Universidade Federal de Minas Gerais
Disciplina de Computação Natural
Gisele L. Pappa

Trabalho Prático 1

Data de Entrega: 15 de outubro de 2012

Um Algoritmo Genético para resolver o problema da p-Mediana com restrições de capacidade

1. Contextualização

O **problema da p-mediana** com restrições de capacidade faz parte de uma classe ampla de problemas conhecidos como problemas de localização. O problema consiste em decidir onde localizar P centros em uma rede composta por vértices e arestas, de forma a minimizar a soma de todas as distâncias de cada vértice ao centro mais próximo. Além disso, no problema da p-mediana com restrições de capacidade, existe uma demanda associada a cada vértice, que restringe a capacidade de atendimento dos centros. O problema da p-mediana é um problema de otimização combinatória NP-difícil.

São diversas as situações práticas reais em que se pretende localizar um centro de modo a minimizar a média das distâncias ponderadas entre cada cliente (vértice do grafo) e o centro mais próximo. Entre elas, podemos citar a localização de serviços públicos (escolas, hospitais, bibliotecas), instalação de novos pontos de ônibus, armazéns, antenas de telecomunicação, entre outros.

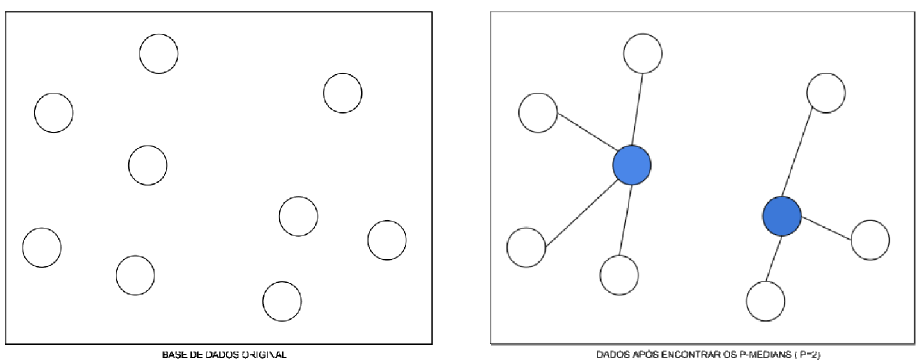

Figura 1: A figura a esquerda mostra os vértices da base, enquanto a da direita mostra a resolução do problema para P=2. Os vértices em azul atendem os em branco.

A seguir definimos formalmente esse problema. Dados um conjunto de pontos *i* aos quais está associada uma demanda (vértices do grafo), as distâncias d_{ij} entre o nó de demanda *i* e um local *j*, queremos encontrar [7]:

$$v(P) = \min \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} x_{ij} \quad (1)$$

onde *n é* o número de vértices na rede, *P* é o número de centros (medianas) a serem localizados; $[d_{ij}]_{nxn}$ é uma matriz de custos (distâncias). Além disso, temos:

$$x_{ij} = \begin{cases} 1, & \text{se o vértice } i \text{ é atendido pelo centro } j,\ i \neq j; \\ 0, & \text{caso contrário.} \end{cases}$$

$$x_{jj} = \begin{cases} 1, & \text{se o vértice } j \text{ é um centro;} \\ 0, & \text{caso contrário.} \end{cases}$$

Além disso, as seguintes restrições devem ser obedecidas:

$$\sum_{j=1}^{n} x_{ij} - 1,\ \forall\ i \in N \tag{2}$$

$$\sum_{j=1}^{n} x_{jj} = p \tag{3}$$

$$\sum_{i-1}^{n} a_i x_{ij} \leq b_j x_{jj},\ \forall\ j \subset N \tag{4}$$

- A restrição (2) garante que, para todos os vértices, ou o vértice é uma mediana, e logo não é atendido por nenhuma outra mediana; ou ele é atendido por uma (e apenas uma) mediana.
- A restrição (3) garante que o número de medianas seja igual a *p*.
- Sendo a_i a demanda do vértice *i* e b_j a capacidade de atendimento do centro *j*, se este for escolhido como mediana, a restrição (4) garante que a capacidade de alocação de uma mediana não seja ultrapassada.

2. Objetivos

O principal objetivo desse trabalho prático é entender e implementar os componentes básicos de um algoritmo evolucionário, fazer uma análise exaustiva dos parâmetros utilizados pelo algoritmo e estudar como esses interferem no resultado obtido pelo algoritmo evolucionário (AE). O trabalho está dividido em duas partes: (1) implementação e (2) análise dos parâmetros e funcionalidade do algoritmo. Na parte de implementação, três pontos principais devem ser levados em consideração:

1. A representação do indivíduo.
2. A fitness dos indivíduos.
3. Os operadores genéticos utilizados

Como estudado anteriormente, esses pontos são fortemente dependentes do problema a ser resolvido, e devem ser escolhidos adequadamente. Quanto ao método de seleção utilizado, recomenda-se utilizar a seleção por torneio, e testar a estratégia de elitismo. Porém, outros conceitos vistos em sala de aula, tais como niching e otimização multi-objetiva, podem ser utilizados para melhorar o desempenho do algoritmo.

Na parte de análise, os seguintes parâmetros deverão ser estudados:

1. Tamanho da população
2. Número de gerações
3. O número de indivíduos utilizados na seleção por torneio (pressão seletiva) - Analisar a diversidade da população gerada e verificar se a população converge e quando converge.
4. Probabilidade de aplicação dos operadores genéticos.

Por ser um método estocástico, o AE deve ser rodado no mínimo 30 vezes. Assim, os resultados apresentados devem sempre considerar estatísticas médias de todas as execuções (seguidas de desvio padrão).

3- Bases de dados

Serão utilizadas quatro bases de dados[1], como diferentes números de nós e valores de p. A Tabela 1 apresenta as 4 instâncias que serão utilizadas, juntamente com o valor do ótimo encontrado. Os valores de solução ótima correspondem a um *lower bound.* Portanto a solução ótima verdadeira pode ser um pouco maior que o apresentado (para maiores detalhes veja [6]). Os resultados encontrados pelo GA devem ser comparados com os das soluções ótimas apresentadas.

Tabela 1: Instâncias do problema da p-mediana a serem executadas

Base	Nós	P	Ótimo
SJC1.dat - http://www.lac.inpe.br/~lorena/instances/pmedian/SJC1.dat	100	10	17246,53
SJC2.dat - http://www.lac.inpe.br/~lorena/instances/pmedian/SJC2.dat	200	15	33225,88
SJC3b.dat - http://www.lac.inpe.br/~lorena/instances/pmedian/SJC3b.dat	300	30	40635,80
SJC4a.dat - http://www.lac.inpe.br/~lorena/instances/pmedian/SJC4a.dat	402	30	61843,23

Os arquivos de entrada seguem o seguinte formato:

1 linha: n Z

i-éssima linha: x y c d

onde:

- n = # de pontos
- p = # de p-medianas
- x = coordenada x do i-éssimo ponto
- y = coordenada y do i-éssimo ponto
- c = capacidade c do i-éssimo ponto
- d = demando d do i-éssimo ponto

4- Guia de Experimentos

1. Escolha do tamanho da população e número de gerações (utilizar elitismo, torneio de tamanho 2 e $p_c = 0.6$ e $p_m = 0.001$) de acordo com o número de arestas do grafo. Inicie com 100, 500 ou 1000.
2. Após alguns testes, defina o tamanho da população e o número de gerações e varie p_c e depois p_m . Os parâmetros escolhidos no passo 1 ainda são apropriados?
3. Definidos o tamanho da população, número de gerações, p_c e p_m , aumente o tamanho do torneio para 5.
4. Escolha os melhores parâmetros anteriores e retire o elitismo. Os resultados obtidos são os mesmos?
5. Se desejar, teste outras características de AEs nesse problema.

Estatísticas importantes

1. Fitness do melhor e pior indivíduos
2. Fitness média da população
3. Número de indivíduos repetidos na população
4. Número de indivíduos gerados por *crossover* melhores e piores que a fitness média dos pais

5- Relatório

O relatório deve ser **claro** e **preciso**, contendo as seguintes seções:

1. O modelo: descreva claramente o seu modelo, as nuanças e as decisões de projetos adotadas. Explique cada decisão e detalhe do mesmo.
2. Metodologia: descreva como foram feitos os experimentos: quantidade de execução, gráficos postados, parâmetros adotados em cada experimento, etc...
3. Experimentos: apresente todos os experimentos feitos.

[1] As bases foram retiradas do site http://www.lac.inpe.br/~lorena/instancias.html.

4. Resultados: explique o resultados dos seus experimentos.
   Comentários óbvios do tipo “Podemos ver pela figura 2 que a função a é maior que a função b.” são desnecessários, apresente conclusões obtidas, facilite o trabalho do leitor: “A figura 2 mostra que no contexto X a função a é sempre superior. Isso comprova nossa tese ...”
5. Conclusão: O que você aprendeu e concluir sobre GA e otimização após fazer esse trabalho?

6- Avaliação

- Código: instruções de como executar; *toy*-testes rodados com sucesso? (peso 2)
- O modelo adotado é coerente? Descreveu corretamente o modelo? (peso 1)
- Descreveu corretamente a metodologia adotada? (peso 1)
- Experimentação: testes relevantes? quantidade boa de testes? testou todas as bases? (peso 3)
- Resultados: A análise apresentada é coerente com os resultados? (peso 2)
- Conclusão: fechou bem o trabalho? (peso 1)

Referências

1. Cristofides, N. *Graph Theory. An Algorithm Approach*. New York: Academic Press. 1975.
2. Hansen, P. and Jaumard, B. Cluster Analysis and Mathematical Programming. *Mathematical Programming* 79 (1997), p. 191-215.
3. Krarup, J. and Pruzan, P. The simple plant location problem: survey and synthesis. *European J. Oper. Res.* 12 (1983), N 1, p.36-81.
4. Mirchandani, P. and Francis, R. (eds.) *Discrete Location Theory*. Wiley-Intersience. 1990.
5. Edson L. F. Senne and Estadual Paulista and Luiz A.N. Lorena and Nacional Pesquisas Espaciais. *Lagrangean/Surrogate Heuristics for p-Median Problems*. 2000.
6. LORENA, Luiz Antonio Nogueira; SENNE, Edson Luiz França; PAIVA, João Argemiro de Carvalho and PEREIRA, Marcos Antonio. Integração de modelos de localização a sistemas de informações geográficas. Gest. Prod. [online]. 2001, vol.8, n.2, pp. 180-195. ISSN 0104-530X.