Expert Systems With Applications 110 (2018) 323–334

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

An aggregation approach for solving the non-linear fractional equality Knapsack problem

Anis Yazidi[a], Tore M. Jonassen[a], Enrique Herrera-Viedma[b,c,*]

[a] Department of Computer Science, Oslo and Akershus University College of Applied Sciences, Oslo, Norway
[b] Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain
[c] Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia

A R T I C L E I N F O

Article history:
Received 17 October 2017
Revised 17 December 2017
Accepted 7 June 2018

Keywords:
Non-linear fractional equality knapsack
Resource allocation
Dynamical system
Rate limiting

A B S T R A C T

In this paper, we present an *optimal*, efficient and yet simple solution to a class of the *deterministic* non-linear fractional equality knapsack (NEFK) problem —a substantial resource allocation problem. The solution is shown to be superior to the state-of-the-art in terms of convergence speed.

We provide a rigorous analysis that proves the optimality of our scheme under general conditions. Our solution resorts to a subtle aggregation procedure that drives the system towards equalizing the derivatives of the material value functions in a similar manner to the Homo Egualis theory. Furthermore, we report experimental results that catalogue the applicability of our solution to the problem of rate limiting in cloud computing, which falls under the *deterministic* NEFK problem.

1. Introduction

The self-optimization aspect of Autonomic Computing (AC) systems (Kephart & Chess, 2003) envisages dynamic allocation of a shared resource pool between applications in order to yield optimal resource usage. In more formal terms, the allocation is viewed as a vector p, where each component p_i represents the share of resources of a consumer i (Loureiro, Nixon, & Dobson, 2012). Thus, the problem reduces to an optimization problem with constraint on the resource capacity. The aim is to find an optimal allocation vector p^* that maximizes a performance function f that depends on p. We will show that the latter problem is an instance of the *deterministic* non-linear fractional equality (Granmo & Oommen, 2010a; 2010b; 2011; Granmo, Oommen, Myrer, & Olsen, 2007). While there are myriad solutions to different classes of knapsack problems (Chabane, Basseur, & Hao, 2017; He, Wang, He, Zhao, & Li, 2016; Lim, Al-Betar, & Khader, 2016), most of them are static and are based on the worst case or average case scenarios. Unlike the realm of convex optimization, NEFK uses a monotonicity assumption, which is common in many real-life problems, including resource allocation problems. In this paper, we will show that, by appropriately defining f, we are able to deal with two distinct problems, namely: a utility optimization problem (Stanojevic & Shorten, 2009b; 2010) and a fairness problem (Stanojevic & Shorten, 2008; 2009b) with constraint on the resource capacity. In order to put our work in the right perspective, we emphasize that the *deterministic* NEFK problem (Granmo & Oommen, 2010a; 2010b; 2011) allows for a large set of applications, including:

- A typical application is the problem of distributed rate limiting for cloud-based services investigated in Stanojevic and Shorten (2008, 2009a) where the resource to be shared among traffic limiters is the bandwidth capacity. An optimal solution in this case aspires to achieve a fairness postulate that states that: “the performance levels at different servers should be (approximately) equal” (Stanojevic & Shorten, 2009a).
- Dynamic speed scaling of processes in cloud computing based on demand and performance constraints in order to minimize energy consumption (Stanojevic & Shorten, 2010).
- A class of utility maximization problems (Loureiro et al., 2012) for resource allocation.

Our model for efficiently solving the NEFK problem is based on the theory of dynamical systems. The model is simple and can be easily implemented in a centralized or decentralized manner thanks to a“dissemination” property of our devised algorithm. Our solution uses a subtle aggregation procedure to drive the system towards equalizing the derivatives of the material value functions in a similar manner to the Homo Egualis theory (De Jong, Tuyls, & Verbeeck, 2008a).

The rest of this paper is structured as follows. Section 2 provides the preliminaries needed for the rest of the paper.

* Corresponding author.
E-mail addresses: anis.yazidi@hioa.no (A. Yazidi), toremj@gmail.com (T.M. Jonassen), viedma@decsai.ugr.es (E. Herrera-Viedma).

https://doi.org/10.1016/j.eswa.2018.06.015
0957-4174/© 2018 Elsevier Ltd. All rights reserved.

Section 3 describes the details of our solution as well as the main theoretical results of the paper. In Section 4, we provide some experimental results that catalogue the convergence of our approach, its conformance to the theoretical findings, and its superiority to the state of the art.

2. Background

The problem we address in this paper is referred to as a *non-linear fractional equality knapsack (NEFK) problem* (Granmo & Oommen, 2010a; 2010b; 2011; Granmo et al., 2007).

Deterministic non-linear equality fractional knapsack (NEFK) problem: involves n materials p_i, $1 \le i \le n$, where each material p_i is available in a certain amount $0 \le p_i \le b_i$. Let $f_i(p_i)$ denote the value of the amount p_i of material i. The problem is to fill a knapsack of fixed volume $c > 0$ with the material mix $\vec{p} = [p_1, \ldots, p_n]$ of maximal value $\sum_1^n f_i(p_i)$ (Black, 2004).

The *nonlinearequality* FK problem is characterized by a separable objective function. The problem can be stated as follows (Kellerer, Pferschy, & Pisinger, 2004):

$$\begin{aligned} &\text{maximize} \quad f(\vec{p}) = \textstyle\sum_{i=1}^n f_i(p_i) \\ &\text{subject to} \quad \textstyle\sum_{i=1}^n p_i = c \ \text{ and } \ p_i \ge 0 \text{ for } i \in \{1, \ldots, n\}. \end{aligned}$$

We suppose that the derivatives of the material value functions $f_i(p_i)$ with respect to p_i, (hereafter denoted f_i'), are non-increasing. In other words, the material value *per unit volume* is no longer constant as in the linear case, but decreases with the material amount, and so the optimization problem becomes:

$$\begin{aligned} &\text{maximize} \quad f(\vec{p}) = \textstyle\sum_{i=1}^n f_i(p_i), \ \text{ where } f_i(p_i) = \int_0^{p_i} f_i'(\xi_i) d\xi_i \\ &\text{subject to} \quad \textstyle\sum_{i=1}^n p_i = c \ \text{ and } \ p_i \ge 0 \text{ for } i \in \{1, \ldots, n\}. \end{aligned}$$

Efficient solutions have been devised to the latter problem based on the principle of Lagrange multipliers. In short, the optimal value occurs when the derivatives f_i' of the material value functions are equal, subject to the knapsack constraints (Bretthauer & Shetty, 2002):

$$\begin{aligned} &f_1'(p_1) = \cdots = f_n'(p_n) \\ &\textstyle\sum_{i=1}^n p_i = c \ \text{ and } \ p_i \ge 0 \text{ for } i \in \{1, \ldots, n\}. \end{aligned}$$

2.1. State-of-the-art for deterministic NEFK and its applications to resource allocation

Palomar and Chiang introduced the concept of decomposition for utility maximization (Palomar & Chiang, 2006). Their approach relies on a distributed algorithm, where the optimization is performed over separable functions that can be regarded as a distributed version of gradient ascent algorithm (Mosk-Aoyama, Roughgarden, & Shah, 2010). Other approaches resort to a form of hierarchy, where the decomposition methods used by Palomar and Chiang (2006) are a representative example. The essence of these methods is to decompose an optimization problem into smaller ones, where each of them can be further decomposed in a hierarchical manner. In order to place our paper within the larger body of scientific research, we briefly present here some representative studies on resource allocation in autonomic systems. Utility is a relatively recent concept in computer science that has been borrowed from the field of economics. Utility originally describes a measure of preferences over some set of goods. Traditionally, Quality of Service (QoS) was defined in a binary manner, i.e., either met or unmet. Thus, many different allocations can be "optimal" in that framework as long as the constraints of the system in terms of QoS are satisfied. By introducing utility functions (Bennani, Menasce et al., 2005; Fulp, Ott, Reininger, & Reeves, 1998), an allocation fitness can be measured in a non-binary manner in order to better quantify the satisfaction of a consumer with his share of a resource.

In many real-life problems, utilities are usually related in a *monotonic* manner to the amount of allocation. This monotonic property of the utility reflects the fact that the performance of a system improves as more resources are allocated to it. For instance, when it comes to a web service, the QoS measured in terms of response time degrades monotonically as less CPU and memory resources are allocated to the web server. Menasce and his collaborators (Bennani et al., 2005) have advocated using performance-dependent utility functions in resource allocation problems in distributed computing environments. Despite the fact that the utility function is usually monotonic when it comes to resource allocation, most of the available optimization paradigms have been developed for convex functions. To the best of our knowledge, the only available work in the literature that operates under the same assumption as our proposed approach is by Stanojevic and colleagues (Stanojevic & Shorten, 2008; 2009a). However, the latter work (Stanojevic & Shorten, 2008; 2009a) requires message exchange between pairs of the components of the allocation vector. In contrast to this, our algorithm is less computationally complex since it does not assume any pairwise message exchange. We will provide some experimental results that show the superiority of our scheme to the approach of Stanojevic and Shorten (2008, 2009a).

Wang, Du, Chen, and Li (2008) employed a constrained non-linear optimization technique, combining both deterministic and stochastic optimization algorithms, to dynamically allocate server capacity with the help of analytical models. Another body of methods for resource allocation that are radically different from the approach presented in this paper include reinforcement learning (Dutreilh et al., 2011) and negotiations between distributed agents (Boutilier, Das, Kephart, Tesauro, & Walsh, 2002). In the realm of resource allocation problems, many combinatorial methods have been used such as simulated annealing, and genetic algorithms. However, these approaches do not fall under the scope of this article. Furthermore, it is worth mentioning that myriad nature-inspired approaches can be found in the literature (Gao & Pan, 2016), but they are also outside the scope of this article. An interesting concept in resource allocation, which emanates from the field of microeconomics, and more particularly from regulation in competitive markets, is tâtonnement (Subramoniam, Maheswaran, & Toulouse, 2002). Tâtonnement aims to balance demand and consumption using smart pricing techniques. The idea behind tâtonnement is simply to increase the price of the resource whenever the demand is less than consumption, in order to encourage consumption, while decreasing the price in the opposite case. This concept was applied in computer systems by assigning fictive budgets to the different applications competing for the shared resources.

Our devised solution in this article tries to equalize the derivatives of the material value functions. The solution is similar to the Homo Egualis theory (De Jong et al., 2008a). An agent in a Homo Egualis society compares his payoff to the payoff of the rest of the agents. Each agent will tend to increase his payoff in proportion to the payoff of agents that have a better payoff than him, and decrease his payoff in proportion to the payoff of agents that have lower payoffs than him (De Jong et al., 2008a; 2008b; de Jong & Tuyls, 2011; Xing & Chandramouli, 2008). Instead, our solution equalizes the derivatives of the material value functions by comparing the derivative of the agent in question to the "mean" of the derivatives of all agents. This simple and subtle principle allows the system to converge towards an optimal solution.

The novel contributions of this paper are listed below:

- We devise an optimal and yet efficient solution to the deterministic NEFK problem under general conditions.
- We provide sound theoretical results that demonstrate the optimality of our solution based on analyzing the resulting dy-