

MODELO HIPERCUBO INTEGRADO A UM ALGORITMO GENÉTICO PARA ANÁLISE DE UM SISTEMA MÉDICO EMERGENCIAL EM RODOVIAS

Ana Paula Iannoni

Departamento de Engenharia de Produção
Universidade Federal de São Carlos
13565-905 São Carlos, SP
papi@iris.ufscar.br

Reinaldo Morabito

Departamento de Engenharia de Produção
Universidade Federal de São Carlos
13565-905 São Carlos, SP
morabito@power.ufscar.br

Resumo:

O objetivo deste trabalho é analisar a configuração e operação de um sistema de atendimento médico emergencial (SAE) em rodovias utilizando o modelo hipercubo. Este modelo é baseado em teoria de filas espacialmente distribuídas, e permite representar as incertezas dos SAE levando em conta a individualidade de seus servidores. Como o modelo hipercubo é um modelo descritivo, combinamos o hipercubo com um algoritmo genético para obter um método prescritivo capaz de otimizar a configuração e operação de SAEs em rodovias. Para analisar o desempenho desta abordagem, realizamos estudos com dados reais do sistema *Anjos do Asfalto* (rodovia Presidente Dutra).

Palavras-chaves: modelo hipercubo, filas espacialmente distribuídas, despacho de ambulâncias.

Abstract:

The purpose of this study is to analyze the configuration and operation of Emergency Medical Systems (EMS) on highways using the hypercube model. This model is based on spatially distributed queuing theory. It allows representing the uncertainties of the EMS, considering the identity of the servers. Since the hypercube model is a descriptive model, we combine the hypercube model with a genetic algorithm to obtain a prescriptive method, which can optimize the configuration and operation of EMS on highways. To analyze the performance of this proposed approach, we conducted studies based on the real data of the system *Anjos do Asfalto* (Presidente Dutra highway).

Keywords: hypercube model, spatially distributed queues, ambulance deployment.

1. Introdução:

Nos sistemas de atendimento médico emergencial em rodovias brasileiras, a rapidez no atendimento a um chamado é uma das principais medidas de desempenho, dado que o atraso no tempo de resposta pode resultar em seqüelas e estado de invalidez dos acidentados. Além disso, em muitas situações, pode também significar a diferença entre a vida e morte das vítimas envolvidas.

Ao analisar os sistemas de atendimento emergencial (SAEs), os fatores probabilísticos relacionados à distribuição temporal e espacial dos servidores e chamadas devem ser considerados, dado que a operação destes sistemas é caracterizada por incertezas com relação à localização e tempo necessário para atender um determinado chamado. O atraso no tempo de resposta está diretamente relacionado ao conflito entre as variáveis aleatórias da demanda por serviço e as restrições de capacidade do sistema. Assim sendo, estes sistemas podem ser analisados por meio de modelos probabilísticos, que consideram as relações entre a demanda por serviços e o tempo de espera para atendimento aos usuários.

Importantes contribuições podem ser encontradas na literatura de estudos sobre métodos de otimização que incorporam os aspectos probabilísticos dos SAEs. No entanto, a maioria destes métodos apenas consideram a aleatoriedade associada à disponibilidade dos servidores e não assumem que há outros aspectos probabilísticos que devem ser considerados na análise. Os estudos de Larson & Odoni (1981), Swersey (1994), Chiyoshi et al. (2000) e Brotcorne et al. (2003), revisam os principais modelos de localização para analisar os sistemas de atendimento emergencial, desenvolvidos nas últimas décadas.

O modelo hipercubo, desenvolvido por Larson (1974) e estendido por vários autores (Swersey, 1994), é baseado nos resultados de Teoria de Filas Espacialmente Distribuídas e aproximações Markovianas. Este modelo tem se mostrado como um método preciso e robusto para analisar os sistemas de emergência, pois permite representar as incertezas de um SAE, considerando a identidade dos servidores, assim como a cooperação e/ou iteração entre os mesmos. O modelo implica na solução de sistema linear de $O(2^N)$ equações, cujas variáveis envolvidas correspondem às probabilidades de estado do sistema em equilíbrio. Através destas probabilidades podem ser estimadas importantes medidas de performance para análise e gerenciamento do sistema, tais como cargas de trabalho, tempo médio de resposta e frações de despacho de cada servidor a cada região. Além de sistemas que consideram fila de espera, o modelo também pode ser aplicado a sistema sem filas como é o caso dos SAEs nas rodovias brasileiras.

Como o modelo hipercubo original de Larson (1974) assume hipóteses que limitam sua aplicação, diversas extensões vêm sendo estudadas para análise de sistemas de emergência, principalmente em patrulhamento policial e despacho de ambulâncias. Alguns estudos podem ser encontrados em Larson (1975), Jarvis (1985), Chelst & Barlach (1981) e Mendonça & Morabito (2000; 2001).

Uma direção promissora na análise dos SAE parece ser a integração do modelo hipercubo em métodos de otimização. As principais publicações com este enfoque são os estudos de Batta et al. (1989), Saydam & Aytug (2003), Chiyoshi et al. (2003), Galvão et al. (2003a) e Galvão et al. (2003b). No Brasil, alguns exemplos de aplicação do modelo hipercubo são: o atendimento a interrupções na distribuição de energia elétrica em Santa Catarina (Albino, 1994) e o Serviço de Atendimento Médico de Urgência (SAMU) da prefeitura de Campinas, SP (Takeda et al, 2004). Em particular, Mendonça & Morabito (2000; 2001) analisaram o SAE “Anjos do Asfalto”, que atendia um trecho da Via Dutra entre as cidades de São Paulo e Rio de Janeiro.

Apesar do recente aprimoramento dos sistemas de atendimento ao usuário, o número de acidentes em rodovias brasileiras ainda é preocupante. Além disso, estes sistemas têm sido muito pouco estudados. Considerando a relevância dos SAEs em rodovias, neste estudo é proposto um método que integra o modelo hipercubo em um algoritmo genético para analisar os sistemas de atendimento médico emergencial em rodovias brasileiras. Através desta

abordagem é possível prescrever qual a configuração ótima do SAE analisado, em termos das principais medidas de desempenho deste sistema. O estudo de caso do sistema *Anjos do Asfalto* na rodovia Presidente Dutra, que foi inicialmente estudado por Mendonça & Morabito (2000; 2001), foi utilizado como base para a implementação inicial da presente abordagem que combina um algoritmo genético com o modelo hipercubo. Um dos aspectos inovadores desta abordagem é tratar a variação dos tamanhos dos trechos de rodovia, como forma de determinar quais as áreas de cobertura de cada servidor. Este artigo está organizado da seguinte forma: a seção 2 apresenta uma breve descrição dos sistemas de atendimento médico emergencial em rodovias, a seção 3 descreve o modelo hipercubo para análise do SAE *Anjos do Asfalto*; a seção 4 descreve o algoritmo genético integrado ao modelo hipercubo, proposto no estudo; a seção 5 apresenta os resultados obtidos por esta abordagem aplicada aos *Anjos do Asfalto* e a seção final discute conclusões deste estudo e apresenta as referências consultadas.

2. O sistema de atendimento médico emergencial em rodovias brasileiras:

Os sistemas de atendimento emergencial (SAEs) em rodovias tem a função de socorrer as vítimas de acidente nas rodovias e, se necessário realizar o transporte das mesmas ao hospital da cidade mais próxima. Estes sistemas também podem realizar outros tipos de atendimento médico em sua base localizada em um ponto da rodovia.

Como descrito a seguir, os SAEs em rodovias são caracterizados por não admitirem fila de espera, pois quando os servidores candidatos estão ocupados, a chamada deve ser transferida a outro sistema como o corpo de bombeiro ou a ambulância da cidade mais próxima, podendo também ser atendida por um outro SAE vizinho. Estes sistemas em geral também admitem políticas particulares de despacho, sendo que nem todos os servidores podem ser despachados para atender um determinado chamado devido às restrições de distância. Alguns destes SAEs também possuem uma política de múltiplo despacho, sendo que, em alguns casos torna-se necessário despachar mais de um veículo para atender um acidente.

2.1 O sistema *Anjos do Asfalto*:

Como descrito em Mendonça & Morabito (2000, 2001), o sistema *Anjos do Asfalto* era uma fundação não governamental sem fins lucrativos que fornecia atendimento médico emergencial em parte da rodovia Presidente Dutra entre as cidades de São Paulo e Rio de Janeiro. Este SAE conta com 6 bases fixas ao longo do trecho da rodovia, sendo que, cada base possui uma ambulância e uma equipe composta por médicos, enfermeiros e motorista, que viajam juntos para o local do chamado. A central de operações, localizada na cidade do Rio de Janeiro, é responsável por receber as chamadas, despachar as ambulâncias e monitorar os movimentos da mesma. Ao receber um chamado, a central imediatamente envia a ambulância disponível mais próxima do local do chamado. A política de despacho consta em despachar a ambulância localizada na base mais próxima, e se esta estiver ocupada a segunda mais próxima é enviada (chamada de *backup*). Se as duas ambulâncias mais próximas estiverem ocupadas, o chamado é transferido para outro sistema (por exemplo, o hospital mais próximo) e a chamada é considerada uma perda para o sistema. Note que, nesta política de despacho particular, cada região pode ser atendida por somente dois servidores (o servidor preferencial e o servidor *backup*). Além disso, neste sistema apenas um servidor pode ser enviado a cada chamada.

A figura 1 ilustra a distribuição das bases dos servidores no sistema *Anjos do Asfalto*. A distância entre duas bases é dividida em duas regiões, cada uma com uma lista de preferência de despacho. De acordo com esta lista, e exceto para as bases das extremidades (1 e 6), todos os servidores são despachados como preferenciais para duas regiões (à esquerda e direita de

sua base) e como *backup* para outras duas regiões (direita e esquerda dos servidores adjacentes à esquerda e direita, respectivamente).

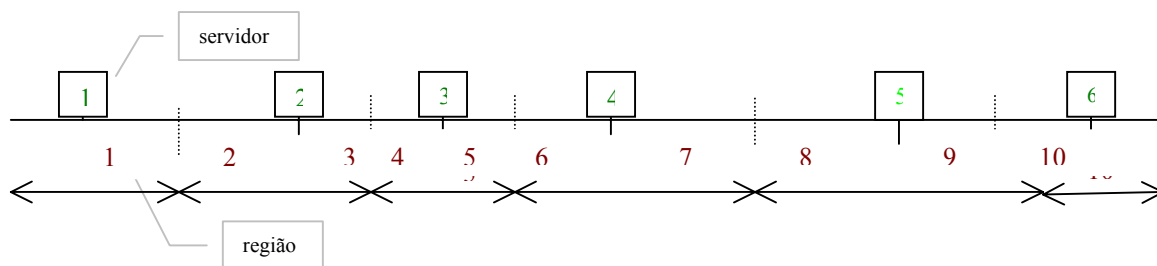


Figura 1 – Simples esquema da distribuição de átomos e servidores ao longo da rodovia no sistema *Anjos do Asfalto*

Por exemplo, para o servidor 5: o lado esquerdo deste servidor (região 8) e o lado direito (região 9) correspondem à sua área primária, ao passo que, o lado direito do servidor 4 (região 7) e o lado esquerdo do servidor 6 (região 10) correspondem à sua área de atendimento *backup*. Os servidores 1 e 6 possuem área primária de uma região (região 1 e região 10, respectivamente), e área *backup* de uma região (região 2 e região 9, respectivamente).

3. O modelo hipercubo aplicado aos SAEs em rodovias:

O nome hipercubo é derivado do espaço de estado do sistema que representa os possíveis estados dos servidores. Dado que há dois estados possíveis para cada servidor: livre (0) ou ocupado (1) em certo instante de tempo, temos então 2^N estados para o sistema. Um estado em particular do sistema é representado pela lista de servidores que estão livres e ocupados. Por exemplo, em um sistema com $N = 3$ servidores, o estado 101 corresponde ao estado em que os servidores 1 e 3 estão ocupados e o servidor 2 está livre. A idéia básica do modelo hipercubo é expandir o espaço de estados de um modelo de filas simples para sistemas com múltiplos servidores (i.e $M/M/N$ ou $M/G/N$, onde N é o número de servidores) para representar cada servidor individualmente e incorporar as complexidades das políticas de despacho.

As principais hipóteses do modelo hipercubo para aplicação em rodovias considerando suas particularidades são:

- A rodovia é particionada em N_a átomos geográficos, e cada átomo corresponde a uma fonte independente de chamadas.
- Processo de Chegadas: chamadas de emergência em cada átomo são geradas de acordo com o processo de Poisson de forma independente dos demais átomos.
- Servidores: há N servidores espacialmente distribuídos ao longo da rodovia, que permanecem fixos em suas bases quando disponíveis. Como mencionado anteriormente, de acordo com a política particular dos SAEs em rodovias, cada ambulância pode somente viajar para átomos de suas áreas preferencial e *backup*.
- Política de despacho: o despacho dos servidores é realizado de acordo com uma lista de preferência. Quando ocorre uma chamada em um determinado átomo, somente os servidores desta lista podem ser despachados. Quando ocorre uma chamada a ambulância mais próxima é despachada e se esta estiver ocupada a segunda mais próxima é enviada

(*backup*). Assim, uma chamada pode ser atendida por um servidor preferencial ou um *backup*.

- Tempo de atendimento: o tempo médio de atendimento para cada servidor j ($1/\mu_j$) é conhecido e inclui o tempo de *set-up* (preparação), o tempo de viagem do servidor ao local da chamada, o tempo em cena e o tempo de retorno do servidor a sua base. O modelo também admite que o desvio padrão dos tempos de atendimento é aproximadamente igual à média (pois o tempo de atendimento é representado por uma distribuição exponencial negativa). Porém, razoáveis aproximações desta hipótese não representam significantes alterações na acuracidade do modelo (Larson & Odoni, 1981). Se o sistema não admite filas, como é o caso presente, esta suposição é ainda mais desnecessária, pois os modelos M/M/N e M/G/N têm a mesma distribuição de equilíbrio.
- Tempo de viagem: o tempo de viagem entre cada par de átomos é conhecido ou pode ser estimado através de conceitos de probabilidade geométrica. As variações no tempo de atendimento devido a variações no tempo de viagem são consideradas de segunda ordem quando comparadas a variações no tempo em cena ou no tempo de *set-up*.

A específica política de despacho dos SAEs em rodovias também requer modificações na determinação das equações de equilíbrio dos estados do sistema com relação ao modelo hipercubo original. Note que, naquele modelo uma chamada somente é perdida se todos os servidores do sistema estão ocupados, pois qualquer servidor pode atender qualquer chamada. No modelo hipercubo aplicado ao SAE em rodovias chamadas podem ser perdidas se os servidores de sua lista de despacho estão ocupados mesmo que há outros servidores disponíveis no sistema.

3.1 Aplicação do modelo hipercubo para um SAE com único despacho (caso dos *Anjos do Asfalto*):

A coleta e análise estatística dos dados do sistema *Anjos do Asfalto* foi realizada por Mendonça & Morabito (2000; 2001). Desta forma, os dados de entrada apresentados a seguir foram analisados e descritos com mais detalhes naquele estudo. O trecho linear da rodovia foi particionado em 10 átomos e para cada átomo atribui-se um servidor preferencial e um servidor *backup*, de acordo com a operação real deste sistema. Como cada servidor, com exceção dos servidores 1 e 6 são preferenciais para o seu lado esquerdo e direito, temos que, de acordo com a figura 1 os servidores preferencial e *backup* de cada átomo são : átomo 1 (1 e 2); átomo 2 (2 e 1); átomo 3 (2;3); átomo 4 (3 e 2); átomo 5 (3 e 4); átomo 6 (4 e 3); átomo 7 (4 e 5); átomo 8(5 e 4); átomo 9 (5 e 6); átomo 10 (6 e 5). A taxa total de chegadas no sistema é $\lambda = 0,01813$ chamadas/min (i.e, 1,0878 chamadas/hora). A taxa total de atendimento considerando na análise que os servidores são não-homogêneos é $\mu = 0,0959$ chamadas/min,

e portanto a taxa de ocupação do sistema é $\rho = \frac{\lambda}{\mu} = 0,1890$. A matriz dos tempos de viagem

foi determinada no presente estudo a partir da distância (base do servidor - centróide do átomo), conforme discutido em Larson & Odoni (1981).

Os dados de entrada e resultados dos testes estatísticos para os processos de chegada e atendimento podem ser encontrados com mais detalhes em Mendonça & Morabito (2000; 2001).

Equações de equilíbrio:

A equação de equilíbrio para cada estado do sistema pode ser construída considerando as transições para fora do estado e as transições para dentro do estado. Como o sistema *Anjos do Asfalto* possui $N = 6$ servidores, há $2^N = 64$ estados possíveis. As equações são determinadas de acordo com a lista de despacho da tabela 1. Por exemplo, para o estado $B = \{110001\}$, a equação de equilíbrio torna-se:

$$p_{110001}((\lambda_3 + \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 + \lambda_8 + \lambda_9 + \lambda_{10}) + \mu_1 + \mu_2 + \mu_6) = p_{100001}(\lambda_1 + \lambda_2 + \lambda_3) + p_{010001}(\lambda_1 + \lambda_2) + p_{110000}(\lambda_{10}) + p_{111001}(\mu_3) + p_{110101}(\mu_4) + p_{110011}(\mu_5)$$

Note que, o termo do lado esquerdo da equação, representando o fluxo para fora do estado $B = \{110001\}$, mostra que o sistema neste estado não pode atender chamadas geradas nos átomos 1 e 2, pois seus dois servidores preferencial e *backup* (1 e 2) estão ocupados. Note que diferentemente do modelo hipercubo de Larson, chamadas são perdidas mesmo com servidores disponíveis.

Calculando as probabilidades de estado do sistema é possível obter interessantes medidas de desempenho para os SAE em rodovias, tais como: probabilidade de perda, carga de trabalho das ambulâncias, frequências de despacho e tempos médios de viagem. Ao revisar aplicação do modelo hipercubo ao sistema *Anjos do Asfalto*, reproduzimos os resultados obtidos por Mendonça & Morabito (2000; 2001) para as medidas analisadas. Um modelo de simulação do sistema foi construído para verificar a acuracidade do modelo hipercubo na análise, e os desvios obtidos foram pouco significativos (média menor que 1%).

4. Desenvolvimento de um algoritmo genético combinado pelo modelo hipercubo

Neste estudo, desenvolvemos um algoritmo que integra o modelo hipercubo com um algoritmo genético para determinar uma configuração ótima (ou perto da ótima), estabelecida pelo tamanho dos átomos do sistema. Os principais componentes considerados na implementação do algoritmo estão descritos nas subseções a seguir. Convém salientar que desconhecemos trabalhos anteriores que tratem a variação no tamanho dos átomos do sistema.

O primeiro passo da implementação do algoritmo genético/hipercubo foi criar um procedimento que permita que novos dados de entrada sejam gerados para o modelo hipercubo de acordo com diferentes configurações estabelecidas pela variação no tamanho de cada átomo do sistema. Para uma dada configuração, o algoritmo calcula as novas taxas de chegada de forma a preservar a distribuição original de demanda ao longo da rodovia e a nova matriz de tempos de viagem de cada servidor ao centróide de cada átomo. A nova taxa de chegada é calculada considerando cada par de átomos adjacentes da seguinte forma:

Seja:

N – número de servidores;

N_a - número de átomos;

l_d - distância entre dois átomos adjacentes, $d = 1 \dots (N-1)$;

y_d - proporção de l_d que é medida do servidor d ao servidor $d+1$ (da esquerda para a direita);

s_j - localização do servidor j , $j = 1 \dots N$;

x_i^0 - tamanho inicial do átomo i , $i = 1 \dots N_a$;

x_j - tamanho final do átomo i , $i = 1 \dots N_a$;

λ_i^0 - taxa de chegada inicial do átomo i , $i = 1 \dots N_a$;

λ_i - taxa de chegada final do átomo i , $i = 1 \dots N_a$;

$$\text{Se } x_i < x_i^0 \text{ então } \begin{cases} \lambda_i = x_i \times \left(\frac{\lambda_i^0}{x_i^0} \right), \\ \lambda_{(i+1)} = \lambda_{(i+1)}^0 + (x_{(i+1)} - x_{(i+1)}^0) \times \left(\frac{\lambda_i^0}{x_i^0} \right) \end{cases}$$

$$\text{Se } x_i > x_i^0 \text{ então } \begin{cases} \lambda_i = \lambda_i^0 + (x_i - x_i^0) \times \left(\frac{\lambda_{(i+1)}^0}{x_{(i+1)}^0} \right), \\ \lambda_{(i+1)} = x_{(i+1)} \times \left(\frac{\lambda_i^0}{x_i^0} \right) \end{cases}$$

onde x_i - novo tamanho do átomo i , x_i^0 - tamanho inicial do átomo i , λ_i^0 - taxa de chegada inicial do átomo i e λ_i - nova taxa de chegada no átomo i , para $i = 1 \dots N_a$ (número de átomos). Note que, admitimos que a demanda está uniformemente distribuída ao longo da rodovia.

4.1 Representação dos cromossomos:

No presente estudo, utilizamos uma representação decimal. Através desta representação, cada gene do cromossomo é representado pela variável y , que é referente à proporção da distância entre dois servidores adjacentes que corresponde a um dos seus átomos primários. Lembre-se que, no sistema *Anjos do Asfalto* exceto para os servidores das extremidades (1 e N), cada servidor possui dois átomos como área primária (sua direita e esquerda). Desta forma, o número de genes em cada cromossomo será $N-1$, onde N é o número de servidores. O exemplo da figura 2 ilustra a idéia geral desta representação.

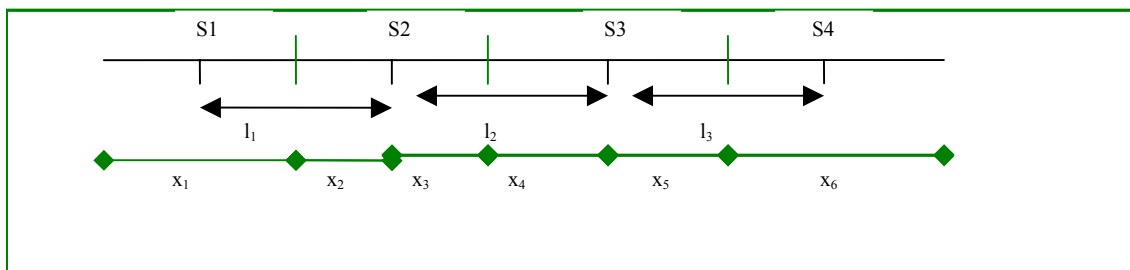
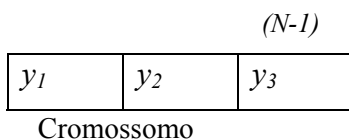


Figura 2 – Exemplo ilustrativo

No exemplo acima, se o cromossomo (solução) é $y_1 = 0.2$, $y_2 = 0.5$, $y_3 = 0.8$ para um sistema com 4 servidores na figura 2 e analisando o servidor 2: l_2 (distância entre os servidores 2 e 3), $y_2 = 0.5$, então $x_3 = l_2 \times y_2$ e $x_4 = (l_2 - x_3)$, onde x_3 é átomo primário à direita do servidor 2 e

x_4 é átomo primário à esquerda do servidor 3. Desta forma, o tamanho dos átomos em cada possível solução gera as taxas de chegada que são utilizadas pelo modelo hipercubo.

4.2 População Inicial:

A população inicial foi criada de forma aleatória, através de duas alternativas: discreta e contínua. No caso discreto, para cada gene há um vetor de valores entre 0.2 e 0.8 que delimitam as partições da rodovia. Desta forma, a extensão de cada átomo primário de um dado servidor é limitada à 20/80 por cento da distância total aos seus servidores vizinhos. O procedimento então determina como a distância entre dois servidores adjacentes deve ser particionada, adicionando à 0.2 um valor de incremento Δ (por exemplo 0.01) multiplicado por um inteiro k , onde $k = 0 \dots M = [0.6/\Delta]$. Portanto, para cada gene, atribui-se um número aleatório entre 1 e M indexando o valor correspondente do vetor. Por exemplo, se o valor aleatório gerado para o gene 1 é 2 e $\Delta = 0.01$, o valor do gene 1 é 0.21 (0.2 + 0.01). O mesmo vetor discreto foi utilizado no processo de substituição dos genes no operador de mutação. No caso contínuo, são gerados valores aleatórios entre 0.2 e 0.8 para cada gene do cromossomo.

4.3 Avaliação e função *fitness*:

O procedimento de avaliação do presente algoritmo genético é baseado no modelo hipercubo. Para cada cromossomo (solução), representando uma dada configuração do sistema, há uma taxa de chegada para cada átomo e uma matriz de tempos de viagem (de cada servidor a cada átomo) que correspondem aos dados de entrada do modelo hipercubo. Através da aplicação deste modelo no procedimento de avaliação, obtemos as medidas de performance de cada solução individualmente na população.

Como mencionado anteriormente, o modelo hipercubo fornece várias medidas interessantes que descrevem o desempenho do sistema. Entretanto, algumas destas medidas podem ser conflitantes em termos dos diferentes interesses envolvidos. Por exemplo, o balanço das cargas de trabalho dos servidores é uma medida de desempenho interna do sistema, que interessa ao gerente do sistema. Este precisa otimizar a utilização dos restritos recursos. Por outro lado, o tempo médio de resposta na região é uma medida de desempenho externa do sistema, que interessa ao usuário do sistema. Devido a este fato, inicialmente realizamos três experimentos com três diferentes funções de aptidão (*fitness*) que representam três objetivos: (i) desvio padrão das cargas de trabalho dos servidores; (ii) tempo médio de resposta para sistema; (iii) fração de chamadas atendidas em mais de 10 minutos.

No primeiro caso, o valor do desvio padrão entre as cargas de trabalho é utilizado para medir o desbalanceamento entre as mesmas, sendo que o objetivo é minimizar esta medida. Por meio da abordagem (algoritmo genético/ hipercubo) procuramos encontrar uma configuração que assegure um melhor balanço das cargas de trabalho. A função *fitness* é dada a seguir :

$$\sigma_{\rho}(c) = \sqrt{\frac{\sum(\rho_j(c) - \bar{\rho}(c))^2}{N-1}}, \text{ onde } \sigma_{\rho}(c) \text{ é o desvio padrão das cargas de trabalho da}$$

solução c e $\rho_j(c)$ representa a carga de trabalho do servidor j na solução c e $\bar{\rho}(c)$ é média das cargas de trabalho na solução c .

Para o segundo caso, o objetivo é oferecer um melhor nível de serviço ao usuário através da minimização do tempo médio de viagem do sistema. Desta forma, a seguinte equação é utilizada como função *fitness* :

$$\bar{T}(c) = \sum_{j=1}^N \sum_{i=1}^{N_A} f_{ji} t_{ji}, \text{ onde } \bar{T}(c) \text{ corresponde ao tempo médio de resposta no sistema}$$

para uma dada solução c , f_{ji} é a fração de despachos do servidor j ao átomo i , e t_{ji} é o tempo de viagem do servidor j ao átomo i .

A terceira função *fitness* corresponde a fração de chamadas que são atendidas em tempo superior à 10 minutos. Como uma das medidas de desempenho fornecidas pelo modelo hipercubo é a fração de despachos de cada servidor a cada átomo e dado que a matriz dos tempos de viagem servidor – átomos é conhecida, podemos então calcular a proporção total de chamadas que requerem mais de 10 minutos para serem atendidas. Novamente, o objetivo é minimizar o valor desta medida. A equação avaliada é:

$$P_{tr>10}(c) = \sum_{j=1}^N \sum_{i=1}^{N_A} f_{ji}^v, \text{ onde } P_{tr>10}(c) \text{ é a fração de chamadas que são atendidas em tempo}$$

superior à 10 minutos para a solução c , f_{ji}^v é a fração de despachos do servidor j – átomo i , que correspondem a um tempo de viagem superior à 10 minutos. Note que, esta formula é aplicada à sistemas que não admitem filas e admite tempos de viagem determinísticos.

4.4 Seleção dos cromossomos pais, *crossover* e mutação:

A seleção de cromossomos pais é baseada no método da roleta de probabilidades. O procedimento funciona como uma roleta, onde cada fatia representa a probabilidade de seleção de cada solução com base no valor de aptidão (*fitness*). Como as melhores soluções apresentam fatias mais largas, ao rodar a roleta (executar a seleção), as melhores soluções têm maiores chances de serem selecionadas que as soluções com menor avaliação. O procedimento de *crossover* utilizado é o *crossover* de um ponto, que consta da seleção aleatória de um gene (posição) no cromossomo e a criação de dois cromossomos filhos através da troca de genes entre dois cromossomos pais a partir deste ponto. Após a escolha de dois cromossomos pais, há uma probabilidade p_c de que os mesmos irão passar pelo procedimento de *crossover*. O procedimento de mutação é aplicado a cada gene após o operador *crossover*, de acordo com uma probabilidade predefinida p_m . Para substituir os genes mutados são escolhidos valores aleatórios através do mesmo procedimento utilizado para a população inicial (veja população inicial).

Esquema básico do algoritmo genético/hipercubo:

Gerar a população inicial

Avaliar (hipercubo), obter melhor solução

Enquanto número de gerações menor que G:

Seleção de cromossomos

Aplicar crossover, mutação e obter nova população

Avaliar (hipercubo), determinar fitness

Atualizar melhor solução e obter melhor solução da geração atual.

5. Resultados (preliminares) obtidos pelo algoritmo genético combinado com o modelo hipercubo:

Esta seção apresenta alguns resultados preliminares obtidos pelo Algoritmo Genético combinado com o modelo hipercubo exato. Os algoritmos foram implementados em linguagem Pascal e executados em um PC processador Pentium de 2.0 GHz.

As respectivas probabilidades de *crossover* e mutação dependem do problema analisado, para a presente análise estes parâmetros foram escolhidos após testes com diferentes combinações de valores dentro de um range. A melhor combinação resultante foi utilizada para a escolha dos demais parâmetros (tamanho da população e número de gerações). Os melhores resultados foram produzidos com a combinação: probabilidade de *crossover* (p_c) em torno de 0.7, probabilidade de mutação (p_m) entre 0.05 e 0.07, tamanho da população $P_{op} = 100$ e número de gerações $G = 1000$ e $G = 2000$. Para o presente problema, um adicional parâmetro a ser determinado corresponde ao intervalo Δ , para o caso de geração discreta de cromossomos. Neste caso os intervalos $\Delta = 0.05$, $\Delta = 0.03$ e $\Delta = 0.01$ foram testados.

A configuração inicial do sistema pode ser representada em um cromossomo de acordo com a representação discutida na seção 4, que é baseada no tamanho dos átomos do sistema. Utilizando os dados dos tamanhos dos átomos da configuração inicial, temos o seguinte cromossomo:

0.50	0.50	0.50	0.50	0.22
------	------	------	------	------

A tabela 1 mostra as três medidas de desempenho obtidas com a aplicação do modelo hipercubo no sistema original. Estas medidas são tratadas como funções de avaliação como descrito na seção 4: σ_ρ - desvio padrão das cargas de trabalho, \bar{T} - tempo médio de viagem no sistema e $P_{tv>10}^-$ - fração de chamadas atendidas em tempo superior à 10 minutos.

Tabela 1 – Três medidas de desempenho da configuração original

Medidas	Conf. original
σ_ρ	0,05507
\bar{T} (min)	7,9121
$P_{tv>10}^-$	0,1281

Como discutido na seção 4, nos procedimentos de geração da população inicial e mutação, consideramos a alternativa de utilizar valores entre 0.2 e 0.8, discretizados por um intervalo Δ . Desta forma, em função do tamanho do presente problema, é possível utilizar um algoritmo enumerativo para realizar todas as combinações possíveis destes valores em cada gene, e escolher a melhor entre elas. Como descrito anteriormente, o algoritmo genético foi testado com três intervalos: $\Delta = 0.05$, $\Delta = 0.03$ e $\Delta = 0.01$. Porém, devido a restrições computacionais, o algoritmo enumerativo foi aplicado apenas para $\Delta = 0.05$ e $\Delta = 0.03$. Inicialmente, conduzimos um conjunto de experimentos utilizando individualmente as três funções objetivo (*fitness*) acima: σ_ρ , \bar{T} e $P_{tv>10}^-$. A seguir são apresentadas as seguintes análises com cada função *fitness*:

- Para todos os intervalos $\Delta = 0.01$, 0.03 e 0.05 e para o caso contínuo foram realizados 2 conjuntos de 20 rodadas do algoritmo, utilizando diferentes sementes para gerar os números aleatórios no algoritmo, com G (número de gerações) igual à 1000 e 2000. Para cada um destes conjuntos, obtemos a melhor solução, a média das soluções e o desvio padrão das soluções encontradas nas 20 rodadas, outras medidas de desempenho da melhor solução e número médio de gerações necessárias para encontrar a melhor solução. As tabelas 2 a 4 apresentam a melhor configuração obtida pelo algoritmo genético em termos das três medidas funções objetivo testadas individualmente. Para os três casos considerados o algoritmo genético encontrou a melhor solução encontrada pelo algoritmo enumerativo.

Função objetivo : Desvio padrão das cargas de trabalho dos servidores (σ_ρ):

Tabela 2 – Três medidas de desempenho da melhor solução encontrada:

Medidas	Melhor solução genético	Conf. original	Desvio p/ original
σ_ρ ←	0,02451	0,05507	-55,49%
\bar{T} (min)	8,9446	7,9121	13,05%
$P_{tv>10}^-$	0,3064%	0,1281	97,61%

Ao observar os resultados da tabela 2, notamos que a medida do desvio padrão das cargas de trabalho (função objetivo que pretendemos minimizar) é reduzida significativamente em 55,5%. Por outro lado, note que o tempo médio de viagem no sistema aumenta em 13,05% com relação ao resultado da configuração inicial e a medida da fração de chamadas atendidas em tempo superior a 10 minutos aumenta de forma significativa em 97,61%.

Função objetivo : Tempo médio de viagem no sistema (\bar{T}):

Tabela 3 – Três medidas de desempenho da melhor solução encontrada:

Medidas	Melhor solução enumerativo	Conf. original	Desvio p/ original
σ_ρ	0,05335	0,05507	-3,12%
\bar{T} (min) ←	7,7781	7,9121	-1,69%
$P_{tv>10}^-$	0,1266	0,1281	-1,17%

A tabela 3 mostra que todas as três medidas de desempenho são reduzidas, porém a redução da medida tempo médio de viagem no sistema (função objetivo) é de apenas 1,69%.

Função objetivo : Fração de chamadas atendidas em tempo superior à 10 minutos ($P_{tv>10}^-$):

Tabela 4 – Três medidas de desempenho da melhor solução encontrada:

Medidas	Melhor solução enumerativo	Conf. original	Desvio p/ original
σ_ρ	0,0445	0,05507	-19,19%
\bar{T} (min)	8,1638	7,9121	3,18%
$P_{tv>10}^-$ ←	0,1201	0,1281	-6,24%

Note na tabela 4 que, a redução obtida para a fração de chamadas atendidas em tempo superior à 10 minutos $P_{tv>10}^-$ (função objetivo) é de 6,24%. Observe também que, nesta configuração há um melhor balanceamento das cargas de trabalho, pois o seu desvio padrão reduz-se em 19,2%. No entanto, o tempo médio de viagem no sistema aumenta em 3,2%.

Desta forma, os resultados obtidos com o algoritmo genético/hipercubo utilizando as três diferentes funções objetivos mostram que estas medidas analisadas são conflitantes, pois a melhor solução obtida no balanceamento das cargas de trabalho, resultou em piores resultados para o tempo médio de viagem no sistema (\bar{T}) e fração de chamadas atendidas em tempo

superior a 10 min ($P_{tv>10}^-$), o que não é o objetivo em termos do nível de serviço oferecido ao usuário.

6. Conclusões e perspectivas futuras:

Neste estudo propomos uma abordagem que integra o modelo hipercubo em um algoritmo genético para analisar os SAE em rodovias brasileiras. O estudo de Mendonça & Morabito (2000;2001) que trata a aplicação do modelo hipercubo ao sistema Anjos do Asfalto foi revisado. Este sistema também foi utilizado como base para a implementação inicial da presente abordagem que combina um algoritmo genético com o modelo hipercubo. O algoritmo genético proposto baseia-se em apontar uma configuração do sistema considerando o tamanho das áreas primárias de cada servidor.

Os resultados preliminares desta pesquisa mostraram que a abordagem foi eficiente em encontrar a solução ótima calculada pelo algoritmo enumerativo. No entanto, como discutido anteriormente, implementações independentes da abordagem genético/hipercubo utilizando diferentes funções *fitness* mostraram que há um conflito entre as medidas de desempenho, especialmente para o balanceamento das cargas de trabalho *versus* minimização do tempo médio de resposta no sistema. Sendo assim, algumas abordagens podem ser integradas ao algoritmo para tratar este tipo de problema tais como: - técnicas de otimização que tratam de problemas multiobjetivo, como por exemplo, geração de curvas de Pareto, usando algoritmos genéticos;- redefinição do problema, introduzindo restrições e utilizando métodos da literatura de algoritmos genéticos para tratá-las, como por exemplo, funções de penalidade.

Referências :

- [1] ALBINO J.C.C. (1994) Quantificação e locação de unidades móveis de atendimento de emergência e interrupções em redes de distribuição de energia elétrica: aplicação do Modelo Hipercubo. *Dissertação* (Mestrado em Engenharia de Produção) - Departamento de Engenharia de Produção, Florianópolis : UFSC.
- [2] BATA R., DOLAN J.M., KRISHNAMURTHY N.N. (1989) The maximal expected covering location problem: Revisited. *Transportation Science* 23, p. 277-287.
- [3] BROTCORNE L., LAPORTE G., SEMET F. (2003). Ambulance location and relocation models. *European Journal of Operational research* 147, p. 451-63.
- [4] CHELST K.; BARLACH Z. (1981) Multiple unit dispatches in emergency services: models to estimate system performance. *Management Science* 27(12), p.1390-1409.
- [5] CHIYOSHI F., GALVÃO R. D., MORABITO R. (2000) O uso do modelo hipercubo na solução de problemas de localização probabilísticos. *Gestão & Produção* 7(2), p.146-174.
- [6] CHIYOSHI F., GALVÃO R. D., MORABITO R. (2003) A note on solutions to the maximal expected covering location problem. *Computers and Operations Research* 30(1), p. 87-96.
- [7] GALVÃO R.D., CHIYOSHI F., ESPEJO L.G.A., RIVAS M.P.A. (2003a) Solução do problema de localização de máxima disponibilidade utilizando o modelo hipercubo. *Pesquisa Operacional* (*in press*).
- [8] GALVÃO R.D., CHIYOSHI F., MORABITO R. (2003b) Towards unified formulations and extensions of two classical probabilistic location models. *Computers & Operations Research*, (*in press*).
- [9] JARVIS J.P. (1985) Approximating the equilibrium behavior of multi-server loss systems. *Management Science* 31, p. 235 - 239.

- [10] LARSON R.C. A. (1974) Hypercube queuing model for facility location and redistricting in urban emergency services. *Computers and operations research* 1, p. 67-95.
- [11] LARSON R.C. (1975) Approximating the performance of urban emergency service systems. *Operations Research* 23, p. 845-868.
- [12] LARSON R.C., ODONI A.R. (1981) *Urban operations research*. Prentice Hall. New Jersey.
- [13] MENDONÇA, F.C.; MORABITO, R. (2000) Aplicação do modelo hipercubo para análise de um sistema médico-emergencial em rodovia. *Gestão & Produção*, v.7, n.1, p.73-91.
- [14] MENDONÇA F.C., MORABITO R. (2001) Analysing emergency service ambulance deployment on a Brazilian highway using the hypercube model. *Journal of the Operational Research Society* 52, p.261- 268.
- [15] SAYDAM C., AYTUG H. (2003) Accurate estimation of expected coverage: revisited. *Socio -Economic Planning Sciences* 37, p. 69-80.
- [16] SWERSEY A.J. (1994) *Handbooks in OR/MS*. Amsterdam: Elsevier Science B.V., v. 6, p. 151-200.
- [17] TAKEDA R. A., WIDMER, J. A., MORABITO, R. (2004) Aplicação do modelo hipercubo de filas para avaliar a descentralização de ambulâncias em sistema urbano de atendimento médico de emergência. *Pesquisa Operacional* (*in press*).