31/03/2021

CSPC43

Operating System

106119100

Rajneesh Pandey

ASSIGNMENT

Topic:

MACHINE LEARNING APPROACH FOR IMPROVING PROCESS SCHEDULING.

Introduction

Improving user experience and interactivity has been challenging task.
One would be improve process scheduling.

Machine Learning is the domain of Computer Science which deals with the capability of computer to learn without specifying direct instruction.
The application of Machine learning are classified as either supervised and unsupervised learning.

Supervised learning:

computer is given the set of data with corresponding label data.

Unsupervised learning:

computer tries to find structure in the dataset based on correlation among variables in the data.

Process Scheduling: is the activity by which the Operating System (OS) selects an available process from the job queue for execution. This selection is performed by the scheduler. An important element of process scheduling is context switching, which takes place when current process is pre-empted.

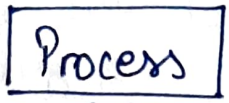
This involved saving of the state of the current process is pre-em before switching the CPU to another process. for each context switch, there is an associated overhead which result in loss of valuable processor time slices

One way of improving a user's experience is to ensure that the process they use are given a longer share of the resources i.e. more priority. These process need to be identified and their performance can then be improved using data from previous execution

Ways / Methods for improving process scheduling

Method - ①

Characteristics of a process, which contribute significantly to the prediction of the amount of required resources, need to be determined. This has been attempted by analysing the data regarding previous execution. —①



"Interactive" "Non Interactive"

* Linux program were used
* 24 attributes were selected.

Part - ①:

(i) data collection phase where programs were run for varying input size.

(ii) Data was then put into 20 classes to used for machine learning WEKA using "Trees, Lazy, Rules" classifier types.

Decision Tree, K-NN & Decision Table used for finding robust & accurate prediction.

Search method : Rank Search, Genetic Search, Best First Search

Evaluation method : CfsSubsetEval and Consistency SubsetEval.

best attribute → "input size", and "page reclaim" was next best.

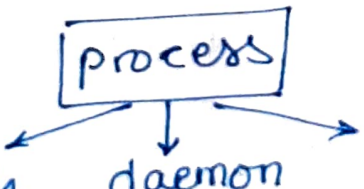
A good prediction rate of (91.04 % - 99.7 %) was achieved using the aforementioned technique

These result were used to improve PBS scheduling. were estimated based on knowledge base developed by keeping track of previous execution of programs.

Method — (11)

Applying Data Mining techniques to the data present in the kernel about each process, to automatically detect and group the processes which have similar behaviour and to classify a new process accordingly.

Here,



* Attribute grouped using unsupervised learning algo.

The performance of each of the algo. is evaluated in the relation to the hit rate and the processing time to achieve at the process was present.

Data was extracted later fed into WEKA for processing.

The Grouping analysis used unsupervised learning algorithm to group the data based on various parameter. Process classified in 6 different ways.

- A (Interactive application) all type of interactive process.
- D (Daemons) run into background.
- f (Desktop features) process that perform tasks. to support the graphical desktop
- N (Network) : network communication.
- C (Text Commands): simple text-mode terminal commands.
- K (kernel threads): inner threads of the core of the operating system.
- O (other) : processes that do not fit into the other groups.

Algorithm used to find the best subsets were genetic search, In Gain : Ranker, CFS : Rank Search and CFS : BFS.

The evaluation procedures were chosen as information Gain and CFS methods.

Through the classification algorithms, four Database

method - 3 :

The current machine learning technique to improve process scheduling by allocating variables time slices for different process. to reduce the overhead of context - switching

Process were associated with a single integer field referred to as special_time_slice (STS). which helped in indicating the best estimate of CPU cycles to be allocated, so as to minimize. their turnaround time.

The process es were categorised into different STS. classes, each having an intervels of 50 ticks.

Mapping was established between process attributes and STS classes. C.45 was found to be the best classifier and 6 characteristic were identified. to help in prediction.

Conclusion

Initial experiments have been fairly successful in pinpointing specific attributes of process that are better suited than other in predicting CPU burst cycles and resource utilising

There is scope for improving scheduling to cater better to the needs of the user.