Probabilistic mathematical model for load balancing in cloud

By

Darshan Patel
13bce071

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
Ahmedabad 382481

Probabilistic mathematical model for load balancing in cloud

MP-1

Submitted in fulfillment of the requirements

For the degree of

Bachelor of Technology in Computer Engineering/Information Technology

By

Darshan Patel
13bce071

Guided By
Prof. Jitendra Bhatia
[DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING]

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
Ahmedabad 382481

CERTIFICATE

This is to certify that the Seminar entitled "Probabilistic mathematical model for load balancing in cloud" submitted by Darshan Patel, towards the partial fulfillment of the requirements for the degree of Bachelor of Technology in Computer Engineering/ Information Technology of Nirma University is the record of work carried out by him/her under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination.

Prof. Jitendra Bhatia
M.Tech(CSE)
Department of Computer Science & Engg.,
Institute of Technology,
Nirma University,
Ahmedabad

Prof. Sanjay Garg
Dept. of Computer Science & Engg.,
Institute of Technology,
Nirma University,
Ahmedabad

ACKNOWLEDGEMENT

Without support of my guide , my friends , my professors , it is impossible to develop knowledge in field of Big data . I thank to Nirma University to give me opportunity for gaining maximum knowledge and exploring enormous amount of sources in campus .

I would like to acknowledge with much appreciation the crucial role of my guide **Prof. Jitendra Bhatia**, without her constant guidance and encouragement would have made my efforts go in vain.

I am grateful to **Department of Computer Science and Engg**, for providing me supportive and congenial environment to work in.

ABSTRACT

To provide valuable information on the task-scheduling process of a load balancing algorithm, maintaining optimal load balancing in cloud or hosted environments , these valuable information contain CPU response time , memory available , server performance etc,. In this project , optimization of load balancing algorithm with input as priority wise rank to individual servers. There are many algorithms of load balancing , for example Predictive Dynamic load balancing algorithm with service differentiation with service types is taken into consideration , Round robin algorithm with session switching , weighted active monitoring load balancing in distributed systems etc,. In this project , rank allocated to servers according to performance of server in past , its memory capacity , its response time and other parameters that effect processing of task allocated to different servers , along with efficient and optimized algorithm of load balancing .

CONTENTS

Chapter 1 Introduction

1.1 General

Many research paper shows assignment problem with different parameters and find the most efficient problem allocation to different severs in cloud or host system .

How this assignment problem arises ? What makes this assignment problem so importance in day to day life ?

1.2 Introduction about assignment problem .

Lets take example to find answer of above question . Before 30-40 years , there was huge industrial revolution throughout the world . At that time , no robots or machines were used in huge amount . Human sources were in huge demand . Labourers were huge in number and here problem arises . Whenever any new labourer joins the organization , higher authority of organization have to allocate work according to their expertise in particular job or field . If there are small numbers of workers then task will be easy as trail and error method will be most efficient to allocate jobs . But the problem arises when number of worker are huge and trial and error method will be difficult one to solve . Another example in this era of information and technology is in cloud computing , taxi applications on mobile(ola cabs , uber cabs) etc,.

In cloud computing , N number of tasks will take N different time in seconds to complete on N different servers . This is the problem called assignment problem in which task is allocated to which server is decided.

If there are 4 taxis and 4 people calls to pick up on their locations , then then solution of assignment problem will decide which taxi will go to which locations . So we see that in day to day life , there can be many examples which have assignment problem .

In this report , overview of different algorithm is explained and my project is to optimize the current most efficient algorithm with different parameters taken as input as explained in chapter 3 .

Chapter 2 Literature Survey

2.1 Introduction of different method to solve assignment problem .

Till now , there are many linear assignment algorithm .

Some of them are :

1. Enumeration Method
2. Transportation method
3. Hungarian Method
4. New alternative method

Out of these , two methods are used now : Hungarian and new alteration method .

Hungarian mathematician Koing develop the most efficient method Hungarian method of optimal solutions of linear assignment problem . It have time complexity of $O(n^4)$. But this method complexity reduced to $O(n^3)$. This was done by Edmonds and Karp .

New alternative method is having same complexity as hungarian method but its complexity of solving assignment problem is much less than hungarian method .

2.2 Brief about Hungarian method .

Hungarian method is based on the principle that if a constant is added to the elements of cost matrix, the optimum solution of the assignment problem is the same as the original problem. Original cost matrix is reduced to another cost matrix by adding a constant value to the elements of rows and columns of cost matrix where the total completion time or total cost of an assignment is zero. This assignment is also referred as the optimum solution since the optimum solution remains unchanged after the reduction.

Chapter 3 Hungarian , New alternative method and its optimization .

3.1 Hungarian mehod

The assignment problem is as follows:

Resources (workers)	Activities(jobs)				Supply
	J_1	J_2	K	J_n	
W_1	c_{11}	c_{12}	K	c_{1n}	1
W_2	c_{21}	c_{22}	K	c_{2n}	1
..	..	..	..	..	..
W_n	c_{n1}	c_{n2}	K	c_{nn}	1
Demand	1	1	K	1	n

Here in above table , associate the problem with workers in an organization .

On the left side of table (in first column) , resources of workers , servers , taxi etc ,. can be taken .

W_1 , W_2 , W_3 , … , W_n are workers or resources .

J_1 , J_2 , J_3 , … , J_n are jobs .

C_1 , C_2 , C_3 , … , C_n are time taken by particular resource to complete particular job in seconds .

In assignment problem , the above table is cost matrix and the required solutions will have only one job allocated to each row . If the resources and number of jobs did not match , or the cost matrix is not n*n matrix then a dummy row or column is added to make it square matrix.

Taking an example , Hungarian method is explained as follows:

Question :

An organization has 5 human source . Time to complete jobs is given in following table . Allocate jobs to minimize the time taken to complete all jobs .

		Employees				
		I	II	III	IV	V
Jobs	A	10	5	13	15	16
	B	3	9	18	13	6
	C	10	7	2	2	2
	D	7	11	9	7	12
	E	7	9	10	4	12

Step 2 :

	I	II	III	IV	V
A	5	0	8	10	11
B	0	6	15	10	3
C	8	5	0	0	0
D	0	4	2	0	5
E	3	5	6	0	8

Step 3 and 4 :

	I	II	III	IV	V
A	5	0	8	10	11
B	0	6	15	10	3
C	8	5	0	0 ×	0 ×
D	0 ×	4	2	0 ×	5
E	3	5	6	0	8

Step 5 :

	I	II	III	IV	V	
A	5	0	8	10	11	
B	0	6	15	10	3	√
C	8	5	0	0 ×	0 ×	
D	0 ×	4	2	0 ×	5	√
E	3	5	6	0	8	√
	√			√		

Step 6 :

	I	II	III	IV	V
A	7	0	8	12	11
B	0	4	13	10	1
C	10	5	0	2	0
D	0	2	0	0	3
E	3	3	4	0	6

Step 7 :

	I	II	III	IV	V
A	7	[0]	8	12	11
B	[0]	4	13	10	1
C	10	5	0 ×	2	[0]
D	0 ×	2	[0]	0 ×	3
E	3	3	4	[0]	6

The allocated jobs is :

Job	Employee	Time (in hours)
A	II	5
B	I	3
C	V	2
D	III	9
E	IV	4
	Total	23 hours

3.2 New Alternative method

The new alternate method of assignment problem discussed here gives optimal solution directly within few steps. It is very easy to calculate and understand. The alternate method developed by us in this investigation seems to be easiest as compare to available methods of assignment problem. Here we explain algorithm for alternate method of solving assignment problem for minimization and maximization cases.

Lets take same example as taken in Hungarian method. :

An organization has 5 human source . Time to complete jobs is given in following table . Allocate jobs to minimize the time taken to complete all jobs .

	Activities(jobs)					
Resources (Employees)		I	II	III	IV	V
	A	10	5	13	15	16
	B	3	9	18	13	6
	C	10	7	2	2	2
	D	7	11	9	7	12
	E	7	9	10	4	12

Solution Consider the data matrix in above table. Now, we select row A and select that column (activity) for which row A has minimum unit cost. In this example, for row A, column II (activity) has the minimum unit cast. So we write resource A under column I and activity II under column II. In the similar way, we select all the rows (resources) and find the minimum unit cost for the respective columns, which are shown in Table given below.

Resource	Activity
A	II
B	I
C	III, IV, V
D	I, IV
E	IV

	I	II	III	IV	V
A	10	5	13	15	16
B	3	9	18	13	6
C	10	7	2	2	2
D	7	11	9	7	12
E	7	9	10	4	12

Here, Activity II is unique as it doesn't occur again and hence assigned resource A to activity II and is shown in Table . Next, delete Row A and Column II.

Again select minimum cost value for the remaining resources, B, C, D and E. which is shown below.

Resource	Activity
B	I
C	III, IV, V
D	I, IV
E	IV

	I	III	IV	V
B	3	18	13	6
C	10	2	2	2
D	7	9	7	12
E	7	10	4	12

Repeating above process , we will get :

Resource	Activity
C	III, IV, V
D	IV
E	IV

	III	IV	V
C	2	2	2
D	9	7	12
E	10	4	12

Since resource D and E have single activity IV. Next we see that resource C has also same activity IV and hence we take the minimum cost difference for resources C, D and E. Here minimum cost difference for resource C is 0, minimum cost difference for resource D is 2 while minimum cost difference for resource E is 6. Since 6 is a maximum difference which represents resource E and hence assign resource E to activity IV and is shown in Table . Further delete row E and Column IV.

Again select minimum cost value for the remaining resources, C and D. which is shown in Table.

Resource	Activity
C	III, V
D	III

	III	V
C	2	2
D	9	12

Finally, different employees have assigned jobs uniquely, which is shown in Table .

Employees	Jobs	Time (in hour)
A	II	5
B	I	3
C	V	2
D	III	9
E	IV	4
	Total	23

The above answer is same as we get in Hungarian method .

3.3 Optimization of new alternative method .

How above algorithms can be optimized ? In this section a new method is proposed to improve complexity of hungarian method and new alteration method . Still , this is only algorithm . This document shows how linear assignment problem is solved using time complexity of O(n*logn) . Here are steps with example which explains this algorithm :

Step 1 :

All the rows in data matrix will be sorted . The new matrix having index is formed as follows :

Table having data matrix :

	Jobs with Rank					
Resources		I	II	III	IV	V
	A	10	5	13	15	16
	B	3	9	18	13	6
	C	10	7	2	2	2
	D	7	11	9	7	12
	E	7	9	10	4	12

Here , the row inputs is taken as rank of particular resource to completer particular job . This rank is calculated as mentioned before .

Now , as first step is to sort (ascending order) each row and create a new matrix having column or job indices as shown in following table :

Jobs with only indices in sorted order(ascending order)					
A	2	1	3	4	5
B	1	5	2	4	3
C	3	4	5	2	1
D	1	4	3	2	5
E	4	1	2	3	5

The advantage of above step is follows :

- We observe that there is no number higher than 5 , or in generalized form if the matrix is n*n , then there will be no number higher than n.
- This step helps to change any number without traversing whole array .
- This step is core part of the algorithm , and important reason for optimization of Hungarian and new alteration method .
- For example if to change value of 5 to -1 , only one step is needed.
- This can be achieved with help of array of pointers in c and reference value on object oriented programming .

Step 2 :

- After sorting each row in given data matrix , a new array is initialized with first non-zero elements of each row . The array size will be n .

Step 3 :

- A new matrix having value of difference between consecutive elements in row of sorted matrix is formed as shown here :

	Difference of consecutive elements			
A	5	3	2	1
B	3	3	4	5
C	0	0	5	3
D	0	2	2	1
E	3	2	1	2

Step 4 :

- Now array obtained in step 2 will have values of first column . The repeated elements in this array is ignored . Other ignored elements are those which having difference 0 respective place in table obtained in step 3 . So the element remained is 2 , which is then changed to -1 in the table . Updated table is as shown below :

	Jobs				
A	-1	1	3	4	5
B	1	5	-1	4	3
C	3	4	5	-1	1
D	1	4	3	-1	5
E	4	1	-1	3	5

Step 5 :

The above process is repeated until all values in data matrix is -1 . So second job is assigned to source A .

The above algorithm is having more space complexity compare to hungarian algorithm.

For worst case , if all values are same , then again one technique should be developed such that there should not be repetitive checking of difference in table obtained in step 3 . So right now this algorithm works only for cases , in which large difference is obtained between consecutive values in row .

References :

1. A new alternate method of problem assignment by from shodganga.infibnet.ac.in
2. Hungarian method from en.wikipedia.org/wiki/Hungarian_algorithm