Learning Automata-based Solutions to the Nonlinear Fractional Bin Packing Problem with Application to distributed web crawling

By Anis Yazidi

Supervisors:

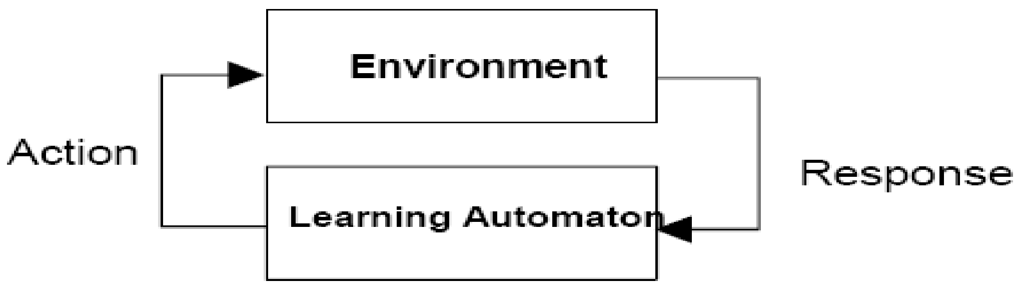
Morten Goodwin Olsen

Ole-Christoffer Granmo

Introduction

- The use of LA to solve the nonlinear fractional bin packing problem and its application to distributed web crawling.
- Recent studies has been performed to solve the the so called nonlinear fractional knapsack problem.
- Extension of the non linear fractional knapsack problem is the non linear fractional bin packing problem.
- Two novel approaches for solving the non linear bin packing problem.
- Mapping of the nonlinear fractional bin packing problem to distributed polling frequency determination problem

Learning automata



Learning automata acting in an environment

Given a finite number of actions that can be performed in a random environment, when a specific action is taken place the environment provides a random response which is either favorable or unfavorable. The choice of the action at any stage should be guided by past actions and responses.

Classical Bin packing problem

Definition:

(Optimization version of)*Bin Packing*: Given a bin capacity B, and a set of items $1,...,n$, with weights $w_1,...,w_n$. Partition the set of items into a minimum number of subsets such that for each subset S, $\Sigma_{i \varepsilon S} w_i \leq B$.

The Nonlinear Bin packing problem

Maximize $\sum_{i=1}^{n} f(x_i)$

subject to $\sum_{j=1}^{m} y_{ij} = 1 \quad , \quad 1 \leqslant i \leqslant n$

$$y_{ij} \in \{0,1\} \quad , \quad 1 \leqslant i \leqslant n \quad , \quad 1 \leqslant j \leqslant m$$

$$\sum_{i=1}^{n} y_{ij} x_i \leqslant C_j \quad , \quad 1 \leqslant j \leqslant m$$

$$x_i \geqslant 0 \qquad 1 \leqslant i \leqslant n$$

The Nonlinear Bin packing problem

- The problem involves *n* materials of different value *vi* per unit volume
- *m bins of different volume Cj,*
- Each material I is available in a certain amount *xi*
- Let *fi*(*xi*) denote the value of the amount *xi*
- The problem is to fill the m bins with the material mix *x* [*x*1,....,*xn*] of maximal value Σ f(xi) without violating the capacity constraint of each bin.

Approach 1

- Consider the problem as a knapsack problem with a capcity Ck
- At each epoch we solve the non linear knapsack problem with a capacity Ck
- If the items can not be packed in the end of the epoch in question reduce the capacity and reloop
- RemarK: An optimal solution to the Knapsack problem with a capacity C=c1+c2+ ..+cm presents an upper boundary solution to the Bin packing problem.

Approach 1

$$C_k = \begin{cases} \sum_{i=1}^{m} c_i & \text{if } k=0 \\ C_k = \eta * C_{k-1} & \text{if } k>0, \text{where} \end{cases}$$

η is a parameter approximately equal to 1, η is typically equal to 0.99.

Approach 2

- Run Oommen scheme for a certain number of epochs
- Each epoch consists of N iterations
- One greedy scheme for moving the web site that intuitively improves the solution is moving the item with the smallest expected value from the bin with the largest expected value to the bin with the lowest expected value.
- The idea is to utilize the capacity in all bins better and thus increase the sum of the estimated values for all bins.

Mapping to distributed Crawling

- *Our problem can be defined as : Given m crawlers with capacities c1,c2,…,cm, allocate the available capacity among web pages so that expected number of pollings uncovering new information per time step is maximized.*
- *Update detection probabilty when polling freqeuncy is x*

$$d_i(x_i) = 1 - q_i^{\frac{1}{x_i}}$$

Problem formulation

Maximize $\sum_{i=1}^{n} x_i \times d_i(x_i)$

subject to $\sum_{j=1}^{m} y_{ij} = 1 \quad , \quad 1 \leqslant i \leqslant n$

$$y_{ij} \in \{0,1\} \quad , \quad 1 \leqslant i \leqslant n \quad , \quad 1 \leqslant j \leqslant m$$

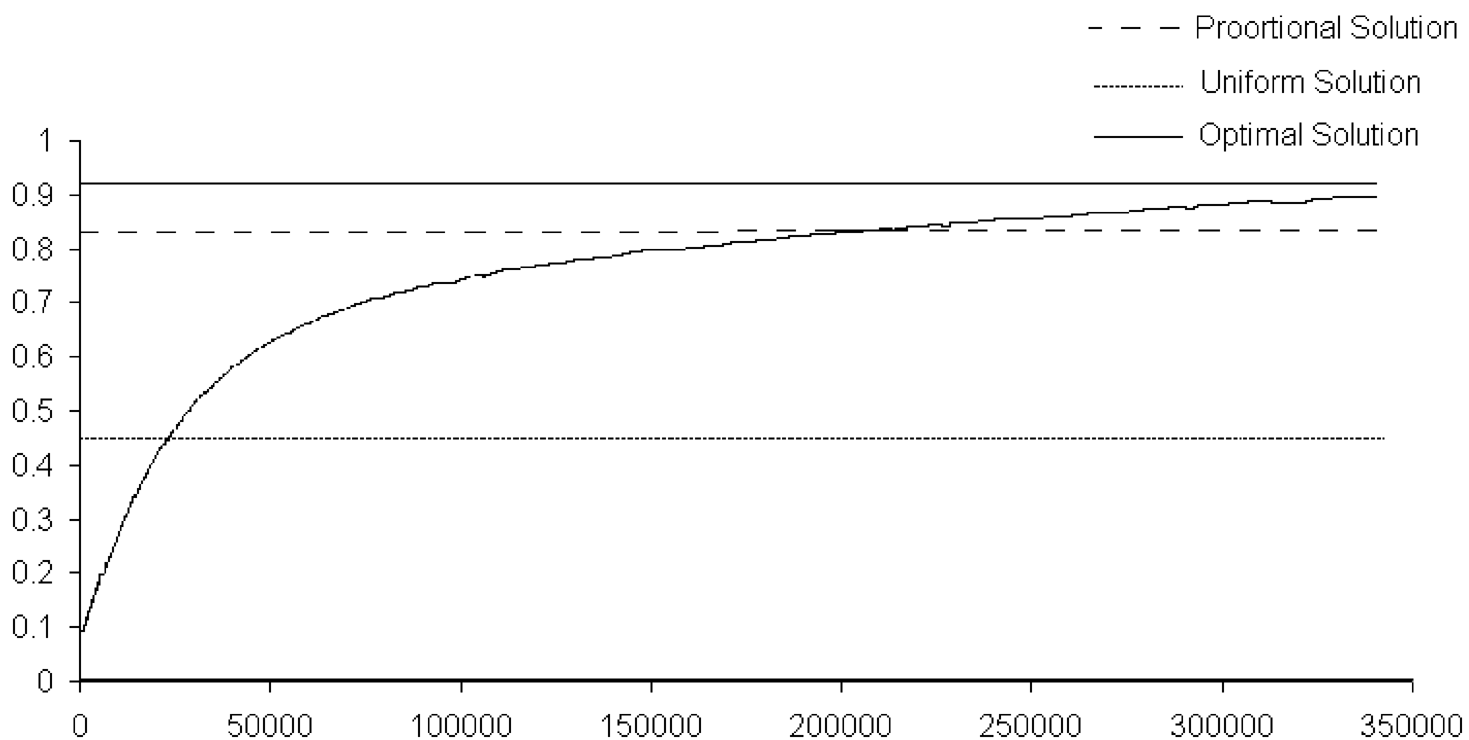
$$\sum_{i=1}^{n} y_{ij} x_i \leqslant c_j \quad , \quad 1 \leqslant j \leqslant m$$

$$x_i \geqslant 0 \qquad 1 \leqslant i \leqslant n$$

Experimental results: Approach 1

(α, β)	Scheme	$\sum x_i \times d(x_i)$
(0.9,1.5)	LA	0.9386
	Uniform	0.23
	Proportional	0.9370
	Optimal	0.943
(0.3,1.5)	LA	0.578
	Uniform	0.122
	Proportional	0.561
	Optimal	0.598
(0.3,1)	LA	0.826
	Uniform	0.452
	Proportional	0.824
	Optimal	0.837

Approach 2


Proortional Solution
Uniform Solution
Optimal Solution
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
50000
100000
150000
200000
250000
300000
350000

Conclusion

- In this project, we introduced a novel variant the bin packing problem namely, the nonlinear bin packing problem.
- Furthermore, we proposed two learning automata based approaches for solving this problem.
- We investigated the optimal solution the problem of resource allocation of a distributed crawler when the polling capacity is restricted by mapping it the non linear fractional bin packing problem.