POLÍTICA DE COMPRA DE MEDICAMENTOS UTILIZANDO SIMULAÇÃO HÍBRIDA E APRENDIZADO POR REFORÇO

David Custódio de Sena
Universidade Federal Rural do Semi-árido (UFERSA)
Av. Francisco Mota, 572 - Bairro Costa e Silva, Mossoró RN
sena@ufersa.edu.br

Rafael de Carvalho Miranda
Universidade Federal de Itajubá (UNIFEI)
Avenida BPS, 1303
rafael.miranda@unifei.edu.br

Alexandre Ferreira de Pinho
Universidade Federal de Itajubá (UNIFEI)
Avenida BPS, 1303
pinho@unifei.edu.br

José Arnaldo Barra Montevechi
Universidade Federal de Itajubá (UNIFEI)
Avenida BPS, 1303
montevechi@unifei.edu.br

Elisa Maria Melo Silva
Universidade Federal de Itajubá (UNIFEI)
Avenida BPS, 1303
lizzbr@gmail.com

RESUMO

Em hospitais, o gerenciamento de estoques de medicamentos é realizado pela farmácia hospitalar, fornecendo uma área apropriada e corpo técnico. O objetivo deste artigo é definir uma política de aquisição de medicamentos periódica em uma farmácia hospitalar, que busque a diminuição conjunta do número de medicamentos não atendidos e expirados e que seja limitado a um orçamento. Para tanto, optou-se pelo uso combinado das simulações a eventos discretos e baseada em agentes com a ferramenta de inteligência artificial aprendizado por reforço, e que consiga lidar com as dimensões falta de medicamentos e perecibilidade mais a limitação orçamentária. Como resultado houve diminuição de 100% do não atendimento em um medicamento e de 100% e 67,89% da expiração em outros dois medicamentos.

PALAVRAS CHAVE. Farmácia hospitalar, simulação híbrida, aprendizado por reforço, gerenciamento de estoques.

Tópicos: Simulação e PO na Área de Saúde

ABSTRACT

When it comes to hospitals, the management of drug stocks is performed by the hospital pharmacy, providing an appropriate area and technical staff. The objective of this research is to define a policy for the purchase of periodic drugs in a hospital pharmacy that seeks to jointly reduce the number of missed and expired drugs and which is limited to a budget. In order to do so, we opted for the combined use of simulations to discrete events and based on agents with the artificial intelligence tool learning by reinforcement, and to be able to deal with the dimensions' lack of drugs and perishability plus the budgetary limitation. As a result, there was a 100% decrease in non-care in one drug and 100% and 67.89% in two other drugs.

KEYWORDS. Hospital pharmacy, hybrid simulation, reinforcement learning, inventory management.

1. Introdução

Em hospitais de médio e grande porte, a farmácia hospitalar exerce um papel fundamental para seu bom funcionamento, sendo a sua cadeia de fornecimento uma das que desempenham um dos papéis mais importantes na área da saúde [Narayana *et al.*, 2014]. Segundo [Yurtkuran e Emel, 2008] a farmácia central no hospital fornece uma área apropriada de armazenagem de remédios e materiais hospitalares. Assim, é imprescindível que ela possua uma gestão baseada em resultados e que consiga proporcionar um bom nível de serviço, pois é um órgão importante para o bom funcionamento hospitalar. Para [Jacobson *et al*., 2006], um dos principais motivos que levaram a esse quadro foi o aumento gradativo dos custos relacionados à saúde, pois é fato que, mesmo em hospitais públicos, o descontrole desses custos acaba causando enormes prejuízos no atendimento aos pacientes.

Segundo [Jiang e Sheng, 2009] é observado que em sistemas de compras na cadeia de suprimento, modelos analíticos são mais utilizados. Para os autores, esses tipos de métodos fornecem deduções estritas, que geralmente envolvem notações complicadas e equações sob premissas. Porém, em situações como as de gerenciamento de cadeia de suprimentos, o ambiente a ser estudado é muito dinâmico, com mudanças que acontecem no decorrer do tempo. Para [Pidd, 2004], a simulação permite utilizar o poder computacional para realizar experimentos em um modelo da realidade, com um mínimo de impacto de custos operacionais.

Outra forma de auxiliar a tomada de decisão é através da inteligência artificial, que tenta representar o processo decisório similar ao do ser humano [Lee, 2007]. Uma dessas ferramentas que apresenta bons resultados combinando seu uso à simulação, principalmente a simulação baseada em agentes, é o aprendizado por reforço (*Reinforcement Learning – RL*) [Kaelbling *et al.*, 1996], [Sutton e Barto, 1998]. Essa combinação tem conseguido abranger as características complexas nativas de sistemas como a farmácia hospitalar, pois ela apresenta elementos que são difíceis de capturar. Dentre essas características convém ressaltar a aleatoriedade de alguns elementos, como a demanda e o prazo de validade dos medicamentos.

O objetivo do trabalho é definir uma política de aquisição de medicamentos periódica em uma farmácia hospitalar utilizando a combinação da simulação híbrida a eventos discretos (SED) e baseada em agentes (SBA) com o RL para diminuir o número de medicamentos não entregues e o número de medicamentos vencidos, dentro de um orçamento limitado. Para tal, a simulação foi importante pois assim consegue-se capturar as informações mais rotineiras do problema, através da SED, e modelar a melhor tomada de decisão para cada situação, através da SBA conjuntamente com o RL.

O trabalho está dividido em 5 seções. A segunda seção apresenta a revisão de literatura; a terceira, o capítulo de descrição do objeto de estudo, que visa apontar as principais características deste. A quarta seção explana sobre a aplicação da modelagem dos dados trabalhados e os resultados encontrados após a simulação. E, por fim, a seção de conclusão.

2. Revisão de literatura

2.1. Gestão de compras, aprendizado por reforço e simulação baseada em agentes

A gestão de estoques é uma atividade crucial nas organizações e pode ser modelada como um problema de decisões sequenciais [Katanyukul e Chong, 2014], pois, em um momento, estas são tomadas observando-se as anteriores e suas consequências. Mais especificamente, essa atividade pode ser visualizada como um *Markov Decision Process* (MDP) [Puterman, 2014]. Assim, segundo [Kaelbling *et al.*, 1996] um MDP consiste em: (i) Um conjunto de estados $\mathcal{S}$; (ii) Um conjunto de ações $\mathcal{A}$; (iii) Uma função de recompensa $\mathcal{R}: \mathcal{S} \times \mathcal{A} \rightarrow \Re$; (iv) Uma função de transição $\mathcal{T}: \mathcal{S} \times \mathcal{A} \rightarrow \Pi(\mathcal{S})$.

O objetivo é maximizar, em um período extenso, a função de recompensa. Uma das técnicas bastante utilizadas para alcançar esse objetivo é o Aprendizado por Reforço, que é uma forma de aprendizado computacional em tempo diferencial [Katanyukul e Chong, 2014] baseado na tentativa e erro e recompensa a longo prazo. Ele pertence a um grupo de aprendizado por máquina, denominado aprendizado não supervisionado. [Kaelbling *et al*., 1996], [Sutton e Barto, 1998] citam que o RL foca, principalmente, no aprendizado direcionado ao objetivo em um contexto geral do

problema a ser enfrentado. Ela é especialmente útil em problemas reais, pois não é necessário modelar analiticamente todo o comportamento do sistema *a priori*, ou seja, conhecer todo o conjunto de transições $\mathcal{T}$; sendo assim, livre do modelo. Neste caso, utilizou-se a simulação a eventos discretos para desempenhar esse papel, pois é adequada e utilizada aos problemas reais.
Os principais componentes presentes do RL são: (i) Uma *política* define o comportamento do agente em um período no tempo, definindo quais as decisões tomar frente à cada situação; (ii) Uma *função de recompensa* representa o retorno imediato para a (s) decisão (ões) tomadas; (iii) Uma *função de valor* representa o quão bom uma política é no longo prazo; (iv) Por fim, o *modelo* que irá representar, dentro do RL, o ambiente em que o agente está inserido para tomar a decisão, no caso desse trabalho, o modelo é representado pela simulação híbrida a eventos discretos e baseada em agentes.
O (s) agente (s) e o ambiente interagem em uma sequência de momentos discretos no tempo, t = {0, 1, 2, 3,...}. Em cada momento, ou episódio, t, o agente recebe alguma representação do *estado* do ambiente, $\mathcal{A}$ é o conjunto de ações disponíveis no estado s_t. Um episódio seguinte, em parte da consequência de suas ações, recebe uma *recompensa* numérica, $r_{t+1} \in \mathbb{R}$, e encontra-se em um novo estado, s_{t+1}. A Figura 1 apresenta esse relacionamento de forma esquemática.

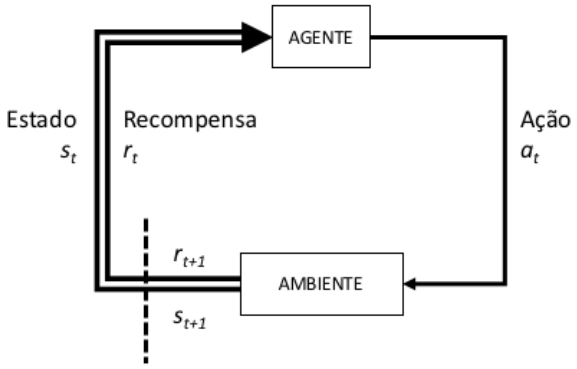

Figura 1 - Interação agente-ambiente

Outra grande vantagem do RL é que ele não exige que haja um conhecimento profundo da estrutura do problema, sendo assim bastante abrangente para um número considerável de problemas de decisão [Katanyukul e Chong, 2014], dentre eles o apoio a tomada de decisão na farmácia hospitalar.
Alguns autores utilizaram o RL no auxílio da tomada de decisão de compras, como em [Valluri *et al.*, 2009], [Jiang e Sheng, 2009] e [Katanyukul e Chong, 2014]. [Valluri *et al*., 2009] compararam três algoritmos de RL, *Q-learning*, *Sarsa* (λ) e *Tile coding with Sarsa* (λ), representando cada estado com uma função de aproximação baseada em redes neurais para avaliar um modelo de cadeia de suprimento simples, linear e com apenas um produto. Nos dois primeiros algoritmos, apenas um agente, o varejista, tem a capacidade de tomada de decisão baseada em RL. Já no último, três deles, varejista, atacadista e distribuidor, apresentam essa característica de aprendizado. Os autores concluíram que as ferramentas se mostraram úteis, porém a demora do aprendizado, mais de mil períodos, foi considerada uma desvantagem.

2.2. Heurística *Q-learning*

A heurística *Q-learning* [Watkins, 1989] é um método de diferença temporal bem utilizada pois consegue representar o valor das ações tomadas sem precisar armazenar os valores históricos. É obtida através da equação (1).

$$\mathrm{Q}(s,a) \leftarrow \mathrm{Q}(s,a) + \alpha\left(r(s,a) + \gamma \min_{a'} \mathrm{Q}(s',a') - \mathrm{Q}(s,a)\right) \quad (1)$$

Em que, nesse trabalho: (i) a é a ação a ser tomada, ou a quantidade de medicamentos a serem compradas; (ii) s é o estado do sistema no momento da compra, aqui representado pela quantidade em estoque; (iii) α é o parâmetro tamanho do passo, que possui o poder de incluir o aprendizado no Qsa; (iv) $r(s,a)$ é o valor da punição; (v) γ é o coeficiente de desconto, com valor entre 0 e 1. Para modelos de gestão de compras, o valor a ser considerado é 1 [Mortazavi, 2015].

Também é necessário distinguir para saber se o agente está no modo exploração aprendiz (*exploration*) ou exploração experiente (*explotation*) [Sutton e Barto, 1998], porque a heurística pode ficar presa em alguns ótimos locais, mas que não necessariamente representem o ótimo global. Esse processo é feito comparando um valor aleatório com outro, ε, pré-determinado. Assim, o algoritmo para implementação do *Q-learning* está representado a seguir.

Início
iteração = 0; α *= valor inicial;* ε *= valor inicial; a = valor aleatório;* $\forall Qsa = 0$
Enquanto a simulação não termina
Simula por um período
iteração = iteração + 1
Escolhe um valor aleatório entre 0 e 1 e compara com ε
Se for menor que ε
Escolhe um valor de a aleatoriamente
Se for maior que ou igual a ε
Escolhe um valor de a de acordo com o menor Qsa

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left(r(s,a) + \gamma \min_{a'} Q(s',a') - Q(s,a)\right)$$

Decrementa α

Cada valor de Qsa relaciona uma ação, que é a quantidade a ser comprada, com o estado atual, ou a quantidade em estoque do medicamento. Para tornar essa informação mais tratável, ela será representada por um código que abrange amplitudes de valores [Mortazavi, 2015], [Kwon *et al.*, 2008].

3. Descrição do objeto de estudo

O objeto em estudo é uma farmácia hospitalar de um hospital público de grande porte, com mais de 200 leitos e vinculado a uma universidade federal. Essa unidade possui duas atividades básicas distintas. A primeira é o atendimento rotineiro e diário das solicitações internas de medicamentos e a segunda é a compra de medicamentos, como mostra a Figura 2.

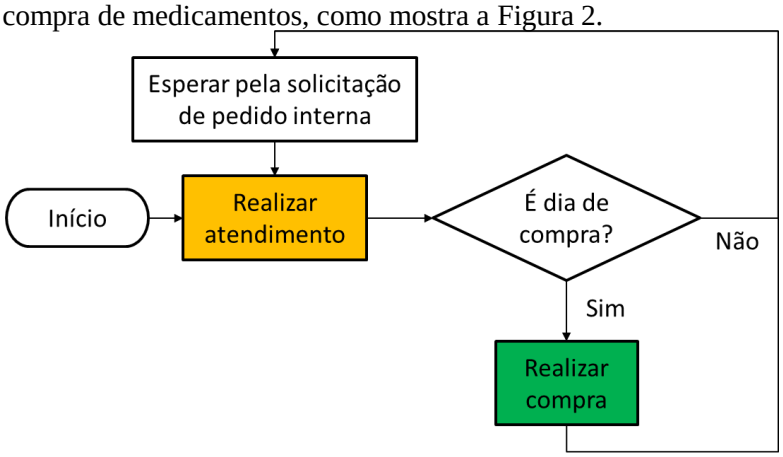

Figura 2 – Processo básico da farmácia hospitalar

As duas atividades básicas estão representadas na Figura 2. Na primeira (cor laranja), o farmacêutico e sua equipe, que a partir daqui serão representados simbolicamente apenas pelo farmacêutico, verifica se a demanda de cada medicamento solicitado possui correspondente em estoque. Essa é denominada de macro atividade aviamento. Na segunda atividade (cor verde), o farmacêutico realiza a solicitação das quantidades de medicamentos. Essa atividade possui uma restrição própria. É necessário que o valor total do pedido realizado caiba dentro de um orçamento pré-definido. Na primeira atividade, por ter características de processos rotineiros, utilizou-se a simulação a eventos discretos, e na segunda, por apresentar características de processo decisório, utiliza-se a simulação baseada em agentes.

3.1. Simulação a eventos discretos

Na Figura 3 é descrito o processo de atendimento de solicitação de pedido interno através da ferramenta IDEF-SIM.

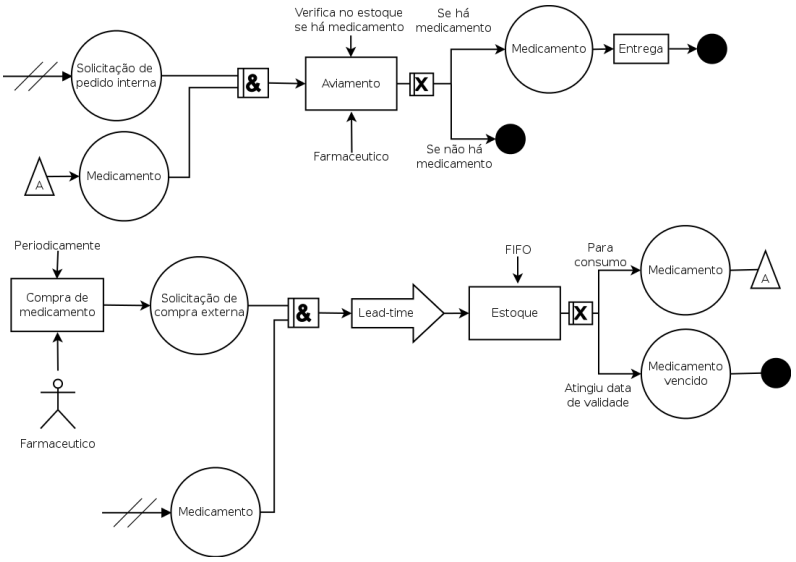

Figura 3 – IDEF-SIM do processo de atendimento

A SED foi utilizada para representar esse comportamento rotineiro do sistema. O processo tem início na chegada de uma entidade "Solicitação de pedido interna". Nela, estão presentes as quantidades individuais de cada medicamento que precisam ser atendidas internamente. Essa entidade, juntamente com a entidade "Medicamento", vindo do estoque, servem como entrada à tarefa aviamento, que verifica se há medicamento em estoque, faz a tradução da receita médica em quantidades de medicamentos e finalmente a distribuição. Para àquela demanda que não havia medicamento em estoque, faz-se um registro desse valor para cálculo posterior de punição.

Na metade inferior da figura, foi modelada a entrada de medicamentos externos à farmácia e possivelmente do próprio hospital, resultado da atividade de compra de medicamentos, que é uma saída da simulação baseada em agentes. Assim, após feito o pedido com as quantidades desejadas, aguarda-se um tempo, variável, para a entrega dos mesmos e entrada no estoque da farmácia hospitalar. Quando o medicamento se encontra no estoque, só há duas formas da saída do mesmo. Ou ele é requerido pela solicitação de pedido interna, explicado anteriormente, ou ele atinge o prazo de validade, sendo então descartado fisicamente, indisponível para o consumo. Também são realizados os registros desses valores de medicamentos expirados.

3.2. Simulação baseada em agentes

Na Figura 4 foi descrito o processo de solicitação de compra externa periódica, onde será utilizada a simulação baseada em agentes, por conseguir representar seu comportamento. O processo é iniciado com a atribuição de todas as variáveis Qsa´s para valor zero e escolha aleatória de uma ação de tamanho de lote para compra inicial. Posteriormente é decidido que, se o farmacêutico irá tomar uma decisão do tipo Aprendiz ou do tipo Experiente. Para isso, definiu-se um valor inicial ε de 0,3 e que decrementa no decorrer da simulação. Com esse valor é feito um teste lógico de comparação com um valor obtido aleatoriamente de uma função Uniforme entre 0 e 1 e, caso esse resultado seja maior, é escolhida a opção aprendiz, do contrário a opção experiente. Na opção aprendiz, a opção de tamanho de lote de compra é feita aleatoriamente, em todos os medicamentos, desde que a soma de seus valores sejam menores que o orçamento disponibilizado. Na opção de compra experiente, a escolha do tamanho de lote de compra é feita através do menor valor Qsa, sendo observado seu estado. Caso a soma dos valores dos lotes excedam o valor de orçamento disponibilizado, é utilizado a heurística gulosa de maior importância.

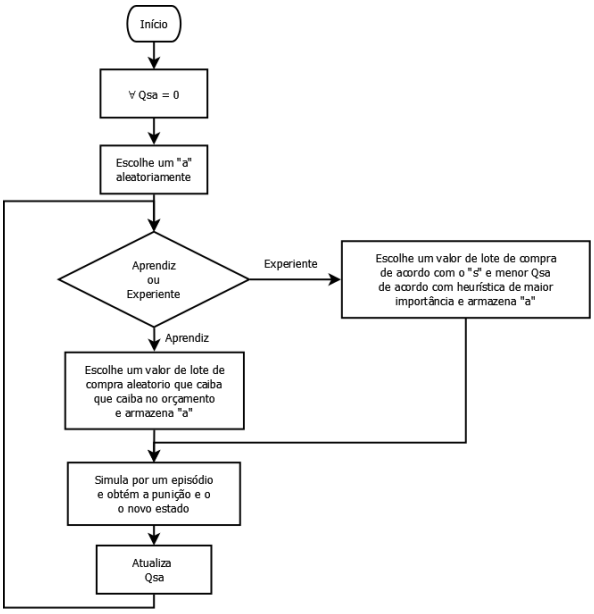

Figura 4 – Processo de compra externa periódica

A heurística funciona observando os valores de importância que são atribuídos a cada medicamento pelo farmacêutico e sua equipe. O seu funcionamento é observado na Figura 5.

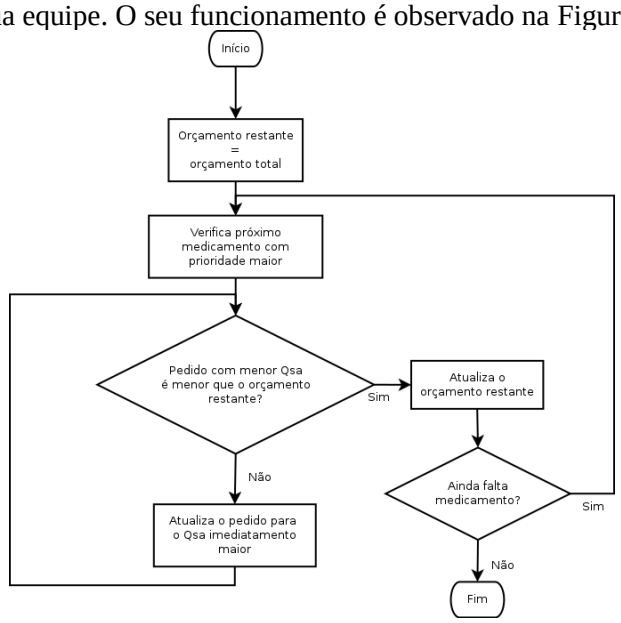

Figura 5 – Heurística de maior importância

Nessa heurística, utilizou-se uma variável denominada “orçamento restante”, que recebe no começo o mesmo valor que o orçamento disponível. Desse montante o medicamento com maior importância tem o seu pedido atendido de acordo com esse valor de orçamento restante. Então ele é atualizado pela diferença entre o seu valor anterior e o valor do pedido do medicamento de maior importância. Posteriormente é verificado o próximo medicamento com importância imediatamente inferior, caso não exista a heurística é encerrada. Do contrário, o mesmo procedimento de adequação feito no medicamento anterior é repetido. Ainda no processo de solicitação de compra externa, a simulação a eventos discretos é alimentada com esses valores de tamanhos de lotes, e são armazenados juntamente com o seu estado. Um período de 30 dias entre compras é simulado e os novos valores de punição e estado são obtidos. Por fim, o valor de Qsa, observando o estado e as quantidades de medicamentos compradas, é calculado. Dessa forma, o processo iterativo é realizado, com o agente farmacêutico aprendendo com as ações e consequências, ou punição, obtido da simulação.

3.3. Punição e valor *Qsa*

A punição será uma função dos valores obtidos da simulação a eventos discretos de quantidades de medicamentos não atendidos e medicamentos expirados. Porém, como os medicamentos apresentam características diferentes, como demanda e custo, optou-se por fazer o cálculo da punição levando em consideração os valores proporcionais de cada dimensão. Assim, a fórmula que calcula a punição é:

$$P_i = w_1 \frac{Q_{ni}}{D_{ni}} + w_2 \frac{Q_{epi}}{E_{li}} \quad (2)$$

Em que: (i) i é o índice do medicamento; (ii) Q_{ni} é a quantidade de medicamentos i não entregues; (iii) Q_{epi} é a quantidade proporcional teórica de medicamento i expirados; (iv) D_{ni} é a demanda do medicamento i para o período; (v) w_1 e w_2 são os pesos para cada dimensão; (vi) E_{li} é o estoque líquido do medicamento i.

Os valores de w_1 e w_2 correspondem respectivamente aos pesos subjetivos às dimensões falta de medicamento e expiração dos medicamentos. Quanto maior for esse valor, maior vai ser o impacto dessa dimensão na punição. A soma de seus valores tem que ser 1.

$$w_1 + w_2 = 1 \quad (3)$$

A dimensão falta de medicamento é um valor proporcional referente ao não atendimento em relação à demanda do período. A equação (4) foi usada para calcular o valor global de expiração para cada medicamento. Esse é o valor da quantidade proporcional de medicamentos expirados, presentes na equação (2).

$$Q_{ep} = \sum_{lotes} E_p * Q_l \quad (4)$$

Em que: (i) E_p é o valor correspondente a quanto um lote expirou proporcionalmente entre a data do início do prazo de solicitação e a data atual; (ii) Q_l é a quantidade de medicamentos em estoque em cada lote.

Assim, a dimensão da punição expiração será ligada diretamente à proporcionalidade da expiração dos medicamentos que não foram consumidos e ainda estão em estoque, em relação as suas parcelas que estavam em estoque no começo do período de solicitação. Essas parcelas são calculadas de acordo com a equação (5).

$$E_l = \overline{Q_{ep_0}} + P_0 \quad (5)$$

Em que: (i) $\overline{Q_{ep_0}}$ é o complemento de expiração do estoque no momento da solicitação; (ii) P0 é a quantidade solicitada para o medicamento.

O complemento da expiração do estoque em todos os lotes é obtido através da seguinte equação:

$$\overline{Q_{ep_0}} = \sum_{lotes} Q_0 * (1 - E_{p0}) \quad (6)$$

Em que: (i) Q_0 é a quantidade em estoque no momento da solicitação; (ii) Ep_0 é o valor correspondente a quanto um lote expirou proporcionalmente igual ao fim do período de solicitação anterior.

Por fim, para melhor representação do medicamento, utilizou-se o conceito de agente simples, ou proto-agente conforme [North e Macal, 2007] (Figura 6).

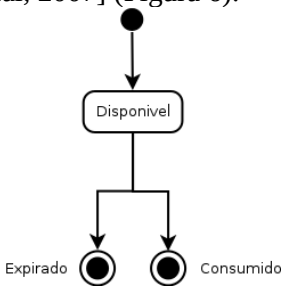

Figura 6 – Agente medicamento

O medicamento foi representado como um agente pois ele possui comportamento autônomo e independente do restante do modelo, além de possuir a capacidade de se comunicar com o sistema,

avisando que o medicamento se encontra expirado, e com o agente farmacêutico, recebendo a informação que o mesmo foi enviado para o consumidor final.

4. Aplicação

O estudo foi desenvolvido no ambiente computacional AnyLogic®. Primeiro, foi simulado por 200 anos utilizando os dados coletados de demanda no período de janeiro de 2012 a dezembro de 2015. A justificativa para esse período de simulação ser tão longo é que o número de pares estado x ação é muito grande. Segundo, os valores de Qsa obtidos são confrontados com dados e situações reais durante o período de janeiro de 2016 a julho de 2016, observando os resultados.

4.1. Dados de entrada

O valor médio das compras durante o primeiro período foi de aproximadamente de R$ 2.000,00. Dessa forma os parâmetros de simulação são:

Orçamento	w1;w2 (Medicamento 1)	w1;w2 (Medicamento 2)	w1;w2 (Medicamento 3)	Importância relativas (Medicamento 1, Medicamento 2 e Medicamento 3)
R$ 2.000,00	0,3;0,7	0,3;0,7	0,95;0,05	2,1,3

Tabela 1 – Parâmetros da simulação

A Tabela 2 apresenta os valores das amplitudes de cada estado e o tamanho de lotes de pedidos.

	Medicamento 1	Medicamento 2	Medicamento 3
Amplitude de cada estado	36	100	400
Tamanhos de lotes dos medicamentos	50	250	100

Tabela 2 – Amplitude de cada estado

O valor de ε será de 0,1 e de α será de 0,4. ε apresentará um valor fixo no decorrer da simulação e o valor de α terá um decréscimo de 0,001 a cada atualização em seu valor de Qsa. A simulação é executada por um período de 20.000 anos para que todos as combinações de valores (estado, ação) tenham sido visitadas em um número suficiente para lhe conferir um valor seguro. A simulação consumiu um tempo aproximado de 25,68 minutos.

4.2. Resultados

4.2.1. Política de compra

A política de compra é a representação de quanto o farmacêutico deve comprar em quantidade e em cada período, observando o estado atual. Os seus valores estão representados na Tabela 3, na Tabela 4 e na Tabela 5 e, organizados pela ordem de prioridade. No cabeçalho estão os estados e nas demais linhas as quantidades que devem ser compradas em cada situação, sendo a primeira linha a primeira tentativa, a segunda linha a segunda tentativa e assim sucessivamente. Convém destacar que nos medicamentos 1 e 2, as compras ficaram limitadas a 350 e 250 medicamentos, respectivamente, quando os mesmos estiverem no estado 0. No primeiro medicamento é possível tentar realizar a compra de tamanhos variados de lotes, para mais de uma tentativa, porém no segundo medicamento ou se compra 250 medicamentos ou não é feita a compra. Esse comportamento é justificado por esses medicamentos apresentarem baixas demandas e uma preocupação maior com a expiração. Já no terceiro medicamento, seu comportamento é o inverso dos demais, logo há uma necessidade de compra maior, mesmo nos estados maiores que 0.

Ainda de acordo com os valores obtidos, é salutar observar que, como o valor que é utilizado para determinar as melhores ações, o Qsa, utiliza em seu cálculo duas dimensões, perecibilidade e não atendimento, que fazem o valor do tamanho do pedido variarem distintamente, não há um comportamento previsível dos valores a serem comprados, como por exemplo o linear, nas sucessivas tentativas dentro de um mesmo estado. Assim, esses valores foram utilizados para orientar as compras nas demandas do ano de 2016, apresentados na sequência.

4.2.2. Dados reais

A Figura 7, a Figura 8 e a Figura 9 apresentam as quantidades das demandas e dos estoques dos medicamentos 1, 2 e 3, respectivamente, durante os meses de janeiro a julho de 2016. O medicamento 1, representado na Figura 7, já iniciou o ano com um estoque bem elevado, fato que pode ser observado pela comparação da quantidade de medicamentos em estoque (coluna verde) com a demanda (coluna azul). Já o medicamento 2 só começou a apresentar demanda a partir do

mês de abril e, como não tinha medicamento em estoque, não foi possível atender a demanda (coluna vermelha), permanecendo até junho sem medicamentos em estoque para poder atender. Já o medicamento 3 apresentou uma demanda baixa nos primeiros meses, o que foi aumentando significativamente no decorrer dos meses, com um período, em março, de não atendimento da demanda.

	Estados						
	0	1	2	3	4	5	
Quantidade	2200	2100	2100	1400	300	0	Quantidade
	2100	1900	1900	200	400	100	
	2000	1700	2000	600	500	200	
	1900	1600	2200	1100	600	300	
	1800	2200	1600	1600	700	400	
	1700	1800	1500	1300	800	500	
	1500	1400	1800	900	900	600	
	1600	1500	1400	2000	1000	700	
	1400	2000	1300	2100	1100	800	
	1300	1300	1200	1200	1200	900	
	1200	1200	1700	1900	1300	1000	
	1100	900	900	2200	1400	1100	
	1000	1100	800	1000	1500	1200	
	900	800	600	1500	1600	1300	
	700	1000	1100	1800	1700	1400	
	800	600	1000	300	1800	1500	
	600	700	500	800	1900	1600	
	500	500	400	700	2000	1700	
	400	400	700	0	2100	1800	
	300	300	300	100	2200	1900	
	200	200	200	400	0	2000	
	100	0	0	500	200	2100	
	0	100	100	1700	100	2200	

Tabela 3 – Política de compra do medicamento 3

	Estados											
	0	1	2	3	4	5	6	7	8	9	10	
Quantidade	350	0	0	0	0	0	0	0	0	0	0	Quantidade
	200	150	250	50	200	150	100	50	50	50	50	
	300	250	100	200	50	200	350	100	100	350	100	
	150	300	150	100	150	350	250	200	200	150	150	
	0	200	300	250	250	250	200	250	300	100	200	

Tabela 4 – Política de compra do medicamento 1

	Estados						
	0	1	2	3	4	5	
Quantidade	250	0	0	0	0	0	Quantidade
	0	250	250	250	3500	250	

Tabela 5 – Política de compra do medicamento 2

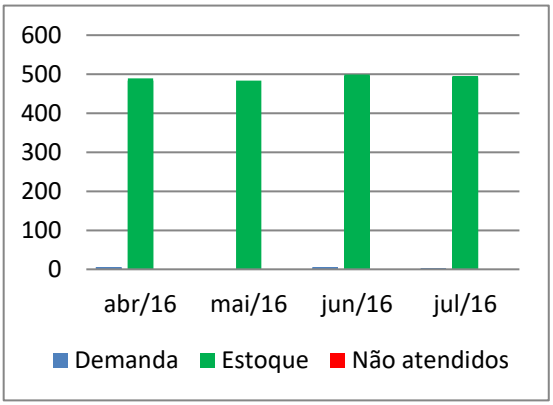

Figura 7 – Comportamento do medicamento 1 para o período janeiro a julho de 2016

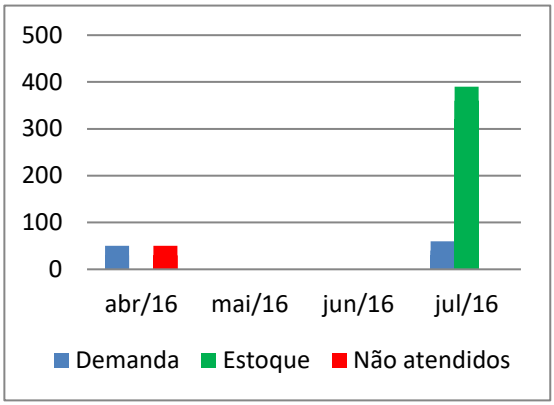

Figura 8 – Comportamento do medicamento 2 para o período janeiro a julho de 2016

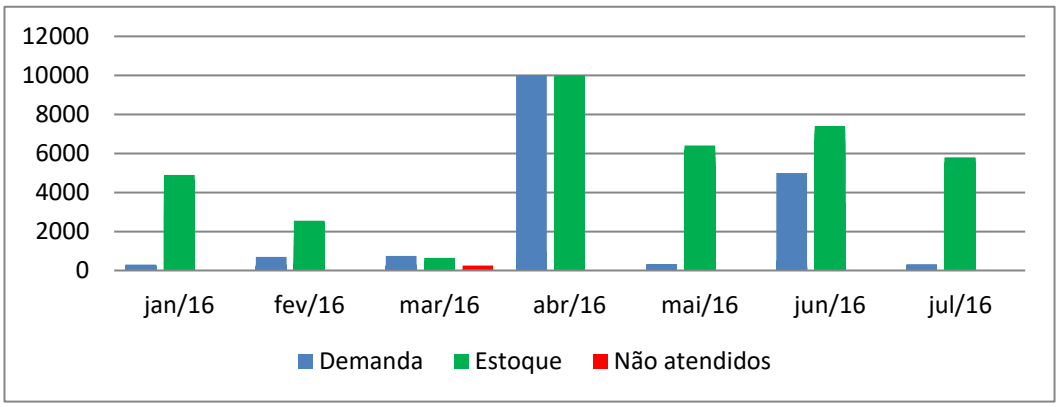

Figura 9 – Comportamento do medicamento 3 para o período janeiro a julho de 2016

Inicialmente foram apresentados os resultados do primeiro cenário e por ordem de importância de cada medicamento. Assim, primeiro será discutido sobre o medicamento 3 e depois do medicamento 2.

4.2.3. Dados simulados

A Figura 10 e a Figura 11 apresentam o comportamento do medicamento 3 e do medicamento 2 para o período de janeiro a julho de 2016, utilizando os valores de Qsa obtidos da simulação. O medicamento 1, devido ao seu comportamento apresentado no período, não teve decisões distintas daquela de não comprar mais unidades. Houve uma diminuição de não atendimentos do medicamento 3, de 400 para 0 e ainda exigiu um valor global de compras para o período. O gráfico de medicamento 2 mostra que o processo de aprendizado vai buscar manter um estoque para poder atender a demanda, e a partir daí, diminuir suas compras, pois ele só irá comprar medicamentos quando tiver até 2105 medicamentos.

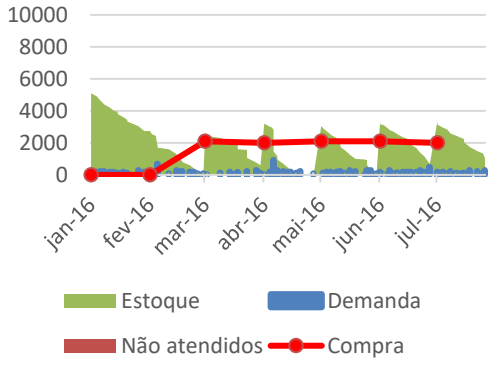

Figura 10 – Comportamento simulado do medicamento 3 para o período janeiro a julho de 2016

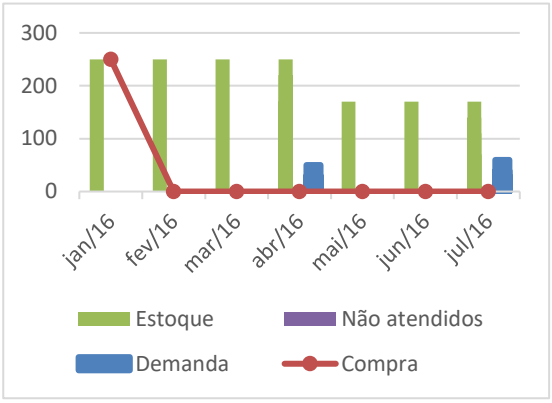

Figura 11 – Comportamento simulado do medicamento 2 para o período janeiro a julho de 2016

Assim, a Tabela 6 apresenta os dados reais e simulados da quantidade e valor de medicamentos expirados.

	Medicamento 1		Medicamento 2		Medicamento 3	
	Quantidade	Valor	Quantidade	Valor	Quantidade	Valor
Real	700	R$2.141,07	703	R$133,57	0	R$ -
Simulado	125	R$687,50	0	R$ -	0	R$ -

Tabela 6 – Quantidade simulada e real de medicamentos expirados

Como é possível observar, o método aqui apresentado conseguiu diminuir a quantidade de medicamentos expirados em relação aos valores reais. Reduzindo em 100% os valores de Medicamento 2 e aproximadamente 67,89 % dos valores de Medicamento 1. A Tabela 7 apresenta uma comparação dos valores reais e simulados do valor de compra mês a mês.

Medicamento 3		Medicamento 2		TOTAL	
Real	Simulado	Real	Simulado	Real	Simulado
R$14.175,00	R$9.540,00	R$90,00	R$47,50	R$14.265,00	R$9.587,50

Tabela 7 – Valores relativos aos gastos reais e simulados

O método utilizado apresentou uma economia de 32,79 % nos gastos em relação à situação real.

5. Conclusão

A utilização da inteligência artificial com a ferramenta aprendizado por reforço apresentou-se útil na capacidade sintética do processo de compra de medicamento em uma farmácia hospitalar de um hospital público. A representação do agente farmacêutico auxiliou nesse processo, pois o uso da simulação baseada em agentes possui como características a modularização e o embarque de ferramentas adicionais, como a RL e a heurística de compra limitada no orçamento.

O resultado desse processo, a política de compras, permite que o farmacêutico e sua equipe tenham subsídios para uma tomada de decisão segura e rápida, bastando dessa forma, apenas observar a quantidade individual de medicamentos em cada estoque. Que torna a sua utilização bem mais acessível do que outros métodos utilizados, como o método “Q, s”, ou até mesmo a otimização matemática. Recomenda-se também que, essa política não seja eterna, exigindo-se estudos frequentes, como por exemplo, anualmente, para fins de controle de sua eficácia.

Os resultados mostram que houve uma melhoria significativa, com 100% de diminuição de não atendimento para o medicamento 3, considerado muito importante, uma diminuição de 100% e 67,89 % da perecibilidade dos medicamentos 2 e 1, respectivamente, em um ano de simulação e uma diminuição de 32,79 % dos gastos simulados em comparação aos reais o ambiente estudado. Ainda assim, acredita-se que há possibilidades de melhorias, já que o foi realizado com apenas três medicamentos, o que pode ser estendido aos demais medicamentos. Mas pelo observado, a inclusão a estrutura aqui apresentada já está pronta para essa possibilidade.

6. Referências

[1] S. A. Narayana, R. Kumar Pati, and P. Vrat, “Managerial research on the pharmaceutical supply chain - A critical review and some insights for future directions,” *J. Purch. Supply Manag.*, vol. 20,

no. 1, pp. 18–40, 2014.
[2] A. Yurtkuran and E. Emel, “Simulation based decision-making for hospital pharmacy,” in *Proceedings of the 2008 Winter Simulation Conference*, 2008, pp. 101–112.
[3] S. H. Jacobson, S. N. Hall, and J. R. Swisher, “Chapter 8 DISCRETE-EVENT SIMULATION OF HEALTH CARE SYSTEMS,” *Patient Flow Reducing Delay Healthc. Deliv.*, pp. 211–252, 2006.
[4] C. Jiang and Z. Sheng, “Case-based reinforcement learning for dynamic inventory control in a multi-agent supply-chain system,” *Expert Syst. Appl.*, vol. 36, no. 3 PART 2, pp. 6520–6526, 2009.
[5] M. Pidd, “Computer Simulation in Management Science,” 2004.
[6] M. Lee, “A study of evolution strategy based cooperative behavior in collective agents,” *Artif. Intell. Rev.*, vol. 25, no. 3, pp. 195–209, 2007.
[7] L. P. Kaelbling, M. L. Littman, and A. W. Moore, “Reinforcement learning: A survey,” *J. Artif. Intell. Res.*, vol. 4, pp. 237–285, 1996.
[8] R. S. Sutton and A. G. Barto, “Reinforcement Learning : An Introduction,” *IEEE Trans. Neural Netw.*, vol. 9, no. 5, p. 1054, 1998.
[9] T. Katanyukul and E. K. P. Chong, “Intelligent Inventory Control via Ruminative Reinforcement Learning,” *J. Appl. Math.*, vol. 2014, 2014.
[10] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley \& Sons, 2014.
[11] A. Valluri, M. J. North, and C. M. Macal, “Reinforcement learning in supply chains.,” *Int. J. Neural Syst.*, vol. 19, no. 5, pp. 331–344, 2009.
[12] C. J. C. H. Watkins, “Learning from delayed rewards,” University of Cambridge England, 1989.
[13] A. Mortazavi, A. A. Khamseh, and P. Azimi, “Designing of an intelligent self-adaptive model for supply chain ordering management system,” *Eng. Appl. Artif. Intell.*, vol. 37, no. JANUARY, pp. 207–220, 2015.
[14] I. H. Kwon, C. O. Kim, J. Jun, and J. H. Lee, “Case-based myopic reinforcement learning for satisfying target service level in supply chain,” *Expert Syst. Appl.*, vol. 35, no. 1–2, pp. 389–397, 2008.
[15] M. J. North and C. M. Macal, *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation*, vol. I, no. 3. New York, NY, USA: Oxford University Press, Inc., 2007.

Agradecimentos

Os autores agradecem à CAPES, ao CNPq e à FAPEMIG pelo apoio financeiro no desenvolvimento desta pesquisa.