

Implementação do Modelo Matemático para Cplex para Identificação de Acadêmicos Promissores de Sucesso Científico através de Programação Linear Inteira

Matheus Breguêz

UFV

Rod. MG 230 Km 08 - Rio Paranaíba/MG

matheusrv@email.com

Tales Nascimento Lúcio

UFV

Rod. MG 230 Km 08 - Rio Paranaíba/MG

tales.lucio@ufv.br

1. Introdução

O trabalho consiste na análise da aplicação do problema da mochila com restrições de cardinalidade na identificação de acadêmicos com potencial de sucesso científico através da programação linear inteira e sua implementação em Simplex. Através do uso da base de dados do ArnetMiner, que mantém um registro acerca de todas as informações de um banco de dados que registra acadêmicos desde 1980 para seleção de profissionais promissores na área acadêmica. “fatores quantitativos que dirigem artigos publicados a aumentarem o índice-h futuro de seus autores” [Dong et al., 2016]. “Colaborações de jovens pesquisadores a pesquisadores importantes impactam na carreira futura de um pesquisador e através de fatores quantitativos é possível destacar o crescimento da carreira desses pesquisadores” [Zhou et al., 2018]. O objetivo do modelo é o de detectar de maneira precoce futuros acadêmicos de sucesso a partir de seus impactos científicos a curto prazo avaliados por meio de indicadores derivados de citação.

2. Descrição do Problema

“A primeira das falhas de todo júri de arte é a precariedade dos critérios de julgamento. Essa precariedade provém de duas heterogeneidades irreduzíveis: a das personalidades julgadoras em presença e a das obras a julgar. Esses julgadores não são só individualidades em diferentes graus de cultura, experiência e saturação, são também pegados ao acaso e reunidos. Dessa maneira, quanto maior o corpo, mais heterogêneo, com toda probabilidade, há de ser o julgamento e mais difícil, por conseguinte, formular de antemão um critério geral – mesmo o mais simples possível – de proceder” [Pereira].

No problema proposto, a função objetiva busca a melhor combinação de acadêmicos, tendo como base o ArnetMiner para os candidatos, sendo assim a função objetiva maximizar esses candidatos utilizando o índice provindo do ArnetMiner e relacionando sua idade acadêmica e número de citações e restringindo o crescimento da função através da idade máxima da combinação proposta pela função objetivo e também pelo número máximo de acadêmicos aceitos.

3. Modelo Matemático

O artigo original utiliza a seguinte modelagem matemática, não foi necessário nenhuma modificação na implementação. A formulação proposta por [Júnior e Chalco].

$$\text{Maximizar } \sum_{j \in P} p_j x_j \quad (1)$$

$$\text{sujeito a } \sum_{j \in P} A_j x_j \leq W \quad (2)$$

$$\sum_{j \in P} x_j \leq K \quad (3)$$

$$x_j = \begin{cases} 1 & , \text{ se } j \text{ é um acadêmico promissor de sucesso científico.} \\ 0 & , \text{ caso contrário.} \end{cases} \quad (4)$$

A Equação 1 representa o somatório dos potenciais vezes o peso para todo acadêmico presente em P.

A Equação 2 demonstra o somatório da idade acadêmica vezes o pesos menor igual a um certo valor inteiro W dado pela entrada da instancia.

A Equação 3 representa o somatório dos pesos menor igual ao número de acadêmicos à serem selecionados que na representação é dado pela variável K.

A Equação 4 demonstra o conjunto de solução em função de j em uma representação binária.

Tabela 1: Descrição das Variáveis

Variável	Descrição
j	Acadêmico (indivíduo)
P	Conjunto de Acadêmicos
p	Potencial
onde c	Número de Citações
h	Índice h
A	Idade Acadêmica
W	Total das Idades dos Selecionados
K	Número de Acadêmicos Selecionados

Na Tabela 1 é possível observar a descrição das variável no modelo matemático, individualmente descrevendo o que cada variável representa.

$$\underbrace{\text{impacto científico futuro}}_{P_j} = \underbrace{\text{impacto a curto prazo}}_{\frac{c_j}{A_j} + \frac{h_j}{A_j}} = \frac{c_j + h_j}{A_j} \quad (5)$$

A Equação 5 demonstra como é calculado o potencial do individuo em questão.

4. Resultados Computacionais

Os resultados obtidos com um computador Intel Core i7 8750H @4,1GHz 12 threads e 6 núcleos, com 16GB de RAM DDR4 2400MHz, HD NVMe 256GB 1500/1000MBs Leitura/Gravação. Utilizando a ferramenta de mensuração de tempo de execução foi a Hyperfine devido ao fato da API permitir a configuração do número de execuções mínima e a entrega dos resultados com tempo médio, desvio padrão e tempo mínimo e máximo.

4.1. Geração das Instancias

Devido ao fato do autor do artigo origina não disponibilizar as instancias e a base de dados AMiner não permitir a filtragem dos acadêmicos disponibilizando, é disponível apenas todos os autores indexados a base foi necessário a criação de um programa em C para que seja possível criar as instancias de maneira aleatória respeitando os critérios já estabelecidos.

A estruturação das instancias segue o seguinte critério:

- O primeiro valor representa o número de acadêmicos naquela instancia;
- O segundo número ainda da primeira linha significa a idade máxima de todos os acadêmicos selecionados;
- Ainda na primeira linha, o terceiro número significa o número de acadêmicos a serem selecionados;
- Já nas linhas subsequentes é possível encontrar o número de citações, o índice h e o ano da primeira publicação respectivamente alinhados na linha.

O programa gerador de instâncias possui a seguinte API:

1. parâmetro: Número da chave aleatório (seed)
2. parâmetro: Número de acadêmicos à serem criados
3. parâmetro: Número de acadêmicos à serem selecionados
4. parâmetro: Nome do arquivo de saída com a extensão

Contudo para orquestrar a execução do software sem grandes dificuldades também foi desenvolvido um orquestrador de execução em *Shell Script*. O orquestrador permite limpar os binários criados e as instancias, gerar os binários com os parâmetros necessários, criar novas instancias inserindo os valores para os parâmetros que desejar.

4.2. Instância do Artigo Original

A instância apresentada no artigo original, serve para entendimento da execução exemplificada do programa.

Tabela 2: Conjunto de Acadêmicos da Instância do Artigo Original

Acadêmico (j)	Citações (c)	Índice h (h)	Ano da Primeira Publicação	Idade Acadêmica (A)	Potencial (p)
1	12	2	2017	2	7
2	15	3	2017	2	9
3	2	1	2018	1	3
4	6	2	2017	2	4
5	8	1	2018	1	9
6	2	2	2018	1	4
7	20	4	2016	3	8

Na Tabela 2 é possível observar as entradas da instância, com o índice do acadêmico, o número de citações, o índice h dado pelo AMiner, o ano da primeira publicação, a idade acadêmica calculada pelo ano atual menos o ano da primeira publicação e o potencial calculado pela Equação 5.

Tabela 3: Características da Instância do Artigo Original

Número de Acadêmicos	W	K
7	4	2

Na Tabela 3 é notado as características de entrada da instância, número total de acadêmicos, somatório máximo das idades do selecionados (W) e número de acadêmicos selecionados (K).

Tabela 4: Resultados da Instância do Artigo Original

Descrição	Valor
Nome do Arquivo	instanciaoriginal.txt
Tempo Médio de Execução	546.2 ms
Desvio Padrão	27.8 ms
Menor Tempo	502.0 ms
Maior Tempo	577.9 ms
Número de Execuções	100
Valor da Função Objetivo	18
Acadêmicos Selecionados	{ 2, 5 }

A Tabela 4 exibe informações pertinentes aos resultado, o nome do arquivo de entrada, o tempo médio de execução, o desvio padrão, o menor e maior tempo de execução, o valor atingido da função objetivo e quais são os acadêmico selecionados.

4.3. Instância Nobel

A instância Nobel, foi retirada manualmente do site AMiner selecionando entre todos os vencedores do premio nobel selecionando apenas o que há alguma relação com a computação seja eles pesquisadores da área da computação ou possuírem experimentos computacionais em suas pesquisas.

Na Tabela 5 é possível observar as entradas da instância, com o índice do acadêmico, o número de citações, o índice h dado pelo AMiner e o ano da primeira publicação.

Tabela 5: Conjunto de Acadêmicos da Instância Nobel

Índice	Citações	Índice h	Ano da Primeira Publicação
1	90986	138	1975
2	21707	65	1988
3	31413	85	1970
4	94125	91	1965
5	16006	43	1969
6	97680	151	1973
7	42164	89	1955
8	36086	92	1970
9	9692	25	1990
10	5015	33	1960
11	29716	81	1968
12	34620	82	1975
13	57371	95	1979
14	9792	47	1977
15	4408	23	1963
16	43840	98	1977
17	21904	54	1971
18	7392	48	1956
19	20684	71	1982
20	55158	112	1967
21	18215	52	1964
22	92095	109	1977
23	49529	54	1957
24	7901	25	1971
25	80874	84	1998
26	4274	30	2003
27	6976	35	1974
28	3700	9	1978
29	37485	99	1964
30	44698	77	1997
31	7302	26	1983
32	4192	12	1976
33	10397	43	1961
34	89439	58	1951
35	28041	11	1948
36	14042	40	1970
37	17375	68	1970
38	1017	15	1956
39	1736	15	1987
40	49327	92	1972

Tabela 6: Características da Instância do Nobel

Número de Acadêmicos	W	K
40	135	5

Na Tabela 6 é notado as características de entrada da instância, número total de acadêmicos, somatório máximo das idades do selecionados (W) e número de acadêmicos selecionados (K).

Tabela 7: Resultados da Instância Nobel

Descrição	Valor
Nome do Arquivo	instancianobel.txt
Tempo Médio de Execução	631.6 ms
Desvio Padrão	33.0 ms
Menor Tempo	555.1 ms
Maior Tempo	714.7 ms
Número de Execuções	100
Valor da Função Objetivo	10211
Acadêmicos Selecionados	{ 6, 22, 25, 30 }

A Tabela 7 exibe informações pertinentes aos resultado, o nome do arquivo de entrada, o tempo médio de execução, o desvio padrão, o menor e maior tempo de execução, o valor atingido da função objetivo e quais são os acadêmico selecionados.

5. Conclusão

Durante o projeto, foram encontradas algumas dificuldades ao comparar os resultados com o artigo escolhido, já que o mesmo não fornece a base de dados para a instância que foi rodada, apenas uma pequena amostra que incluem sete acadêmicos é fornecida. Os resultados encontrados na instância no simplex foram idênticos aos fornecidos pelo artigo, corroborando para conclusão de que o algoritmo proposto pelo grupo funciona. Por esse motivo, o grupo optou por criar um gerador de instâncias para a realização de testes e também um menu que possibilita a inserção e execução de qualquer instância que o usuário possa vir a desejar, como visto na instância dos ganhadores de Nobel por exemplo.

Em relação ao algoritmo utilizado para executar as instâncias, as dificuldades encontradas, relacionadas às particularidades do modelo simplex, foram rapidamente sanadas através dos algoritmos de exemplos fornecidos pelo professor Dr. Pablo Munhoz em sala de aula. O software se mostrou muito rápido e prático na inserção de fórmulas matemáticas e restrições dos problemas por sua simplicidade o que também ajudou o grupo.

A tomada de decisão, ao se tratar de contratação de recursos humanos, é extremamente importante para a realização dos objetivos dos envolvidos nessas decisões de contratação e alocação de recursos, logo, informações de relevância e impacto de um acadêmico se mostram como dados extremamente importantes e que ajudam nesse tipo de tomada de decisão.

6. Execução do Programa

Existem duas formas de executar o programa, a primeira é através do main.sh que fornece a interface para que qualquer usuário consiga rodar o programa, a segunda e executando o comando `make ./cplex <verboso> <nomedoarquivodeentrada>` onde <verboso> significa exibir os dados de entrada e saída utilizando uma variável binária e <nomedoarquivodeentrada> é o nome do arquivo

da instancia. Exemplo: `./cplex 1 instancia10.txt` onde 1 irá exibir os dados de entrada na tela e será carregado a instancia10.

Referências

- Dong, Y., Johnson, R. A., e Chawla, N. V. (2016). Can scientific impact be predicted? *IEEE Transactions on Big Data*, 2(1):18–30.
- Júnior, A. d. A. B. e Chalco, J. P. M. Identificação de acadêmicos promissores de sucesso científico através de programação linear inteira.
- Pereira, V. C. O processo de seleção nos primeiros anos das bienais de São Paulo: protocolos e embates.
- Zhou, J., Zeng, A., Fan, Y., e Di, Z. (2018). Identifying important scholars via directed scientific collaboration networks. *Scientometrics*, 114(3):1327–1343.