



International Journal of Control Theory and Applications

ISSN : 0974–5572

© International Science Press

Volume 10 • Number 27 • 2017

Survey on Query Optimization for Declarative Crowdsourcing Systems

W.C.Cincy^a and J.R. Jeba^b

^aResearch Scholar, Department of Computer Applications, Noorul Islam Centre for Higher Education, Kumaracoil, Tamilnadu, India.
E-mail: cincyaji@gmail.com

^bAssociate Professor & HOD, Department of Computer Applications, Noorul Islam Centre for Higher Education, Kumaracoil, Tamilnadu, India.
E-mail: jrjeba@rediffmail.com

Abstract: In recent years query optimization is the challenge for researchers. Query optimization is the most concerning issue for crowdsourcing framework. Crowdsourcing has a valuable device for solving some problems that computer cannot do well. In the crowdsourcing marketplace, the customer is simply give an sql like query and then the system takes the responsibility to assembling the given query, then generating the arrangement of execution plan in the Human Intelligence Task to the crowdsourcing market place. The cost based query optimization is the future work for crowdsourcing systems. The objectives of cost and latency explain its balance of the query Plan. To optimizing queries such as Selection, Joins and Complex selection Join queries developed in first stage and generate efficient algorithms, in the second stage and validate our approach through suitable experiment by simulation as well as with the actual crowd. The query optimization objectives latency and cost create a query plans. The query plans give a balance between the cost and latency.

Keywords: Crowdsourcing, Query Optimization, HIT, Survey.

1. INTRODUCTION

1.1. Crowdsourcing

Data mining is the procedure of extracting and mining knowledge from large amount of facts. Gold mining from rock or sand is equal as the knowledge mining from data. Data mining is also known as knowledge discovery in data.

Crowdsourcing is a request to the total group in the people to create, discuss, and refine meaningful thoughts or tasks or payments via the web. Crowdsourcing is an online, distributed problem solving and production of model. Crowds of people can answer some problems quicker than people or small groups. The crowd can also sort through the solutions, decision the super ones. Crowdsourcing is used to clarify problems in hand writing recognition, translation and audio transcription. Information can be collected together quickly and efficiently. Crowdsourcing requires no or little expertise from participants. Participant's supervision is not acceptable. Nowadays many industries attracted by crowdsourcing it act as an effective tool for use of an human

knowledge to solve many problems that pc and smart systems cannot present well, such as, calligraphy finding, paraphrase, photo labeling and audio transcript [19]. The term crowdsourcing was presented by Howe in his article *The Rise of Crowdsourcing*. Among the past couple of years crowdsourcing has ended up in style among firm, foundations and colleges. Crowdsourcing is the new method of using the power of people in projects, which usually requires a large number of people. Crowdsourcing is also meant to reach a wide range of people, which may sometimes be required to obtain a solution correctly and efficiently.

1.2. Query Optimization

Many Relational Database Systems have Query Optimization function. Selecting the Most Relevant Query Plan is the Process of Query Optimization. Query optimization is the main one function of many relational database management systems. The Query Optimizer workout to determine the most active way to carry out a given query by considering the Probable query strategy. Basically the query optimizer cannot be getting straightforwardly by the customer once queries are obtain to database server, and parsed by the parser, they are then accepted to the query optimizer where optimization happen. In the database, query is a request for information. Queries results are produced by accessing relevant database and manipulating it in a way that yields the important information [17]. Since in most particular cases database structures are complex, for not very simple queries, the needed data for a query can be collected from a database by accessing it in different ways, through different data-structures, and in different orders. Each different way normally requires different processing time. Finally Query optimization is used to find the best query plan in terms of estimated monetary cost.

2. SURVEY ON DECLARATIVE CROWDSOURCING

Many authors have tried to develop crowdsourcing algorithm for reduce cost and latency. Some highlights of the related work are outlined here.

2.1. Crowdop Query Optimization

Ju Fan et al [1] formally studied the problem of cost and latency. They found, query plan that is suitable for balance between cost and latency. They proposed query optimization algorithm that generate a query plan and evaluate in the market place, Amazon Mechanical Turk. It is used to generate a query Plan and also it is a good query plan for balance between the cost and Latency [16].

2.2. K-Anonymity

Sai Wu et al [2] proposed the problem of privacy of published data. Published records in Human Intelligence Task (HIT) may contain sensitive or secreat attributes that cause the privacy leakage. They proposed K-Anonymity algorithm, that algorithm control the information failure and the accurateness of crowdsourcing. K-anonymity techniques are proposed to keep the privacy of published Data. K-Anonymity approach needs to solve the transaction between the privacy and the accuracy. K-Anonymity can be used to reduce crowd cost and maintain high accuracy.

2.3. Species Estimation Algorithm

Beth Trushkowsky et al [3] discusses the query progress. Monitoring the query progress is very difficult due to irregularly arrival of data. They studied about the problem of irregularly arrival of data. Statistical tools that help to drive query execution and crowdsourcing strategies. Species estimation algorithm is used to manage the execution of set enumeration queries, in crowdsourcing database systems that also used to develop the accuracy of cardinality estimation.

2.4. Online, Off Line and Incremental Algorithm

Eleonora Ciceri et al [4] states that the problem of uncertain data in Top-K Queries. A well known class of applications in the system is commonly mentioned to as Top-K queries. The main objective of Top-K queries is finding the best k-objects. Online, offline and incremental algorithm are used to find out the uncertain data. In the minimal set questions introduced uncertainty resolution is the problem for trust-based intrusion detection determines the optimal trust threshold for minimizing false positive and false negative.

2.5. Smart Trace

Demetrios Yazti et al [5] studied about the problem of tracing the query in distributed smart phones. Crowdsourced trace similarity frame work that is “Smart Trace”. The smart trace frame work performs well both in query response time and energy consumption. This is a novel iterative algorithm. Author tested the algorithm on both synthetic and real data set. The results show less energy but returning result faster.

2.6. Dishonest Workers

Luke Gottlieb et al [6] discusses the problem about dishonest workers. Crowdsourcing is used for range of applications. The author studied about the problem of dishonest workers in AMT. The author try to improve the quality of a crowd sourced labour. Author proposed to find out the experts workers on AMT. But this is not a systematic Technique to find out the experts in AMT

2.7. Crowd optimization Algorithm

G.Archana et al [7] explained about the problem of extracting the data from the crowdop. Author proposed preprocessing method, which is used to extracting information from the crowdop. Preprocessing methods are divided into data cleaning, data integration, data transformation and data reduction. Crowd optimization algorithms named optimization structure, optimization select, optimization join, create parse tree, latency. Bound optimization used to extract the information. The user selects the latency rating for product. Customer can purchase the best product with the help of proposed method.

2.8. Crowdsearch

Akshay.S.Patil et al [8], [18] studied about the problem of best query. Author proposed novel improvement approach, which is used to find out the most effective query. In this paper author used the multiple data bases. Crowdsourcing structures can be extensively setup into principle based and expense based. From this principle based approach is difficult to actualize.

2.9. Crowdsourced learning Mechanism

Bere sachin, Mane Kush [9] proposed a new approach to social Science. Human experience need to select the best information. Participants are motivated to find out the correlates of some human behavior result. Here two models are implemented here. The first model forecast users monthly electric charge consumption the second one is user’s body mass index. This method is useful to answer many difficult questions.

2.10. CrowdCE

Tarek Ali, Eman Nasr [10] proposed a CrowdCE model.CrowdCE is an SDLC model intended for use by crowds. SDLC using the mesh style model. A collaboration model structurally decomposes the overall computing element into human based based computing elements and atomic machine computing elements. CrowdCE is a model for managing the collaboration and coordination of MBCE and HBCE.

2.11. Top Crowd

Christian Nieke et al [11] studied about the problem of incomplete Data. A top crowd, a novel and efficient algorithm works on incomplete data and also controlling query processing cost in terms of response point in time and money. This algorithm drastically reduced number of crowd access. Building crowd enabled query processing operators for Top-K query processing with the help of top crowd algorithm. Correct query results are generated by Safe pruning rules. Real and synthetic data set shows the benefit of our Technique.

2.12. CDAS

This work was presented by X. Liu, M. Lu, B. C. Ooi, Y. Shen, S. Wu, and M. Zhang [20]. CDAS is a quality control sensitive answering Model, that is used to guide the Crowdsourcing engine, to monitor the Human Task. Computer Oriented Complex Job may be divided in to Computer Oriented Task and Human Oriented Task. The Author proposed two sub models Prediction model and verification model. In the Prediction model, estimate how many workers are required to achieve specific accuracy. To select the correct answer based on Probability is the verification model. Malicious workers can provide Random answers to all questions .This can degrade the quality of query results.

3. DISCUSSION ON ALGORITHMS

Table 1
Comparison of the Algorithms

<i>S. No</i>	<i>Algorithm</i>	<i>Features</i>	<i>Issues</i>
1.	Query Optimization Algorithm	Balance between cost and latency	Not Support Sql operator such as sort and aggregation
2.	K-Anonymity Algorithm	Reduces crowd cost, maintain high accuracy	Sometimes not guarantee for privacy.
3.	Species Estimation Algorithm	Algorithm improves query result. statistical tools is help to drive query completeness	The quality of query result not specified.
4.	Incremental Algorithm	Ordering probability achieve better performance	Computationally impractical
5.	Smart trace	Less energy but returning result faster	Algorithm not fully coverage the problem
6.	Dishonest workers	Achieved better result	Not a systematic technique
7.	Crowd optimization Algorithm	Customer can purchase best product	Sort and aggregation sql operators not supported
8.	Crowdsearch	Discover the most effective query	Difficult to actualize
9.	Crowd Sourced learning Mechanism	Useful method to answer difficult questions	Limited behavior processed
10.	CrowdCE	Feasible model	Not efficient
11.	Top Crowd	Reduce crowd cost	Difficult to actualize
12.	CDAS	Improve query Result. Reduce Processing Cost	Malicious Worker submits Random answers. This can degrade the quality of Query Result.

In declarative crowdsourcing systems [1] using the query optimization algorithms, that is used to solve the problem of cost and latency. Algorithm generates a query plan and evaluate in the real crowd market place.

The comparison of the related work is in Table 1. The existing techniques do not get systematic result. Malicious workers also submit random answers to all questions. This can degrade the quality of query result. Published data privacy can be affected by malicious behaviors. So some times this does not guarantee for published data. The cost and latency are the important objectives in the crowdsourcing systems. The present study contains different algorithms, which obtain cost and latency not fully balanced in the existing papers.

4. PROPOSED METHODOLOGY

The proposed methodology uses both optimization framework algorithm and optimization selection for query optimization. This efficient algorithm optimizes optimal query, connection query and advance (selection join query). Data pre-processing method describes any type of processing performed on rare data to arrange it for another processing procedure. Data cleaning, data integration, data transformation and data reduction are the classifications of Preprocessing. The data set consists of brand, style, make, price and model of car for processing. The crowdsourcing environment for the dataset has been simulated using the MATLAB as a platform.

5. PROCESS FLOW

The appropriate process flows of proposed algorithm for generating optimal query are as follows.

1. Analyzing the system from the dataset.
2. Select query for the demand of brands and submit query.
3. Fill query to complete the expectation
4. Join query to combine all the results
5. After acquiring results from the crowd rating for brand, style, prize etc.

6. CONCLUSION

Special approaches of query optimization for crowdsourcing have been discussed in briefly. Suitable and proper query optimization algorithms have been developed for select, join, complex query. In the present development, simulated and real crowd data convey the usefulness of query optimizer and take review of query optimization objective and generates query plans. The query plan provides a good balance between the cost and latency. The efficiency of the optimization framework and optimization selection algorithm reduces the latency and monetary cost that improves the accuracy of query optimization. Proposed work is used to find out the most efficient query plan for answering query within the monetary cost. And also minimize no of queries and find out the exact Value and cluster.

REFERENCES

- [1] J.Fan, M.Zhang, S.Kok, M.Lu and B.C.Ooi, "Crowdop:Query Optimization for Declarative CrowdSourcing Systems", IEEE Trans. Knowl. Data Eng., Vol. 27, No. 8, Aug 2015.
- [2] S.Wu,X.Wang,S.Wang,Z.Zhang,A.Tung, "K-Anonymity for Crowdsourcing Database", IEEE Trans. Knowl. Data Eng., Vol. 26, No. 9, Sep 2014.
- [3] B.Trushkowsky,T.Kraska, M.Franklin,P.Sarkar,V.Ramachandran,"Crowdsourcing Enumeration Queries:Estimators and Interfaces", IEEE Trans. Knowl. Data Eng.,Vol.27 ,No. 7,July 2015.
- [4] E.Ciceri, P.Fraternali, D.Martinenghi and M.Tagliasacchi, "Crowdsourcing for Top-K Query Processing over Uncertain Data", IEEE Trans. Knowl. Data Eng., Vol. 27, No. 2, Aug 2015.

- [5] D.Yazti,Member,C.Laoudias,C.Costa, M.Vlachos M.Andreou, D.Gunopulos,“Crowdsourced Trace Similarity with Smart Phone”, IEEE Trans. Knowl. Data Eng., Vol. 25, No. 6, June 2013.
- [6] L.Gottlieb, G.Friedland,Jaeyoung Choi, Pascal Kelm, and T.Sikora, “Creating Experts From the Crowd: Techniques for Finding Workers for Difficult Tasks”, IEEE Trans.on Multimedia., Vol. 16, No. 7, Nov 2014.
- [7] Archana,P.Srinivasan,“Efficient query optimization for Easy Retrieval of Crowd Resources”,International Journal of Innovative Research in Computer and Communication Engineering.,Vol.4,Issue2,Feb 2016.
- [8] Akshay S. Patil,Ankit D. Katiyar,Satyaprakash A. Singh, Prashant P. Kachhava, Prof. D. B. Bagul,“ Crowd Search:Generic Crowdsourcing Systems Using Query Optimization”,International Journal On Rcent and Innovation Trends in Computing and Communication,Vol.3,Issue 9,Seb 2015.
- [9] M.K.Chandrakant,B.S.Sukhadeo,“A Implementation On Forecasting Behavioural Outcomes through Crowdsourcing Mechanism”, International Journal of Innovative Research in Computer and Communication Engineering,Vol.3, Issue6, June 2015.
- [10] T.Ali,E.S.Nasr,“CrowdCE:A Collaboration Model for Crowdsourcing Software with Computing Elements”, International Journal of Innovative Research in Computer and Communication Engineering,Vol.4, Issue 2, Feb 2016.
- [11] C.Nieke,U.Guntzer,W.Balke,“Top Crowd-enabled Top-K Retrieval on Incomplete Data”, Springer,2014.
- [12] J.R.Jeba, Dr.S.P.Victor, “A Novel approach for finding Frequent Item Sets with Hybrid Strategies”,International Journal Of Computer Applications.,Vol17, No.5, 2011.
- [13] J.R.Jeba,Dr.S.P.Victor,“Comparison of frequent item set mining algorithms”,International Journal Of Computer Applications.,Vol.2, No 6, 2011.
- [14] D.S.Mispha,Dr.J.R.Jeba,“Scheduling Effective Cloud Updates in Streaming Data Warehouses using RECSS Algorithm”, International Journal of Applied Engineering Research.,Vol. 11, No . 5, 2016.
- [15] Jincymol Joseph,Dr.J.R.Jeba,“ Survey on web content extraction”, International Journal of Applied Engineering Research.,Vol. 11, No .7, 2016.
- [16] Priyanka D Sarode, Imran R Shaikh,“Automatic Query Generation and Query Optimization for Declarative Crowdsourcing System”,International Journal of Scientific & Engineering Research., Vol. 6, No. 11, Nov 2015.
- [17] Rohini Pingle, Prof. Mrs. Rucha Samant ,”A Review Paper on Query Optimization for Crowdsourcing Systems”,International Research Journal of Engineering and Technology.,Vol .2 No.9 Dec 2015.
- [18] Akshay S. Patil,Ankit D. Katiyar,Satyaprakash A. Singh, Prashant P. Kachhava, Prof. D. B. Bagul,“ Easy Rerieval of Crowd Resources Usig Query Optimization “,International conference on Emerging Trends in Engineering and Management Research.,Mar 2016.
- [19] Vinay V. Mandhare, Vinod Nayyar,“A Survey on Crowdsourcing and Behavioral Outcome”, International Journal of Advanced Research in Computer Science and Software Engineering., Vol. 4, Issue 1, Jan 2014.
- [20] X. Liu, M. Lu, B. C. Ooi, Y. Shen, S. Wu and M. Zhang,“CDAS: A Crowd sourcing data analytics system”, VLDB Endowment,Vol. 5, No. 10, pp. 1040–1051, 2012.