Paper Info

*Improving interactivity and user*

*experience has always been a challenging task.*

*One aspect of this could be to improve process*

*scheduling. This paper is a detailed survey about*

*the attempts that have been made to incorporate*

*machine learning techniques to improve process*

*scheduling. Various approaches to find the*

*appropriate attributes of a process for predicting*

*resource utilization have been discussed here.*

Important points

1. An important element of process scheduling is context switching,

which takes place when the current

process is pre-empted. This involves saving of the

state of the current process before switching the

CPU to another process. For each context switch,

there is an associated overhead which results in

loss of valuable processor time slices.

2. One way of improving a user’s experience is to

ensure that the processes they use are given a larger

share of the resources i.e. more priority. These

processes need to be identified and their

performance can then be improved using data from

previous executions

3. Prediction of the amount of required resources that happens during scheduling of processes, need to be determined:-

a) The first part of this work was the data collection

phase where the programs were run for varying

input sizes. The collected data was then put into 20

classes to be used for machine learning in WEKA

using the “Trees, Lazy, Rules” classifier types

supplied. Decision Trees, K-NN and Decision

Tables were used for finding robust and accurate

predictions.

b) Various subsets of

the attributes were tested in order to determine

which subset produced the best predictions for

“total execution time”. An “attribute evaluator”,

which assigned a weight to each subset, and a

“search method”, which determined the kind of

search to be performed, were used in the process.

Search methods used were: Genetic Search, Best

First Search and Rank Search. Evaluation methods

employed were: CfsSubsetEval and Consistency

SubsetEval.

4. The three major groups into which the processes

are usually divided are batch, daemon and

interactive. The attributes of the processes present

in the Linux context were extracted and used to

generate a training base. The attributes were

grouped using an unsupervised learning algorithm.

The groups formed were manually analysed. Each

process was identified by the group and a base was

generated with the labelled data. The

dimensionality of the attribute vector was reduced

through algorithms to generate new training bases.

The classification algorithms were then used to

perform training samples on the reduced bases.

5. The performance of each of the classification algorithm

was evaluated in relation to the hit rate and the

processing time to achieve the proposed objective.

6. The data was provided in two phases -

learning and testing

7. The proposed approach was to identify attributes of

processes that could help determine their CPU

burst lengths. These were then used by machine

learning techniques for prediction. This was done

in the following manner:-

● Data sets of real workloads containing process

attributes were utilized. These were divided

into training and testing sets.

● Relevant attributes were selected and divided

into two categories based on whether historical

data was available or not.

● Different ML algorithms, namely SMOreg,

MLP, Decision tree and K-NN, were trained

on both the categories of training data sets and

models were generated and evaluated using the

testing datasets.

● Two criteria for evaluation were used, namely

correlation coefficient (CC) and relative

absolute error (RAR). These were used to

determine how close the predicted values of

CPU bursts were to the actual values. For a

good result, a high value of CC accompanied

by a low RAR percentage was required.

● Relief selection technique, implemented on the

WEKA tool, was then used to rank the

attributes according to their significance in

determining the CPU burst lengths.

● From all of the above, tables were constructed

for both variations (with and without history)

of the datasets and analysed to determine the

best set of attributes as well as the best ML

technique.

Conclusion:-

In conclusion, it is seen that machine learning

techniques can be efficiently integrated into

existing operating systems to deliver a seamless

user experience. Initial experiments have been

fairly successful in pinpointing specific attributes

of processes that are better suited than others in

predicting CPU burst cycles and resource

utilisation. Classification of processes has been

extensively employed in previous works to

simplify decision making. Granting extra time

slices based on analysis of previous executions of a

process was an innovative approach to improve

performance.

Reference:-

[1] Siddharth Dias1, Sidharth Naik2, Sreepraneeth K3, Sumedha Raman4, Namratha M5.

“A Machine Learning Approach for Improving Process Scheduling: A Survey” , In International Journal of Computer Trends and Technology (IJCTT) – Volume 43 Number 1 – January 2017