



MODELO HIPERCUBO INTEGRADO A UM ALGORITMO GENÉTICO PARA ANÁLISE DE SISTEMAS MÉDICOS EMERGENCIAIS EM RODOVIAS

Ana Paula Iannoni
Reinaldo Morabito

Departamento de Engenharia de Produção,
Universidade Federal de São Carlos, CEP 13565-905, São Carlos, SP,
e-mail: papi@iris.ufscar.br, morabito@power.ufscar.br

Recebido em 24/5/2005
Aceito em 18/1/2006

Resumo

O modelo hipercubo, conhecido na literatura de problemas de localização de sistemas servidor para cliente, é um modelo baseado em teoria de filas espacialmente distribuídas e aproximações Markovianas. O modelo pode ser modificado para analisar os sistemas de atendimentos emergenciais (SAEs) em rodovias, considerando as particularidades da política de despacho destes sistemas. Neste estudo, combinou-se o modelo hipercubo com um algoritmo genético para otimizar a configuração e operação de SAEs em rodovias. A abordagem é efetiva para apoiar decisões relacionadas ao planejamento e operação destes sistemas, por exemplo, em determinar o tamanho ideal para as áreas de cobertura de cada ambulância, de forma a minimizar o tempo médio de resposta aos usuários e o desbalanceamento das cargas de trabalho das ambulâncias. Os resultados computacionais desta abordagem foram analisados utilizando dados reais do sistema Anjos do Asfalto (rodovia Presidente Dutra).

Palavras-chave: modelo hipercubo, sistemas médicos emergenciais, despacho de ambulâncias, rodovias.

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 significar a diferença entre a vida e morte das vítimas envolvidas. A perda de chamadas e o atraso no tempo de resposta estão diretamente relacionados ao conflito entre as variáveis aleatórias da demanda por serviço e as restrições de capacidade do sistema. Dado que, devido a restrições de orçamento, os SAEs não podem ser planejados de forma a trabalhar com um número muito grande de servidores, há claramente um importante *trade-off* a ser considerado entre a qualidade de atendimento e os custos de investimento e operação nestes sistemas. Ao se analisarem 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 a um determinado chamado. Apesar disso, a maioria dos estudos na literatura de problemas de localização de instalações considera apenas a aleatoriedade relacionada à disponibilidade dos servidores. Swersey (1994), Owen e Daskin (1998), 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 (Larson, 1974; 1975; Larson e Odoni, 1981), baseado nos resultados de teoria de filas espacialmente distribuídas e aproximações Markovianas, é um dos métodos mais efetivos para analisar os sistemas emergenciais. A idéia básica é expandir o espaço de estados de modelos de filas simples com múltiplos servidores (p.e., modelos $M/M/N$ ou $M/G/N$, em que N é o

número de servidores), de forma a tratar cada servidor individualmente e incorporar as complexidades das políticas de despacho. O modelo envolve a solução de um sistema linear de $O(2^N)$ equações, cujas variáveis envolvidas correspondem às probabilidades de estado do sistema em equilíbrio. Por meio destas probabilidades, podem ser estimadas importantes medidas de desempenho para análise do sistema, tais como cargas de trabalho dos servidores, tempos médios de resposta aos usuários, frações de despacho de cada servidor a cada região e frações de atendimento fora da área primária, entre outras medidas.

Este modelo vem sendo largamente estudado e entendido, por vários autores, para análise de sistemas de emergência, principalmente em patrulhamento policial e despacho de ambulâncias. Por exemplo, Larson (1975), Jarvis (1985), Halpern (1977), Chelst e Barlach (1981), Burwell et al. (1993) e Swersey (1994) estenderam o modelo original de forma a eliminar algumas hipóteses limitantes ou melhorar a eficiência computacional do modelo. Outros estudos têm sido dedicados a combinar o modelo hipercubo com métodos de otimização. Algumas publicações com este enfoque são os estudos de Batta et al. (1989), Saydam e Aytug (2003), Chiyoshi et al. (2003), Galvão et al. (2003) e Galvão et al. (2005). Exemplos de aplicações do modelo hipercubo em SAEs nos Estados Unidos podem ser encontrados em Larson e Odoni (1981), Chelst e Barlach (1981), Brandeau e Larson (1986), Burwell et al. (1993) e Sacks e Grief (1994). Recentemente, o modelo hipercubo vem sendo estudado para aplicação a sistemas de emergência que atuam em casos de ataques terroristas e catástrofes naturais de grande escala (Larson, 2004). No Brasil, alguns exemplos de aplicação do modelo hipercubo são: a análise dos SAEs urbanos (Takeda et al., 2000; Costa, 2004, Takeda et al., 2004; 2005) e os SAEs em rodovias do Estado de São Paulo e do Rio de Janeiro (Mendonça e Morabito, 2000; 2001; Iannoni et al., 2004; Iannoni, 2005).

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. Enquanto os principais estudos que combinam o modelo hipercubo com métodos de otimização buscam determinar a localização ótima para as ambulâncias do sistema (*location problem*), a presente abordagem trata do dimensionamento das áreas de cobertura de cada servidor (*districting problem*). O método permite considerar diferentes objetivos conflitantes do sistema, tais como o tempo médio de resposta ao usuário, o desbalanceamento

das cargas de trabalho e a fração de chamadas não atendidas dentro de certo período de tempo (p.e., 10 minutos). Além disso, esta abordagem também permite uma análise de *trade-off* entre estes objetivos. Os resultados computacionais são analisados aplicando-se a abordagem para o estudo de caso de um SAE que operava em trechos da rodovia Presidente Dutra (o sistema **Anjos do Asfalto**).

Este artigo está organizado da seguinte forma: a seção 2 apresenta uma breve descrição do SAE em rodovias, a seção 3 descreve como o modelo hipercubo pode ser adaptado para analisar estes SAEs, a seção 4 apresenta o algoritmo genético combinado com o modelo hipercubo (AG/hipercubo), a seção 5 descreve como esta abordagem pode ser utilizada para gerar as curvas de *trade-off* (curvas de Pareto) entre as medidas conflitantes, a seção 6 apresenta e analisa os resultados computacionais e, finalmente, a seção 7 discute conclusões deste estudo e perspectivas para pesquisas futuras.

2. SAE em rodovias brasileiras

O SAE em rodovia, utilizado neste estudo para avaliar a abordagem AG/hipercubo, foi inicialmente estudado por Mendonça e Morabito (2000, 2001). Este SAE, chamado Anjos do Asfalto, é similar a outros SAEs em diferentes rodovias brasileiras. O sistema prestava atendimento médico emergencial em parte da rodovia Presidente Dutra entre as cidades de São Paulo e Rio de Janeiro, contando com 6 bases fixas ao longo do trecho da rodovia, sendo que, cada base possuía uma ambulância e uma equipe composta por médicos, enfermeiros e motorista, que viajavam juntos para o local do chamado. A central de operações, localizada na cidade do Rio de Janeiro, era responsável por receber as chamadas, despachar as ambulâncias e monitorar os movimentos.

Ao receber um chamado, a central imediatamente enviava a ambulância disponível mais próxima do local do chamado e, se esta estivesse ocupada, a segunda mais próxima era enviada (chamada de *backup*). Se as duas ambulâncias mais próximas estivessem ocupadas, a chamada era considerada perdida para o sistema (mesmo se houvesse outras ambulâncias disponíveis) e era transferida para outro sistema (por exemplo, o hospital ou SAE mais próximo). Nesta política de despacho particular, cada região podia ser atendida por somente dois servidores (o servidor preferencial e o servidor *backup*), dado que a terceira ambulância nunca era despachada, devido às limitações de distância. Além disso, neste sistema, apenas um servidor podia ser enviado a cada chamada. Esta política particular de despacho é chamada *backup* parcial. Portanto, este sistema é um sistema sem filas com *backup* parcial, que não respeita uma das principais hipóteses do modelo hipercubo original de Larson (1974), o qual admite que qualquer servidor do sistema pode via-

jar a qualquer região. Uma descrição detalhada sobre a operação dos **Anjos do Asfalto** pode ser encontrada em Mendonça e Morabito (2000, 2001).

A Figura 1 ilustra a distribuição das seis bases de ambulâncias do SAE **Anjos do Asfalto**. A distância entre duas bases adjacentes é dividida em duas regiões (chamadas átomos), cada uma com uma lista de preferência de despacho. De acordo com esta lista, e exceto para as ambulâncias 1 e 6, todas as ambulâncias são despachadas como preferenciais para dois átomos (à esquerda e à direita de sua base) e como *backup* para outros dois átomos (direita e esquerda das ambulâncias adjacentes à esquerda e à direita, respectivamente). Por exemplo, para a ambulância 5, seu lado esquerdo (átomo 8) e seu lado direito (átomo 9) correspondem às suas regiões preferenciais, ao passo que, o lado direito da ambulância 4 (átomo 7) e o lado esquerdo da ambulância 6 (átomo 10) correspondem às suas regiões *backup*. Note que as ambulâncias 1 e 6 possuem apenas uma região preferencial (átomos 1 e 10, respectivamente), e uma única região *backup* (átomos 2 e 9, respectivamente).

3. O modelo hipercubo para SAEs em rodovias

O nome hipercubo é derivado do espaço de estado do sistema, sendo que cada estado corresponde a um vértice do hipercubo. Um estado (vértice) em particular do sistema é representado pela lista de servidores que estão livres e ocupados. Dado que há dois estados possíveis para cada servidor: livre (0) ou ocupado (1) em certo instante de tempo, temos então 2^N estados (vértices) do sistema. 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.

3.1 Principais hipóteses do modelo e equações de equilíbrio do sistema

As principais hipóteses do modelo hipercubo para aplicação aos SAEs em rodovias são:

- A rodovia é particionada em N_A átomos geográficos, os quais correspondem a fontes independentes de chamadas. Na Figura 1, $N_A = 2N - 2$, em que N é o número de bases de ambulâncias. Uma área preferencial de uma ambulância corresponde aos átomos para os quais a ambulância é despachada se disponível, mesmo que todas as outras ambulâncias estejam disponíveis;
- As chamadas de emergência em cada átomo j são geradas de acordo com um processo de Poisson com taxa de chegada λ_j , independente dos demais átomos;
- Há N ambulâncias espacialmente distribuídas ao longo da rodovia, que permanecem fixas 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 de *backup*;
- O despacho dos servidores é realizado de acordo com uma lista de preferência para cada átomo. De acordo com esta lista, a primeira ambulância da lista (a mais próxima) é despachada e, se esta estiver ocupada, a segunda da lista é enviada (*backup*). Se a ambulância *backup* estiver ocupada, a chamada é perdida para o sistema mesmo se houver outras ambulâncias disponíveis (*backup* parcial);
- O tempo médio de atendimento para cada ambulância inclui o tempo de *set-up* (preparação), os tempos de viagem e o tempo em cena. Em geral, cada ambulância i do sistema possui um tempo médio de serviço distinto $\left(\frac{1}{\mu_i}\right)$. O modelo também admite que o tempo de atendimento é representado por uma distribuição exponencial negativa. Porém, razoáveis aproximações desta hipótese não representam significantes alterações na acuracidade do modelo. Dado que os SAEs em rodovias não admitem filas, esta suposição é ainda mais desnecessária, pois os modelos $M/M/N$ e $M/G/N$ têm a mesma distribuição de equilíbrio (Larson e Odoni, 1981; Chiyoshi et al., 2000); e

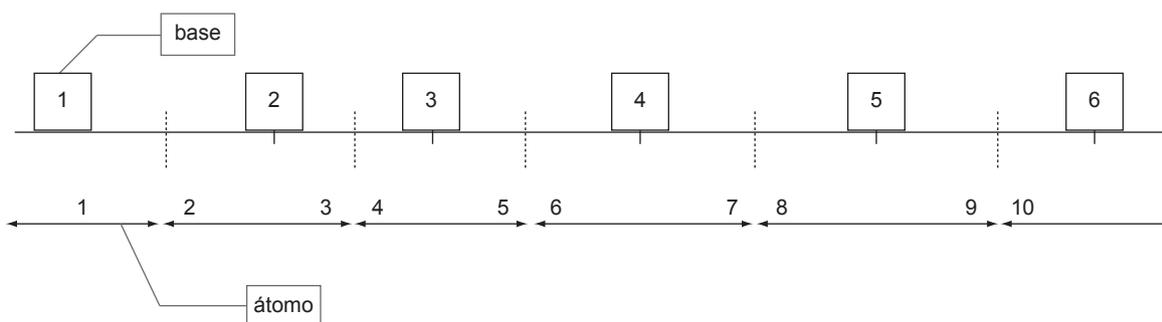


Figura 1. Bases das ambulâncias e átomos ao longo da rodovia.

- O tempo de viagem entre cada par de átomos é conhecido ou pode ser estimado utilizando os conceitos de distribuição de probabilidade geométrica. As variações no tempo de atendimento devido às variações no tempo de viagem são consideradas de segunda ordem, quando comparadas às variações no tempo em cena ou no tempo de *set-up*.

A análise estatística dos dados do sistema **Anjos do Asfalto** foi realizada por Mendonça e Morabito (2000, 2001). Para aplicação do modelo hipercubo, o trecho linear da rodovia foi particionado em 10 átomos, de acordo com a área preferencial das ambulâncias do sistema descrita na Figura 1. Com exceção das ambulâncias 1 e 6, como as ambulâncias são preferenciais para seus lados esquerdo e direito, temos que, de acordo com a Figura 1, as ambulâncias preferencial e *backup* de cada átomo são: átomo 1 - ambulâncias 1 e 2; átomo 2 - ambulâncias 2 e 1; átomo 3 - ambulâncias 2 e 3; átomo 4 - ambulâncias 3 e 2; átomo 5 - ambulâncias 3 e 4; átomo 6 - ambulâncias 4 e 3; átomo 7 - ambulâncias 4 e 5; átomo 8 - ambulâncias 5 e 4; átomo 9 - ambulâncias 5 e 6; e átomo 10 - ambulâncias 6 e 5. A taxa total de chegadas no sistema é $\lambda = 0,01813$ chamadas/min, e a taxa total de atendimento, considerando na análise que os servidores são não-homogêneos, é $\mu = 0,0959$ chamadas/min. Em Mendonça e Morabito (2000, 2001), pode ser encontrada uma descrição mais detalhada sobre os dados de entrada e os resultados da análise estatística dos processos de chegada e serviço deste sistema.

As equações de probabilidade de equilíbrio são definidas supondo que o sistema atinge o estado de equilíbrio. Como o sistema **Anjos do Asfalto** possui $N = 6$ servidores, há $2^6 = 64$ estados possíveis. Para cada estado, o fluxo para dentro do estado (probabilidade de o sistema estar em outros estados, vezes a taxa de transição para o presente estado) deve ser igual ao fluxo para fora do estado (probabilidade de o sistema estar no presente estado, vezes a taxa de transição para outros estados do sistema). As equações são determinadas de acordo com a lista de despacho descrita acima. Por exemplo, para o estado 110001 (ambulâncias 1, 2 e 6 ocupadas e ambulâncias 3, 4 e 5 disponíveis), a equação de equilíbrio é:

$$P_{110001} ((\lambda_3 + \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 + \lambda_8 + \lambda_9 + \lambda_{10}) + \mu_1 + \mu_2 + \mu_6) = P_{100001} (\lambda_1 + \lambda_2 + \lambda_3) + P_{010001} (\lambda_1 + \lambda_2) + P_{110000} (\lambda_{10}) + P_{111001} (\mu_3) + P_{110101} (\mu_4) + P_{110011} (\mu_5) \quad (1)$$

Por exemplo, no lado direito da Equação 1 (fluxo para fora), o sistema sai do estado 110001 quando chega uma chamada em um dos átomos de 3 a 10 (cujas ambulâncias, preferencial ou *backup* estão disponíveis) ou quando ocorre o término de atendimento das ambulâncias ocupadas (ambulâncias 1, 2 e 6). O sistema também não atende a chamadas que chegam nos átomos 1 e 2, pois os

servidores de sua lista de despacho (ambulâncias 1 e 2, que são preferencial e *backup* para estes átomos) se encontram ocupados no estado 110001. Desta forma, diferentemente do modelo hipercubo original, chamadas são perdidas mesmo quando há servidores disponíveis no sistema. Substituindo uma das 64 equações por uma equação de normalização (a soma de todas as probabilidades de estado deve ser igual a 1), obtemos um sistema determinado com 64 equações linearmente independentes.

3.2 Medidas de desempenho

Calculando as probabilidades de equilíbrio de estado do sistema é possível obter interessantes medidas de desempenho para o sistema, tais como: tempo médio de resposta, cargas de trabalho dos servidores, fração de despachos de cada servidor a cada átomo e medidas agregadas de tempos de viagem. Por exemplo, o tempo médio de resposta do sistema corresponde simplesmente ao tempo de *set-up*, mais o tempo médio de viagem, dado que o sistema não admite fila de espera. Assim, o tempo médio de viagem é definido por:

$$\bar{T} = \sum_{i=1}^N \sum_{j=1}^{N_A} f_{ij} t_{ij} \quad (2)$$

em que f_{ij} é a fração de despachos do servidor i ao átomo j (calculada em função das probabilidades de equilíbrio de estado; Mendonça e Morabito, 2001), e t_{ij} é o tempo de viagem do servidor i ao átomo j (calculado utilizando os dados de tempo de viagem entre os átomos). Lembrando que N é o número de servidores e N_A é o número de átomos.

A fração de chamadas que requerem mais de 10 minutos para serem atendidas por uma ambulância é:

$$P_{t > 10} = \sum_{i=1}^N \sum_{j=1}^{N_A} f_{ij} p(t_{ij} > 10) \quad (3)$$

em que o termo $f_{ij} p(t_{ij} > 10)$ é a fração de todos os despachos do servidor i ao átomo j , para os quais o tempo de viagem excede 10 minutos, e p é a probabilidade de o tempo de viagem do servidor i para o átomo j ser maior que 10 minutos. Neste estudo, como os dados de tempos de viagem não estão disponíveis, calculou-se $p(t_{ij} > 10)$ determinando para qual fração de cada átomo j uma chamada de emergência é atendida em tempo de viagem superior a 10 minutos pela ambulância i .

O desbalanceamento das cargas de trabalho dos servidores pode ser estimado pelo desvio-padrão das cargas de trabalho, definido como:

$$\sigma_\rho = \sqrt{\frac{\sum_{i=1}^N (\rho_i - \bar{\rho})^2}{N}} \quad (4)$$

em que ρ_i é a carga de trabalho do servidor i (i.e., a soma das probabilidades de equilíbrio dos estados nos quais o

servidor i está ocupado) e $\bar{\rho} = \sum_{i=1}^N \rho_i / N$. Outras medidas de desempenho também podem ser obtidas em função das probabilidades de equilíbrio (Larson e Odoni, 1981; Chiyoshi et al., 2000).

4. Modelo hipercubo integrado a um algoritmo genético

Como mencionado anteriormente, existem estudos na literatura combinando meta-heurísticas com o modelo hipercubo para apoiar decisões de localização dos servidores nos SAEs, tais como em Saydam e Aytug (2003); Chiyoshi et al. (2003) e Galvão et al. (2005). Nesta seção, propõe-se um algoritmo genético integrado ao modelo hipercubo modificado para análise dos SAEs em rodovias. Esta abordagem (chamada de AG/hipercubo) apoia decisões de dimensionamento das áreas preferenciais de cada servidor (*districting problem*), de forma a determinar uma configuração ótima (ou perto da ótima), estabelecida pelo tamanho dos átomos do sistema. Neste estudo, utiliza-se um algoritmo genético básico, tal como descrito em Holland (1975), Goldberg (1989), Michalewicz (1996) e Beasley (2002). A seguir, são descritos brevemente os principais componentes considerados na implementação do algoritmo AG/hipercubo.

4.1 Geração e representação dos cromossomos

Define-se um procedimento que varia o tamanho dos átomos do sistema, produzindo diferentes configurações viáveis (ou cromossomos). A Figura 2 ilustra como ocorre a variação do tamanho dos átomos 1 e 2 entre as duas bases adjacentes 1 e 2. Na Figura 2, o tamanho dos átomos 1 e 2 variam de z_1^0 e z_2^0 para $z_1 = x_1 d_1$ e $z_2 = d_1 - z_1$, respectivamente, em que x_1 é a porcentagem de d_1 (distância entre as bases 1 e 2).

Cada cromossomo do sistema pode ser representado por um vetor $x = (x_1, x_2, x_{N-1}, \dots)$, em que cada x_i é a proporção da distância entre duas ambulâncias adjacentes i e $i + 1$, que determina o tamanho do átomo preferencial da ambulância i . Após variar o tamanho de cada par de átomos j e $j + 1$ entre as ambulâncias adjacentes i e $i + 1$, calculam-se as novas taxas de chegada λ_j e λ_{j+1} , a partir das taxas iniciais de chegada λ_j^0 e λ_{j+1}^0 , da seguinte forma:

$$\text{Se } z_j < z_j^0 \text{, então, } \begin{cases} \lambda_j = z_j (\lambda_j^0 / z_j^0) \\ \lambda_{j+1} = \lambda_{j+1}^0 + (z_{j+1} - z_{j+1}^0) (\lambda_j^0 / z_j^0) \end{cases} \quad (5)$$

$$\text{Se } z_j > z_j^0 \text{, então, } \begin{cases} \lambda_j = \lambda_j^0 + (z_j - z_j^0) (\lambda_{j+1}^0 / z_{j+1}^0) \\ \lambda_{j+1} = z_{j+1} (\lambda_{j+1}^0 / z_{j+1}^0) \end{cases} \quad (6)$$

em que, de acordo com a Figura 2, z_j^0 é o tamanho inicial do átomo j , z_j é o novo tamanho do átomo j , e d_i corresponde à distância entre duas bases de ambulâncias

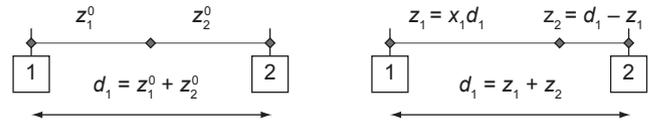


Figura 2. Exemplo da variação do tamanho dos átomos 1 e 2 entre as bases 1 e 2.

adjacentes i e $i + 1$. Utiliza-se um procedimento para gerar aleatoriamente a população inicial. Conforme sugerido pelos gerentes e operadores do sistema **Anjos do Asfalto**, admite-se que $0,2 \leq x_i \leq 0,8$, limitando a área preferencial de cada ambulância i de 20% a 80% da distância d_i . Para gerar possíveis configurações do sistema, este procedimento simplesmente adiciona 0,2 (o limite inferior do intervalo) a um incremento $k\Delta$, em que Δ é fixo e k é um inteiro sorteado aleatoriamente dentro do intervalo $[0, M = (0,8 - 0,2)/\Delta]$. Portanto, há $M + 1$ valores possíveis para cada x_i . Aplica-se este procedimento discreto, ao invés de um procedimento que simplesmente sorteasse um número aleatório contínuo entre 0,2 e 0,8, devido ao seu bom desempenho computacional (i.e., em encontrar mais rapidamente a solução ótima) sob uma precisão de Δ .

4.2 Avaliação e função de aptidão (fitness)

O procedimento de avaliação do AG/hipercubo proposto utiliza o modelo hipercubo da seção 3 para descrição das medidas de desempenho de cada configuração (representada por um cromossomo). De acordo com o procedimento para variar o tamanho dos átomos do sistema proposto na seção 4.1, recalcula-se a matriz de tempos de viagem com base na distância de cada base de ambulância ao centróide de cada átomo (como proposto em Larson e Odoni, 1981). Portanto, cada nova configuração gera uma nova matriz de tempos de viagem devido à nova posição do centróide dos átomos.

É importante ressaltar que algumas medidas de desempenho podem ser conflitantes em termos dos diferentes interesses das partes envolvidas na operação dos SAEs. Por exemplo, o tempo médio de resposta numa região é uma medida de desempenho externa do sistema, que interessa principalmente ao usuário do sistema. Por outro lado, o balanço das cargas de trabalho dos servidores é uma medida de desempenho interna do sistema, que interessa particularmente ao gerente do sistema. Como mostrado adiante na seção 6, em alguns casos, ao se melhorar o valor do primeiro, piora-se o valor do segundo, e vice-versa. Neste estudo, inicialmente são conduzidos diferentes experimentos otimizando separadamente três diferentes medidas de desempenho (funções de aptidão $f(x)$ do cromossomo x) para avaliar cada nova configuração, conforme descrito a seguir.

Em um experimento, o objetivo é minimizar o tempo médio de viagem no sistema, ou $\min f(x) = \bar{T}(x)$ seja, (expressão (1) na seção 3.1). Em outro experimento, o objetivo é minimizar a fração de chamadas atendidas em mais que 10 minutos, isto é, $\min f(x) = P_{t>10}(x)$ (expressão (2)) ou o desvio padrão das cargas de trabalho dos servidores para avaliar o desbalanceamento das cargas de trabalho das ambulâncias, isto é, $\min f(x) = \sigma_p(x)$ (expressão (3)).

4.3 Seleção de cromossomos, crossover e mutação

A seleção de cromossomos pais é baseada no método da roleta de probabilidades (Goldberg, 1989; Michalewicz, 1996). O procedimento funciona como uma roleta, em que cada fatia representa a probabilidade de seleção de cada solução com base no valor de aptidão. 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. Após a escolha de um par de cromossomos pais, aplica-se com probabilidade p_c o conhecido **crossover de um ponto** (e com probabilidade $1-p_c$ o par selecionado é preservado na próxima população). O procedimento de mutação é aplicado a cada gene do cromossomo com uma probabilidade p_m (Goldberg, 1989; Michalewicz, 1996). Para substituir aleatoriamente o gene, utilizou-se o mesmo procedimento discreto de inicialização descrito no último parágrafo da seção 4.1. Por exemplo, para cada x_i (com probabilidade p_m) este valor é mutado para $x_i = k\Delta$, em que k é uniformemente sorteado no intervalo $[0, \dots, M]$.

4.4 Escolha dos parâmetros para o AG/hipercubo

Para uma adequada escolha dos principais parâmetros do algoritmo AG/hipercubo, tais como probabilidades de *crossover* (p_c) e mutação (p_m), número de gerações (G) e tamanho da população (P_{op}), foram realizados inicialmente testes extensivos com diferentes combinações de valores dentro de intervalos. Os melhores resultados foram obtidos com: $p_c = 0,7$, $p_m = 0,05$, $G = 1000$ gerações e $P_{op} = 100$ indivíduos. Um parâmetro adicional é a precisão Δ para a geração discreta de cromossomos (seção 4.1). Os intervalos $\Delta = 0,05$ e $\Delta = 0,03$ (i.e., $M = 12$ e $M = 20$, respectivamente) foram testados para cada uma das três funções de aptidão.

A Figura 3 apresenta o esquema geral do algoritmo AG/hipercubo. Vários estudos, como Hertz e Kobler (2000), Beasley (2002), Jaszkiwicz (2002) e Arroyo e Armentano (2005), têm ressaltado a superioridade dos algoritmos genéticos híbridos, os quais incluem um procedimento de busca local para melhorar as soluções geradas após o *crossover* e a mutação. Porém, no presente estudo, esta

alternativa não foi adotada porque o processo de avaliação envolvendo a solução do modelo hipercubo é computacionalmente bem custoso (i.e., a solução de um sistema linear de $O(2^N)$ para cada cromossomo gerado).

5. Geração das curvas de *trade-off*

Como discutido anteriormente, algumas medidas de desempenho interessantes para a operação dos SAEs que se deseja otimizar podem ser conflitantes. Neste estudo, utilizou-se o conceito de dominância de soluções (chamado dominância de Pareto) para analisar a relação entre estes objetivos conflitantes. Um problema multiobjetivo (como definido em Cohon, 1978; Steuer, 1986; Poulos et al., 2001; Jaszkiwicz, 2002; Arroyo e Armentano, 2005) consiste em determinar um vetor de variáveis de decisão que otimize um vetor de funções objetivo e satisfaça o conjunto de restrições envolvidas. Uma solução x é chamada Pareto ótima (ou eficiente ou não-dominada),

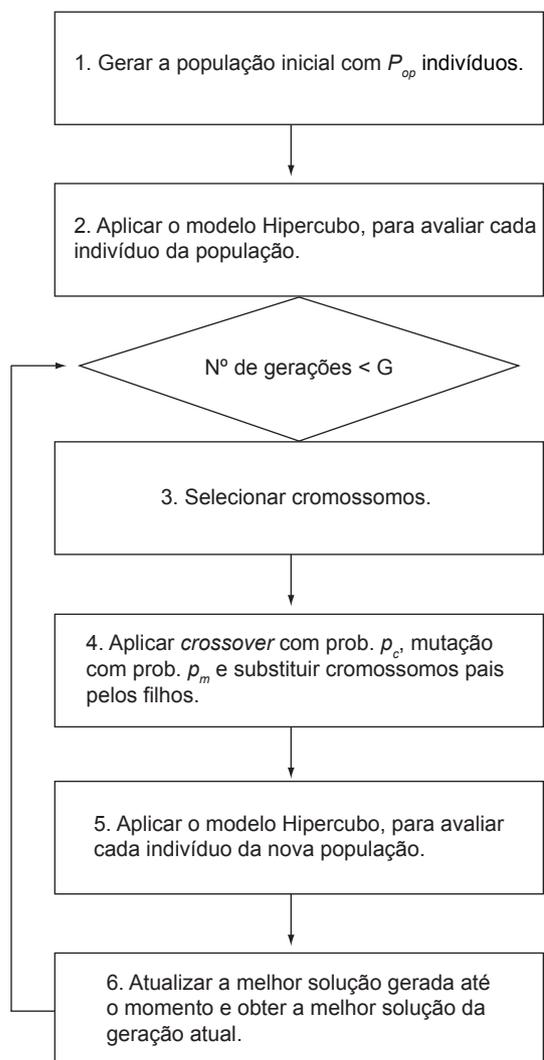


Figura 3. Passos do Algoritmo AG/hipercubo.

se o valor de alguma função objetivo $f_i(x)$ não pode ser melhorada sem piorar ao menos uma das outras funções objetivo. Por exemplo, o problema biobjetivo com objetivos $f_1(x)$ e $f_2(x)$ pode ser formulado da seguinte forma:

$$\text{Min } Z = (f_1(x), f_2(x)) \quad (7)$$

$$\text{s.a } x \in X^* \quad (8)$$

em que x corresponde ao vetor solução, Z corresponde à imagem de x (ou espaço objetivo) e X^* corresponde ao conjunto de soluções viáveis do problema. Se $x_1, x_2 \in X^*$ são dois vetores solução do problema acima, considerando o modo como estão selecionados, há três possibilidades:

$$\text{i) } f_1(x_1) < f_1(x_2) \text{ e } f_2(x_1) < f_2(x_2); (x_1 \text{ domina } x_2) \quad (9)$$

$$\text{ii) } f_1(x_1) > f_1(x_2) \text{ e } f_2(x_1) > f_2(x_2); (x_2 \text{ domina } x_1) \quad (10)$$

iii) nenhuma das alternativas acima (x_1 e x_2 são indiferentes entre si).

Por exemplo, na Figura 4, para os pontos x_3 e x_4 , a relação é: $f_1(x_3) < f_1(x_4)$ e $f_2(x_3) < f_2(x_4)$, assim a solução x_3 domina a solução x_4 (condição i). Considerando os pontos x_2 e x_3 , tem-se $f_1(x_2) > f_1(x_3)$ e $f_2(x_2) < f_2(x_3)$, então, as soluções x_2 e x_3 são indiferentes entre si (condição iii). A solução x^* é chamada eficiente ou Pareto ótima se não há outra solução $x \in X^*$ que domine x^* . O conjunto de soluções eficientes determina a curva de *trade-off* chamada curva Pareto ótima (linha sobre os pontos em negrito da Figura 4). Ao melhorar continuamente uma população de soluções, um algoritmo genético pode encontrar as soluções não-dominadas em uma simples rodada, o que o torna uma ferramenta promissora para solucionar problemas de otimização multiobjetivo.

Um simples método de gerar soluções ótimas de Pareto é o método ϵ -restrito (Cohon, 1978; Arroyo, 2002). Este consiste em otimizar um objetivo, enquanto os demais

objetivos são limitados por um valor ϵ . Por exemplo, considerando os objetivos: $\min f_1(x) = \bar{T}(x)$ e $\min f_2(x) = \sigma_p(x)$, como definidos na seção 4.2, tem-se a função objetivo: $\min Z = (f_1(x), f_2(x)) = (\bar{T}(x), \sigma_p(x))$. Utilizando o método ϵ -restrito, para minimizar o tempo médio de viagem no sistema, e considerando o desvio padrão das cargas de trabalho como uma restrição, o problema pode ser formulado da seguinte forma:

$$\text{Min } Z = f_1(x) = \bar{T}(x) \quad (11)$$

$$\text{s.a } f_2(x) = \sigma_p(x) \leq \epsilon \quad (12)$$

$$x \in X^* \quad (13)$$

em que ϵ corresponde ao valor do limite superior de σ_p . No problema (4)-(6), alternativamente, poder-se-ia optar por otimizar σ_p (ao invés de \bar{T}) e limitar \bar{T} (ao invés de σ_p).

O algoritmo AG/hipercubo da seção 4 pode ser adaptado para solucionar o problema (4)-(6), em que a função de aptidão $f_1(x) = \bar{T}(x)$ é otimizada e as soluções eficientes são selecionadas de acordo com as condições i), ii) e iii) definidas acima. Para isso, o algoritmo foi modificado da seguinte forma: durante uma simples rodada, os valores de ϵ são variados a cada $G = 1000$ gerações, e o algoritmo carrega (guarda na memória) as melhores soluções encontradas até as últimas G gerações para as próximas G gerações, atualizando os valores de ϵ e os melhores valores da função objetivo \bar{T} para cada valor de σ_p . Além disso, modificações adicionais são feitas nos passos 1 e 2 e nos passos 4 e 5 do algoritmo (Figura 3), para garantir que sejam gerados apenas cromossomos viáveis para o problema (4)-(6). Por exemplo, nos passos 4 e 5, um cromossomo pai é substituído por um cromossomo filho x somente se este filho x for viável, isto é, se satisfizer $\sigma_p(x) \leq \epsilon$. Para se verificar isso, é preciso aplicar o modelo hipercubo para avaliar cada filho x . Após variar todos os diferentes valores de ϵ , o algoritmo obtém uma curva aproximada de *trade-off* entre \bar{T} e σ_p (ou seja, as funções não-dominadas encontradas durante toda a rodada considerando os possíveis valores de ϵ).

6. Resultados computacionais

O algoritmo AG/hipercubo foi implementado em Pascal e executado em um microcomputador Pentium IV 2.0 GHz. A configuração original do sistema, com base nos dados dos tamanhos dos átomos reportados por Mendonça e Morabito (2000, 2001), é representada pelo cromossomo $x = (x_1, x_2, x_3, x_4, x_5) = (0,50, 0,50, 0,50, 0,50, 0,22)$ (como ilustrado na Figura 1). Por exemplo, na Figura 1, sendo $x_1 = 0,50$, a distância entre as bases 1 e 2 é dividida ao meio correspondendo aos átomos 1 e 2 da configuração original. Esta configuração resulta

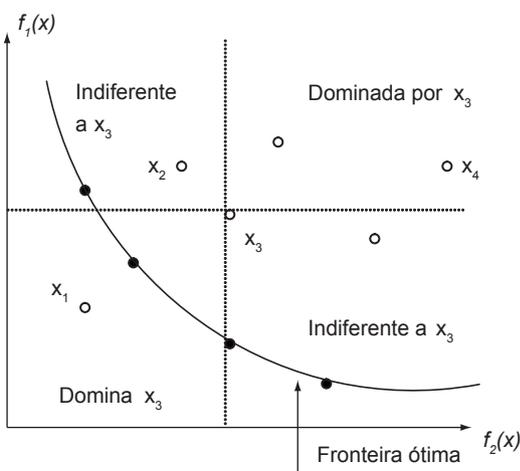


Figura 4. Dominância de Pareto.

em $\bar{T} = 7,912$ minutos, $P_{t > 10} = 0,299$ e $\sigma_p = 0,0551$ minutos (em que as cargas de trabalho são: $\rho_1 = 0,1352$; $\rho_2 = 0,1928$; $\rho_3 = 0,1612$; $\rho_4 = 0,3026$; $\rho_5 = 0,1833$ e $\rho_6 = 0,1490$).

Para verificar a qualidade das soluções produzidas pelo algoritmo AG/hipercubo, foi desenvolvido um algoritmo enumerativo simples que incorpora o modelo hipercubo da seção 3. Este procedimento exaustivo determina a configuração ótima em termos do tamanho dos átomos (sob uma precisão de Δ), testando todas as configurações possíveis para o sistema (utilizando o método discreto descrito na seção 4.1). Dado que há $M + 1$ valores possíveis para cada x_i , o procedimento considera $(M + 1)^{N-1}$ combinações possíveis. Este algoritmo enumerativo é tratável computacionalmente somente para problemas de tamanho moderado, como o SAE **Anjos do Asfalto** que tem apenas $N = 6$ ambulâncias e considerando uma precisão de $\Delta = 0,05$ ou $\Delta = 0,03$ (o que resulta em 13^5 e 21^5 combinações, respectivamente).

6.1 Resultados para o SAE Anjos do Asfalto

Inicialmente são realizados três experimentos otimizando separadamente cada uma das funções de aptidão discutidas na seção 4.2: $\min f(x) = \bar{T}(x)$, $\min f(x) = P_{t > 10}(x)$ ou $\min f(x) = \sigma_p(x)$. Aplicando o algoritmo enumerativo descrito acima, as melhores configurações para os três objetivos (usando $\Delta = 0,03$) são: $(x_1, x_2, x_3, x_4, x_5) = (0,41, 0,44, 0,44, 0,41, 0,29)$, $(x_1, x_2, x_3, x_4, x_5) = (0,41, 0,77, 0,53, 0,32, 0,41)$ e $(x_1, x_2, x_3, x_4, x_5) = (0,80, 0,47, 0,80, 0,20, 0,20)$, respectivamente. Os correspondentes valores ótimos de $\bar{T}(x)$, $P_{t > 10}(x)$ e $\sigma_p(x)$ em cada experimento são apresentados na Tabela 1 (valores em negrito).

Considerando cada função separadamente, o algoritmo AG/hipercubo obteve os mesmos resultados que o algoritmo enumerativo, ou seja, encontrou a solução ótima para cada objetivo. Foram aplicados os parâmetros definidos na seção 4.4 e conduzidas 20 rodadas, usando diferentes sementes para $G = 1000$ gerações (usando $\Delta = 0,03$). Além dos valores ótimos (em negrito) para cada objetivo, a Tabela 1 também apresenta os valores das outras duas medidas de desempenho computadas pelo AG/hipercubo em cada experimento, assim como o desvio relativo (porcentagem de melhoria) para a configuração original.

No primeiro experimento (Tabela 1), \bar{T} , $P_{t > 10}$ e σ_p são todos melhorados quando comparados com a configuração original do sistema, mas \bar{T} (sendo minimizado) é reduzido em apenas 1,69%. No segundo experimento, $P_{t > 10}$ (sendo minimizado) é reduzido em 14,86%, enquanto \bar{T} aumenta 1,63%, e no terceiro experimento, σ_p (sendo minimizado) é reduzido significativamente em 55,53%, enquanto \bar{T} e $P_{t > 10}$ aumentam 13,06% e 26,21%, respectivamente. Portanto, como já era esperado, os resultados destas análises (considerando cada objetivo separadamente) mostram que estas medidas são conflitantes. Por exemplo, uma melhor solução em termos de σ_p resulta numa pior solução em termos de \bar{T} e $P_{t > 10}$, o que não é desejável em termos do nível de serviço ao usuário. Considerando os tempos computacionais, usando $\Delta = 0,03$, uma simples rodada do AG/hipercubo (com $N = 6$ ambulâncias e $G = 1000$ gerações) consome em média 192 segundos, enquanto o algoritmo enumerativo consome em média 8,340 segundos.

É importante enfatizar que os resultados apresentados acima mostram que \bar{T} , $P_{t > 10}$ e σ_p podem ser melhorados apenas variando-se o tamanho dos átomos do sistema, sem necessitar realocar ambulâncias e sem requerer investimentos adicionais de capacidade.

6.2 Resultados para instâncias maiores que o SAE Anjos do Asfalto

Nesta seção, foi verificado o desempenho desta abordagem para instâncias maiores que o SAE **Anjos do Asfalto**, considerando problemas testes com $N = 6, 8, 10$ e 12 ambulâncias e $N_A = 2N - 2$ átomos. Estas instâncias foram geradas aleatoriamente com base nos dados de entrada da SAE dos **Anjos do Asfalto**. Por exemplo, a taxa de chegada λ_j de cada átomo j foi gerada aleatoriamente sorteando um valor no intervalo $(\lambda_{\min}, \lambda_{\max})$, em que $\lambda_{\min} = 0,00008$ e $\lambda_{\max} = 0,00375$ correspondem às taxas de chegada mínima e máxima no estudo de caso. Similarmente, a taxa de atendimento μ_i para cada ambulância i foi sorteada aleatoriamente no intervalo (μ_{\min}, μ_{\max}) , em que $\mu_{\min} = 0,0101$ e $\mu_{\max} = 0,0241$ correspondem às taxas de atendimento mínima e máxima no estudo de caso.

Para $N = 6$ ambulâncias, o modelo hipercubo envolve um sistema linear com apenas $2^6 = 64$ equações, que é facilmente resolvido por métodos exatos, como o método de eliminação de Gauss. Entretanto, para sistemas maio-

Tabela 1. Resultados do algoritmo AG/hipercubo para as três funções fitness (usando $\Delta = 0,03$).

Medida	Sistema original	Objetivo $\min \bar{T}(x)$	Melhoria (%)	Objetivo $\min P_{t > 10}(x)$	Melhoria (%)	Objetivo $\min \sigma_p(x)$	Melhoria (%)
$\bar{T}(x)$	7,912	7,778	1,69	8,041	- 1,63	8,945	- 13,06
$P_{t > 10}(x)$	0,300	0,275	8,18	0,255	14,86	0,378	- 26,21
$\sigma_p(x)$	0,0551	0,0533	3,09	0,0511	7,26	0,0245	55,53

res (por exemplo, $N \geq 10$), os métodos exatos tornam-se computacionalmente pouco tratáveis para a abordagem AG/hipercubo (dado que o número de equações aumenta exponencialmente). O uso (aproximado) de métodos iterativos, como o método de Gauss-Siedel, parece ser mais apropriado, como discutido em Chiyoshi et al. (2001). No entanto, a convergência torna-se sensível para o valor $\tau = \lambda/\mu$ e para a ordem em que os coeficientes aparecem na matriz do sistema linear. Nestes experimentos, a convergência ocorreu somente gerando a matriz de coeficientes na ordem $y_n \rightarrow y_1$ (ao invés, $y_1 \rightarrow y_n$, em que y é vetor solução do sistema linear, e no modelo hipercubo corresponde ao valor das probabilidades de equilíbrio de estado), em um número razoável de iterações para os problemas testes (menor que 20 iterações, usando como critério de convergência um erro relativo máximo menor que 0,0001).

Similarmente à Tabela 1 da seção 6.1, a Tabela 2 apresenta os resultados obtidos com a aplicação do algoritmo AG/hipercubo para instâncias maiores que o SAE **Anjos do Asfalto**. O método de Gauss-Siedel foi utilizado para resolver todas as instâncias. Nestes experimentos, otimizaram-se separadamente duas das funções de aptidão discutidas na seção 4.2: $\min f(x) = \bar{T}(x)$ e $\min f(x) = \sigma_p(x)$. Como observado na seção 6.1, na Tabela 2, as medidas de desempenho também estão em conflito.

Finalmente, é importante ressaltar que, mesmo usando um método iterativo para resolver o sistema linear do modelo hipercubo, os tempos computacionais do algoritmo AG/hipercubo aumentam significativamente para $N \geq 10$ servidores. Por exemplo, uma simples rodada do algoritmo para uma instância com $N = 12$ servidores consome mais de 100 horas de tempo computacional! No entanto, há poucos SAEs em rodovias brasileiras

com $N \geq 10$ servidores (atualmente, o maior SAE possui $N = 11$ bases). Além disso, se esta abordagem for utilizada para apoiar decisões no nível estratégico, parece ser razoável que o tomador de decisão disponha de maior tempo para defini-la. A presente versão do algoritmo AG/hipercubo torna-se menos promissora para apoiar decisões tomadas de forma mais dinâmica, por exemplo, de acordo com as condições operacionais de uma semana ou de um dia do sistema.

6.3 Curvas de trade-off

Como discutido na seção 5, o algoritmo AG/hipercubo pode ser utilizado para gerar curvas de *trade-off* (curvas de Pareto) entre as diferentes medidas de desempenho do sistema. A Figura 5 mostra o gráfico das soluções eficientes (não-dominadas) obtidas com a aplicação do algoritmo para solucionar o problema (4)-(6), para diferentes valores de ϵ variando-o de 0,026 a 0,050 (usando $\Delta = 0,03$). Conforme descrito na seção 5, os valores da restrição σ_p foram variados para atualizar os valores de \bar{T} em cada conjunto de gerações, obtendo os valores finais para as soluções eficientes mostrados na Figura 5. Na Figura 5, por exemplo, para $\sigma_p = 0,037$, o tempo médio de viagem no sistema \bar{T} é menor que 8 minutos.

7. Conclusões

Neste estudo, foi apresentada uma abordagem que integra o modelo hipercubo em um algoritmo genético para otimizar a configuração e operação dos SAEs em rodovias. O algoritmo AG/hipercubo pode ser aplicado para apoiar decisões no nível estratégico e operacional, por exemplo, determinar qual o tamanho ótimo das áreas de cobertura de cada ambulância, de forma a minimizar o tempo médio de resposta ao usuário do sistema

Tabela 2. Resultados do algoritmo AG/hipercubo para duas funções de aptidão (usando $\Delta = 0,03$).

Instância	Medida	Sistema original	Objetivo $\min \bar{T}(x)$	Melhoria (%)	Objetivo $\min \sigma_p(x)$	Melhoria (%)	Tempo médio de CPU (horas)
N = 6	$\bar{T}(x)$	6,735	6,197	7,98	6,635	1,48	0,05
	$P_{t>10}(x)$	0,181	0,184	- 1,66	0,252	- 39,27	
	$\sigma_p(x)$	0,0576	0,0295	48,78	0,0024	95,83	
N = 8	$\bar{T}(x)$	6,585	6,449	2,06	7,392	- 12,26	1,38
	$P_{t>10}(x)$	0,167	0,173	- 3,59	0,302	- 80,84	
	$\sigma_p(x)$	0,0554	0,0513	7,40	0,0358	35,38	
N = 10	$\bar{T}(x)$	7,083	6,991	1,30	7,873	- 11,15	28,90
	$P_{t>10}(x)$	0,213	0,239	- 12,21	0,348	- 63,38	
	$\sigma_p(x)$	0,0699	0,0758	- 8,44	0,0454	35,05	
N = 12	$\bar{T}(x)$	6,910	6,841	0,10	7,341	- 6,24	150,15
	$P_{t>10}(x)$	0,197	0,219	- 11,17	0,301	- 52,79	
	$\sigma_p(x)$	0,0725	0,0633	12,69	0,0443	38,90	

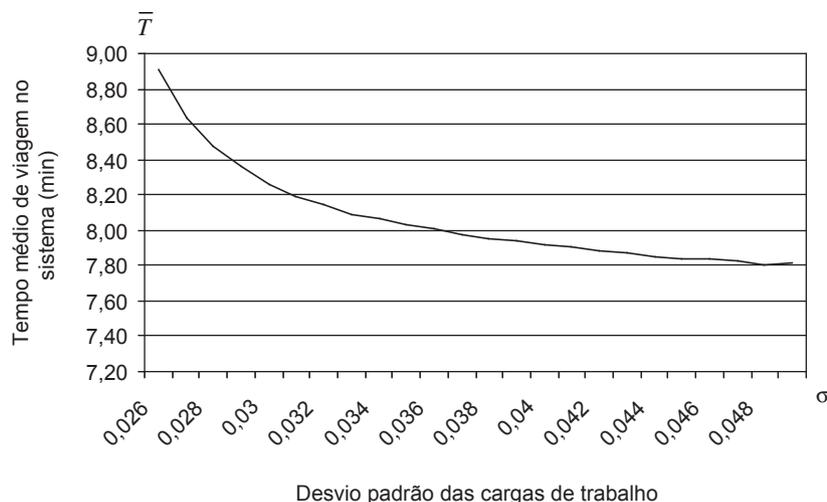


Figura 5. Curva de *trade-off* entre \bar{T} e σ_p .

e o desbalanceamento das cargas de trabalhos entre as ambulâncias. Dado que estas medidas podem ser conflitantes (como discutido na seção 6), foi mostrado como esta abordagem pode ser adaptada para realizar uma análise de *trade-off* e gerar uma fronteira eficiente de Pareto entre estas medidas.

Resultados computacionais foram analisados aplicando-se o algoritmo AG/hipercubo para o estudo de caso do SAE **Anjos do Asfalto**, estudado por Mendonça e Morabito (2000, 2001). Para verificar a qualidade das soluções encontradas pelo AG/hipercubo, foi desenvolvido um algoritmo enumerativo simples para encontrar a solução ótima do problema (sob uma precisão Δ). Em particular, foi mostrado que diferentes medidas de desempenho, tais como tempo médio de resposta ao usuário, fração de chamadas atendidas em mais que 10 minutos e desbalanceamento das cargas de trabalho das ambulâncias, podem ser melhoradas simplesmente variando-se o tamanho dos átomos do sistema, sem necessidade de realocação de ambulâncias ou investimentos adicionais de capacidade.

Uma desvantagem do algoritmo AG/hipercubo (na presente versão) é o tempo computacional exigido para analisar instâncias maiores (i.e., com $N > 10$ ambulâncias), mesmo utilizando métodos iterativos para resolver os sistemas li-

neares do modelo hipercubo. O uso de métodos aproximados do modelo hipercubo (não baseados em sistemas com 2^N equações lineares; ver Larson, 1975; Jarvis, 1985) na abordagem AG/hipercubo pode ser uma alternativa para reduzir os tempos computacionais. Porém o modelo teria menor acuracidade em estimar as medidas de desempenho do sistema. Isto motiva pesquisas futuras baseadas na extensão dos métodos aproximados para analisar SAEs em rodovias.

Outra interessante linha de pesquisa é a combinação de meta-heurísticas, tais como busca tabu (ver p.e., Glover e Laguna, 1997; Gendreau et al., 1997) e *simulated annealing* (ver p.e., Chiyoshi e Galvão, 2000; Galvão et al., 2005), com o modelo hipercubo, ao invés de um algoritmo genético, para otimizar o tamanho dos átomos dos SAEs em rodovias. Em particular, uma extensão interessante do presente estudo é combinar as decisões de localização e dimensionamento das áreas de cobertura das ambulâncias em uma mesma abordagem.

Agradecimentos

Os autores agradecem aos revisores anônimos pelos úteis comentários e sugestões e o apoio financeiro do CNPq (processos: 140178/01-5 e 522973/95-4) e da CAPES (processo BEX2468/02-6).

Referências Bibliográficas

ARROYO, J. C. **Heurísticas e Metaheurísticas para otimização combinatória multiobjetivo**. 2002. 256p. Tese (Doutorado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, 2002.

ARROYO, J. C.; ARMENTANO, V. A. Genetic local search for multi-objective flowshop scheduling. **European Journal of Operational Research**, v. 167, p. 717-738, 2005.

BATTA, R.; DOLAN, J. M.; KRISHNAMURTHY, N. N. The maximal expected covering location problem: Revisited.

- Transportation Science**, v. 23, n. 4, p. 277-287, 1989.
- BEASLEY, J. E. **Population heuristics**. In: Handbook of Applied Optimization. Pardalos, P. M., Resende, M. G. C. (Eds). Oxford: Oxford University Press, 2002. p. 138-157,
- BRANDEAU, M.; LARSON, R. C. **Extending and applying the hypercube queuing model to deploy ambulances in Boston**. In: Delivery of Urban Services. TIMS Studies in the Management Science. Swersey, A. J, Ingnall, E. J. (Eds). Elsevier v. 22, p. 121-153, 1986.
- BROTCORNE, L.; LAPORTE, G.; SEMET, F. Ambulance location and relocation models. **European Journal of Operational Research**, v. 147, p. 451-63, 2003.
- BURWELL, T. H.; JARVIS, J. P.; MCKNEW, M. A. Modeling co-located servers and dispatch ties in the hypercube model. **Computers and Operations Research**, v. 20, n. 2, p. 113-119, 1993.
- CHELST, K.; BARLACH, Z. Multiple unit dispatches in emergency services: models to estimate system performance. **Management Science**, v. 27, n. 12, p. 1390-1409, 1981.
- CHIYOSHI, F.; GALVÃO, R. D. A statistical analysis of simulated annealing applied to the p-median problem. **Annals of Operations Research**, v. 96, p. 61-74, 2000.
- CHIYOSHI, F.; GALVÃO, R. D.; MORABITO, R. Modelo hipercubo: análise e resultados para o caso de servidores não-homogêneos. **Pesquisa Operacional**, v. 21, n. 2, p. 199-218, 2001.
- _____. O uso do modelo hipercubo na solução de problemas de localização probabilísticos. **Gestão & Produção**, v. 7, n. 2, p. 146-174, 2000.
- _____. A note on solutions to the maximal expected covering location problem. **Computers and Operations Research**, v. 30, n. 1, p. 87-96, 2003.
- COHON, J. L.; **Multiobjective programming & planning**. New York: Academic Press, 1978.
- COSTA, D. M. B. **Uma metodologia iterativa para determinação de zonas de atendimento de serviços emergenciais**. 2004. 120 p. Tese (Doutorado em Engenharia de Produção) - Departamento de Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2004.
- GALVÃO, R. D.; et al. Solução do problema de localização de máxima disponibilidade utilizando o modelo hipercubo. **Pesquisa Operacional**, v. 23, n. 1, p. 61-78, 2003.
- GALVÃO, R. D.; CHIYOSHI, F.; MORABITO, R. Towards unified formulations and extensions of two classical probabilistic location models. **Computers & Operations Research** v. 32, n. 1, p. 15-33, 2005.
- GENDREAU, M.; LAPORTE, G.; SEMET, F. Solving an ambulance location model by tabu search. **Location Science**, v. 5, n. 2, p. 75-88, 1997.
- GLOVER, F. W.; LAGUNA, M. **Tabu Search**. Dordrecht: Kluwer Academic Publishers, 1997.
- GOLDBERG, D. E. **Genetic Algorithms in Search, Optimization and Machine Learning**. New York: Addison Wesley, 1989.
- HALPERN, J. Accuracy of estimates for the performance criteria in certain emergency service queuing systems. **Transportation Science**, v. 11, n. 3, p. 223-242, 1977.
- HERTZ, A.; KOBLER, D. A framework for the description of evolutionary algorithms. **European Journal of Operational Research**, v. 126, p. 1-12, 2000.
- HOLLAND, J. H. **Adaptation in Natural and Artificial Systems**, Cambridge: MIT Press, 1975.
- IANNONI, A. **Otimização da configuração e operação de sistemas médico emergenciais em rodovias utilizando o modelo hipercubo**. 2005. 213 p. Tese (doutorado em Engenharia de Produção) - Departamento de Engenharia de Produção, Universidade Federal de São Carlos, São Carlos, 2005.
- IANNONI, A. P.; MORABITO, R.; SAYDAM, C. Analyzing the configuration and operation of emergency medical systems on highways using the hypercube model. In: XII CLAIO – Congresso Latino-Iberoamericano de Investigación de Operaciones y Sistemas, Outubro, 2004, Havana, Cuba. **Anais - Resumos XII CLAIO 2004**. 2004, p. 55-55,
- JARVIS, J. P. Approximating the equilibrium behavior of multi-server loss systems. **Management Science**, v. 31, n. 2, p. 235-239, 1985.
- JASZKIEWICZ, A. Genetic local search for multi-objective combinatorial optimization. **European Journal of Operational Research**, v. 137, p. 50-71, 2002.
- LARSON, R. C. Hypercube queuing model for facility location and redistricting in urban emergency services. **Computers and Operations Research**, v. 1, n. 1, p. 67-95, 1974.
- _____. Approximating the performance of urban emergency service systems. **Operations Research**, v. 23, n. 5, p. 845-868, 1975.
- _____. OR models for homeland security. **OR/MS Today**, v. 31, n. 5, p. 22-29, 2004.
- LARSON, R. C.; ODONI A. R. **Urban operations research**. New Jersey: Prentice Hall. 1981.
- MENDONÇA, F. C.; MORABITO R. Aplicação do modelo hipercubo para análise de um sistema médico-emergencial em rodovia. **Gestão & Produção**, v. 7, n. 1, p. 73-91, 2000.

- _____. Analyzing emergency service ambulance deployment on a Brazilian highway using the hypercube model. **Journal of the Operation Research Society**, v. 52, p. 261-268, 2001.
- MICHALEWICZ, Z. **Genetic Algorithms + Data Structures = Evolution Programs**, 3. ed., Berlin: Springer-Verlag, 1996.
- OWEN, S. H.; DASKIN, M. S. Strategic facility location: A review. **European Journal of Operational Research**, v. 111, p. 423-447, 1998.
- POULOS, P. N.; *et al.* A Pareto-optimal algorithm for warehouse multi-objective optimization. **Engineering Applications of Artificial Intelligence**, v. 14, p. 737-749, 2001.
- SACKS, S. R.; GRIEF S. Orlando Police Department uses OR/MS methodology, new software to design patrol districts. **OR/MS Today**, p. 30-32, 1994.
- SAYDAM, C.; AYTUG, H. Accurate estimation of expected coverage: revisited. **Socio-Economic Planning Sciences**, v. 37, p. 69-80, 2003.
- STEUER, R. E. **Multiple criteria optimization: theory, computation & application**. New York: John Wiley, 1986.
- SWERSEY, A. J. **Handbooks in OR/MS**. Amsterdam: Elsevier Science B. V., v. 6, p. 151-200, 1994.
- TAKEDA, R. A.; WIDMER, J. A.; MORABITO, R. Uma proposta alternativa para avaliação do desempenho de sistemas de transporte emergencial de saúde brasileiros. **Transportes**, v. 9, n. 2, p. 9-27, 2000.
- _____. Aplicação do modelo hipercubo de filas para avaliar a descentralização de ambulâncias em um sistema urbano de atendimento médico de urgência. **Pesquisa Operacional**, v. 24, n. 1, p. 39-72, 2004.
- _____. Analysis of ambulance decentralization in an urban medical emergency service using the hypercube queueing model. Aceito para publicação em **Computers & Operations Research**, 2005.

THE HYPERCUBE QUEUING MODEL INTEGRATED TO A GENETIC ALGORITHM TO ANALYZE EMERGENCY MEDICAL SYSTEMS ON HIGHWAYS

Abstract

The hypercube model, well-known in the literature on problems of server-to-customer localization systems, is based on the spatially distributed queueing theory and Markovian analysis approximations. The model can be modified to analyze Emergency Medical Systems (EMSs) on highways, considering the particularities of these systems' dispatching policies. In this study, we combine the hypercube model with a genetic algorithm to optimize the configuration and operation of EMSs on highways. This approach is effective to support planning and operation decisions, such as determining the ideal size of the area each ambulance should cover to minimize not only the average time of response to the user but also ambulance workload imbalances, as well as generating a Pareto efficient boundary between these measures. The computational results of this approach were analyzed using real data Anjos do Asfalto EMS (which covers the Presidente Dutra highway).

Keywords: hypercube queueing model, emergency medical systems, ambulance deployment, highways.