Journal of Biomedical Informatics xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Improving hospital bed occupancy and resource utilization through queuing modeling and evolutionary computation

Smaranda Belciug [a], Florin Gorunescu [b,*]

[a] Department of Computer Science, University of Craiova, Craiova 200585, Romania
[b] Department of Biostatistics and Informatics, University of Medicine and Pharmacy of Craiova, Craiova 200349, Romania

ARTICLE INFO

Article history:
Received 21 July 2014
Accepted 18 November 2014
Available online xxxx

Keywords:
Queuing system
Compartmental model
Genetic algorithm
Bed allocation optimization
Resource utilization optimization
What-if analysis

ABSTRACT

Scarce healthcare resources require carefully made policies ensuring optimal bed allocation, quality healthcare service, and adequate financial support. This paper proposes a complex analysis of the resource allocation in a hospital department by integrating in the same framework a queuing system, a compartmental model, and an evolutionary-based optimization. The queuing system shapes the flow of patients through the hospital, the compartmental model offers a feasible structure of the hospital department in accordance to the queuing characteristics, and the evolutionary paradigm provides the means to optimize the bed-occupancy management and the resource utilization using a genetic algorithm approach. The paper also focuses on a "What-if analysis" providing a flexible tool to explore the effects on the outcomes of the queuing system and resource utilization through systematic changes in the input parameters. The methodology was illustrated using a simulation based on real data collected from a geriatric department of a hospital from London, UK. In addition, the paper explores the possibility of adapting the methodology to different medical departments (surgery, stroke, and mental illness). Moreover, the paper also focuses on the practical use of the model from the healthcare point of view, by presenting a simulated application.

1. Introduction

A hospital department may face the situation when patients are turned away because all beds are occupied, and the corresponding healthcare service is thus postponed due to the insufficient number of available beds. An insufficient financial support or a poor resource management often causes this situation. On the other hand, an over-provision of hospital beds or an unrealistic health service time is a waste of the already limited resources. Accordingly, there is need for a complex involvement bringing together under the same umbrella advanced analytical methods and machine learning techniques to help make better decisions regarding the allocation and use of hospital beds in order to improve patient care and save money.

A wide range of different techniques have been used and reported in the literature. [1] presents a model of the cost of treating stroke patients within a healthcare facility using a mixture of Coxian phase type model with multiple absorbing states. A non-homogeneous discrete time Markov chain incorporating time-dependent covariates is developed in [2] to model the patient flow in a cost or capacity constrained healthcare system. A multi-objective comprehensive learning particle swarm optimization with a representation scheme based on binary search for bed allocation problem in general hospital is presented in [3]. [4] developed a semi-closed migration network to capture patient flow into the clinic, and between the clinic and hospital.

Although queuing models are widely used in industry to improve customer service, the number of applications in healthcare, however, is relatively small. This is probably due to the different nature of the two domains, the client-patient equivalence being however difficult to be generally accepted. Previous works [5,6] have introduced M/PH/c and M/PH/c/N queuing models in order to optimize the use of hospital resources both in a loss model and in an extended model providing an extra waiting room. A multi-objective decision aiding model based on queuing theory and goal programming is introduced in [7] for allocation of beds in a hospital. A queuing approach based on non-homogeneous arrival patterns, non-exponential service time distributions, and multiple patient types along with a spreadsheet implementation of the resulting queuing equations is used in [8] to increase the capacity

* Corresponding author at: Department of Biostatistics and Informatics, University of Medicine and Pharmacy of Craiova, 2 Petru Rares Str., Craiova 200349, Romania. Fax: +40 251 412 673.

E-mail addresses: smaranda.belciug@inf.ucv.ro (S. Belciug), gorunef@gmail.com, fgorun@rdslink.ro, florin.gorunescu@webmail.umfcv.ro (F. Gorunescu).

http://dx.doi.org/10.1016/j.jbi.2014.11.010
1532-0464/© 2014 Published by Elsevier Inc.

of an Emergency Department. In [9] a decision support system based on the Erlang loss model is developed to evaluate the size of nursing units.

Compartmental models have previously been shown to provide a suitable description of the patient flow through a hospital department, especially for geriatric medicine. Starting with a deterministic two-compartment mathematical model [10], further progress occurred when stochastic models along with mixed exponential distributions, continuous-time Markov model and Bayesian belief networks have been proposed [11–14].

This paper proposes a flexible strategy to improve the hospital management regarding its two main aspects: (*a*) bed allocation policy, and (*b*) financial resource utilization. First, it uses results from queuing theory to model the patient flow, where a Poisson process describes the patients' arrivals, hospital beds are servers, and the length of stay is modeled using a phase-type distribution. Second, in conjunction with the queuing system, a compartmental model describes the hospital department. Finally and most importantly, the previous approach has been enriched with the support of the evolutionary paradigm used to optimize both the bed allocation policy and the resource utilization. In addition, a "What-if" analysis has been performed to explore in depth the various possible options available for the hospital management. The main contributions of the paper are twofold: first, the evolutionary-based optimization of the hospital management, and, secondly, the "What-if" analysis allowing the evaluation of different available options.

2. Materials and methods

2.1. The queuing model

The theoretical model refers to a M/PH/*c* queuing system in which M denotes Poisson (Markov) arrivals, the service distribution is phase-type [15], the number of servers is *c*, and no queue is allowed. In such a *loss model* in which the customers that find all the servers busy are lost for the system, λ represents the Poisson arrival rate, and the phase-type service has the probability density function given by:

$$f(t) = \sum_{i=1}^{l} \alpha_i \rho_i e^{-\alpha_i t}, \tag{1}$$

with the corresponding mean $\tau = \sum_{i=1}^{l} \rho_i/\alpha_i$, where l represents the number of phases/compartments, α_is the mixing proportions, and the ρ_is the transition rates with $\sum_{i=1}^{l} \rho_i = 1$.

The parameters defining the above queuing model, λ, τ, and c are considered as variable entities being subject to an optimization process enabling the improvement of the bed occupancy and resource utilization.

The average number of arrivals occurring during a time interval of length t is given by $\lambda \cdot t$; thereby, the *offered load* of the system, i.e., the average number a of arrivals during an average length of stay τ is $a = \lambda \cdot \tau$. Since the probability of having j occupied servers is given by:

$$P_j = \frac{a^j/j!}{\sum_{k=0}^{c} a^k/k!}, \tag{2}$$

the probability that all the c servers are occupied is given by:

$$P_c = B(c,a) = \frac{a^c/c!}{\sum_{k=0}^{c} a^k/k!} \ (\textit{Erlang's loss formula}) \tag{3}$$

In other words, $B(c,a)$ represents the fraction of customers that is lost by the system [16,17]. Note that the above results apply when the system is in statistical equilibrium, i.e., after a sufficiently long period of time, P_j being referred as steady-state or statistical equilibrium probabilities.

2.2. Fundamental queuing characteristics

Basically, there are three fundamental quantities of interest for queuing models:

- L – the average number of customers in system.
- W – the average time spent in system by an arbitrary customer.
- ρ – the server occupancy.

Among useful relationships between the above characteristics, we mention:

- The *carried load* $L = a \cdot [1 - B(c,a)]$, representing the average number of customers in system, also known as *Little's formula*.
- The average time spent in system by an arbitrary customer $W = \tau \cdot [1 - B(c,a)]$.
- The server occupancy $\rho = \frac{L}{c}$ (with $\rho \leqslant 1$ for steady-state).

One of the two main goals of this study is an evolutionary-based optimization of the bed occupancy management by estimating the model's parameters c, λ and τ, in order to obtain:

- An acceptable threshold for the delay probability $B(c,a)$, seen as the suitable proportion of refused patients which the system is prepared to tolerate.
- The corresponding average time spent in system.
- The corresponding average number of customers in system.

2.3. The associated cost model

A main concern in proposing a model to solve real-world issues, especially in healthcare, is to provide the best service to customers with minimum costs by using the maximum utilization of existing resources. In queuing models, this could be "translated" by maintaining the lost requests (lost potential customers) at a minimum level with minimum costs. Following [5], a base-stock policy approach [18] is used to set up an associated cost model to balance the fraction of customers that is lost by the system against the service costs.

As it was stated above, the model's parameters c, λ and τ are supposed to be variable. This study focuses on finding their (near) optimal values providing a trade-off between serving costs and penalty costs corresponding to unsatisfied demands.

In order to define the associated cost model, let us consider that the number c of servers comprises both the number of occupied beds and the number of idle beds, ready to be used in emergencies.

In a similar fashion to the newsvendor model [18], the cost model envisages the two following parameters:

- A *holding cost* of h units per day per empty (non used) server.
- A fixed *penalty cost* of π units per unsatisfied orders.

With the aim of improving servers occupancy and resource utilization in the long-run department activity, the cost per day under the base-stock policy with server level c can be expressed as a function of the queuing system parameters c, λ, τ, and the cost model parameters h, π, by:

$$g(c,\lambda,\tau,h,\pi) = \pi \cdot \lambda \cdot B(c,\lambda \cdot \tau) + h \cdot \{c - \lambda \cdot \tau \cdot [1 - B(c,\lambda \cdot \tau)]\} \tag{4}$$

Based on the cost function $g(c, \lambda, \tau, h, \pi)$, the issue of optimizing the inventory level, in other words, the resource utilization, is equivalent to a minimization problem, i.e., to find c, λ, τ, h, and π in order to minimize the cost (fitness) function g.

A controversial method still in use in healthcare to measure the inpatients activity is based on the turnover per allocated bed per