Applied Soft Computing 62 (2018) 571–578

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Full Length Article

A new approach based on particle swarm optimization algorithm for solving data allocation problem

Mostafa Mahi [a], Omer Kaan Baykan [b], Halife Kodaz [b,*]

[a] Department of Computer Engineering and Information Technology, Payame Noor University, PO Box 19395-3697 Tehran, Iran
[b] Department of Computer Engineering, Faculty of Engineering, Selcuk University, Konya, Turkey

ARTICLE INFO

Article history:
Received 1 December 2016
Received in revised form 18 October 2017
Accepted 12 November 2017

Keywords:
Data allocation problem
Particle swarm optimization
Distributed databases system
Site-fragment dependency

ABSTRACT

The effectiveness distributed database systems highly depends on the state of site that its task is to allocate fragments. This allocation purpose is performed for obtaining the minimum execute time and transaction cost of queries. There are some NP-hard problems that Data Allocation Problem (DAP) is one of them and solving this problem by means of enumeration method can be computationally expensive. Recently heuristic algorithms have been used to achieve desirable solutions. Due to fewer control parameters, robustness, speed convergence characteristics and easy adaptation to the problem, this paper propose a novel method based on Particle Swarm Optimization (PSO) algorithm which is suitable to minimize the total transmission cost for both the each site – fragment dependency and the each inter – fragment dependency. The core of the study is to solve DAP by utilizing and adaptation PSO algorithm, PSO-DAP for short. Allocation of fragments to the site has been done with PSO algorithm and its performance has been evaluated on 20 different test problems and compared with the state-of-art algorithms. Experimental results and comparisons demonstrate that proposed method generates better quality solutions in terms of execution time and total cost than compared state-of-art algorithms.

1. Introduction

One of the most crucial applications that recently have attracted attentions is distributed database such as Data Allocation Problem (DAP). The purpose of the DAP is to determine a placement of fragments at the best different sites in order to minimize the total transaction cost when a query is taken from one site to another. DAP refers to a standard test problem used in performance analysis of optimization algorithms with certain constraints [1]. Nowadays it is extremely important to efficiently allocate data to sites. In the real world, the data used by search engines or mail servers are very large and disorganized. The locations of those who request data can change. In this case, it is necessary to efficiently organize the data. For example some issues such as parallel query execution, network load and servers load balancing are design factors that need to be handled. In the real world without considering of mentioned parameters, DAP is an NP-hard problem. Two types of static and dynamic algorithms are presented to solve DAP. Static algorithm refers to data allocation based on static transaction execution pattern in the target environment whereas these patterns can be changed in dynamic algorithms [2–5].

There are some similarities between a Quadratic Assignment Problem (QAP) with a DAP, such that, it keeps track of the resource locality. A QAP is developed by Koopmans and Beckman [6] and is one of the fundamental combinatorial optimization problems. QAP refers to the dependency of sites and fragments, the sites are pointed as locations and the fragments as facilities.

There is several solutions to solve DAP and QAP problems. Some solution methodologies are detailed in related works (Section 2). Among the previous solution methodologies developed for solving DAP and QAP, Particle Swarm Optimization (PSO) algorithm has been used due to its speed convergence characteristics, solution quality, robustness and easy adaptation to solve the problem in this study. There is no overlapping and mutation calculation over PSO algorithm. In the PSO-DAP method, search operation is performed by the speed of particle and during several generations, only the most optimist particles can transmit information to others. In this case researching speed will be so high [7,8].

In this study, we focus on solving DAPs by utilization and adaptation of PSO. According to the literature review, PSO has not been used to solve DAPs. Due to fact that PSO has fewer control parameters, speed convergence characteristics and lower consuming time,

* Corresponding author.
E-mail address: hkodaz@selcuk.edu.tr (H. Kodaz).

https://doi.org/10.1016/j.asoc.2017.11.019
1568-4946/

robustness against to solution space of the optimization problems, it is frequently used to solve optimization problems with different characteristics by the practitioners and researchers [7,8]. In this study, DAP which is NP-hard problem is solved by PSO and the performance of the proposed algorithm is compared with the performance of genetic algorithm [1], tabu search [1], ant colony algorithm [1] and simulated annealing [9] on solving 20 problems with different dimensionalities. The execution time is important indicator and factor for the problem like total cost. The proposed algorithm produced the acceptable and comparable results in a reasonable time when we compare the execution time of PSO-DAP with other algorithms, the best results belong to proposed PSO-DAP. Additionally, a new dataset is prepared for comparison of future studies.

In Section 1, the study is introduced and the rest of the paper is organized as follow: The related works and methodologies for solving DAP and QAP are given in Section 2. Section 3 refers to material and methods that include background information for PSO algorithm, and presents the proposed method (PSO-DAP). Experiments and comparison are conducted in Section 4 and finally the study is concluded in Section 5.

2. Related works

In this section, some studies proposed for solving DAP and QAP are reviewed and summarized. As mentioned earlier, DAP and QAP are known as NP-hard problem and various meta heuristic methods such as Genetic Algorithm (GA) [10,11] and Ant Colony Optimization (ACO) algorithm [1,2] for solving these problems. Other studies are as follow:

Sen et al. proposed a Simulated Annealing (SA) method to solve data allocation problem [12]. They have analyzed the SA approach with benchmarks that are obtained from CPLEX benchmarks.

Wang et al. have modeled the Radio-Frequency Identification (RFID) tag oriented data allocation problem as a nonlinear knapsack problem [13]. They have utilized the artificial immune network (DA − aiNet) to address it. Some numerical assessments have been done to illustrate the effect of memory capacity and correlation matrix. Also, they have compared it with other present methods.

A set of Teaching Learning Based Optimization (TLBO) methods based on hybrid algorithms to solve the challenging combinatorial optimization QAP is suggested by Dokeroglu. Recombination operators and later a Robust Tabu-search engine that processing them, were used to train individuals. Sequential and parallel TLBO-based hybrid algorithms efficiency was compared with those of the mentioned heuristics in terms of the best solution and computational effort. The performance of proposed algorithms demonstrate that they are competitive with the best algorithms for the solution of the QAP with which many real life problems could be modeled [14].

A novel algorithm for replicated fragment allocation during distributed database design for static environment using Biogeography-Based Optimization (BBO) is presented by Singh et al. Mentioned algorithm targeted to design a replicated fragments method to minimize the sum of data transmission cost and storage cost of fragments. The results of biogeography – based optimization algorithm for data allocation were compared with GA to consider the effectiveness of the technique [15].

A new recombination operator based on Order-1 crossover algorithm was suggested by Tosun which runs the quick sort partitioning algorithm to produce various chromosomes for partitions. Offspring's with the other chromosomes are generated with the minimum cost partition [16]. The GA, the fast ACO and the robust Tabu-search for solving the DAP are used by Tosun. The presented method efficiency for sample size smaller than 50 with respect to the optimal results proposed in QAPLIB is outstanding.

Li and Wong have used time series modeling approach to predict the shot-term load [17]. In this approach the node number adjustment and fragment reallocation is considered to omit node overloading's and fragment reallocation originated from fragment mitigations.

In 2013, Tosun et al. have proposed a collection of SA, GA and Fast ACO to solve DAP [9]. Some GA crossover operator for QAP is explained by Tosun et al. and their effectiveness is assessed using well-known samples of QAPLIB library. The parameters of GA (such as the number of processing units, size of initial population and generations) are set similar to the best solutions for large QAP samples described in the library. The solution batches on each processing clusters are separated to design an efficient parallel GA using an island model. At the first level of optimization process, each processors solutions trade ten percent rate. By using the island Parallel GA, there are improvement in both run times and goodness of the solutions for large QAP samples [10].

Abdalla, by bringing a change to data access pattern, proposed a new data re-allocation model for replicated and non-replicated constrained Distributed Database Systems (DDBSs). Mentioned method assumption is based on a distribution of fragments over network sites that were done according to a properly forecasted set of query frequency values which could be employed over sites that takes sites' restrictions into account in the re-allocation phase. The proposed approach is an efficient plan to re-allocate data fragments across sites based on communication and update cost values for each fragment individually. The re-allocation process was founded on selecting the maximal update cost value for each fragment and making the re-allocation accordingly. Results demonstrate that mentioned algorithm act as an efficient algorithm in solving fragments re-allocation problem in a dynamic distributed relational databases environment [18].

Another data re-allocation model for DDBSs by altering procedure of data access over sites is proposed by Amer and Abdalla, which assumed that the scattering of fragments over network sites was initially achieved according to an appropriately predicted collection of query frequency values that could be employed over sites. In their assumption a plan to re-allocate data fragments based on communication costs between sites and update cost values for each fragment is very important. The re-allocation process was based on selecting the maximum update cost value for each fragment and deciding on the re-allocation accordingly. It is confirmed that the proposed technique will effectively contribute in solving dynamic fragments re-allocation problem in the context of a distributed relational database systems [19].

In 2009, Adl and Rankoohi have proposed ACO-DAP model based on the ACO and a local search [2]. Another heuristic GA for solving Resource Allocation Problems (RAPs) was proposed by Lee et al. in 2003 [20]. Its objective was to overcome RAPs which are appeared in practice. Their studies were based on several genetic algorithms and simulation results confirmed that the proposed algorithm had a good performance. A novel dynamic DDBS, called the threshold algorithm was proposed by Ulus and Uysal in 2003 [21]. This algorithm performs data reallocation via changing data access procedure. Simulation was a means to analyze a fragment. Where data access pattern changed dynamically, threshold algorithm was suitable for the DDS.

In 2002, Ahmad and Karlapalem have modeled the dependency between the fragments achieved by a query using a dependency graph. The graph has been utilized to formulate and tackle DAPs for distributed database systems based on query-site and move-small query execution strategies [22]. Evolutionary algorithms for data allocation in distributed database systems were designed and evaluated in their algorithm.

As a result of the literature review, new methods are needed to increase the cost and time efficiency of the methods proposed for