Uma formulação de dispersão discreta para o problema de rotulação cartográfica de pontos

Sóstenes Pereira Gomes[1], Luiz Antonio Nogueira Lorena[2], Glaydston Mattos Ribeiro[3]

[1]Programa de Doutorado em Computação Aplicada – CAP
Instituto Nacional de Pesquisas Espaciais – INPE

[2]Laboratório Associado de Computação e Matemática Aplicada – LAC
Instituto Nacional de Pesquisas Espaciais – INPE

[3]Departamento de Engenharias e Computação – DECOM
Universidade Federal do Espírito Santo – CEUNES/UFES

{sostenes.gomes,lorena,glaydston}@lac.inpe.br

Abstract. *This work presents a brief review on the Point Feature Label Placement Problem and its modeling in conflict graphs. In this context, a formulation of discrete dispersion for the Point Feature Label Placement Problem is presented as a new approach to the problem. Preliminary results are also presented.*

Resumo. *Neste trabalho é apresentada uma breve revisão sobre o Problema de Rotulação Cartográfica de Pontos (PRCP) e sua modelagem em grafos de conflitos. Neste contexto, uma formulação de dispersão discreta para o PRCP é apresentada como uma nova abordagem do problema. Resultados preliminares são também apresentados.*

Palavras-chave: *rotulação cartográfica, problema de dispersão, otimização combinatória.*

1. Introdução

O posicionamento de rótulos é uma tarefa essencial no desenvolvimento de mapas, diagramas e objetos gráficos em geral. A atividade de rotular mapas automaticamente, em especial, é considerada um problema de grande complexidade, pois é necessário considerar diversos critérios quanto à escolha da localização, orientação, dimensões, característica a ser rotulada, etc. Um problema surge quando ao rotular características cartográficas, rótulos são sobrepostos, dificultando a leitura do mapa. Este problema é denominado Problema de Rotulação Cartográfica na literatura. Exemplos de sobreposições de rótulos são apresentados na Figura 1.

Este trabalho tem por objetivo apresentar um estudo sobre o problema da rotulação automática de características de pontos, evitando a sobreposição dos rótulos e suas principais abordagens, conhecido na literatura como o Problema de Rotulação Cartográfica de Pontos (PRCP). Além disto, é introduzida outra abordagem do problema, que considera a dispersão no posicionamento dos rótulos como um fator que pode aumentar a legibilidade do mapa.

Figura 1. Sobreposições de rótulos indicadas por setas.
Fonte: [Ribeiro e Lorena 2005].

2. Problema de Rotulação Cartográfica

Rotulação cartográfica refere-se ao posicionamento de rótulos em mapas, diagramas, grafos ou qualquer objeto gráfico a ser rotulado. Em geral, as principais características cartográficas a serem rotuladas são linhas (cidades, rios, etc.), polígonos (lagos, estados, construções em escalas menores, etc.) e pontos (cidades, construções isoladas, etc.). Este trabalho foca-se no problema de rotulação de características de pontos, que é melhor explicado a seguir.

2.1. Representação do PRCP em Grafos de Conflitos e Principais Abordagens

No PRCP, ao rotular um conjunto de pontos (em um mapa, diagrama, grafo ou figura), utiliza-se um conjunto de posições candidatas correspondentes a cada ponto.

O termo "posições candidatas" no PRCP refere-se ao conjunto de todas as posições possíveis de localização de um rótulo, para um determinado ponto. Em [Christensen *et al.* 1995] o autor apresenta padronizações para a escolha de posições candidatas utilizando números para representar a prioridade de escolha de uma posição, de forma que o menor número representa a prioridade mais alta. A Figura 2 exemplifica dois conjuntos de posições candidatas de acordo com esse padrão, com 4 e 8 posições respectivamente.

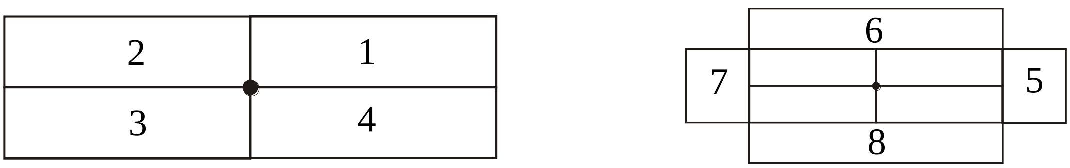

Figura 2. Padronização cartográfica proposta por [Christensen *et al.* 1995].

Quando uma característica cartográfica é rotulada em uma posição candidata que sobrepõe um rótulo posicionado no mapa é dito que estes rótulos são "conflitantes". O problema de selecionar posições candidatas de maneira que a quantidade de conflitos seja minimizada é um problema NP-Difícil e tem sido abordado na literatura como um problema de Otimização Combinatória. Em geral o problema é modelado através de um Grafo de Conflitos.

No grafo de conflitos para o PRCP, cada vértice $v_{i,j} \in V$ representa uma posição candidata j do ponto i. Grafos de conflitos são utilizados para gerar inequações de cliques, onde uma solução viável é uma seleção de vértices de V com a restrição de que no máximo um nó em cada clique do grafo deve ser selecionado. A Figura 3 apresenta um exemplo com quatro pontos a serem rotulados, com cada ponto possuindo 4 posições candidatas e o grafo de conflitos gerado a partir dos possíveis conflitos entre estas posições.

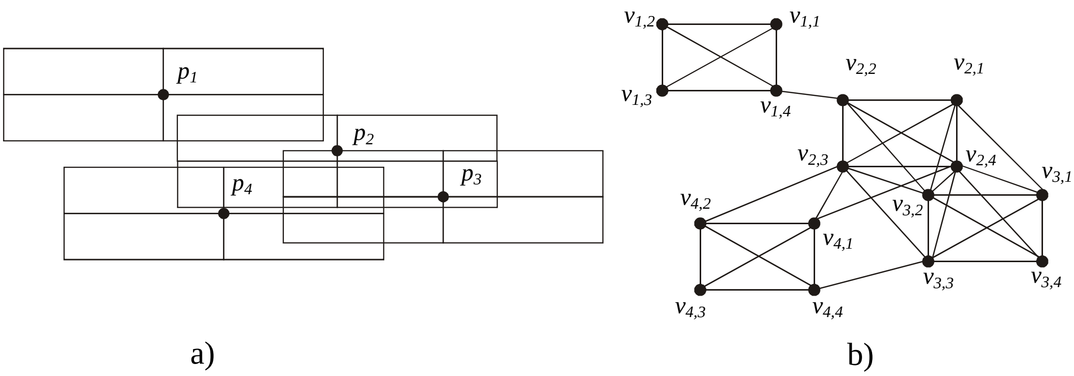

Figura 3. Exemplo de modelagem do PRCP em grafo de conflitos: a) Instância do problema com 4 pontos, b) Grafo de conflitos gerado.

Existem atualmente diversas abordagens do PRCP e, dentre elas, destacam-se na literatura o Problema de Rotulação como um Problema do Máximo Conjunto Independente de Vértices (PMCIV) [Zoraster 1990], o Problema do Máximo Número de Rótulos Sem Conflitos (PMNRSC) [Mauri *et al.* 2010] e o Problema da Minimização do Número de Conflitos (PMNC) [Ribeiro e Lorena 2004].

Como um PMCIV, o objetivo é rotular a maior quantidade de pontos possível, e como problemas reais são em geral muito complexos, alguns pontos podem não ser rotulados com esta abordagem.

No PMNRSC, o objetivo é rotular todos os pontos do mapa de forma que a quantidade de pontos rotulados sem conflitos seja a máxima possível, não levando em consideração que algumas áreas do mapa podem ficar ilegíveis. Já o PMNC visa minimizar a quantidade de conflitos gerados no processo de rotular todos os pontos do mapa. Na Figura 4, a diferença entre as duas abordagens é ilustrada.

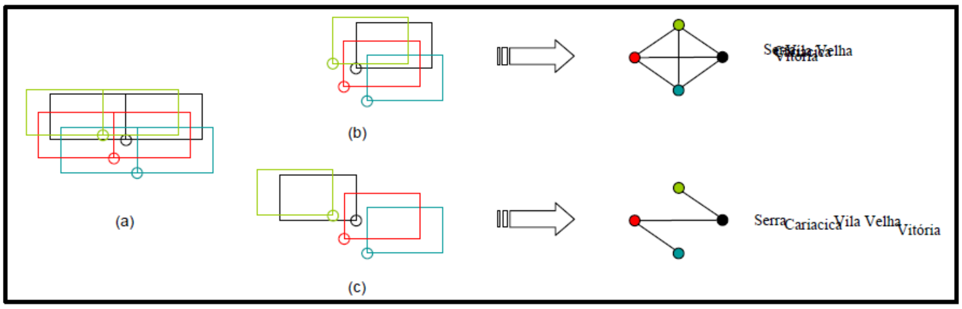

Figura 4. Diferença entre PMNRSC e PMNC: a) Posições candidatas de uma instância. b) Solução como PMNRSC. c) Solução como PMNC.

A Figura 4 apresenta duas soluções possíveis para uma instância do PRCP (a). Ambas as soluções são idênticas do ponto de vista do PMNRSC, pois as duas soluções possuem quatro rótulos em conflito. Considerando as soluções do ponto de vista do PMNC, a solução (b) é pior que a solução (c), pois possui maior quantidade de conflitos. Observa-se que nas abordagens dos problemas como PMNRSC e PMNC, ao contrário do PMCIV, todos os pontos são rotulados.

Para comparação com o modelo apresentado neste trabalho é utilizada a Formulação Matemática Baseada em Posições Candidatas (FMBPC) proposta por Ribeiro e Lorena [Ribeiro e Lorena 2008], que obteve bons resultados de rótulos livres de conflitos na literatura. O FMBPC é descrito a seguir:

$$Min \sum_{i=1}^{N} \sum_{j=1}^{P_i} \left(w_{i,j} x_{i,j} + \sum_{(k,t) \in S_{i,j}} y_{i,j,k,t} \right) \quad (2.1)$$

Sujeito a:

$$\sum_{j=1}^{P_i} x_{i,j} = 1, \forall\, i = 1 \dots N \quad (2.2)$$

$$\begin{aligned} x_{i,j} + x_{k,t} - y_{i,j,k,t} \leq 1, \forall\, i = 1 \dots N \\ \forall\, j = 1 \dots P_i \\ (k,t) \in S_{i,j} \end{aligned} \quad (2.3)$$

$$\begin{aligned} x_{i,j}, x_{k,t}, y_{i,j,k,t} \in \{0,1\}, \forall\, i = 1 \dots N \\ \forall\, j = 1 \dots P_i \\ (k,t) \in S_{i,j} \end{aligned} \quad (2.4)$$

Na formulação (2.1) – (2.4) é considerado que cada ponto i={1, 2,...,N} tem um conjunto P_i de posições candidatas $j \in P_i$ e para cada posição candidata $j \in P_i$, existe uma variável binária $x_{i,j}$ para a decisão de posicionar ou não um rótulo (0 ou 1). Esta decisão é ponderada por um custo $w_{i,j}$ que refere-se à preferência cartográfica definida por [Christensen *et al.* 1995], como ilustrado na Figura 2. Cada posição candidata j de um ponto i – representado por um par de índices (i, j) – possui um conjunto $S_{i,j}$ de pares de índices (k, t): $k > i$ de posições candidatas em possível conflito com (i, j). A restrição (2.2) assegura que somente uma posição candidata de cada ponto será escolhida para o posicionamento de rótulo. A restrição (2.3) assegura que se posições candidatas em possível conflito são escolhidas para a rotulação, a função objetivo será penalizada, atribuindo o valor 1 à variável de conflito $y_{i,j,k,t}$. A restrição (2.4) indica todas as variáveis como binárias.

3. Dispersão de rótulos no PRCP

Nesta seção é apresentada uma nova formulação para o PRCP que considera a distância euclidiana entre posições candidatas que são potencialmente conflitantes entre si. Esta formulação aborda o PRCP como um Problema de Dispersão Discreta [Curtin e Church 2006].

Modelos de dispersão têm sido aplicados a diversos problemas na literatura e principalmente ao Problema de Localização de Facilidades. Problemas de Dispersão de Facilidades surgem em diversas situações como: dispersão de centros de reabilitação criminal de centros populacionais, dispersão na localização de usinas nucleares para maximizar a segurança, dispersão na localização de lojas de uma mesma franquia, etc.

No contexto do PRCP, é desejável que ao rotular um conjunto de pontos os rótulos sejam posicionados dispersamente de maneira que conflitos sejam evitados. Neste trabalho apresenta-se um Modelo de Dispersão de Rótulos (MDR) para o PRCP. A formulação é descrita a seguir:

$$Max\ z \tag{3.1}$$

Sujeito a:

$$\sum_{j=1}^{P_i} x_{i,j} = 1, \forall\ i = 1 \dots N \tag{3.2}$$

$$z \leq\ d_{i,j,k,t}(2 - x_{i,j} - x_{k,t}\ ), \forall\ i = 1 \dots N \quad \forall\ j = 1 \dots P_i \quad (k,t) \in S_{i,j} \tag{3.3}$$

$$x_{i,j}, x_{k,t} \in \{0,1\}, \forall\ i = 1 \dots N \quad \forall\ j = 1 \dots P_i \quad (k,t) \in S_{i,j} \tag{3.4}$$

$$z \geq 0 \tag{3.5}$$

A restrição (3.2) assegura que apenas uma posição candidata de cada ponto será escolhida para o posicionamento de rótulo. A restrição (3.3) utiliza a distância $d_{i,j,k,t}$ entre a posição candidata j do ponto i e a posição candidata t do ponto k como recompensa para sobreposições evitadas. Como a função objetivo é maximizar o valor de z, é mais vantajoso tentar atribuir 0 a $x_{i,j}$ e $x_{k,t}$ para que z aumente até $2d_{i,j,k,t}$, evitando assim a sobreposição entre (i, j) e (k, t).

A abordagem do PRCP como Problema de Dispersão Discreta consta então de obter a máxima menor distância entre rótulos posicionados. Na seção a seguir são apresentados alguns resultados preliminares destas formulações.

4. Resultados

Nesta seção são apresentados resultados obtidos através de testes computacionais da formulação definida na seção 3, realizados utilizando o *solver* de Programação Matemática CPLEX 12.1. Não foi possível obter instâncias na literatura com os valores das distâncias entre posições candidatas para a execução de testes comparativos, já que a abordagem da dispersão de rótulos é nova e sendo assim, para os testes, foram geradas instâncias com dimensões de rótulos variadas para um conjunto de 505 pontos com 4 posições candidatas cada, somando um total de 2020 posições candidatas.

Para comparação, foram utilizados resultados da Formulação Matemática Baseada em Posições Candidatas (FMBPC), que obteve bons resultados para o PRCP abordado como um PMNC.

Seguem os dados considerados nos testes:

- A e C – Altura e comprimento dos rótulos respectivamente;
- Tmp. (*seg.*) – Tempo computacional em segundos;
- Rot. liv. (%) – Proporção de rótulos livres de conflitos na solução;
- $min(d_{i,j,k,t})$ – Menor distância entre rótulos conflitantes em uma solução.

A seguir apresenta-se uma tabela comparativa entre os resultados das duas formulações.

Tabela 1. Resultados das formulações para 10 instâncias.

	Dim.		**Tmp. (*seg.*)**		**Rot. Liv (%)**		**$min(d_{i,j,k,t})$**	
	A	**C**	**FMBPC**	**MDR**	**FMBPC**	**MDR**	**FMBPC**	**MDR**
1	2	24	0.09	0.12	100	100	-	-
2	2	32	0.28	1.31	100	100	-	-
3	2	42	1.59	3.85	99,2	99.2	**10.04**	9.05
4	2	48	2.02	4.29	99,2	**99.6**	**30.01**	9.05
5	3	16	0.14	1.17	100	100	-	-
6	3	24	2.41	2.29	98.4	**99.6**	**11.04**	9.21
7	3	28	4.28	4.41	97.2	**99.2**	3.16	**9.05**
8	3	32	6.01	34.62	95.6	**98**	3.16	**9.05**
9	4	16	15.97	5.79	96.4	**98.4**	6.08	**9.05**
10	4	18	6.92	4.04	95.2	**98.4**	3.16	**9.21**

A formulação MDR visa dispersar ao máximo os rótulos quando não é possível evitar conflitos e para as instâncias com rótulos maiores e maior quantidade de possíveis conflitos o MDR obteve melhor resultado de dispersão. O MDR obteve também melhores resultados de proporção de rótulos livres, empatando apenas nos resultados 100% livres de conflitos e na instância 3.

5. Conclusões e trabalhos futuros

A modelagem de dispersão para rótulos é nova e a pesquisa está ainda em estágio inicial. Com a formulação proposta foi possível obter bons resultados tanto quanto à legibilidade da solução, quanto com relação à quantidade de rótulos posicionados livres de conflitos.

Algumas melhorias na formulação podem ser feitas, como por exemplo, a reformulação da função objetivo para considerar as posições candidatas rotuladas sem conflitos (problema bi-objetivo). Outra alteração na formulação a ser feita é a reformulação da restrição (3.3) para considerar conjuntos de posições potencialmente conflitantes, ao invés de considerar os possíveis conflitos par a par, que possibilitará reduzir o número de restrições nos modelos. Além disso, é possível explorar outras vertentes da abordagem dispersiva do PRCP, como por exemplo, a modelagem de dispersão total, que visaria maximizar as distâncias dos rótulos posicionados par a par, ao invés da distância mínima.

Referências

Christensen, J., Marks, J., Shieber, S. (1995), An empirical study of algorithms for point-features label placement. In: ACM TRANSACTIONS ON GRAPHICS, v. 14, n. 3, p. 203–232.

Curtin, K. M., Church, R. L. (2006), A Family of Location Models for Multiple-Type Discrete Dispersion. GEOGRAPHICAL ANALYSIS, n. 38, p. 248–270.

Mauri, G.R., Ribeiro, G.M., Lorena, L.A.N. (2010), A new mathematical model and a lagrangean decomposition for the point-feature cartographic label placement problem. COMPUTERS & OPERATIONS RESEARCH, v. 37, n. 12, p. 2164-2172.

Ribeiro, G. M., Lorena, L. A. N. (2004), Modelagem matemática e relaxações Lagrangeana e Lagrangeana/Surrogate para o problema de rotulação cartográfica de pontos. In: Simpósio Brasileiro de Pesquisa Operacional (SBPO), 36., São João del Rei. Anais... Rio de Janeiro: SOBRAPO, 2004. 1 CD-ROM.

Ribeiro, G. M., Lorena, L. A. N. (2005), Lagrangean relaxation with clusters for point-feature cartographic label placement problems. ECCO XVII post conference special issues.

Ribeiro, G. M., Lorena, L. A. N. (2008), Lagrangean relaxation with clusters for point-feature cartographic label placement problems. Computers & Operations Research 35(7), p. 2129-2140.

Zoraster, S. (1990), The solution of large 0-1 integer programming problems encountered in automated cartography. OPERATIONS RESEARCH, v. 38, n. 5, p. 752–759.