IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-2, Feb 2019 E-ISSN: 2347-2693

A Novel Scheduler for Task scheduling in Multiprocessor System using Machine Learning approach

Nirmala H[1*], Girijamma H A[2]

[1,2]Dept. of CSE, RNSIT, VTU, Bangalore, India

[*]*Corresponding Author: hnirmala@sjbit.edu.in*

DOI: https://doi.org/10.26438/ijcse/v7i2.140143 | Available online at: www.ijcseonline.org

Accepted: 26/Feb/2019, Published: 28/Feb/2019

Abstract In today's computing world scheduling of real time task in a multiprocessor environment is very crucial. To do the scheduling, how the scheduler is implemented? what parameters are considered ? and how those parameters affect? Is also very important. Using the realistic parameters of the task the scheduling can be done and predict the resource requirement and analysis of the resource utilization factor can be done. Based on the tasks parameter it is necessary to classify them into dependent and independent, which is very important for the scheduler to assign them to the processors. For this prediction process machine learning algorithms are applied like logistic regression, decision tree, K-means and k-NN. In this paper initially classification of tasks into two categories dependent and independent is done later the same sets can be assigned to the processors for their execution.

Keywords— Multiprocessor scheduling,Global scheduling, Partitioned scheduling, Machine Learning.

I. INTRODUCTION

In recent years major advancements have been made in multiprocessor scheduling, so that tasks need to be allocated among them optimally. This is a serious issue in real time scheduling [1]. Typically Multiprocessor real time systems scheduling approaches are categorized into global scheduling or partitioning scheduling [2]. In global scheduling, there will be a single queue of ready tasks that are allocated among number of available processors. In partitioning scheduling each task is statically allocated to a single processor while using established uniprocessor scheduling approach. Global scheduling approaches offer better processor utilization for some set of tasks, but they can also incur runtime overhead and cache affinity problems. The multiprocessor scheduling is more complicated than uniprocessor because the operating system is responsible for allocating resources to tasks in a controlled environment. Multiprocessors systems are classified as symmetric multiprocessing and asymmetric multiprocessing. In symmetric multiprocessing execution of tasks are done on multiple processors that share a common operating system and memory. Where as in asymmetric multiprocessing one machine is in hot standby mode while the other is running the applications. Real time tasks are categorized as periodic, aperiodic and sporadic tasks. For the periodic tasks arrival time is known and aperiodic or sporadic tasks are event based and their arrival time is unknown. Along with arrival time, deadline, worst case execution time are known for the tasks. Apart from these characteristics, tasks precedence relations and resource requirements are also considered in this paper for allocation of tasks to processor. In multiprocessor system or cloud computing assigning the requests to the processors or clouds without concerning the scheduling policy makes task scheduling very inefficient. Moreover, mapping the tasks to the processors has two-phases, namely matching followed by scheduling. The matching finds the processors for the tasks whereas scheduling arranges the execution order of all the tasks [3].

Rest of the paper is organized as follows Background work is covered in section II, Proposed Methodology and Architecture explained in section III and IV respectively. Conclusion of the paper is given in section

II. RELATED WORK

Machine-learning methods and their applications can be applied in various fields with algorithms. For decision making we have algorithms such as Neural Networks, Support Vector Machine, Genetic Algorithms, Fuzzy Logics, Bayesian Networks and Decision Tree [4].

Machine Learning is an application of artificial intelligence that provides system the ability to automatically learn and improve from experience without being exactly programmed. Machine learning algorithms are classified into different levels, based on the desired outcome of the algorithm. Machine learning employs the following two strategies [5] supervised learning and unsupervised learning. In supervised learning the algorithm contains training data set, generates a function that maps input to the desired outputs. One standard formulation of the supervised learning is the classification problem the learner is required to learn or to approximate the behaviour of a function which maps input into one of different classes by looking at several input-output examples of the function. In case of unsupervised learning models, a set of inputs labelled examples are not available. Semi-supervised learning uses both labelled and unlabeled examples to generate an appropriate function or classifier.

Reinforcement learning here the algorithm learns a policy of how to act in a given situation based on an observation of the world. Every event has some impact in the environment, and the environment will respond with feedback that guides the learning algorithm. Transduction is similar to supervised learning, it does not use any explicit function instead, tries to estimate new outputs based on training inputs, training outputs, and new inputs. Learning to learn algorithm learns by previous experience using its own inductive bias based on previous experience.

Supervised learning algorithms further classified into: Classification, Regression, and Ranking. In classification, binary classification algorithms are categorized as Logistic Regression classification, Support Vector Machines, Decision Trees and Neural Networks. To classify the tasks Decision tree approach is used and how it is done is explained in the next section. Decision trees are trees that segregate the input instances by arranging them in some order based on characteristic values. Each node in a decision tree represents a characteristic classified, and each branch represents a value that the node can assume. Instances are classified starting at the root node and sorted based on their feature values.

In logistic regression binary logistic model is used to estimate the probability of a binary response based on one or more predictor variables or features. Model allows the presence of a risk factor, which increases the probability of a given outcome by a specific percentage. It is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). It is used to predict a binary outcome (1 / 0, Yes / No, True / False) given a set of independent variables. To represent binary / categorical outcome, a dummy variable can be used. Logistic regression as a special case of linear regression when the outcome variable is categorical, where log of odds are used as dependent variable.

K-means clustering is non-complicated. Here in the beginning K number of clusters are determined and assume the centre of these clusters. Then choose any random objects as the initial centre or the first K objects in sequence can also be served as the initial centre. Then the K means algorithm will perform the three steps below until convergence occur. Iterate until *stable* (= no object move group):
1. The centre coordinate is determined.
2. The distance of each object to the centre is determined.
3. The object are grouped based on minimum distance.

The k-NN algorithm [6] is an instance based machine learning algorithm which is used to classify objects according to various features that the objects possess and makes use of training samples to do the actual classification. The objects are represented in a so-called feature space as N dimensional vectors whose entries consist of numerical values. These are the features chosen to characterize the objects (hence the name features space). The dimension of the feature space is equal to the number of features which are selected to represent an object. The k-NN algorithm classifies objects as belonging to a particular category by simply determining the majority category among an objects' K-nearest neighbours. The metric for determining the distance between objects is often the Euclidean distance from the between the objects

III. METHODOLOGY

A Novel scheduler is proposed for task scheduling in the Figure 1 which uses machine learning algorithms for tasks categorization as dependent and independent. Further the tasks are stored in dependent task queue and independent task queue respectively. The tasks from the queue are later assigned to processor by applying genetic algorithms [7] for dependent tasks and for independent tasks based on deadline or laxity can be allocated to processor. Also tasks from the independent task queue are scheduled to processors using the algorithm based on the priority assigned to the tasks. In cloud computing resource utilization is very important [8][9] hence if requirement is predicted in advance helps in assignment of task to processor.

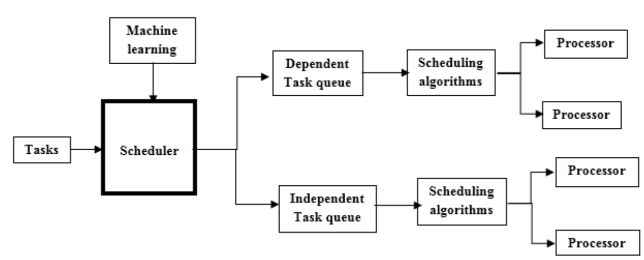

Figure 1. A Novel scheduler for task scheduling

IV. RESULTS AND DISCUSSION

The Table 1 gives different characteristics of task set [10]. Execution time of a task is the maximum time required to complete its execution. Deadline is the time within which task has to complete its execution. Laxity is the elapsed time between the execution time and deadline [11]. The parameter resource dependency gives the information about resources requirement either it can be shared or exclusive [12] of a task, contain only 0 (exclusive) & 1(shared) value as discussed in the logistic regression algorithm. This value classifies the set of task into dependent and independent. After applying this algorithm to the dataset the tasks $T_{1,}$ $T_{4,}$ T_8 & T_{10} are classified as independent based on resources, and T_2, T_3, T_5, T_6, T_7, T_9 are classified as dependent. Using this information a decision tree is constructed which is a Directed Acyclic Graph (DAG) [13] as shown in the Figure 2.

Table1: Tasks with their characteristics.

Set of Task	Execution time(msec)	Deadline (msec)	Laxity (msec)	Resource dependency
T_1	1	3	2	0
T_2	4	5	1	1
T_3	9	10	1	1
T_4	8	11	3	0
T_5	6	10	4	1
T_6	7	9	2	1
T_7	4	6	2	1
T_8	3	5	2	0
T_9	2	4	2	1
T_{10}	1	2	1	0

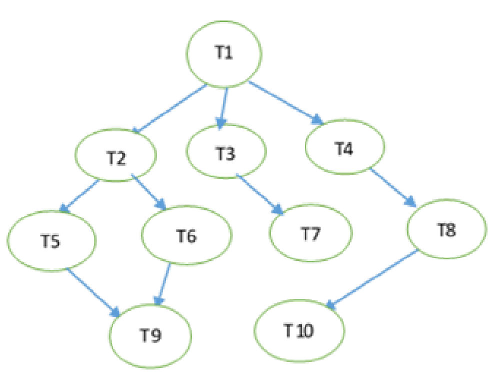

Figure 2. Decision Tree for the give task set

From the decision tree tasks can be allocated to processors using genetic algorithms [7] [8] [13] [14]. A schedule shows the sequences of tasks determined for each processor, and it is expressed by a two- dimensional array that its length is equal to the total number of tasks in the task graph and a processor is randomly selected for each task .Hence number of processors required for the execution of tasks can be predicted as two. If the same set of task is executed on a single processor it takes 45 msec. According to the classification based on decision tree constructed in the Figure 1 takes 13msec and 32msec on two processors P_1 and P_2 respectively.

K-means clustering will be applied for the task scheduling as follows: The algorithm is considered on the deadline parameter. Here some centroid point is considered and then the data is classified into groups as one to the left of the centroid and the other to the right of the centroid. If deadline 6msec is considered as centroid , the tasks that need to be executed within 6msec are grouped as $T_{1,}$ $T_{2,}$ $T_{8,}$ $T_{9,}$ T_{10} and the tasks that need to be executed in 11msec are groups as $T_{3,}$ $T_{4,}$ $T_{5,}$ $T_{6,}$$T_{7,}$

In Table 2, the precedence relation for the task is considered. There are two values in that column one for decision tree and other for K-means clustering the decision tree will work based on the 0 &1 value as discussed above and K-means algorithm will work for the other value in the Table 2.

Table 2: The Precedence Relation for task set.

Set of Task	Resource dependency	Precedence relation
T_1	0	0(0)
T_2	1	1(1)
T_3	1	1(1)
T_4	0	1(1)
T_5	1	2(1)
T_6	1	2(1)
T_7	1	2(1)
T_8	0	2(1)
T_9	1	4(1)
T_{10}	0	3(1)

The decision tree classifies the tasks into dependent and independent, further K-means clustering is applied only on the independent tasks [15]. Let us assume centroid as 3, the algorithm is going to take the input as precedence relation value and it will check for distance of centroids after that it classifies tasks into,

Partial scheduling tasks:T_2,T_3,T_4,T_5,T_6,T_7&T_8
Global scheduling tasks: T_9 &T_{10}

After applying K-means clustering, we are applying K-nearest neighbour to our classified dataset, when the K-means classifies the tasks into partial and global scheduling task set, then the K-nearest neighbour algorithm is applied on the parameter called "laxity". It takes the input from partial and global scheduling task set and checks for the deadline, if deadline is in the trained data/threshold that task will be selected and allocated to processor.

For example in Table 1 the independent task,T_1 is allocated to processor $P_{1,}$then in the set of partial scheduling task set it checks for the threshold which matches first it will be allocated to processor $P_{2,}$at last it will go for global scheduling task set it checks the threshold which task matches the threshold then it will be allocated to $P_{3.}$

V. CONCLUSION AND FUTURE SCOPE

Machine learning algorithms can be applied for almost all categories of problems at hand. In this paper a Novel scheduler is proposed for the categorization of tasks into independent and dependent .The main is to minimize the execution time, increase resource utilization and high availability. Also can predict number of processors required for a given set of tasks in a multiprocessor environment either in distributed computing or cloud computing where decision making is very important.

REFERENCES

[1] Mehdi Akbari, Hassan Rashidi, "A multi-objectives scheduling algorithm based on cuckoo optimization for task allocation problem at compile time in heterogeneous systems", Expert Systems With Applications, Elsevier Vol.60, pp.234-248, 2016.
[2] Baruah Bipasa Chattopadhyay, Haohan Li Insik Shin, "Mixed-criticality scheduling on multiprocessors", Vol 50, Issue 1, pp 1–4, 2014.
[3] Sanjaya K. Panda, Indrajeet Gupta, Prasanta K. Jana, "Allocation-Aware Task Scheduling for Heterogeneous Multi-Cloud System", Procedia Computer Science, Elsevier Vol 50, pp.176-184, 2015.
[4] Yang Xin, Lingshuang Kong, Zhi Liu , Yuling Chen, Yanmiao Li, Hongliang Zhu, Mingcheng Gao,Haixia Hou, And Chunhua Wang, "Machine Learning and Deep Learning Methods for Cybersecurity", IEEE Access, Vol.6, pp. pp.35365-35381, 2018.
[5] Panos Louridas and Christof Ebert, "Machine Learning", IEEE Computer Society, Vol.4, Issue.11, pp.110-115, 2016.
[6] Chouhan Kumar Rath , Shashank Sekhar Suar ,Prasanti Biswal, "A Comparative Study On Dynamic Task Scheduling Algorithms", Journal on Information Technology, Vol.71, Issue.1, pp.1-6, 2018
[7] Sri Raj Pradhan, Sital Sharma, Debanjan Konar, Kalpana Sharma, "A Comparative Study on Dynamic Scheduling of Real-Time Tasks in Multiprocessor System using Genetic Algorithms", International Journal of Computer Applications Network Security and Communication, Vol.120, Issue.1, pp.0975-8887, 2015.
[8] Sakshi kathuria., "A Survey on Security Provided by Multi-Clouds in Cloud Computing", IJSRNSC, Vol.6, Issue.1, pp.23-27, 2018.
[9] Pradeep K.Sharma, Vaibhav Sharma and Jagrati Nagdiya, "A proposed Method for Mining High Utility Itemset with Transactional Weighted Utility using Genetic Algorithm Technique (-GA)," IJSRCSE, Vol.5, Issue.1, pp.31-35, 2017.
[10] Marko Bertogna, Michele Cirinei, Giuseppe Lipari Member "Schedulability analysis of global scheduling algorithms on multiprocessor platforms", IEEE Transactions on Parallel and Distributed System,Vol X,No. X 2008.
[11] Rekha A Kulkarni ,Suhas H Patil, N.Balaji, "Fuzzy Real Time Scheduling on Distributed Systems to Meet the Deadline of Applications", International Journal of New Technology and Research, Vol.2, Issue.4, pp.56-58, 2016.
[12] Wei Zhao and Krithi Ramamritham, "," Simple and Integrated Heuristic Algorithms for Scheduling Tasks with Time and Resource Constraints", Journal of Systems and Software,Vol 7, pp.195-205, 1987.
[13] Ashish Sharma and Mandeep Kaur," An Efficient Task Scheduling of Multiprocessor Using Genetic Algorithm Based on Task Height", JITSE, Vol 5, Issue 2 1000151, ISSN: 2165-7866,2015.
[14] Nirmala H, Girijamma H A," Aperiodic task Scheduling Algorithms for Multiprocessor systems in Real Time environment", International Journal of Engineering and Computer Science, ISSN 2319-7242, Vol 4, Issue 8, pp 13838-13841, 2015.
[15] Christos Gogos, Christos Valouxis, Panayiotis Alefragis , George Goulas, Nikolaos Voros ,Efthymios Housos," Scheduling independent tasks on heterogeneous processors using heuristics and Column Pricing", Future Generation Computer Systems, Issue 60, pp 48-66, 2016.

Authors Profile

Mrs. Nirmala H pursed Bachelor of Computer Science & Engineering from Kuvempu University Davangere, in 1998 and Master of Computer Science & Engineering from Visvesaraya Technological University Davangere, in 2005. She is currently pursuing Ph.D. and currently working as Associate Professor in Department of Computer Science & Engineering, Bangalore since She is a Life Member of Indian Society for Technical Education (ISTE), Member of Computer Society of India (CSI), IEE. She published research papers in reputed international journals and conferences including IEEE and it's also available online. Her main research work focuses on Multiprocessor scheduling Algorithms, Embedded systems, Cloud computing, IoT .She has 17 years of teaching experience.

Mrs Dr.Girijamma H A pursed Bachelor of Computer Science and Engineering from Mysore University, BIET Davangere, Master of Computer Science & Engineering from Visvesvaraya Technological University ,Gulbarga and P.hD from Kuvempu University, BIET, Davangere of Karnataka, India in year 2012. She is currently working as Professor in Department of Computer Science & Engineering RNSIT, Bangalore since 2008. She has published more than 20 research papers in reputed international journals and conferences including IEEE and it's also available online. Her main research work focuses on Fuzyy Automata, machine learing, Finite Automata & Formal Languages, Embedded Systems.She has more than 20 years of teaching experience and 2 years of industry Experience.

