International Journal of Soft Computing 11 (3): 145-154, 2016
ISSN: 1816-9503

Query Plan Generation in DDS Using Non Dominant Based Teacher-Learner Optimization (ND-TLBO) Algorithm

[1]S. Venkata Lakshmi and [2]Valli Kumari Vatsavayi
[1]Department of Information Technology, Gitam University, Visakhapatnam, India
[2]Department of CS and SE, College of Engineering, Andhra University, Visakhapatnam, India

Abstract: The query optimization issue in large-scale distributed datasets are NP and difficult to resolve. The demand for the optimizer upsurges as the number of relations and positions in a distributed database query processing rises. Investigation are performed so as to obtain a suitable methodology to pursue an optimum output particularly when the dimension of the dataset upsurges. The foremost issue for the need of query optimization in any disseminated datasets systems is to diminish the total query processing cost for the specified queries. This study, proposed a novel approach known as Non Dominant Based Teacher-Learner Optimization algorithm (ND-TLBO) for the multi objective test functions in the distributed query processing as to obtain the optimized query plans. TLBO is a parametric less optimization algorithm which is enhanced in such a way that in order to employ the multi objective test function into the approach, the dominance property is used between the individuals. Fast non dominated ranking algorithm and crowding measure are the two dominance property employed in this approach. The performance of the suggested methodology is matched with the prevailing TLBO query plan generation technique and multi objective GA based query plan generation technique.

Key words: Distributed query processing, query optimization, fast non dominant sorting approach, dominance rule, TLBO

INTRODUCTION

Now a days, lots of data is being generated and are freely available a nd accessible through internet. These data stream assets gives auspicious obtainable data for global wide individuals. Advancement in technology has made it possible today to gather timely and effective information from vast sources of data (sites) distributed geographically across a network. With the introduction of great speed communication systems, substantial exploration is dedicated to build extremely effective procedures for dispensation of multifaceted queries in a price efficient way in a distributed database atmosphere.

Distributed systems are a group of autonomous collaborating systems that permits storage of information at physically distributed places, depending on the frequency of access by individual's site to the distributed site. Distributed datasets can be described as a cluster of manifold, rationally interconnected database scattered over a computer network (Kumar *et al.*, 2010; Ceri and Pelagati, 1984; Ozsu and Valduriez, 1991). These reasonably interconnected datasets are controlled by Distributed DBMS that assists the formation and preservation of distributed dataset. Query optimization is a challenging job in distributed atmosphere due of several reasons such as information gathering, quickness of communicating channel, indexing, accessibility of memory, dimension of the datasets, accumulation of intermediary outcomes, pipelining and dimension of information broadcasting (Ioannidis and Kang, 1990) that has stages such as world wide optimization and in digenous optimization (Aljanaby *et al.*, 2005; Ghaemi *et al.*, 2008). The world wide optimization is a significant stage whenever numerous locations are included and in digenous optimization is included inside the locations. The role of query optimizer is to generate Query Execution Plans (QEP) that signifies an implementation approach of the query with lower price.

Query optimization comprises of two significant jobs, exploration domain production and detecting an optimum strategy from the exploration domain, employing exploration approaches and cost model. The exploration domain is a group of alternate implementation strategy for the input query. The specified query is denoted by means of query strategies in an exploration domain by means of numerous alteration guidelines. If the specified query consists of numerous operations and numerous relations, formerly the exploration domain will be very huge, since it contains a number of correspondent query strategies for the specified query. Examining a huge exploration domain might rises optimization time and price. Therefore, query

Corresponding Author: S.Venkata Lakshmi, Department of Information Technology, Gitam University, Visakhapatnam, India

optimization enforces certain constraint on the dimension of the exploration domain. The issue of discovering an optimum query strategy from group of replacements is resolved by employing numerous exploration approaches. The exploration strategy is employed to discover the search domain so as to obtain an optimum plan from group of replacements by means of a cost model (Ghaemi *et al*., 2008). The user can just define what information is essential to be obtained and the datasets invade the job of discovering the utmost effective methodology of recollecting that information. It is the work of the query optimizer to estimate substitution approaches of performing a query and choosing the finest alternate (Aljanaby *et al*., 2005).

The distributed dataset systems permits merging of information from these disseminated locations using queries. The queries given to a distributed datasets might need processing at diverse locations of the network. The execution of a distributed dataset query hinges on how swift and effectively information is obtained from numerous locations. Quicker recovery of information in a distributed database system is a complicated issue as several locations are includes. Numerous aspects influence the execution of distributed query processing such as the picking of suitable locations, (whenever similar information is duplicated at numerous locations), dimension of the relations in a query, dimension of the datasets in the corresponding sites etc. Owing to the huge number of issues included, there are numerous processing plans for a single query. Every plan is accompanied with a cost and the goal of distributed query optimizer is to obtain a strategy with lowermost conceivable price (optimum strategy).

In query optimization for distributed systems, the price is related with every query execution strategy which is the summation of local cost such as I/O and CPU cost at every location and the rate of transmitting information amongst the sites. The complication and the cost upsurges with the increased number of associations in the query. Additionally, diminishing the quantity of information broad casting is significant to diminish the query processing rate. Owing to the huge number of constraints influencing query execution cost, a particular query can be accomplished in numerous dissimilar methods. A query execution strategy is essential to diminish the cost of query processing.

Need of parameter-less naturally inspired optimization algorithms: Optimized evaluation in maximum evolutionary and swarm intellectual depending approaches are probabilistic methods and need communal regulating constraints like population dimension and number of repetitions, elite size, etc. Besides the mutual regulating constraints, diverse procedures necessitate its individual algorithmic definite regulating constraints. GA employs mutation and crossover rate. Likewise, PSO employs inertia weight, community and intellectual limitations. The appropriate alteration of the algorithmic definite constraint is a very vital aspect for the effective functioning of the evolutionary and swarm intellect dependent methodology as it influences the global performance of the methodology. The inappropriate alteration of the algorithmic definite constraints either augments the computational strength or generates the local optimum solution. Considering this aspects, currently (Kumar and Panicker, 2011) suggested the TLBO approach which do not need any algorithmic definite constraints. TLBO entails merely common regulating constraints like population size and number of repetitions for its functioning. Other evolutionary approaches need the regulation of common regulating constraints in addition to the controller of algorithmic definite constraints. The burden of altering regulating constraint is reasonably lower in the TLBO, consequently the TLBO is humble, efficient and comprises minimum evaluation effort.

Motivation for the study: The foremost issue for the need of query optimization in any distributed dataset systems is to diminish the complete query processing cost for the specified queries. The distributed systems suffer from numerous challenges. The utmost common challenge is the situation where the systems frequently run on erratic and instable atmospheres. Thus it is challenging to generate effective dataset query optimization depending on data accessible at execution time. Specifically, optimality on query plans depends on the approaches of diverse cost models that are presumed on characteristics of query processing. In numerous practical issues, optimization on two or more objective functioning concurrently are essential. These are named as multi-objective optimization issues and results includes detecting no single, however a group of outcomes which signifies the finest conceivable trade-offs amongst objective functions being augmented.

On the other hand, evolutionary approaches have obtained immense popularity owing to their applications in resolving the complex scientific and engineering optimization problems that is are stimulated by the Darwinian evolution which emphasizes the notion of “Survival of the Fittest”. The evolutionary methods are shown to be finely suitable for this group of problems since they can concurrently enhance the dissimilar objectives and obtain effective tradeoffs different from traditional approaches, where the objectives were separately optimized and weighed based on the prior knowledge about the problem in hand. But the Major drawback of the evolutionary algorithms is that the algorithm is highly dependent on the parameter set where the self-adaptive parameters are commonly used. Thus, to overcome all the above mentioned issues of multi objective problem and parameters-less method in combined, this study proposed a novel approach known as Non Dominant based Teacher-Learner Optimization algorithm (ND-TLBO) for the multi objective test functions in the distributed query processing as to obtain the optimized query plans. TLBO is a parametric less optimization algorithm which is enhanced in such a way that in order to employ the multi objective test function into the approach, the dominance property is used between the individuals. Fast non dominated ranking algorithm and crowding measure are the two dominance property employed in this approach. The efficiency of the suggested approach is matched with the existing TLBO query plan generation technique and multi objective GA based query plan generation technique.

Literature rivew: Numerous methodologies have been suggested to enhance conventional query optimization. These approaches comprises of improved measure ments (Poosala *et al.*, 1996), novel approaches for optimization, and adaptive designs for implementation (Avnur and Hellerstein, 2000). Wide spread exploration has been performed for query optimization in distributed datasets. Abundant optimization approaches such as fixed, vibrant and arbitrary approaches are suggested for defining an optimum plan. Never the less, these optimization approaches necessitate understanding of the complete system

Iterative Dynamic Programming (IDP) (Kossmann and Stocker, 2000) is dynamic programming procedure that generates worthy optimization strategies, nevertheless its huge complexity could be expensive if composite queries are essential to be treated. In a novel optimization method known as IDP is recommended which is an amalgamation of dynamic programming and greedy approach. Arbitrarily optimization approaches (Ioannidis and Kang, 1990) construct one or additional initial plans using greedy approach and formerly it attempts to enhance this initial plan by arbitrarily meeting anther nodes. These approaches comprises of simulated annealing and repetitive Enhancement (Nahar *et al.*, 1986) methodologies.

In the researcher suggested a novel GA dependent query optimizer named as NGA that enhances the specified QEP by means of join order, join site and semi join minimizers (Sevinc and Cosar, 2011). The approach was capable to diminish the local processing cost and network broadcasting cost by means of novel values for mutation and crossovers and outstripped comprehensive exploration. The amalgamation of GA and learning automata (Naohoghdem, 2010) for generating optimum QEP on the origin of the join order execution and join site election in disseminated dataset was likewise suggested by the researcher. In researcher suggested an amalgamation of GA and heuristics for resolving join ordering issues since travelling salesman problem in huge database where the computational experimentations demonstrated it to be a feasible result for Distributed Systems (Zhou, 2007). Hybrid of GA and best-worst ACO (Zhou *et al.*, 2009) was executed to obtain the optimum QEP and join order by minimizing query implementation Time for multi-join query optimization in datasets. The procedure performed employed positive response approach of ACO with world wide exploration ability of GA.

Hybrid of GA and ACO (Kadhkhodaei and Mahmoudi, 2011) was executed on Join Ordering issues in Databases by overwhelming the inadequacies of both the approaches. The methodology considers GA to provide pheromone to allocate. And formerly it uses of ant colony procedure to obtain the accuracy of the result. The ability of Hybrid GA-ACO to explore widespread amplitude as to respond for join queries in relational Datasets can be prolonged to enhance the join queries in distributed dataset where the utmost significant complexity is to produce the finest QEP for optimum solutions. In a GA assisted result approach to rapidly define optimum QEP is suggested (Rho and March, 1997). This model comprises of duplicate documentation, advantageous semi joins identification, join site selection, join order implementation and local processing cost and broadcasting cost.

The dispensation of distributed query plans tends to the issues of an comprehensive exploration and not computationally achievable. Additionally, the paper being a combinatorial optimization problem (Lin and Kernighan, 1973), it can be addressed by methods depending on methods such as greedy, evolutionary and arbitraries (Dong and Liang, 2007). Never the less, efficacy of these practices are influenced by the unusual performance, in precise occurrences, of the issue (Stillger and Spiliopoulon, 1996). In a methodology that produces "close" query plans pertaining to the number of sites included and the attention of relations in the locations for a disseminated relational query is defined. As per query processing on smaller number of locations will be further effective and therefore query plans consisting lesser locations required to produce. This kind of query plans, denoted as "close" query plans, are developed by means of GA (Vijay Kumar *et al.*, 2011), deprived of partaking the communication and local processing cost on optimization of QEP's. None of the prevailing methodology considered the rudimentary cost models for optimization.

Teacher Learning Based Optimization (TLBO) approach: TLBO is suggested to attain worldwide outcomes for uninterrupted non-linear functions with fewer evaluation struggle and more reliability. This methodology depends on the result of impact of a teacher on the outcome of students in a session. Now, result is obtained in terms of marks or scores. Similar to other naturally motivated methods, TLBO is a population dependent methodology that employs a population of outcomes to progress towards universal results. The population is deliberated like a cluster of students or a class of students. In optimization procedures, the population comprises of diverse design variables that are equivalent to diverse subjects presented to learners and the learners' output is similar to the 'fitness', similar to other population dependent optimization methodologies. The teacher is considered as the finest result attained till now. The approach of TLBO is separated into two fragments. The initial fragment comprises of the 'teacher stage' and the subsequent fragment comprises of the 'learner stage'. The 'teacher stage' specifies educating from the teacher and the 'learner stage' specifies educating by means of communication amongst learners.

Initialization: Subsequent are the symbolizations employed for illustrating the TLBO:

- N: Number of learners in class i.e. "class size"
- D: Number of courses offered to the learners
- MAXIT: Maximum number of allowable iterations

The population X is arbitrarily initialized by an exploration domain constrained by matrix of N rows and D columns. The j^{th} parameter of the learner is the allocated values arbitrarily by means of the equation:

$$x_{i,j}^{0} = x_j^{min} + rand \times (x_j^{max} - {}_{xj}^{min}) \quad (1)$$

here rand signifies a consistently disseminated arbitrary variable amongst 0 and 1 and signify the lowest and highest value for jth constraint. The constraints of i^{th} learner for the iteration are specified by:

$$x_{(i)}^{g} = [x_{(i,1)}^{g}, x_{(i,1)}^{g}, x_{(i,3)}^{g}, x_{(i,4)}^{g}, \ldots\ldots\ldots x_{(i,d)}^{g} \quad (2)$$

Teacher phase: A worthy teacher is one who conveys its learners nearer to its stage in terms of understanding. However in reality this is not conceivable and a teacher can merely shifts the average of a class to certain degree reliant on the ability of the class. The average constraint M^g of every subject of the students in the class at iterations g is specified as:

$$M^g = \left[M_1{}^g, M_2{}^g, M_3{}^g, \ldots M_D{}^g\right] \quad (3)$$

The learner with the minimized fitness function value is deliberated as the teacher $X^g_{Teacher}$ for individual repetition. The Teacher stage makes the approach proceeded by fluctuating the average of the learners towards its teacher. To attain a novel group of enhanced learners an arbitrary weighted difference vector is constructed from present average and preferred average constraints and summed to the prevailing population of learners:

$$Xnew_i{}^g = X_i{}^g + rand \times (X^g_{Teacher} - T_F.M^g) \quad (4)$$

T_F is the teaching factor that chooses the value of average to be altered. Value of T_F can be either 1 or 2. The value of T_F is defined arbitrarily with identical likelihood as:

$$T_F = round\left[1 + rand(0,1)\{2-1\}\right] \quad (5)$$

Here T_F is not a constraint of the TLBO approach. The value of T_F is not specified as an input to the approach and is arbitrarily defined using Eq 5. Once performing numerous experimentations on several benchmark functions it is decided that the approach accomplishes well if the value of T_F is amongst 1 and 2. Nevertheless, the procedure is discovered to execute more enhanced if the value of is either 1 or 2 and henceforth to streamline the procedure, the teaching factor is proposed to be either 1 or 2 reliant on the turning up standards specified by Eq. 5. If is identified to be a greater learner compared to in iteration, formerly it substitutes lesser learner in the matrix.

Learner phase: Learners augments its awareness by means of two diverse ways: one via input from the teacher and another via communication amongst themselves. A learner interrelates arbitrarily with one another with the assistance of group consultations, demonstrations, proper communications, etc and acquires some novel things if the other student has higher understanding compared to him or her. In this stage, the collaboration of learners with each other is performed. The method of communal communication leads to upsurge the information of the learner and the arbitrary communication amongst learners progresses his or her understanding. For the specified learner X_i^g, the other learner X_i^g is haphazardly designated and. The constraint of the matrix in the student stage is defined as:

$$Xnew_i{}^g = \begin{cases} x_i{}^g + rand \times (x_i{}^g - x_r{}^g) if\ F(X_i{}^g) F(X_r{}^g) \\ x_i{}^g + rand \times (x_i{}^g - x_r{}^g) \quad otherwise \end{cases}$$

Discovering the universal optimal value(s) of an issue comprising of more than single objective with contradictory environment rises in numerous technical appliances. The issue of optimization comprising more than single objective function with contradictory landscape is named as Multi-Objective Optimization (MOO) issue. MOO has been demonstrated as defining a vector of conclusion variables whereas enhancing numerous objectives concurrently with a specified group of limitations. The Multiple Objective TLBO is a novel multi objective evolutionary approach suggested for resolving the utmost vital optimization issue in query plan generation in DDS. The suggested approach is a multi-objective alteration of the TLBO approach, a population dependent optimizer that specifies a group of individuals with the intention of accumulating its information through dissimilar learning stages. Hence, for resolving Multiple Objective optimization issue, novel attributes are integrated to the method. In this study, the Multi Objective TLBO (MO-TLBO) approach is presented. This novel multiple objective evolutionary approach integrates famous multiple objective concepts like the dominance concept.

MATERIALS AND METHODS

Proposed non dominant based teacher learning optimization algorithm for query plan generation: In this study, a novel methodology is proposed for development of optimum query plans using query optimization approach in the distributed database systems. The novel query optimization approach that is employed in this paper is different from the most of the well-known evolutionary optimization approaches that overcomes the flaws in these Existing approaches. The introduced technique for query optimization is a naturally and heuristically inspired algorithms which is a parameter less methodology employed and popularly known as Teacher-Learner Based Optimization (TLBO). For this approach the dominant rule is given in Algorithm 1 and the proposed approach is specified through Algorithm 2 followed by the diagrammatic representation of the approach in Fig. 1. This study proposed a novel and enhanced teacher-learner based optimization approach that generates the query plans by considering a bi-objective for the optimization of total query processing time.

Multi objective Distributed Query Plan Generation (DPQG) problem formulation: Traditionally for the query plan generation in the genetic algorithm based distributed database systems has single objective that is to minimize the total query processing cost. It is clearly, witnessed that the total processing cost is the combination of two other costs that is local processing cost present in the participating sites known as Total Local Processing Cost (TLPC) and Communication cost in between the sites known as Total Communication Cost (TCC). Consequently, it can be viewed as bi-objective DPQG problem where, it is essential to minimize the total local processing cost and total communication cost simultaneously to achieve the acceptable trade off amongst the two objectives. The TLPC and TCC is mathematically given as:

$$\mathrm{TCC} = \sum_{i=1, j=i+1}^{i=u-1, j-u} \mathrm{CC}_{ij} \times b_i$$

$$\mathrm{TLPC} = \sum_{i=1}^{u} \mathrm{LPC}_i \times a_i$$

Where:

- u = The number of locations accessed through the query plan in ascending order of cardinality per sites
- Cc_{ij} = The communication cost per bytes amongst locations and j
- LPC_i = The local processing cost per byte at location I
- b_i = The bytes to be broadcasted from location i
- a_i = The bytes to be processedat site i

If a site contains a single relation, its LPC is considered to be zero here:

$$\mathrm{Card}_i = \frac{\mathrm{Card}(R_t) \times \mathrm{Card}(R_s)}{\mathrm{Dist}_{ts} \min(\mathrm{Card}(R_t), \mathrm{Card}()R_s)}$$

where, Dist_{ts} is the number of distinct tuples in the minimum relation amongst R_t and R_s. The dimension of the resultant relation Ri at site is given as:

$$\mathrm{Size}_i = \mathrm{Size}(R_t) + \mathrm{Size}(R_s)$$
$$\alpha_i = \mathrm{Card}\, d_i \times \mathrm{size}_i,\; b_j = \mathrm{Card}\ d_j \times \mathrm{size}_j$$

This multi objective DPQG problem in this approach is solved by using the non-dominated ranking approach for every individual in the population and assigning them accordingly into the non-dominated fronts. While ranking the individuals, if there occurs a tie in between any individuals then a crowding distance measure is used for the selection which is a parameter-less

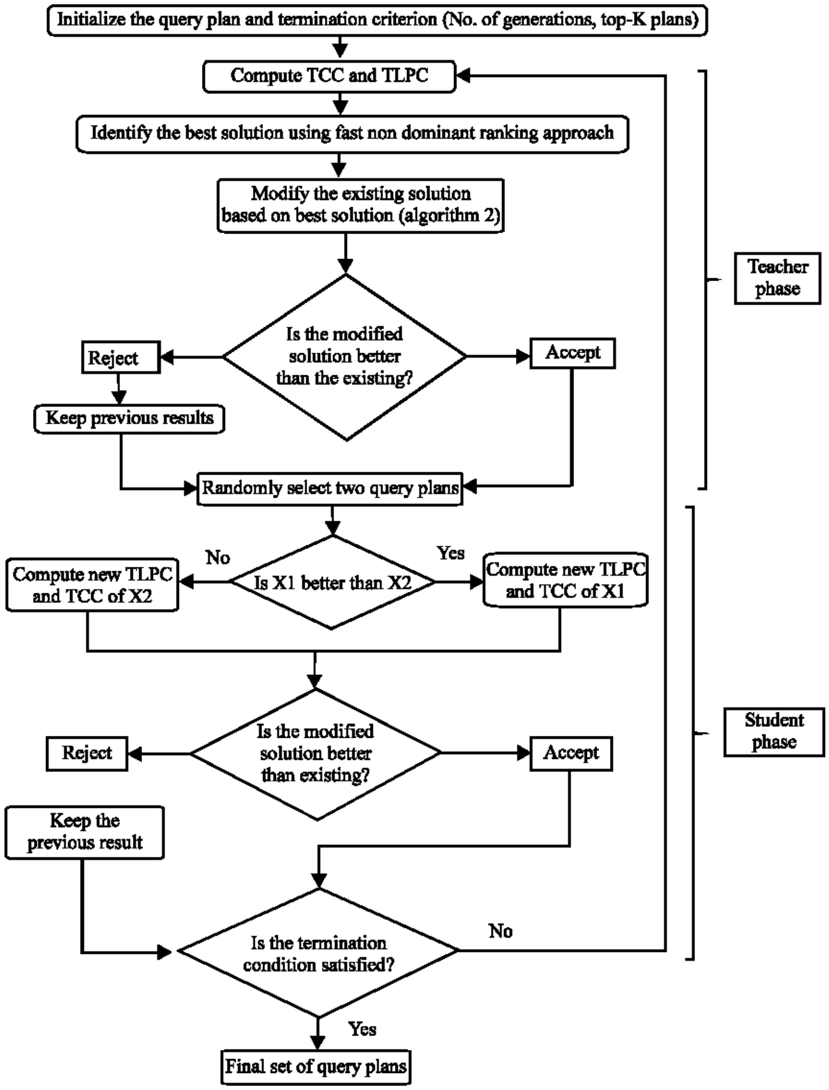

Fig. 1: Diagrammatic representation of proposed approach

sharing approach. The crowding distance is an evaluation of the thickness of results adjoining a specific result in the objective region. The crowding distance is given as:

$$I(d_k) = I(d_k) + \frac{I(K+1)_m - I(K-1)_m}{f_m^{max} - f_m^{min}}$$

Here, $I(K)_m$ is the value of the mth objective function of kth query plan in the rank and f_m^{max} and f_m^{min} are the maximal and minimal values obtained for the objective function m.

The non dominated sorting solution is preferred while ranking the individuals and between the solutions having the similar rank, a solution that exist in the lesser congested area is desired that is with higher crowding distance. In order to perform a non-dominated ranking, each query plan is matched with every other query plan in the population to detect if it is dominated.

The fast non dominated sorting procedure takes the current population as input and produces a list of non-dominated ranks as output. For every query plan "i", the following entities need to considered:

- n_i: The number of query plans that dominate the query plan i
- $Site_i$: The set of query plans that query plan I dominates

Fast non dominated sorting algorithm for query plan generation (Algorithm 1):

1. Consider the initial population size with number of Query Plan initialized.
2. Compute the TPLC and TCC for each and every Query Plan in the Population.
3. Now check for the condition
   a. Initialize $site_i = \{\phi\}$ and $n_i = 0$ and rand = 0
   b. for I = = 1: pop-size
      for j = 2: pop-size
      if $TCC_i < TCC_j$&& $TPLC_i < TPLC_j$
      $Site_i = Site_i \cup \{j\}$
      $n_j = n_j+1$
   c. For j =1: pop-size
      for all n_j =, rank = 0,
      then reduce all n_j by 1 and prioritiz the rank to the next increased number when every n_j = 0 or length $\{Site_i\}$ = 1untill all values are rank with the numbers
4. Finally the ranked query plans are stored

Enhanced teacher-learning based optimization approach: The non-dominated based Teacher-Learner Optimization Algorithm (ND-TLBO), initially begins through initializing the complete group of query plans for the specified individual query and employing pre-defined Relation-Site Matrix (RMS), these query plans in solution domain are comparable to learners of TLBO. A Decision maker or query optimizer delivers values of stopping condition for TLBO in the course of the initialization phase, eg. No. of iterations, values of K for Top-K query plans, average query cost etc. In the teacher stage, students study via the teacher, a teacher efforts to augment the average outcome of the class in the subject that a teacher teaches pertaining to its own ability, diverse design objective are equivalent to the diverse subjects imparted by teacher and the corresponding subject values signify the rank of certain student in the subject. The finest of complete outcome considering the complete subjects combined in the population of learners could be the solution of the finest learner. Therefore, in this approach, the best query plan is identified from the entire population which is obtained by employing fast non dominant ranking technique initially, then the existing solution space is modified using the best query plan and the newly obtained solutions is checked whether it is better than existing or not. Once the modified solution is accepted, this finishes the teacher stage and moves to the learner or student stage.

In this second phase, any two random query plans are selected from the modified solution space and best plan is identified by means of fast non dominant ranking technique once again and checked whether the obtained result is better than the existing. If the result is best then solution space is updated and checked for the termination condition. The termination condition that is employed in this approach is the maximum number of iterations. At the situations while applying non dominant approach, if any tie occurs in between the obtained best query plans, then the crowding distance measure is evaluated amongst them.

Non Dominant Based Teacher-Leaner Optimization Algorithm (Algorithm 2): This Approach works on two phases. They are:

- Teacher Phase:

1. Initialize the number of Query Plans as the population dimension and the termination criteria as maximum number of generations, the top-k query plans.
2. Compute Bi-Objective test functions such as Minimum TLPC and Minimum TCC as different design variables.
3. Identify the best Query plan amongst the initialized Query Plans by applying the Fast Non Dominated Sorting Algorithm on TLPC and TCC. If any tie occurs, then compute Crowding Distance Measure on the TLPC and TCC of the respective query plans.
4. Modify the existing query plans based on the obtained best solution by means of Computed TLPC and TCC as:

$$TPLC_{modified} = r\ (TPLC\ best - TF(TPLC_{existing}))$$

The random value r is in between [0, 1] and TF = round [1+rand (0, 1), 2-1]

5. Now the modified TLPC and TCC values of the Corresponding Query Plans are compared with the Existing TLPC and TCC values by using Fast Non Dominated Sorting Algorithm.
6. If the modified values are better than the existing then existing values are replaced the new ones
Else the existing solutions remains similar

Learner phase:

1. Any two Query Plans and are selected randomly amongst the modified solution space and by means using Fast Non Dominated Sorting Algorithm, the Query plan having the best TLPC and TCC values is selected.
2. If QP_1 then compute

$$TPLC_1^{new} = TLPC_1^{old}+r\ (TLPC_2-TLPC_1^{old}) \text{ and}$$
$$TCC_1^{new} = TCC_1^{old}+r\ (TCC_2-TCC_1^{old})$$

3. Else compute

$$TPLC_2^{new} = TLPC_2^{old}+r\ (TLPC_1-TLPC_2^{old}) \text{ and}$$
$$TCC_2^{new} = TCC_2^{old}+r\ (TCC_1-TCC_2^{old})$$

4. If the modified values are better than the existing then existing values are replaced the new ones
Else the existing solutions remains similar.

Once the single iteration of Teacher Phase and Learning Phase is completed, then check for the termination criteria with maximum number of iterations and select the Top-K Query plans from the Obtained Solution Space.

RESULTS AND DISCUSSION

The proposed Non Dominated Based Teacher-Learner Optimization (ND-TLBO) algorithm was implemented in Matlab 14a in Windows 8 professional 64 bit OS. The Experiment was carried out for population size of 20 query plan with every plan comprising 10 relations disseminated on 10 different sites. The experimental analysis for the proposed approach is carried by constructing a Site Relation Matrix (SRM) for the distributed database and the initial population of different query plans of size 20 are considered. Along with this, the necessary computations for the query optimization approach are done like calculating the total local processing cost and total communication cost for query plan where the best query plan is selected by using the fast non-dominated ranking approach for the query plan.

The performance analysis of the anticipated methodology is matched with the existing Single Objective Teacher-Learner Based Optimization technique (SO-TLBO) for distributed query plan generations and with the Multi Objective Genetic Algorithm (MO-GA) for distributed query plan generation. The average query cost for all the three methodologies are computed and compared against the number of generations in the respective Methodologies. The experiment is carried out for the Top-5 and Top-10 query plans. Table 1 and Fig. 2 demonstrate that the Average Query Cost for the Top-5 Query Plan and it is shown that the Query cost of the Proposed ND-TLBO is less compared to the other two existing techniques for increased number of generations. Similarly, Table 2 and Fig. 3 shows that the average query cost for the top-10 query cost for the proposed approach is minimum when compared to other two techniques. However, the query cost for top-5 query plan is minimum when compared with top-10 query plans.

Table 3 and Fig. 4 demonstrate that the number of generations employed for the top-k query plan with average query cost = 0.45 and it is shown that the number of generation of the proposed ND-TLBO is less compared to the other two existing techniques for different top-K query plans. Similarly, Table 4 and Fig. 5 shows that the number of generation employed for the

Table 1: Performance comparison of proposed ND-TLBO, MO-GA and MO-TLBO for top-10 query plan

	Average query cost		
No. of generations	MO-GA	SO-TLBO	ND-TLBO
20	0.76	0.75	0.74
40	0.57	0.55	0.54
60	0.47	0.49	0.32
80	0.33	0.32	0.25
100	0.24	0.23	0.2

Table 2: Performance comparison of proposed ND-TLBO, MO-GA and MO-TLBO for top-5 query plan

	Average query cost		
No. of generations	MO-GA	SO-TLBO	ND-TLBO
20	0.73	0.7	0.68
40	0.64	0.62	0.6
60	0.5	0.47	0.45
80	0.36	0.35	0.3
100	0.31	0.3	0.29

Table 3: Performance comparison of proposed ND-TLBO, MO-GA and MO-TLBO for average query cost = 0.45

	No. of generations		
Top-K query plan	MO-GA	SO-TLBO	ND-TLBO
3	85	80	70
5	100	100	80
8	105	110	100
10	115	120	110
15	125	130	120

Table 4: Performance comparison of proposed ND-TLBO, MO-GA and MO-TLBO for average query cost = 0.25

	No. of generations		
Top-K query plan	MO-GA	SO-TLBO	ND-TLBO
3	65	60	55
5	75	70	60
8	87	80	70
10	98	90	85
15	107	100	95

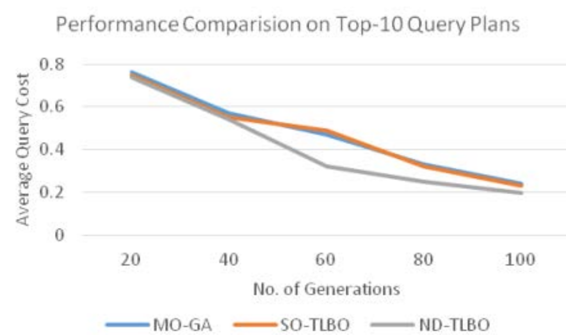

Fig. 2: Performance Comparison of Proposed ND-TLBO, MO-GA and MO-TLBO for Top-10 Query Plan

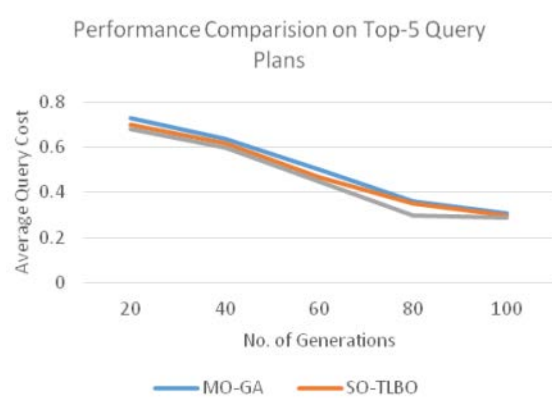

Fig. 3: Performance comparison of proposed ND-TLBO, MO-GA and MO-TLBO for Top-10 query plan

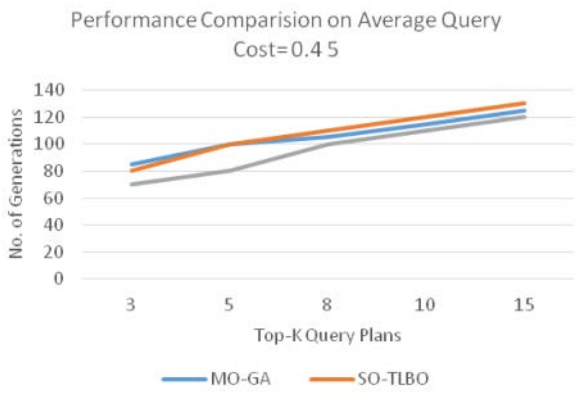

Fig. 4: Performance comparison of proposed ND-TLBO, MO-GA and MO-TLBO for average query cost = 0.45

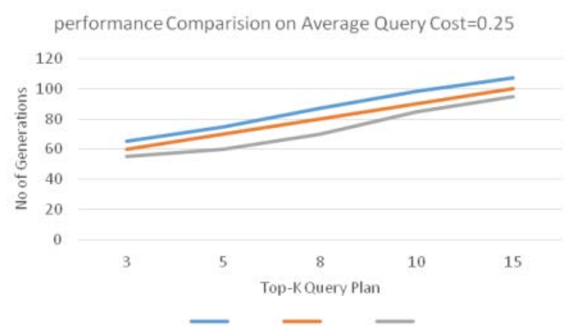

Fig. 5: Performance comparison of proposed ND-TLBO, MO-GA and MO-TLBO for average query cost = 0.25

Top-k query cost with average query cost = 0.25 for the proposed approach is minimum when compared to other two techniques. However, number of generations essential for average query cost 0.25 is minimum compared to average query cost 0.45.

CONCLUSION

Query optimization is one of the utmost significant step in the DDS for disseminated query processing. Distributed systems are a group of self-determining collaborating systems that permits accumulation of information at geologically discrete positions, depending on the occurrences of access through individual's local site to a global site. The performance of a distributed dataset query hinge on how quick and professionally information is obtained from numerous sites. Quicker accessing of information in a DDS is a complicated issue as manifold sites are included. This study proposed a novel bi-objective test function based query optimization in the distributed databases where, the parametric less naturally inspired algorithm is employed and is entirely different from the well-known evolutionary algorithms. The performance evaluation of the suggested approach is matched with the existing multi objective and parametric less methodologies and is shown that the obtained results that is average query processing cost is minimized for the different top-k query plans. This shows that the suggested method is efficient and effective for the multi objective query optimization in distributed systems.

REFERENCES

Aljanaby, A., E. Abuelrub and M. Odeh, 2005. A survey of distributed query optimization. Int. Arab J. Inf. Technol., 2: 48-57.

Avnur, R. and J.M. Hellerstein, 2000. Eddies: Continuously adaptive query processing. ACM. SIGMoD. Rec., 29: 261-272.

Ceri, S. and G. Pelagati, 1984. Distributed Database: Principles and System. McGraw Hill, New York, USA.,.

Dong, H. and Y. Liang, 2007. Genetic algorithms for large join query optimization. Proceedings of the 9th Annual Conference On Genetic And Evolutionary Computation, July 7-11, 2007, ACM, London, England, UK, ISBN: 978-1-59593-697-4, pp: 1211-1218.

Ghaemi, R., A.M. Fard, H. Tabatabaee and M. Sadeghizadeh, 2008. Evolutionary query optimization for heterogeneous distributed database systems. World Acad. Sci., 43: 43-49.

Ioannidis, Y.E. and Y. Kang, 1990. Randomized algorithms for optimizing large join queries. ACM. Sigmod Rec., 19: 312-321.

Kadkhodaei, H. and F. Mahmoudi, 2011. A combination method for join ordering problem in relational databases using genetic algorithm and ant colony. Proceedings of the 2011 IEEE International Conference on Granular Computing (GrC), November 8-10, 2011, IEEE, Tehran, Iran, ISBN: 978-1-4577-0372-0, pp: 312-317.

Kossmann, D. and K. Stocker, 2000. Iterative dynamic programming: A new class of query optimization algorithms. ACM. Trans. Database Syst. TODS., 25: 43-82.

Kumar, T.V. and S. Panicker, 2011. Generating Query Plans for Distributed Query Processing using Genetic Algorithm. In: Information Computing and Applications. Baoxiang, L. and C. Chunlai (Eds.). Springer Berlin Heidelberg, Berlin, Germany, ISBN: 978-3-642-25254-9, pp: 765-772.

Kumar, T.V., V. Singh and A.K. Verma, 2010. Generating distributed query processing plans using genetic algorithm. Proceedings of the 2010 International Conference on Data Storage and Data Engineering (DSDE), February 9-10, 2010, IEEE, New Delhi, India, ISBN: 978-1-4244-5678-9, pp: 173-177.

Lin, S. and B.W. Kernighan, 1973. An effective heuristic algorithm for the traveling-salesman problem. Operat. Res., 21: 495-516.

Nahar, S., S. Sahni and E. Shragowitz, 1986. Simulated annealing and combinatorial optimization. Proceedings of the 23rd ACM/IEEE Conference on Design Automation, June 29-July 2, 1986, IEEE Press, Piscataway, New Jersey, USA, ISBN: 0-8186-0702-5, pp: 293-299.

Ozsu, M.T. and P. Valduriez, 1991. Principles of Distributed Database System. 2nd Edn., Prentice Hall, Englewood Cliffs, New Jersey, USA., ISBN: 9780136916437, Pages: 562.

Poosala, V., P.J. Haas, Y.E. Ioannidis and E.J. Shekita, 1996. Improved histograms for selectivity estimation of range predicates. ACM. SIGMOD. Rec., 25: 294-305.

Rho, S. and S.T. March, 1997. Optimizing distributed join queries: A genetic algorithm approach. Ann. Oper. Res., 71: 199-228.

Sevinc E. and A. Cosar, 2010. An evolutionary genetic algorithm for optimization of distributed database queries. Comput. J., Vol. 54, 10.1093/comjnl/bxp130.

Stillger, M. and M. Spiliopoulou, 1996. Genetic programming in database query optimization. Proceedings of the 1st Annual Conference on Genetic Programming, July 28-31, Stanford, California, pp: 388-393.

Zhou, Y., W. Wan and J. Liu, 2009. Multi-joint query optimization of database based on the integration of best-worst ant algorithm and genetic algorithm. Proceedings of the IET International Communication Conference on Wireless Mobile and Computing (CCWMC 2009), December 7-9, 2009, IEEE, Shanghai, Chona, ISBN: 978-1-84919-138-8, pp: 543-546.

Zhou, Z., 2007. Using heuristics and genetic algorithms for large scale database query optimization. J. Inf. Comput. Sci., 2: 261-280.