Implementation and Comparative Analysis of k-means and Fuzzy c-means Clustering Algorithms for Tasks Allocation in Distributed Real Time System

Harendra Kumar, Gurukula Kangri Vishwavidyalaya, Haridwar, India
Isha Tyagi, Gurukula Kangri Vishwavidyalaya, Haridwar, India

ABSTRACT

Distributing tasks to processors in distributed real time systems is an important step for obtaining high performance. Scheduling algorithms play a vital role in achieving better performance and high throughput in heterogeneous distributed real time systems. To make the best use of the computational power available, it is essential to assign the tasks to the processor whose characteristics are most appropriate for the execution of the tasks in a distributed processing system. This study develops two algorithms for clustering the heavily-communicating tasks to reduce the inter-tasks communication costs by using k-means and fuzzy c-means clustering techniques respectively. In order to minimize the system cost and response time, an algorithm is developed for the proper allocation of formed clusters to the most suitable processor. The present algorithms are collated with problems in literature. The proposed algorithms are formulated and applied to numerous numerical examples to demonstrate their effectiveness.

KEYWORDS


1. INTRODUCTION

Distributed real time system (DRTS) consists of a set of nodes (computer) interconnected by a real-time communication network in some fashion so that a program or procedure runs on a system with multiple processors. DRTS has many applications in various fields such as telephone networks, cellular networks, computer games, image processing, cryptography, industrial process monitoring simulation of nuclear reactor, air-plane, banking system, transportation problems and network flow problem etc. Tasks allocation is a process of partitioning a program into a number of tasks where each task executes on a separate processor. Tasks partitioning and tasks allocation are two major steps in the formation of DRTS. In order to make best use of the resources, it becomes essential to maximize the overall throughput by allocating the tasks to processors in such a way that the allocated load on all the processors should be balanced and also minimize the inter-tasks communication by assigning tasks to same processor as much as possible. These conditions contrary to each other as an increase in the number of processors may actually decrease the total throughput of the system. This degradation effect is known as saturation effect, which occurs due to heavy communication traffic induced by data transfers between tasks that reside on separate processors. An allocation policy can
be static or dynamic, depending upon the time at which the allocation decisions are made. In static allocation once a task is assigned to a processor, it remains there statically during the execution of a distributed program.

It is found that if the fulfillments of these steps are not properly completed then an increase in the number of processors may result in a decrease of total throughput of the system (Chu et al., 1978; Crespo et al., 2016). Graph theoretic approach is widely used for tasks allocation problem with the objective of minimizing total execution and communication costs (Ma et al., 1982; Bokhari, 1987; Shen and Tsaı, 1985). A heuristic algorithm (Lo, 1988) has been developed for stone’s classic model to minimize the total execution and communication costs. In recent years, several heuristic models have been developed (Efe, 1982; Tindell et al., 1992; Woodside and Monforton, 1993; Sharma et al., 2011; Martinez et al., 2015; Akbari and Rashidi, 2016; Kumar et al., 2018) for tasks allocation with the goal of minimizing system cost and maximizing the system throughput. Branch and bound method technique are efficiently used for optimal tasks assignments to minimize the sum of task execution and communication costs (Peng et al., 1997). A heuristic algorithm has been derived from well-known simulated annealing (SA) technique with the goal of maximizing the system reliability (Attiya and Hamam, 2006). Artificial Neural Network has been used for finding the optimal assignment of the tasks in DRTS (Yadav et al., 2011). In order to make the best use of resources in a DRTS, it is necessary to reassign tasks dynamically during program execution (Kong et al., 2011; Kumar et al., 2016). Recently, a two-phase scheduling technique is presented for DRTS in which it has two distinct phases; the first phase is in charge of producing a scheduling sequence whereas the second phase aims to dispatch tasks to computing nodes of a DRTS (Jiang et al., 2017).

Clustering is one of the standard workhorse techniques in field of DRTS. Its intention is to systematize a dataset into a set of groups or clusters, which contain similar data items, as measured by some distance function. Clustering algorithms are developed to solve a particular problem in a specialized field (Jain et al., 1999). Fuzzy clustering technique has been used by many researchers (Yadav et al., 2011; Lemos et al., 2013; Velmurugan et al., 2014; Zhou et al., 2014; Nayak, 2016).

Recently, the developments of clustering theory have demonstrated successful applications in diverse fields. They are able to provide quality solutions with reasonable time. Inspired by the success of clustering techniques, in this paper two algorithms are designed for clustering the heavily communicating tasks to reduce the inter-tasks communication costs by using k-means clustering (KMC) technique and fuzzy c-means clustering (FCMC) technique, respectively. For the mapping of formed clusters to the most suitable processor to reduce the system cost and response time of the system, an allocation algorithm is also developed in the present paper. The main contributions of the present study are as follows:

- Formation of tasks clusters on the basis of inter-tasks communication costs.
- Development of a heuristic model for the optimal assignment of task clusters onto the processors.

The remainder of this paper is organized as follows. Some basic preliminaries are described in section 2. Section 3 formulates the DRTS for response time and system cost optimization. Section 4 describes the proposed model. Section 5 reports the comparative performances. Finally, Section 6 concludes this work.

2. BACKGROUND INFORMATION

2.1. K-Means Clustering (KMC)

K-means is one of the popular unsupervised learning algorithms that have been most popularity used to solve the clustering problems. K-means was first used by Macqueen (1967). K–means algorithm is very simple and can be easily implemented in solving many practical problems. This k-means clustering algorithm aims at minimizing an objective function known as squared error function given by-
Buffer Managed Multiple Replication Strategy Using Knapsack Policy for Intermittently Connected Mobile Networks
www.igi-global.com/chapter/buffer-managed-multiple-replication-strategy-using-knapsack-policy-for-intermittently-connected-mobile-networks/105669?camid=4v1a

Subscription Policy Control Framework for IMS-Based Networks
www.igi-global.com/article/subscription-policy-control-framework-ims/34038?camid=4v1a