

Rafael Barcellos Pessanha Crespo

*Título do Meu Trabalho*

Campos dos Goytacazes/RJ

2012

Rafael Barcellos Pessanha Crespo

## *Título do Meu Trabalho*

Monografia apresentada ao Curso de Graduação em Ciência da Computação da Universidade Estadual do Norte Fluminense Darcy Ribeiro como requisito para obtenção do título de Bacharel em Ciência da Computação, sob orientação do Prof<sup>o</sup>. Luis Antonio Rivera Escriba.

Tutor: Luis Antonio Rivera Escriba.

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY RIBEIRO

Campos dos Goytacazes/RJ

2012

# *Sumário*

<b>1</b>	<b>Introdução</b>	<b>2</b>
1.1	Introdução . . . . .	2
1.2	Objetivos e justificativas . . . . .	3
1.3	Estrutura do trabalho . . . . .	3
<b>2</b>	<b>Mineração de Uso Web</b>	<b>4</b>
2.1	Descrição do Problema de Programação Linear . . . . .	4
2.2	Métodos de Solução . . . . .	6
2.2.1	Método Simplex . . . . .	6
2.2.2	Método de Pontos Interiores . . . . .	7
2.3	Aplicações Práticas . . . . .	7
2.4	Ferramentas de Programação Linear . . . . .	10
	<b>Referências Bibliográficas</b>	<b>12</b>

# *1 Introdução*

## **1.1 Introdução**

A Mineração Web é uma metodologia de recuperação da informação, que usa ferramentas de mineração de dados para extrair informações tanto do conteúdo das páginas da internet e de sua estrutura de relacionamentos (links), quanto dos registros de navegação dos usuário.

Este segmento da mineração vem chamando a atenção principalmente das empresas de e-commerce nos últimos anos, pois tem sido visto como uma grande arma para “conhecer” seu cliente, que neste caso é o usuário que está navegando no site. Essas empresas visam investir em pesquisa e desenvolvimento de métodos de mineração para aplicar essa prática e assim conseguir reverter os dados obtidos ao seu favor, direcionando para uma estratégia de marketing mais eficaz, por exemplo.

Todas as empresas que vendem produtos ou serviços, dependem de sua fama e de seus clientes, portanto uma boa relação entre a empresa e o cliente é de extrema importância para que este torne-se fiel à determinada empresa, o problema é que com a globalização, que trouxe à tona os serviços remotos, como vendas online e pelo telefone, veio junto uma maior dificuldade de entender e “analisar” o comportamento do cliente, saber o que ele procura, qual a abordagem utilizar, o que oferecer, já que agora este não está mais fisicamente presente, ou seja, não há um contato direto entre vendedor e cliente, e é a partir deste problema que a mineração web passa a ser uma ferramenta essencial.

Para começarmos a falar do assunto principal deste trabalho, que é a Mineração Web, teremos que introduzir alguns termos e conhecimentos mínimos de entendimento necessário para facilitar o entendimento.

## 1.2 Objetivos e justificativas

- Item 1 Bla bla bla
- Item 2
- Item 3

## 1.3 Estrutura do trabalho

## 2 *Mineração de Uso Web*

### 2.1 Descrição do Problema de Programação Linear

O modelo de um problema de programação linear normalmente é apresentado em uma das formas a seguir (PASSOS, 2009):

$$\text{Max } z = c^T x$$

$$s.a. \begin{cases} Ax \leq b \\ x \geq 0 \end{cases}$$

ou

$$\text{Min } z = c^T x$$

$$s.a. \begin{cases} Ax \geq b \\ x \geq 0 \end{cases}$$

Um problema de programação linear com até três variáveis pode ser representado graficamente utilizando três eixos cartesianos. Os problemas com duas variáveis podem ainda ser facilmente resolvidos por meio da representação gráfica (PASSOS, 2009).

A seguir é apresentado um problema com duas variáveis e sua representação. Apesar de, na prática os problemas de programação linear possuir um número de variáveis muito maior que dois ou três, a visualização gráfica de modelo, mesmo que simples, contribui para o entendimento dos métodos de resolução apresentados nas seções a seguir. No problema exemplo, uma empresa, que fabrica vários produtos, deseja maximizar o lucro na venda de 2 desses produtos (HILLIER; LIEBERMAN, 2006).

$$\text{Maximize } z = 3x_1 + 5x_2$$

*Sujeito a*

$$1x_1 \leq 4(a)$$

$$2x_2 \leq (b)$$

$$3x_1 + 2x_2 \leq 8(c)$$

$$x_1 \geq 0, x_2 \geq 0$$

Onde,

- **x1** representa a quantidade do produto 1 produzido em uma semana
- **x2** representa a quantidade do produto 2 produzido em uma semana
- **Z** representa o lucro total por semana de produção desses dois produtos (em milhões de dólares), sendo o lucro do produto 1 de 3 milhões e o do produto 2 de 5 milhões.

E as restrições representam as restrições de tempo de cada máquina utilizada no processo de produção,

- **(a)** representa que, durante o processo de produção, cada produto 1 necessita de 1 hora na máquina 1, e a máquina só tem disponível 4 horas por semana
- **(b)** representa que, durante o processo de produção, cada produto 2 necessita de 2 horas na máquina 2, e a máquina só tem disponível 12 horas por semana
- **(c)** representa que, durante o processo de produção, cada produto 1 necessita de 3 horas na máquina 3, e cada produto 2 necessita de 2 horas na máquina 3, e a máquina só tem disponível 8 horas por semana

Graficamente representado o problema ficaria da seguinte forma:

Figura 1: Representação gráfica de um Problema de Programação Linear de duas variáveis

Onde cada reta representa uma restrição do modelo, e a área cinza representa a região viável, ou seja nessa área estão contidas os valores ótimos de  $x_1$  e  $x_2$  para a maximização do lucro.

Os métodos para resolução de problemas de programação linear buscam a esses valores de  $x_1$  e  $x_2$  para determinação da solução ótima.

## 2.2 Métodos de Solução

Entre os métodos mais famosos para a resolução de problemas de programação linear estão o método simplex e o método de pontos interiores. Depois da apresentação do método simplex, outros métodos com diferentes abordagens foram propostos (TODD, 2002). Porém, dentre os métodos existentes apenas o método de pontos interiores é atualmente competitivo em relação ao método simplex (BIXBY, 1992 apud JUNIOR, 2009). A principal diferença entre esse dois métodos é o que o método simplex caminha pelos vértices da região viável, enquanto o método de pontos interiores caminha pelo interior da região viável (MACULAN; FAMPA, 2006). Além disso, uma outra diferença é que o simplex exige muitas iterações com cálculos simples, enquanto o método de pontos interiores poucas iterações são exigidas, porém com cálculos mais elaborados. Apesar das vantagens do método de pontos interiores em relação ao método estudado neste trabalho, o método simplex possui melhor desempenho na resolução de problemas de pequeno porte em relação ao método de pontos interiores, tornando-se um método indispensável em ferramentas de programação linear.

### 2.2.1 Método Simplex

O método simplex é um dos algoritmos mais populares para a resolução de problemas de programação linear. Surgiu a mais de 60 anos atrás e foi proposto por George Dantzig.

É um método iterativo, e sua ideia principal consiste no fato de que a cada iteração uma nova solução é encontrada, sempre melhor que a anterior até o ponto em que a solução ótima é obtida. Outra característica do método é o fato de ser matricial, ou seja, os dados a serem calculados são armazenados em matrizes.

Com a utilização do método foi percebido que a cada iteração do método simplex eram requeridos muitos cálculos sobre valores que nem sempre importavam para a iteração seguinte, fato que do ponto de vista computacional tornaria o método ineficiente. A partir desse fato foi desenvolvido o método simplex revisado visando a resolução de problemas de programação linear computacionalmente.

O presente trabalho tem foco no método simplex revisado e suas variantes.



## 2.2.2 Método de Pontos Interiores

Em 1984, Karmarkar revolucionou a área de programação linear com a publicação de um algoritmo de complexidade polinomial que apresentou bom desempenho quando aplicado a problemas práticos (MACULAN; FAMPA, 2006). Essa publicação deu origem a um novo campo de pesquisa, chamado de método dos pontos interiores.

O método de pontos interiores tem como principal característica a de realizar a busca por soluções no interior da região viável do problema, até encontrar a solução ótima (MENEZES, 2008). Em teoria, o método de pontos interiores é melhor que o método simplex, principalmente quando se leva em conta o critério de complexidade de pior caso. No entanto, na prática ambos os métodos concorrem até hoje. Já que o sucesso do método depende da estrutura dos problemas, da esparsidade <sup>1</sup> e da arquitetura dos computadores (MACULAN; FAMPA, 2006).

## 2.3 Aplicações Práticas

Um problema de programação linear, como já dito anteriormente, busca a otimização na distribuição de recursos sujeitos a restrições. Por isso é considerada uma poderosa ferramenta de apoio a decisão (FROSSARD, 2009) e com utilização em diversas áreas, como: indústria, saúde, computação, produção, etc. As empresas, por exemplo, devem estar constantemente atentas a competitividade e as restrições existentes com o objetivo de alcançar suas metas, para isso é necessário otimizar os recursos disponíveis (FROSSARD, 2009). Daí a importância da utilização da programação linear empregada em seu exemplo mais geral: maximizar o lucro e minimizar custos. Na definição de modelos desse tipo deve-se considerar o preço de venda e o custo de produção, além de restrições do tipo: quantidade de matéria-prima e mão-de-obra disponíveis, máquinas disponíveis para produção, entre outros (FROSSARD, 2009).

“Administrar com eficiência os recursos disponíveis na empresa, através do planejamento, controle e execução das atividades relacionadas à utilização destes, é fator fundamental na busca da otimização do resultado global da empresa. A programação linear juntamente com as técnicas de pesquisa operacional, permite identificar o resultado ótimo, considerando todas as restrições impostas no modelo adotado.” (FROSSARD, 2009, p. 31)

---

<sup>1</sup>Quando uma matriz possui uma grande proporção de elementos nulos diz-se que é uma matriz esparsa (JUNIOR, 2009).

A decomposição wavelet tem sido abordada mediante diferentes métodos. Dentre eles destacam-se os métodos que tem como base a programação linear. Um dos primeiros trabalhos nesta área foi proposto por Saunders (2001) utilizando métodos de pontos interiores. Devido ao fato de que os problemas lineares obtidos a partir da decomposição wavelet possuem matrizes muito densas, os autores restringiram a pesquisa apenas ao caso de problemas com dicionários com uma estrutura especial. Vale resaltar também que o método gera problemas lineares de grande porte. Por exemplo, um problema de sinal de onda típico de comprimento 8192, se traduz em um problema linear de tamanho 8192 por 212.992.

Em seu trabalho Tziritas (2007) diz que problemas de análise de imagens podem ser formulados como problemas de rotulagem, onde a partir de uma imagem segmentada, são realizadas comparações entre pontos vizinhos a fim de rotular um grupo de pontos afins. Uma questão que vem incentivando pesquisas nessas áreas, é como resolver problemas desse tipo de forma eficiente e precisa. O autor propõe a utilização do esquema primal-dual da programação linear, e diz que essa utilização revelou excelentes resultados

Em Slyke (2009) e Yarmish (2001) demonstram a eficiência da utilização do método simplex padrão de forma distribuída. Esse tipo de implementação é útil para a computação distribuída. Já que nesse tipo de implementação as colunas das matrizes, onde são armazenados os dados, são distribuídas entre os processadores, é mais intuitivo o uso do método simplex padrão. Além disso, esse método é mais utilizado e mais eficiente que o método revisado na resolução de problemas de grande porte, onde existe um maior volume de dados, fazendo mais sentido a utilização da programação distribuída. De acordo com Slyke (2009) e Yarmish (2001) a eficiência do método simplex revisado é afetada pela densidade do problema, ou seja, é um método mais eficiente para problemas esparsos, os quais são mais comuns. Porém existem aplicações que exigem modelos mais densos, como em processamento de imagens. Nesse caso a implementação de um algoritmo do método simplex padrão distribuído se mostra mais eficiente. Além disso não existem algoritmos eficientes do método simplex revisado distribuído. Em conclusão o autor diz que a implementação apresentou bons resultados, especialmente para problemas grandes e densos.

Em Allgöwer (2011) é proposto um algoritmo simplex distribuído para problemas de programação linear degenerados e atribuições multi-agentes. Um problema de programação linear é dito degenerado se na solução uma das variáveis valer zero. O objetivo do trabalho é propor um algoritmo simplex para ser implementado em uma distribuição

multi- agente, onde os agentes devem concordar com uma solução ótima ou declarar o problema como inviável, caso não haja solução ótima.

No trabalho de Yarmish (2006) o autor resolve um conjunto de problemas de decomposição wavelet, utilizando tanto o método simplex revisado quanto o método simplex padrão. O autor mostra que embora o método simplex revisado seja superior no caso do problema mais geral (problema esparso), a resolução de problemas de decomposição wavelet resulta em um problema denso, onde o método simplex padrão tem melhor desempenho. O termo wavelet refere-se basicamente a um conjunto de funções com forma de pequenas ondas (BLITZKOW, 2008). A decomposição wavelet é uma metodologia de decomposição de uma função ou sinal em um domínio de frequências e espaço, sendo possível investigar a ocorrência de fenômenos localizados no espaço e frequência simultaneamente (PEIXOTO, 2009). A decomposição wavelet em sua versão discreta é utilizada na compreensão de dados sendo útil no processamento de imagens. No processamento de imagens a utilização da decomposição wavelet se torna eficiente na extração de informações de um sinal e na análise de frequências, principalmente quando ocorrem rápidas variações na frequência (LEITE, 2007)

A programação linear também se aplica na área da saúde, em Goldberg (2006) é feita uma proposta de utilização da programação linear no tratamento do câncer. Em um determinado tipo de tratamento são inseridos cateteres na área afetada para introduzir o medicamento necessário, porém o medicamento acaba afetado células saudáveis além das cancerígenas. Como os problemas de programação linear podem ser resolvidos como problemas determinísticos e com solução exata, Goldberg (2006) propõe que a formulação de um problema de otimização para determinar o tempo de permanência dos cateteres, minimizando os desvios na quantidade da dose necessitada pelo paciente. Nos testes realizados, os resultados obtidos não foram mostraram uma vantagem significativa em relação ao método atualmente utilizado.

De acordo com Nagel (2008) dentre as inúmeras ferramentas de programação linear existentes, existe a carência de uma ferramenta que resolva problemas de programação linear de forma integrada a um banco de dados, utilizando procedimentos armazenados em sql pre compiladas em armazenadas no banco de dados juntamente com os dados. Essa características em uma ferramenta se torna importante, principalmente, em problemas de grande porte, suprimindo o tráfego de dados já que toda alógica é executada internamente no banco de dados, e apenas o resultado final é enviado para o usuario. Na implementação apresentada por Nagel (2008) os dados do modelo a ser resolvido são armazenados como

modelos relacionais, ou seja, em matrizes bi-dimensionais. Apesar das vantagens desse tipo de implementação, o autor chegou a conclusão que o tempo de execução ainda deve ser melhorado.

Outros exemplos bastante realistas da utilização da programação linear são os problemas de mistura e os de planejamento de produção e controle de estoque. No primeiro, basicamente trata-se da mistura de diferentes matérias para a produção de produtos, que devem obedecer a algumas especificações e, ao mesmo tempo, minimizar o custo ou maximizar o lucro. Já os problemas de planejamento de produção e controle de estoque possuem inúmeras aplicações, desde a alocação de máquinas para atender determinada demanda, até a utilização do estoque para atender a uma mudança imprevisível na demanda e necessidades de contratação de demissão para enfrentar mudanças nas necessidades de mão de obra (TAHA, 2008). A programação linear além de estar presente, é fundamental em diversas áreas, tornando-se uma ferramenta de apoio a decisão e contribuindo para o sucesso de projetos nas áreas em que se aplica.

## 2.4 Ferramentas de Programação Linear

Durante a pesquisa foram utilizadas e encontradas algumas ferramentas para a resolução de problemas de programação linear, além de bibliotecas com problemas testes. Uma das ferramentas é o Tora (TAHA, 2008), que é software com uma interface simples, mas com algumas características que não há tornam muito amigável, além disso é um software mais educacional, inclusive com o recurso que mostra o passo a passo na resolução do problema utilizando o método simplex tabular.

Figura 2: Telas de entrada e saída de dados do software Tora

Já o Lingo disponibiliza uma versão demo gratuita. A maior dificuldade na utilização dessa ferramenta é o fato do usuário precisar digitar o modelo sem nenhuma orientação em relação a sintaxe, além disso a forma como o resultado é exibido dificulta a interpretação. O Cplex é uma ferramenta mais robusta, com mais recursos e uma interface mais elaborada. Não é gratuita, mas uma versão com licença acadêmica é disponibilizada. Permite a modelagem e resolução de problemas de otimização de diversa natureza, como por exemplo: programação inteira, programação linear e programação inteira mista. Utilizando para isso diferentes métodos de solução. O Glpk é um conjunto de rotinas destinadas a

resolução de problemas de programação linear, além de problemas de programação inteira mista. As rotinas são escritas em ANSI C e podem ser utilizadas como biblioteca.

## *Referências Bibliográficas*

- ALLGÖWER, M. B. G. N. F. B. F. A Distributed Simplex Algorithm for Degenerate Linear Programs and Multi-Agent Assignments. *Automatica*, Maio 2011.
- BIXBY, R. E. Implementing the simplex method: The initial basis. *ORSA Journal*, v. 4, p. 267–284, 1992.
- BLITZKOW, A. C. B. B. D. *Ondaletas: Histórico e Aplicação*. Maio 2008.
- FROSSARD, A. C. P. Programação Linear: Maximização de Lucro e Minimização de Custos. *Revista Científica da Faculdade Lourenço Filho*, v. 6, n. 1, 2009.
- GOLDBERG, R. A. E. L. J. P. I.-C. J. H. J. F. O. K. Optimization of HDR brachytherapy dose distributions using linear programming with penalty costs. *The International Journal of Medical Physics Research and Practice*, v. 33, n. 11, 2006.
- HILLIER, F.; LIEBERMAN, G. *Introdução à Pesquisa Operacional*. 8<sup>o</sup>. ed. [S.l.]: Mc-Graw-Hill, 2006.
- JUNIOR, P. A. M. *Técnicas Computacionais para uma implementação eficiente e estável de métodos tipo simplex*. Dissertação (Mestrado) — Instituto de Ciências Matemáticas e de Computação, ICMP-USP, 2009.
- LEITE, F. E. A. *Análise Estatística de Padrões Sísmicos: Decomposição em Multiescala*. Tese (Doutorado) — Universidade Federal do Rio Grande do Norte Centro de Ciências Exatas e da Terra Departamento de Física Teórica e Experimental, Dezembro 2007.
- MACULAN, N.; FAMPA, M. H. C. *Otimização Linear*. [S.l.]: Editora Universidade de Brasília, 2006.
- MENEZES, L. de L. P. M. A. F. Implementação de algoritmos simplex e pontos interiores para programação linear. *Estudos*, v. 15, n. 2, p. 225–246, 2008.
- NAGEL, A. K. A. New Developments in Robotics, Automation and Control. In: \_\_\_\_\_. [S.l.]: InTech, 2008. cap. Linear Programming in Database, p. 339–354.
- PASSOS, A. N. *Estudos em Programação Linear*. Dissertação (Mestrado) — Instituto de Matemática, Estatística e Computação Científica, UNICAMP, 2009.
- PEIXOTO, P. da S. *Resolução numérica de EDPs utilizando ondaletas harmônicas*. Dissertação (Mestrado) — Instituto de Matemática e Estatística da Universidade de São Paulo, Junho 2009.
- SAUNDERS, S. S. C. D. L. D. M. A. Atomic Decomposition by Basis Pursuit. *SIAM Review*, v. 43, n. 1, p. 129–159, 2001.

SLYKE, G. Y. R. A Distributed, scaleable simplex method. *The Journal of Supercomputing*, v. 49, n. 3, p. 373 – 381, Setembro 2009.

TAHA, H. A. *Operation Research: An Introduction*. 8. ed. [S.l.]: Pearson, 2008.

TODD, M. J. The many facets of linear programming. *Mathematical Programming*, v. 91, p. 417–436, 2002.

TZIRITAS, N. K. G. Approximate Labeling via Graph Cuts Based on Linear Programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 29, n. 8, p. 1436–1453, 2007.

YARMISH, G. *A Distributed Implementation of the Simplex Method*. Tese (Doutorado) — Polytechnic University, Março 2001.

YARMISH, G. The Simplex Method Applied to Wavelet Decomposition. In: SCIENCE, U. o. T. a. D. G. R. Dattatreya Department of C. (Ed.). *MATH'06 Proceedings of the 10th WSEAS International Conference on Applied Mathematics*. [S.l.: s.n.], 2006. p. 226–228.