

Cobertura Máxima e Custo Mínimo para Alocação de Recursos no Tempo

Felipe Whitaker de A. M. Tavares

Pontifical Catholic University of Rio de Janeiro, Department of Industrial Engineering
Rua Marquês de São de Vicente, 225 – Gávea, Rio de Janeiro, RJ – Brazil – 22541-900
felipewhitaker@aluno.puc-rio.br

Bruno Fanzeres

Pontifical Catholic University of Rio de Janeiro, Department of Industrial Engineering
Rua Marquês de São de Vicente, 225 – Gávea, Rio de Janeiro, RJ – Brazil – 22541-900
bruno.santos@puc-rio.br

RESUMO

Modelar decisões sob incerteza depende da compreensão das decisões disponíveis e da incerteza a ser tratada. O presente trabalho modela o problema de cobertura e alocação de recursos como um processo de decisão em duas fases com incerteza do local e momento da ocorrência. Utilizando dados de Crimes em Chicago [Currie32, 2017] e aplicando o modelo FLEET-ICt [de Oliveira, 2020], foi possível determinar quais instalações escolher e como alocar especialistas. Cinco amostras de 1.000 linhas foram geradas e otimizadas. Para todas as amostras o modelo foi capaz de atingir aproximadamente 90% de cobertura em poucos segundos. O modelo também foi adaptado para minimizar o custo a partir de uma cobertura mínima como parâmetro, o que se provou computacionalmente mais complexo, tendo levado até 20 minutos. Com esse outro modelo foram encontradas soluções com a mesma cobertura que a cobertura máxima com um custo menor.

PALAVRAS CHAVE. Cobertura Máxima, Custo Mínimo, Dados Espaço-Temporais. Tópicos: Otimização Combinatória, Defesa Pública, Big Data

ABSTRACT

Modeling decisions under uncertainty depend on understanding the available decisions and the uncertainty to be addressed. The present work models the problem of coverage and allocation of resources as a two-phase decision process with uncertainty of the place and time of occurrence. Using Crime data in Chicago and applying the FLEET-ICt [de Oliveira, 2020] model, its objective is to determine which facilities to use and how to allocate specialists. Five samples of 1,000 lines were generated and optimized. For all samples the model was able to achieve about 90% coverage in a few seconds. The model was also adapted to minimize the cost given a minimum coverage, which proved more computationally more complex, taking up to 20 minutes. It was found that there are solutions with the same coverage as the maximum coverage that have a lower cost.

KEYWORDS. Maximum Coverage. Minimum Cost. Spatio-Temporal Data. Paper topics: Combinatorial Optimization, Public Defense, Big Data

1. Introdução

O custo de tomar uma decisão sob incerteza pode ser minimizado utilizando técnicas de programação matemática. Este trabalho tem o objetivo otimizar o problema de tomada de decisão em duas etapas de cobertura e alocação de recursos.

O problema a ser tratado é o de resposta à diferentes crimes na cidade de Chicago. Assim, a primeira etapa de decisão é a escolha de instalações para agentes de defesa pública de diferentes especialidades trabalharem. A segunda decisão é onde e quando colocá-los. O modelo então procura maximizar a cobertura dado que a localização e turno dos crimes é incerto. O modelo de decisão é discutido na seção 2.

O estudo foi feito considerando dados espaço-temporais para cinco diferentes tipos de crime na cidade de Chicago (US) em Julho dos anos 2012 à 2017 [Currie32, 2017]. A partir da cobertura máxima encontrada, também é feito um estudo para entender como o custo varia em função da cobertura. A descrição dos dados é apresentada na seção 2.1.

Assim, a primeira modelagem explorada é a de cobertura máxima.

1.1. Cobertura Máxima

O problema de cobertura máxima já é bastante explorado na literatura. O modelo possui o objetivo de maximizar a demanda coberta dado que há um limite para a quantidade de instalações e que a demanda só pode ser considerada coberta caso esteja a uma distância máxima da instalação. Uma modelagem possível encontra-se abaixo [Church e ReVelle, 1974].

$$\begin{aligned} \text{MAX } z &= \sum_{i \in I} a_i \cdot y_i \\ \sum_{i \in N_i} x_i &\geq y_i \forall i \in I \\ \sum_{j \in J} x_j &\leq P \\ x_j &= (0, 1) \forall j \in J \\ y_i &= (0, 1) \forall i \in I \end{aligned}$$

onde

I = nós de demanda

J = nós de instalação

S = distância a partir da qual o ponto é considerado não coberto

$d_{i,j}$ = menor distância entre i e j

$x_j = 1$ se instalação existe em j e 0, caso contrário

$N_i = \{j \in J | d_{i,j} \leq S\}$

a_i = população a ser servida no ponto de demanda i

P = número de possíveis instalações

Posteriormente o modelo acima foi revisitado para considerar a necessidade de equipamentos específicos (aplicado ao batalhão de bombeiros de Baltimore) para cada demanda, recebendo o nome de FLEET [Schilling et al., 1979]. A partir dessa consideração, o modelo foi aplicado

ao desafio de localização de ambulâncias em Minas Gerais, Brasil, permitindo também que diferentes quantidades e tempos de resposta fossem considerados para o recurso em questão, ambulâncias [Guimarães, 2018].

Por fim, também aplicado à máxima cobertura de ambulâncias, uma nova dimensão para a variável de decisão do material foi adicionada: o tempo. Assim o modelo, FLEET-ICt (*Time Facility Location and Equipment Emplacement Technique with Independent Coverages*), encontra-se abaixo, procura escolher o local das instalações e onde alocar seus recursos no tempo [de Oliveira, 2020].

$$\begin{aligned}
 \text{MAX} \quad & \sum_{i \in I, u \in U, t \in T} d_{i,u}^t \cdot y_{i,u}^t \\
 \text{s.a.} \quad & \sum_{j \in N_{i,j}^u} x_{j,u}^t \geq y_{i,u}^t \quad \forall i \forall u \forall t \\
 & \sum_{j \in J} x_{j,u}^t \leq P_u \quad \forall u \forall t \\
 & \sum_{u \in U} x_{j,u}^t \leq C_j \cdot z_j \quad \forall j \forall t \\
 & x_{j,u}^t \leq z_j \quad \forall j \forall u \forall t
 \end{aligned}$$

onde

I = conjunto de pontos de demanda

J = conjunto de candidatos a instalação

T = conjunto de períodos

U = conjunto de tipos de recurso

P_u = quantidade de recurso $u \in U$ disponível

S_u = distância máxima para recurso $u \in U$

C_j = capacidade que instalação $j \in J$ pode receber

$r_{j,i}$ = distância entre instalação $j \in J$ e ponto de demanda $i \in I$

$N_{i,j}^u$ = conjunto de instalações j que podem servir ponto de demanda $i \in I$ dentro da distância máxima do recurso $u \in U$

$d_{i,u}^t$ = demanda no ponto $i \in I$ pelo recurso $u \in U$ no tempo $t \in T$

com variáveis

$z_j = 1$ se a instalação $j \in J$ estiver ativa e 0, caso contrário

$x_{j,u}^t = 1$ se há na instalação $j \in J$ o recurso $u \in U$ no momento $t \in T$ e 0, caso contrário

$y_{i,u}^t = 1$ se o ponto de demanda $i \in I$ está coberto pelo recurso $u \in U$ no tempo $t \in T$ e 0 caso contrário

Além da localização da instalação e da demanda, podemos considerar a dependência da demanda por um recurso específico (por exemplo, um especialista em crimes relacionados a narcóticos) no momento no tempo em que essa demanda é feita.

1.2. Custo Mínimo

Um outro lado para o problema é minimizar o custo enquanto uma cobertura mínima arbitrária é mantida. Isso porque entende-se que os recursos disponíveis - capital para a construção de instalações e para alocação de especialistas - são limitados.

Para isso a função objetivo é alterada para minimizar o custo da operação, dado que uma cobertura mínima é alcançada.

$$\begin{aligned}
 \text{MIN} \quad & \sum_{j \in J} c_j \cdot z_j + \sum_{u \in U, j \in J, t \in T} x_{j,u}^t \cdot c_u \\
 \text{s.a.} \quad & \sum_{j \in N_{i,j}^u} x_{j,u}^t \geq y_{i,u}^t \quad \forall i \forall u \forall t \\
 & \sum_{j \in J} x_{j,u}^t \leq P_u \quad \forall u \forall t \\
 & \sum_{u \in U} x_{j,u}^t \leq C_j \cdot z_j \quad \forall j \forall t \\
 & x_{j,u}^t \leq z_j \quad \forall j \forall u \forall t \\
 & \frac{\sum_{i \in I, u \in U, t \in T} d_{i,u}^t \cdot y_{i,u}^t}{\sum_{i \in I, u \in U, t \in T} d_{i,u}^t} \geq \text{cov}
 \end{aligned}$$

onde

I = conjunto de pontos de demanda

J = conjunto de candidatos a instalação

T = conjunto de períodos

U = conjunto de tipos de recurso

P_u = quantidade de recurso $u \in U$ disponível

S_u = distância máxima para recurso $u \in U$

C_j = capacidade que instalação $j \in J$ pode receber

$r_{j,i}$ = distância entre instalação $j \in J$ e ponto de demanda $i \in I$

$N_{i,j}^u$ = conjunto de instalações j que podem servir ponto de demanda $i \in I$ dentro da distância máxima do recurso $u \in U$

$d_{i,u}^t$ = demanda no ponto $i \in I$ pelo recurso $u \in U$ no tempo $t \in T$

com variáveis

$z_j = 1$ se a instalação $j \in J$ estiver ativa e 0, caso contrário

$x_{j,u}^t = 1$ se há na instalação $j \in J$ o recurso $u \in U$ no momento $t \in T$ e 0, caso contrário

$y_{i,u}^t = 1$ se o ponto de demanda $i \in I$ está coberto pelo recurso $u \in U$ no tempo $t \in T$ e 0 caso contrário

c_j = custo de abrir a instalação $j \in J$

c_u = custo de alocar um especialista no crime $u \in U$

É importante notar que trata-se de um problema combinatório e, portanto, existem diversas soluções para uma mesma cobertura. Isso significa que, utilizando o resultado de cobertura

máxima encontrada, outra solução com a mesma cobertura pode ser encontrada tendo um custo inferior. Isso será explorado na seção de 3.2

Finalmente, para resolver o sistema apresentado, FLEET-ICT [de Oliveira, 2020], foi utilizada a linguagem Julia [Bezanson et al., 2017] com o pacote JuMP [Dunning et al., 2017] e o solver GLPK [Makhorin, 2010].

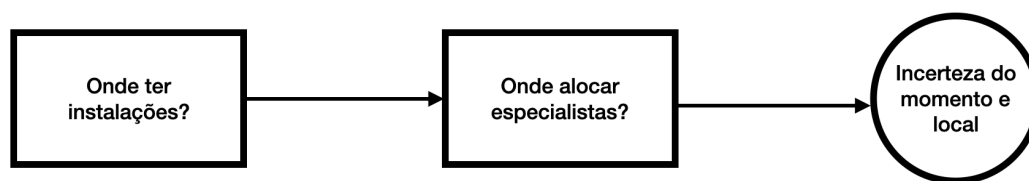
Por fim, o presente trabalho não se propõe a explorar padrões nos dados, mas sim compreender como modelar a tomada de decisão a partir de programação matemática.

2. Modelagem

Após explorar os modelos de programação matemática, é importante também entender quais variáveis existem e como se relacionam. Há, para o problema, cinco principais aspectos para a decisão:

1. Quais são as localizações viáveis;
2. Quantos e quais especialistas estão disponíveis;
3. Onde o crime ocorreu;
4. Quando o crime ocorreu; e
5. Qual é o tipo do crime.

Assim, o processo de decisão sob incerteza tratado nesse trabalho ocorre em duas fases: a primeira é decidir onde serão feitas as instalações enquanto a segunda é onde e quando alocar os especialistas. Essa estrutura de decisão é então modelada para maximizar a cobertura de crimes que ocorrem de forma incerta, explorado na seção 1.1, e posteriormente para minimizar o custo de uma dada cobertura, seção 1.2.



2.1. Dados

Os dados utilizados nesse trabalhos vêm de uma base com quase 1.5 milhão de entradas, e suas primeiras linhas podem ser observadas na tabela 1.

Os dados foram então filtrados seguindo os pontos abaixo:

- não resultaram em prisão (Arrest=False);
- foram no mês de Julho (Month=7), por ser o mês com mais ocorrências, figura 1, dos anos 2012 à 2017; e

	Date	Primary Type	Arrest	Location
0	2016-05-03 23:40:00	BATTERY	True	(41.864, -87.707)
1	2016-05-03 21:40:00	BATTERY	False	(41.783, -87.604)
2	2016-05-03 23:31:00	PUBLIC PEACE VIOLATION	False	(41.895, -87.758)
3	2016-05-03 22:10:00	BATTERY	False	(41.886, -87.750)
4	2016-05-03 22:00:00	THEFT	False	(41.886, -87.762)

Tabela 1: Cinco primeiras linhas do conjunto de dados

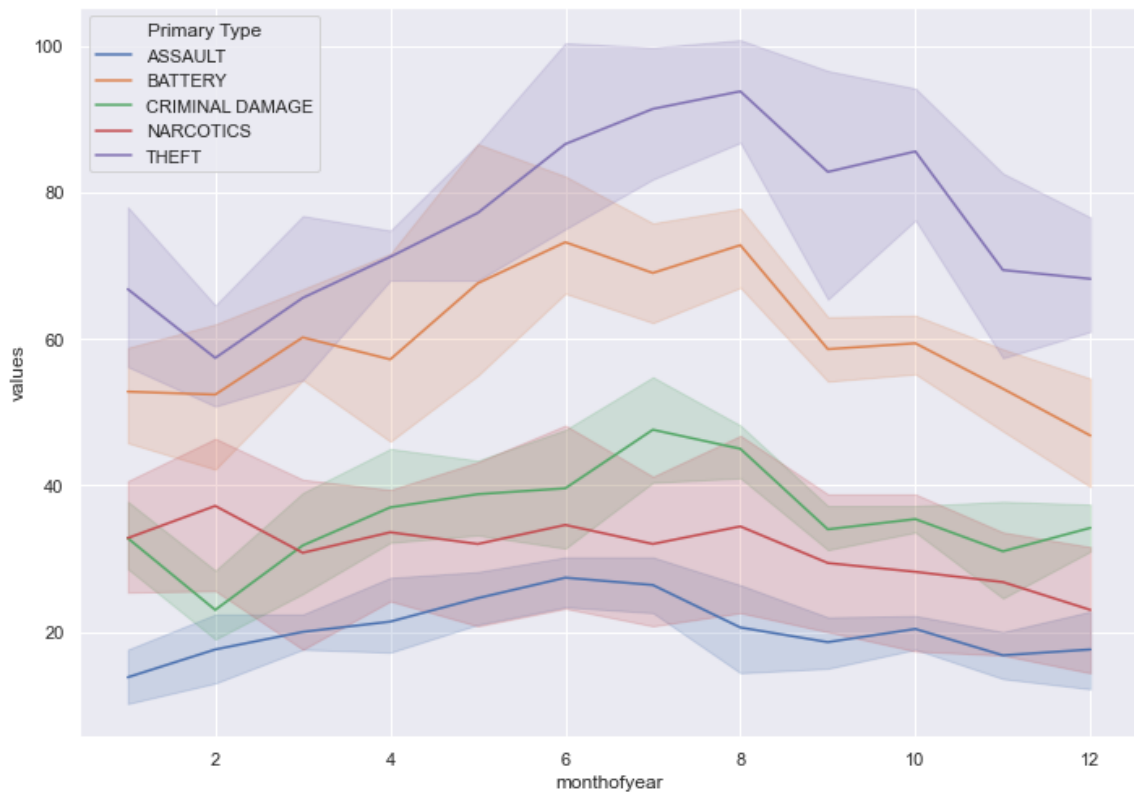


Figura 1: Série temporal mensal dos crimes de 2012 à 2017

- que façam parte dos mais cometidos, cuja proporção na amostra está na tabela 2.

Após a filtragem inicial dos dados, foram selecionadas 10.000 linhas para a geração de cinco sub amostras de 1.000 entradas com substituição utilizando a função de *sample* da biblioteca de análise de dados PANDAS [pandas development team, 2020] com python [Van Rossum e Drake, 2009], com o argumento *weights* para manter a proporção entre crimes .

Além disso, para colocar os dados no formato requerido, a demanda foi considerada binária, isto é: $d_{i,u}^t$ é uma matriz esparsa, indicando 1 para um crime em $i \in I, u \in U, t \in T$ e 0, caso contrário.

Primary Type	Percentage
THEFT	41.52%
BATTERY	28.66%
CRIMINAL DAMAGE	20.26%
ASSAULT	9.47%
NARCOTICS	0.09%

Tabela 2: Proporção de crimes na amostra

Já os valores de especialistas disponíveis, custo por turno, distância máxima de resposta e custo por localização não estavam disponíveis, foram todos escolhidos arbitrariamente.

Para isso, procurou-se estimar diferenças de experiência dada a ocorrência dos crimes, além do preço de imóveis na cidade.

Assim, os valores procuram representar uma certa ordem: sendo roubo o crime mais comum, há mais especialistas do que de outros tipos; sendo assalto um tipo de crime cujo tempo de resposta é crítico, seu valor é inferior ao de dano de propriedade, que exige menos rapidez; etc..

tipo	disponível	distância	custo
THEFT	50	3	300,00
BATTERY	20	30	500,00
CRIMINAL DAMAGE	30	15	800,00
ASSAULT	30	5	500,00
NARCOTICS	15	10	1.000,00

Tabela 3: Disponibilidade, distância máxima e custo por especialidade

Para estimar o custo de uma instalação, entendeu-se que quanto mais distante do centro de Chicago, menos custoso é. Assim, um valor fixo de 10.000,00 foi escolhido - equivalente a 10 vezes o custo de um especialista em narcóticos -, e 1.000 para cada quilômetro de distância entre a instalação e o centro da cidade. No entanto, os resultados aqui presentes foram encontrados com custos seguindo o oposto do defendido acima, dado um problema encontrado no código.

Para a escolha dos possíveis locais para instalações, foi feita uma amostra de 50 localizações de crimes para serem as localizações das instalações. As capacidades, por sua vez, são aleatórias, sendo um inteiro uniformemente distribuído entre 30 e 100.

Por fim, o tempo foi discretizado em 3 turnos de 8 horas cada, sendo eles de 0h às 8h, 9h às 16h e 17h às 23h.

É importante considerar que os valores dos custos, embora utilizados no modelo, não possuem valor real, sendo todos valores arbitrários para permitir que o problema seja otimizado.

3. Resultados

3.1. Cobertura

Abaixo está a relação entre a amostra, cobertura atingida e custo. Observa-se que há várias instalações são utilizadas repetidamente para a solução da amostra.

amostra	cobertura	custo	instalações
1	92.30	682,082.05	0, 1, 3, 4, 5, 6, 8, 9, 10, 11, 13, 14 17
2	91.20	760,336.31	0, 1, 2, 3, 4, 8, 9, 10, 11, 13, 14, 16, 17
3	91.60	693,510.67	3, 4, 8, 9, 10, 11, 13, 14, 16, 17, 18, 21, 22, 25
4	91.40	668,539.55	1, 3, 4, 7, 8, 9, 10, 11, 12, 13, 14, 16, 18
5	89.40	691,892.83	1, 2, 3, 4, 5, 8, 9, 10, 13, 14, 19, 20, 21

A imagem 2 mostra a localização escolhida para a amostra número 1, com os pontos de demanda para o primeiro turno.

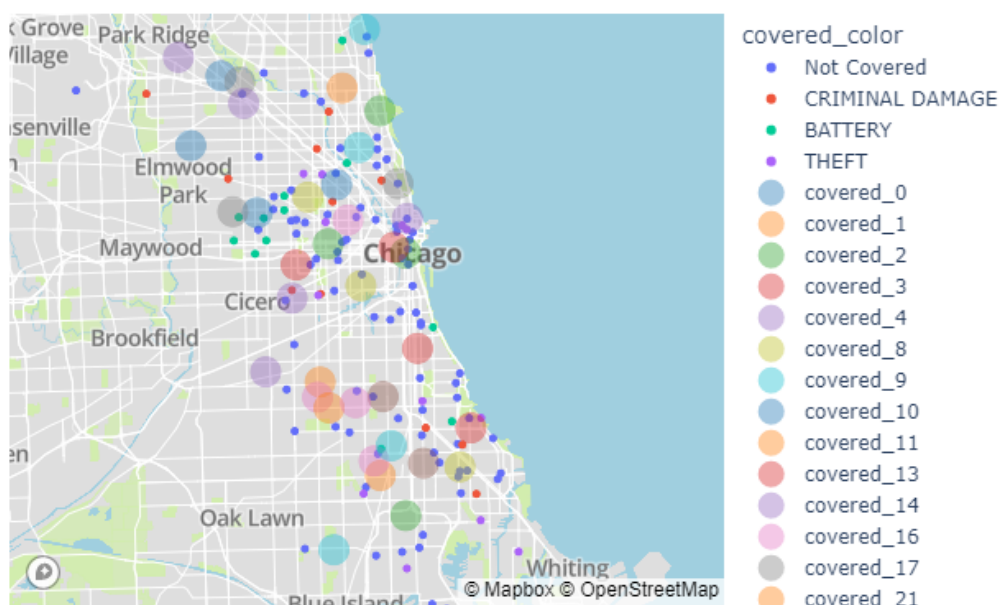


Figura 2: instalações e demandas da amostra 1

3.2. Custo

Após encontrada a cobertura máxima, foi também feito um estudo de como o custo é influenciado por ela. A imagem 3 mostra essa relação. Há dois pontos interessantes a serem notados: o primeiro é que o custo parece aumentar não linearmente em função da cobertura; a segunda é que, encontrada a melhor cobertura, é possível encontrar uma solução com a mesma cobertura por um custo inferior.

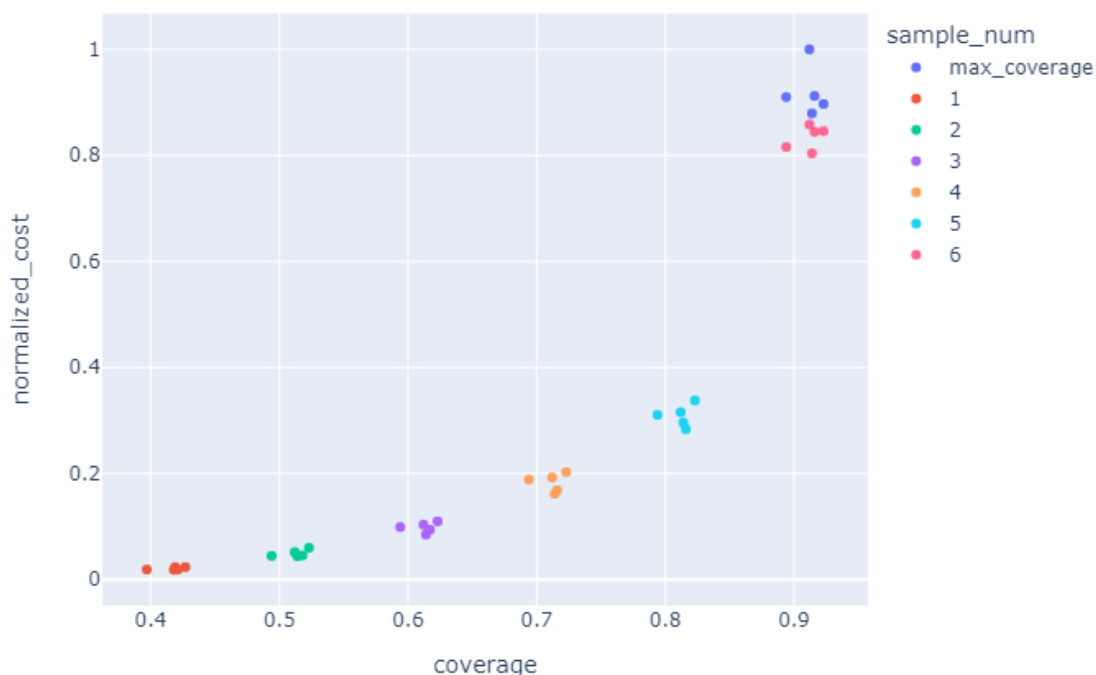


Figura 3: custo por cobertura

4. Conclusão

Um estudo sobre otimização da decisão em duas etapas dada a incerteza do local e momento do crime foi apresentado. Para isso foi utilizado o modelo FLEET-ICt [de Oliveira, 2020] com dados de Crimes de Chicago (US) [Currie32, 2017].

O problema foi otimizado para cobertura máxima - tendo encontrado uma solução extremamente rápido, possivelmente porque a matriz de demanda é uma matriz esparsa e binária - e seu resultado foi utilizado para entender como custo de uma solução varia de acordo com a cobertura.

Algumas limitações desse projeto são importantes de serem destacadas:

- Não foi considerado o tempo de deslocamento entre uma instalação e outra;
- Não foi considerado o descanso dos especialistas (o que pode ser levado em conta dividindo o número total de especialistas pelo número de turnos);
- Não foi alterado o tamanho da amostra, que poderia ser utilizado para entender quais instalações são colocadas e onde está o ponto de convergência; e
- O custo computacional para otimizar custo mínimo é bastante superior ao custo necessário para cobertura máxima, limitando a quantidade de cenários que puderam ser utilizados.

Trabalhos futuros podem explorar como estimar os valores discutidos na seção 2.1. Também há valor em prever o momento e espaço do objeto de estudo, trazendo luz para padrões nos dados que não foram explorados no presente trabalho.

5. Anexo

Todo código utilizado no projeto está disponível em https://github.com/felipewhitaker/decision_and_risk_analysis.

Referências

- Bezanson, J., Edelman, A., Karpinski, S., e Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98. URL <https://epubs.siam.org/doi/10.1137/141000671>.
- Church, R. e ReVelle, C. (1974). The maximal covering location problem. *Papers of the Regional Science Association*, 32(1):101–118. URL <https://doi.org/10.1007/bf01942293>.
- Currie32 (2017). Crimes in chicago. Retrieved 26/06/2022 from <https://www.kaggle.com/datasets/currie32/crimes-in-chicago>.
- de Oliveira, C. P. (2020). Modelos de otimização aplicados ao problema de máxima cobertura: Estudo de caso do samu-bh. Master's thesis, CEFET-MG. Accessed 18/06/2022 at https://sig-arquivos.cefetmg.br/arquivos/2020157098a09e2498076fd7bdf5ac24e/DISSERTAO_CHARLES_PAULINO_DE_OLIVEIRA.pdf.
- Dunning, I., Huchette, J., e Lubin, M. (2017). Jump: A modeling language for mathematical optimization. *SIAM Review*, 59(2):295–320.
- Guimarães, M. M. (2018). Modelos mono e multiobjetivo para o problema de localização de máxima cobertura: estudo de caso samu-bh. Master's thesis, CEFET-MG. Accessed 27/06/2022 at https://sig-arquivos.cefetmg.br/arquivos/20181021070a25203757545d56286a9f8/dissertacaoVersaoFinal_MarianaMGuimaraes.pdf.
- Makhorin, A. *GNU Linear Programming Kit*. GNU Software Foundation, 2010. URL <http://www.gnu.org/software/glpk/glpk.html>.
- pandas development team, T. (2020). pandas-dev/pandas: Pandas. URL <https://doi.org/10.5281/zenodo.3509134>.
- Schilling, D., Elzinga, D. J., Cohon, J., Church, R., e ReVelle, C. (1979). The team/fleet models for simultaneous facility and equipment siting. *Transportation Science*, 13(2):163–175. URL <https://doi.org/10.1287/trsc.13.2.163>.
- Van Rossum, G. e Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA. ISBN 1441412697.