



Modelagem Computacional no Sertão Mineiro



PPGMCS
Programa de Pós-Graduação em
Modelagem Computacional e Sistemas

Distribuição de Recursos para Minimização do Coeficiente de Gini entre Localidades e Critérios

Felipe Costa¹ – costinha.fca@gmail.com

Aline de L. Z. Lunkes² - alinelunkesazl@gmail.com

Marcos Flávio Silveira Vasconcelos D'Angelo³ - marcos.dangelo@unimontes.br

Matheus Pereira Libório³ – m4th32s@gmail.com

Petr Iakovlevitch Ekel¹ – petr.ekel2709@gmail.com

¹ PUC Minas, Belo Horizonte, MG, Brazil

² Universidade Federal do Pampa, UNIPAMPA - Alegrete, RS, Brazil

³ Universidade Estadual de Montes Claros, UNIMONTES - Montes Claros, MG, Brazil

Abstract. Esta pesquisa avalia a viabilidade de uma distribuição minuciosa de recursos para o problema da desigualdade. O problema consiste em dividir uma quantidade fixa de recursos arrecadados entre um conjunto de localidades, de modo que essas localidades possam investir em áreas como saúde e educação. Propõe-se uma forma de distribuir esses recursos priorizando o acesso a eles, de modo que o planejamento de gastos pós-distribuição seja considerado já no momento da alocação para cada localidade.

Essa proposta envolve a resolução de um problema multicritério que, devido à natureza da vetorização exigida, cresce de forma impraticável em tamanho real. Por esse motivo, são estudados métodos heurísticos alternativos capazes de resolver o problema em uma base de tempo razoável; essas soluções são comparadas por meio de métricas de qualidade e confiabilidade, considerando a quantidade de recursos disponível e o nível de detalhamento dos dados.

Discute-se também uma característica do problema — a independência entre as funções objetivo — que pode ser explorada para reduzir o tempo de execução da resolução completa, aproximando-o do tempo necessário para resolver um problema tradicional de menor ordem.

Keywords: Coeficiente de Gini, Otimização Multicritério, Alocação de Recursos, Heurísticas, Desigualdade

1. Introdução

Em diversos cenários nos quais é possível quantificar recursos, observa-se desigualdade em sua distribuição. O coeficiente de Gini é uma métrica tradicionalmente utilizada para quantificar

esse fenômeno (Gini, 1912; Lorenz, 1905). Com a arrecadação de recursos públicos, torna-se um objetivo do Estado distribuir essa renda de forma eficiente para melhorar a qualidade de vida da população — objetivo que, classicamente, é aplicado à distribuição de renda.

Na proposta apresentada, o Gini é utilizado para o cálculo da desigualdade na distribuição de recursos entre localidades, o que torna o problema essencialmente diferente da aplicação clássica. Como o problema envolve várias métricas que precisam ser otimizadas simultaneamente — normalmente com interesses conflitantes —, ele é caracterizado como um problema de otimização multicritério.

Os critérios utilizados são decididos a partir do propósito dos recursos a serem distribuídos, sendo necessário encontrar uma resposta harmônica em que os objetivos sejam otimizados de forma balanceada. No caso de saneamento básico, por exemplo, poderiam ser escolhidos critérios como acesso à rede de esgoto, acesso à água limpa e investimento familiar em produtos de higiene pessoal, entre outros. Já no caso de escolaridade, poderiam ser utilizados critérios como o nível de escolaridade de cada membro da família e o número de filhos, entre outras métricas.

Após a definição desses critérios, é necessária uma estimativa de custo para a provisão dos novos recursos, que pode ser obtida a partir de pesquisas e planejamentos governamentais. Com essas informações, é possível calcular o coeficiente de Gini para cada critério e prever a quantidade de recursos necessária para influenciar cada critério individualmente. Como resultado, torna-se possível uma otimização mais precisa, capaz de distribuir recursos de forma específica para cada critério e localidade, atendendo aos diferentes objetivos sociais do projeto.

2. Formulação do Problema

Devido às características do problema, combinadas à necessidade de vetorização dos dados de entrada, o custo computacional cresce com o produto entre o número de localidades estudadas e o número de critérios escolhidos: o tempo de execução das otimizações é quadrático em relação a esse produto, o que corresponde a uma ordem de quarto grau em relação às dimensões originais do problema.

Por esse motivo, considera-se a utilização de métodos heurísticos, com o objetivo de encontrar uma resposta satisfatória em uma ordem de complexidade menor.

A proposta deste estudo é, portanto, testar e comparar diferentes formas de simplificar a resolução do problema e avaliar sua efetividade. Os métodos aplicados são a otimização multicritério e métodos heurísticos de distribuição, seguidos de otimizações locais tradicionais.

2.1 Definição Matemática

Considerando uma região (estado ou país) dividida em localidades $L = \{l_i\}$, cada localidade é avaliada em relação a um conjunto de critérios j , sendo t o número total de locais pesquisados em cada localidade. O dado coletado l_{ij} representa o número de locais que possuem acesso ao critério j . Com esses dados, o critério normalizado é dado por:

$$D = \{d_{ij} \mid d_{ij} = l_{ij}/t\}$$

Seja p_{ij} o custo estimado para prover o critério j a uma pessoa na localidade i . A matriz de custo total de universalização é:

$$P = \{p_{ij} \mid p_{ij} = t \cdot c_{ij}\}$$

Seja R o total de recursos disponíveis e $X = \{x_{ij}\}$ a matriz de alocação de recursos a ser determinada (mesma dimensão de D e P); X/P representa a fração de universalização financiada em cada célula. Seja $G(\cdot)$ o coeficiente de Gini aplicado a um vetor. Os objetivos, em ordem lexicográfica de prioridade, são:

$$\forall j : \operatorname{argmin}_X G_j\left(D + \frac{X}{P}\right) \rightsquigarrow G_j\left(D + \frac{X}{P}\right) = 0 \quad (1)$$

$$\forall i : \operatorname{argmin}_X G_i\left(D + \frac{X}{P}\right) \rightsquigarrow G_i\left(D + \frac{X}{P}\right) = 0 \quad (2)$$

$$\forall i, j : \operatorname{argmax}_X d_{ij}\left(D + \frac{X}{P}\right) \rightsquigarrow d_{ij}\left(D + \frac{X}{P}\right) = 1 \quad (3)$$

sujeito às restrições:

$$\sum_{i,j} x_{ij} \leq R \quad (4)$$

$$0 \leq x_{ij} \leq p_{ij} \cdot (1 - d_{ij}) \quad (5)$$

Esses objetivos correspondem, respectivamente: ao coeficiente de Gini aplicado a cada critério entre todas as localidades, dado que o problema está relacionado à desigualdade entre os locais; ao coeficiente de Gini aplicado a cada localidade entre os critérios definidos, dada a importância de não negligenciar nenhum critério escolhido para a pesquisa; e à maximização de cada sub-indicador individualmente, o que permite ao método melhorar a qualidade de vida geral após a redução da desigualdade nos dados.

Uma vez encontrada a resposta, é possível gerar previsões sobre os resultados esperados. Essas previsões podem ser usadas para a criação de políticas públicas, bem como para conferir posteriormente a porcentagem do objetivo efetivamente cumprida pelas localidades beneficiadas.

A melhora esperada M e a previsão final F são calculadas por:

$$M = \frac{X}{P}, \quad F = D + M$$

3. Metodologia

3.1 Avaliação Multicritério

Para aplicar o algoritmo de otimização multicritério, os dados são vetorizados: os objetivos de Gini precisam operar sobre subconjuntos de um vetor único que representa toda a tabela $i \times j$. Define-se o *Gini vertical* $G_j(X)$, calculado sobre o intervalo de índices $[j \cdot i, j \cdot (i + 1)]$, e o *Gini horizontal* $G_i(X)$, calculado sobre índices espaçados $\{\dots, i + kj, \dots\}$, $\forall k \in \mathbb{Z}$.

Como o objetivo horizontal (Equação 2) só é relevante após o objetivo vertical (Equação 1) ser satisfeito, a otimização é feita apenas com G_j . A inclusão de G_i precisa de modificações no método como descrito já que só podemos considerar estes objetivos após garantirmos G_j como concluído.

O terceiro objetivo (Equação 3) não precisa ser explicitamente especificado, pois todas as otimizações realizadas já distribuem os recursos em sua totalidade.

3.2 Avaliação Heurística

Como alternativa de menor custo computacional, calculam-se pesos iniciais $W = \{w_k\}$ e distribuem-se os recursos por meio do método de pesos ponderados, adaptado da abordagem de decisão multicritério de Ekel et al. (2020), em que o total recebido por cada critério ou localidade é dado por:

$$S_k = R \cdot \frac{w_k}{\sum_k w_k}$$

Após o cálculo da quantidade de recursos para cada objeto, são realizadas k otimizações locais limitadas a S_k .

Heurística Vertical. Realiza a distribuição por critério; em seguida é feita a otimização vertical do Gini. Assume-se que os valores faltantes para cada critério são proporcionais à quantidade de recursos recebida. Os pesos iniciais verticais, calculados incluindo o custo, são dados por:

$$W = \{w_j\} = \sum_i p_{ij} \cdot (1 - d_{ij})$$

Heurística Horizontal. Realiza a distribuição por localidade; em seguida é feita a otimização horizontal do Gini. Assume-se que os valores faltantes para cada localidade são proporcionais à quantidade de recursos recebida. Os pesos horizontais iniciais, calculados incluindo o custo, são dados por:

$$W = \{w_i\} = \sum_j p_{ij} \cdot (1 - d_{ij})$$

4. Experimentos

A base de dados é gerada sinteticamente: cada localidade recebe um nível latente de desenvolvimento (sorteado de duas distribuições normais inversamente correlacionadas), a partir do qual a cobertura D_{ij} e o custo P_{ij} de cada critério são derivados com ruído independente por critério. Isso reproduz a correlação negativa entre acesso e custo esperada na prática: localidades isoladas apresentam simultaneamente menor cobertura e maior custo de universalização, por estarem distantes da fonte dos recursos.

O objetivo de longo prazo é resolver o problema em até 300 localidades e 30 critérios; nos experimentos preliminares deste artigo, usou-se uma escala reduzida ($m = 22$, $n = 6$), suficiente para comparar os três métodos sem o custo computacional da escala máxima. A otimização foi testada em diferentes frações do máximo de recursos possível, $R_t = \sum_i \sum_j (1 - D_{ij}) \cdot P_{ij}$ (valor que levaria toda a tabela D à cobertura total, inviável na prática): orçamentos de 10%, 30% e 60% de R_t , com 5 repetições (seeds distintas) cada, totalizando 45 execuções (3 métodos $\times$ 3 orçamentos $\times$ 5 repetições).

Os três métodos foram implementados em Python: o multicritério utiliza otimização do relógio também implementado em Python; as heurísticas utilizam uma estratégia de pesos ponderados — distribuindo recursos inicialmente proporcional a quantidade de recursos

faltantes. Métricas coletadas: Gini horizontal e vertical médios, cobertura média final e tempo de execução.

Os dados gerados inicialmente vem de uma única distribuição normal por critério o que acaba causando algumas estranhezas quando otimizadas pelas heurísticas,

5. Resultados

Os resultados indicam a existencia de 3 regiões notáveis, sendo elas:

1. Região 1 - Caracterizada pela falta de recursos para resolver o problema do gini vertical. Nesta região a qualidade da heurística vertical cresce rapidamente.
2. Região 2 - Caracterizada pela falta de recursos para resolver o segundo objetivo principal. Nesta região ambas as heurísticas começam a apresentar uma melhora lenta.
3. Região 3 - Caracterizada pela capacidade de resolver o problema do gini vertical e horizontal. Nesta região as heurísticas melhoram rapidamente.

Table 1: Resultados comparativos entre os métodos.

Orçamento	Método	$\tilde{G}_j$	$\tilde{G}_i$	Cobertura	Tempo (minutos)
0%	Todos	0,1218	0,3472	0,538	00,00
10%	Multicritério	0,0161	0,3109	0,576	287,75
	Heurística Vertical	0,0871	0,3353	0,580	13,91
	Heurística Horizontal	0,0780	0,3472	0,622	6,43
30%	Multicritério	0,0018	0,2435	0,661	688,10
	Heurística Vertical	0,0608	0,3136	0,667	27,45
	Heurística Horizontal	0,0504	0,3472	0,726	78,72
60%	Multicritério	≈ 0	0,1691	0,787	511,62
	Heurística Vertical	0,0511	0,2843	0,799	32,56
	Heurística Horizontal	0,0474	0,3466	0,858	52,35

Como observado no teste de 10% de recursos, nem sempre alcança um resultado adequado. Principalmente quando há poucos recursos há uma tendência da normalização não conseguir otimizar o gini completamente mesmo utilizando todos os recursos em uma coluna; Com isso durante a execução essa coluna não procura uma solução satisfatória.

6. Discussão

Uma característica notável do problema é a independência entre as funções objetivo, que pode ser aproveitada para modificar o processo de otimização multicritério, desde que se garanta que todos os objetivos continuem sendo otimizados. O fato dos objetivos serem independentes permite a otimização pela quantidade de passos obviamente inúteis. Devido ao objetivo da

otimização multi critério olhar o objetivo mínimo¹ qualquer direção que não inclua o objetivo mínimo não precisa ser testada.

Mesmo conseguindo ajudar consideravelmente o problema da desigualdade utilizando métodos heurísticos observa-se pelos resultados da Seção 5 ajudam a explicar por que a Heurística Vertical, apesar de rápida é inadequada, quando o objetivo prioritário é a equidade *entre localidades* (Gini vertical): ao distribuir os recursos inicialmente fornece-se muitos recursos para critérios com magnitude baixa, o que a torna incapaz de resolver o Gini vertical para todos os critérios mas sistematicamente pior no horizontal. A Heurística Horizontal, que ataca diretamente o objetivo de segunda maior prioridade lexicográfica (Equação 2), entregou Ginis comparáveis à heurística vertical a um custo computacional maior.

Ainda assim a otimização multi critério é claramente superior o que motiva trabalhos futuros — onde a melhor opção é reinventar o método multi critério onde incluindo-se a independência das funções objetivo para ignorar direções irrelevantes e a remoção da normalização dos critérios pois o gini já é uma variável que existe no intervalo de 0 a 1 e pode ser utilizada como é, tornando-o competitivo também em tempo de execução à medida que m e n crescem em direção à escala-alvo de 300 localidades e 30 critérios.

7. Conclusão

A principal contribuição deste artigo é a formulação do problema de alocação de recursos como uma otimização multicritério lexicográfica baseada no coeficiente de Gini (horizontal e vertical), acompanhada de uma comparação empírica entre essa formulação de referência e duas heurísticas de menor custo computacional. Nos experimentos preliminares ($m = 22$, $n = 6$), a Heurística Vertical e Horizontal são comparáveis, com tempo de execução ordens de grandeza menor, enquanto o método multi critério obtém um resultado claramente superior.

As principais limitações deste trabalho são: (i) experimentos em escala reduzida e com dados sintéticos, não em base real de indicadores sociais; (ii) o multicritério foi implementado com solver de propósito geral e pesos fixos, o que pode não representar seu desempenho ótimo; e (iii) apenas 5 repetições por configuração. Como trabalhos futuros, pretende-se aplicar a redução de complexidade da Seção 6 ao método multicritério, testar a formulação com dados reais de indicadores sociais municipais, utilizar uma linguagem mais eficiente e ampliar a escala dos experimentos em direção ao tamanho-alvo do problema.

Acknowledgements

Os autores agradecem à CAPES, e a FAPEMIG (grant n OET-00243-26).

REFERENCES

Gini, C. (1912), *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche*, Studi Economico-giuridici, Facoltà di Giurisprudenza della R. Università di Cagliari, Bologna, Tipografia di P. Cuppini.

¹[https://soclibrary.futa.edu.ng/books/WitoldPedrycz,PetrEkel,RobertParreiras\(auth\(BookZZ.org\).pdf](https://soclibrary.futa.edu.ng/books/WitoldPedrycz,PetrEkel,RobertParreiras(auth(BookZZ.org).pdf)

- Lorenz, M.O. (1905), “Methods of Measuring the Concentration of Wealth”, *Publications of the American Statistical Association*, vol. 9, n. 70, 209-219.
- Ekel, P., Pedrycz, W. and Pereira Jr., J. (2020), *Multicriteria Decision-Making Under Conditions of Uncertainty: A Fuzzy Set Perspective*, John Wiley & Sons, Chichester.
- Libório, M.P., Fusco, E., Diniz, A.M.A., Martinuci, O.S. and Ekel, P.I. (2024), “A Novel Approach for Multispatial and Multitemporal Analysis of Composite Indicators”, *Social Indicators Research*, vol. 173, n. 3, 783-800.