_q .

Algorithm 1. Naive Algorithm

1: G := R; Preds = ϕ; bestrank := qual(G);
2: bestG := G; bestPreds :ϕ;
3: procedure NAIVERANGE(G, Preds)

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 12, December 2015

```
    if all attributes are in Preds then
        if qual(G) > bestrank then
            bestrank :=qual(G);
            bestG := G; bestPreds :=Preds;
    else
        Pick any attribute A not in Preds
        A:preds := all possible predicates on A for G
            such that |σpredG| ≥sup .|R|;
    for each pred ϵA.preds do
        G` = σpredG;
        Preds` :Preds ᴜ {pred};
        NAIVERANGE(G`, Preds`);
return {bestG, bestPreds};
```

VII.PROPOSED METHODS

In a view of hardness of SMP, there are no of greedy approaches.

i) **Browsing algorithm.**

This algorithm extracts the classification rules from set of records. it iteratively selects the subrange which

(i) includes tq,

(ii) includes at least sup.|R| records when applied together with the predicates Selected so far, where sup is the minimum support constraint, and

(iii) Maximizes the ratio of positive to all tuples covered by the rule (i.e., range).

Browsing algorithm compares the records with the t_q whether it is less than or equal to or it is greater than equal to. With the help of this it may decide that whether it is positive dimension or negative dimension. It works definitely; but the working of BA is slow.

Hence there is another solution that works relatively faster than the BA. This algorithm is known as Diversified –Path Browsing Algorithm (DBA). The DBA works on Diversified Predicate for the single attribute.

ii)**Enumerating Diversified –Path Browsing (EDBA)**

In BA (Browsing Algorithm it works very iteratively. At each Iteration it picks the most useful attribute. while in this EDB Algorithm, it examines all permutations of the predicate attribute. This algorithm is based on the current percentile of the record. It increases the finding a better percentile rank. By examining all possible permutations of predicate attribute. EDBA takes the prioritized attributes for the work. It arranges according to the improvement made at the each record. Hence the time required is less as compared to the previous method. EDBA gives the optimum solution.

VIII.CONCLUSION

When we are going to maximise the rank of given tuple in selection the selection result, there is a problem in finding the set of selection predicate on relation. It seems NP-hard. Fast and approximate solutions can be found using proposed greedy methods. For this we proposed the methods named BA, DBA and EDBA. Among these all EDBA is the most effective one which gives the optimum output.

REFERENCES

1. R. Fagin, A. Lotem, and M. Naor, "Optimal aggregation algorithms for middleware," in Proc. ACM Symp. Principles Database Syst., 2001, pp. 102–113.
2. S. B€orzs€onyi, D. Kossmann, and K. Stocker, "The skyline operator,"in Proc. Int. Conf. Data Eng., 2001, pp. 421–430.
3. G. Das, V. Hristidis, N. Kapoor, and S. Sudarshan, "Ordering the attributes of query results," in Proc. ACM SIGMOD Int. Conf. Manage.Data, 2006, pp. 395–406.
4. T. Wu, Y. Sun, C. Li, and J. Han, "Region-based online promotion analysis," in Proc. Int. Conf. Extending Database Technol.: Adv. Database Technol., 2010, pp. 63–74.

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 12, December 2015

5. T. Wu, D. Xin, Q. Mei, and J. Han, "Promotion analysis in multidimensional space," Proc. VLDB Endowment, vol. 2, no. 1, pp. 109– 120, 2009.
6. Y. Zhang, Y. Jia, and W. Jin, "Promotional subspace mining with EProbe framework," in Proc. ACM Conf. Inf. Knowl. Manage., 2011,pp. 2185–2188.
7. C. Mishra and N. Koudas, "Interactive query refinement," in Proc.12th Int. Conf. Extending Database Technol.: Adv. Database Technol.,2009, pp. 862–873.
8. Sultana, N. Hassan, C. Li, J. Yang, and C. Yu, "Incremental discovery of prominent situational facts," in Proc. Int. Conf. Data Eng., 2014, pp. 112–123.
9. T. Lappas, G. Valkanas, and D. Gunopulos, "Efficient and domain-invariant competitor mining," in Proc. ACM SIGKDD Int.Conf. Knowl. Discovery Data Mining, 2012, pp. 408–416.