ARTICLE IN PRESS
Simulation Modelling Practice and Theory xxx (2015) xxx–xxx
Contents lists available at ScienceDirect
Simulation Modelling Practice and Theory
journal homepage: www.elsevier.com/locate/simpat

# A reinforcement learning methodology for a human resource planning problem considering knowledge-based promotion

Amir-Mohsen Karimi-Majd [1], Masoud Mahootchi *, Amir Zakery [1]

*Amirkabir University of Technology, Industrial Engineering and Management Systems Department, 424 Hafez Ave., Tehran 15875-4413, Iran*

ARTICLE INFO

*Article history:*
Received 25 January 2015
Received in revised form 23 June 2015
Accepted 16 July 2015
Available online xxxx

*Keywords:*
Reinforcement learning
Production-inventory control
Human resource planning
Stochastic dynamic programming
Knowledge-intensive

ABSTRACT

This paper addresses a combined problem of human resource planning (HRP) and production-inventory control for a high-tech industry, wherein the human resource plays a critical role. The main characteristics of this resource are the levels of "knowledge" and the learning process. The learning occurs during the production process in which a worker can promote to the upper knowledge level. Workers in upper levels have more productivity in the production. The objective is to maximize the expected profit by deciding on the optimal numbers of workers in various knowledge levels to fulfill both production and training requirement. As taking an action affects next periods' decisions, the main problem is to find the optimal hiring policy of non-skilled workers in long-time horizon. Thus, we develop a reinforcement learning (RL) model to obtain the optimal decision for hiring workers under the demand uncertainty. The proposed interval-based policy of our RL model, in which for each state there are multiple choices, makes it more flexible. We also embed some managerial issues such as layoff and overtime-working hours into the model. To evaluate the proposed methodology, stochastic dynamic programming (SDP) and a conservative method implemented in a real case study are used. We study all these methods in terms of four criteria: average obtained profit, average obtained cost, the number of new-hired workers, and the standard deviation of hiring policies. The numerical results confirm that our developed method end up with satisfactory results compared to two other approaches.


## 1. Introduction

The main goal of a production plant or a service supplier in a developed and developing country is to have more shares in the internal or external markets especially where different competitors exist. To reach this valuable goal, a company should efficiently utilize all different resources such as workforces and facilities such that it could meet the required satisfaction of customers. As the level of satisfaction is usually changing with the enhancement of technology, all different operations accomplished need to be based on up-to-date knowledge. This is called knowledge intensive operations. Among different important resources employed in knowledge intensive operations, human resource is more critical because people and their knowledge are the most strategic resource for firms [2].

One of the main issues in human resource planning (HRP) is staffing and recruitment decision-making to provide enough qualified manpower for producing high quality products or giving superior services. Recruitment is usually a mid-term or

* Corresponding author. Tel.: +98 21 64545387.
*E-mail addresses:* mmahootchi@aut.ac.ir, mmahootchi@gmail.com (M. Mahootchi).
[1] Tel.: +98 21 64545387.

http://dx.doi.org/10.1016/j.simpat.2015.07.004

even long-term decision which can really affect the near future of the company and its success. Furthermore, human resources, as a strategic and valuable asset, possess the knowledge and skills which are substantially necessary to move a company toward its predefined goals. In other words, one of the important aspects of HRP is to determine the required number of workers in different knowledge levels (e.g., new-hired, semi-skilled, and skilled workers) that should be utilized in various parts of the production process in a company. This is in fact a way of improving the utilization of knowledge resources toward a better efficiency.

There are few quantitative approaches employed to cope with staffing problems for knowledge-intensive operations. One of the pioneering works in the area of human resources planning in a knowledge-based situation has been proposed by Bordoloi and Matsuo [7]. They proposed a model obtaining the number of different needed knowledge level. They also embedded employee's learning and turnover rate into their optimization model to find the better recruitment decisions where demand is non-deterministic. The learning occurs during the production process in which a worker can be transferred from a lower knowledge level to the upper one (e.g., from the first level to the second one) after some periods. Furthermore, turnover in a company is defined as the rate of losing its workers in each knowledge level (semi-skilled or skilled level) at the end of each period [20]. When a company loses skilled workers in the upper levels, it cannot be directly compensated. This means that the company is only able to do the demand satisfaction by recruiting workers in the first level (new-hired workers). They used the chance-constraint method to tackle the high uncertainty of demand and the high volatility of knowledge workers in the last two levels. Their method will fail if we want to address the production-inventory control problem. This also ends up with the static hiring policy (i.e., the hiring rate is constant for all periods in the real-time decision-making process) which is very conservative (i.e., the policy is obtained for a pessimistic situation). Furthermore, the layoff has not been considered in their model.

Given this fact that the demand is stochastic and unknown at the time of decision making, there is a possibility of not satisfying demand (we called this slack or shortfall hereafter) which is assumed to be lost sale in our paper. There is also another choice to construct a physical *buffer* to store the remaining goods for a situation in which the demand is more than the production level so that the extra demand can be met using the stored stock. Of course, it might be possible to compensate the stock-out using an *overtime* working shift with existing workers. By considering these managerial issues (i.e., overtime working hours, slack/shortfall, surplus, and layoff to the mathematical optimization model), the planning model based on the knowledge-intensive workers would be more compatible to what happens in reality and the final hiring policy would be more useful for managers and beneficial for the respective company.

To address all the aforementioned issues, this paper contributes three important goals. First, this paper proposes a new optimization model in which the inventory level is also taken into consideration. This consideration makes our model more compatible to the reality, so it would lead to more proper decisions. Second, in order to efficiently solve this model, we develop a reinforcement learning (RL) method. Furthermore, to have a more applicable decision policy we achieve optimal interval decisions instead of single ones for every state using a modified version of the value iteration technique as a well-known approach in stochastic dynamic programming (SDP). This makes the optimization model more flexible as it gives multi choices to the decision-maker. It is worth mentioning that all the respective information about the demand are used to find the optimal hiring policy while in the chance-constraints approach (i.e., the basic model proposed by Bordoloi and Matsuo [7]) only the mean and the standard deviation of the stochastic demand are used. We will refer to their proposed model in more details in the rest of the paper and compare the results of our two models and their model using the data obtained from a semiconductor equipment manufacturing company.

This paper is organized as follows: Section 2 provides a review on the related literature and introduces some papers related to the subject topic. Section 3 includes basic definitions of SDP and RL. In Section 4, we describe the production-inventory control problem and our proposed methods. Sections 5 and 6 present numerical results and conclusion, respectively.

## 2. Related work

Workforce planning determines the required level of workers by which the strategic goals of an organization/company could be achieved. Bulla and Scott [9] defined it as a process in which the required level for the human resource in an organization is properly identified and efficient plans for satisfying those requirements are designed. Khoong [27] specified manpower planning as a core of HRP, which is supported by other aspects of HRP.

Different mathematical approaches have been used in HRP (for a review, interested readers can see [5,17]). These approaches can be generally categorized in three parts: optimization, Markovian and computer simulation models. Most of research works in the area of HRP are devoted to staffing or recruitment decisions as an important and popular area of research [2]. Although all types of the mentioned categories have been applied in staffing, the main focus here is to make a suitable stochastic optimization model in an uncertain environment.

Demand is mostly considered as a stochastic parameter in staffing problems. The demand could be either product demand [1,7,8,21], workforce demand [14,28,29,32,13,34] or service demand which can be, for example, the amount of call arrivals in a call center [3,4,19].

In some researches the staff knowledge as a critical resource is under focus, either generally from knowledge elicitation and knowledge management view [18], or particularly in the form of different knowledge levels of workers. Entering the