

Ferramenta para suporte à decisão de frentes de corte de cana-de-açúcar usando algoritmos genéticos

Carlos Eduardo Rodrigues de Alencar¹, Renato Fernandes Corrêa^{1,2},
Fernando Buarque de Lima Neto¹

¹ Escola Politécnica, Universidade de Pernambuco

Rua Benfica, 455, Madalena, Recife - PE, Brasil, 50.750-410

² Centro de Informática, Universidade Federal de Pernambuco

P.O. Box 7851, Cidade Universitária, Recife - PE, Brasil, 50.732-970

cadualencar@yahoo.com.br, rfc@cin.ufpe.br, fbln@dsc.upe.br

Abstract

This work presents a solution for the decision problem of choosing a set of plots cultivated with sugar cane to be harvested in a sugar cane plantation. For this, a software architecture was developed; it uses Genetic Algorithms in the optimization of the function that computes the agronomical return involving the variables: PCC (Apparent Percentage of sucrose in the sugar cane); TCH (Ton of sugar cane for hectare); and Fiber (Total of dry matter after the milling). The tool possess a friendly interface and supplies for an agricultural manager the functionality to adjust the function of agronomical return and to establish restrictions in order to get a set of optimized solutions capable of improving the agronomical and industrial return during harvest. These solutions aim at to assist in a more efficient way the decision of which lots to cut, and thus to allow the increase of the agricultural and industrial performance of sugar cane plantations.

1. Introdução

Dentre os muitos aspectos gerenciais, a escolha do momento ideal para colheita de lotes da cana-de-açúcar é uma das decisões mais difíceis de tomar numa usina canavieira. Esta decisão deve levar em consideração um conjunto de critérios relacionados a esta prática agrônômica, como também outros aspectos industriais, com o intuito de reduzir custos operacionais e de demandas energéticas.

O planejamento de colheita na cultura de cana-de-açúcar busca otimizar o retorno agrônômico, baseado no conceito que a cana tem uma época, durante o ano, onde ocorre a máxima concentração de sacarose nos colmos. Essa época pode ser visualizada na Figura 1, na fase de maturação.

Na fase de maturação, a cana-de-açúcar atinge seu tamanho máximo. É quando se inicia o período seco. A cana quando estimulada pela diminuição de água no solo passa a acumular energia, açúcares, a fim de subsidiar o período seco.

Após a época da maturação é desencadeado o processo de reprodução onde as reservas de açúcares serão utilizadas, fazendo com que haja perda de produtividade na colheita.

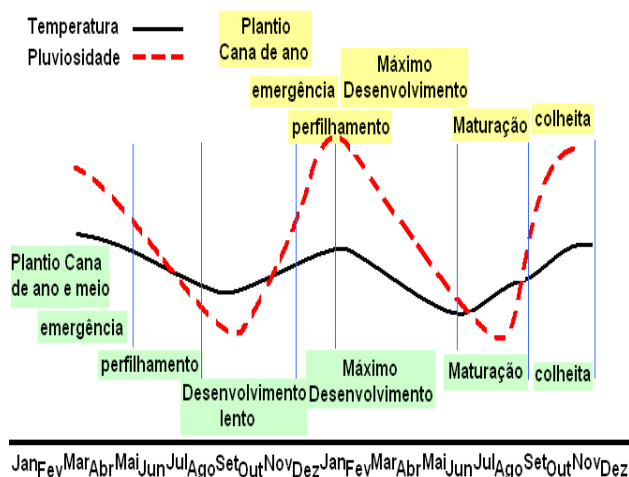


Figura 1. Ciclo da cana-de-açúcar [1].

Dessa forma, a definição da melhor época para o corte da cana de cada lote, isto é, do momento de acúmulo máximo de açúcares, é de grande importância econômica.

Outros fatores importantes a serem levados em consideração na decisão de colheita são: a limitação quanto a tonelagem de cana a ser cortada e transportada até o pátio da usina; e a quantidade de bagaço de cana disponível para alimentar as caldeiras da usina.

De acordo com [2], três indicadores podem ser utilizados para a avaliação do desempenho econômico em função de suas efetivas contribuições para a decisão de colheita:

- TCH (Tonelada de Cana por Hectare) – Para garantia de volumes de produção;
- PCC (Porcentagem aparente de açúcar no Caldo da Cana) – Para medição da qualidade da matéria prima (quantidade de açúcares);
- Fibra – Para medição da qualidade do potencial calorífico produzido pela queima do bagaço da cana nas caldeiras após sua moagem.

Definindo uma função de avaliação do retorno econômico que recebe como parâmetros as estimativas dos indicadores TCH, PCC e Fibra para cada lote, bem como a tonelagem de cana a ser cortada, pode-se avaliar o quão boa é a decisão de cortar um conjunto de lotes em detrimento de outro conjunto de lotes.

Entretanto, o problema de seleção do melhor conjunto de lotes a cortar é NP-difícil [3], uma vez que para se obter a solução ótima, todas as combinações possíveis de lotes selecionados teriam que ser geradas e avaliadas.

Uma técnica de otimização que seja capaz de convergir para soluções próximas da solução ótima em tempo hábil se torna então necessária. O presente trabalho apresenta de forma pioneira uma solução do problema de decisão de seleção de lotes de cana-de-açúcar a cortar utilizando Algoritmos Genéticos [4].

O objetivo deste trabalho consiste na aplicação de Algoritmo Genético (AG) na seleção dos melhores lotes de cana-de-açúcar de uma unidade produtora de açúcar ou álcool a cortar, de forma a maximizar uma função de retorno agrônomo que envolve os indicadores PCC, TCH e Fibra estimados para cada lote na pré-colheita. As soluções obtidas com AG podem ser utilizadas pelo gerente agrícola como sugestões na elaboração do planejamento de corte, propiciando assim um resultado mais eficaz e eficiente.

2. Algoritmos Genéticos

Algoritmos Genéticos (AGs) [4] são métodos de otimização global, baseados nos mecanismos de seleção natural. Nos AGs, um indivíduo ou cromossomo codifica um ponto num espaço de soluções. A cada indivíduo é associado um valor de uma função objetivo a ser maximizada. De acordo com o valor da função objetivo alcançado por cada indivíduo, este é avaliado com um nível de aptidão.

Com base no mecanismo de seleção natural e operadores genéticos de cruzamento e mutação, os AGs empregam uma estratégia de busca paralela e estrutu-

rada, mas aleatória, que reforça a busca na vizinhança de pontos de “alta aptidão”, ou seja, indivíduos nos quais a função a ser maximizada tem valores relativamente altos.

Apesar de aleatórios, os AGs não utilizam passos aleatórios não direcionados, pois exploram informações obtidas no processo de busca para encontrar novos indivíduos onde são esperados melhores resultados. Isto é feito através de processos iterativos, onde cada passo da iteração é chamada de geração. O conjunto de indivíduos em uma dada geração é chamado população. Em uma dada geração, a população é obtida através dos indivíduos da população na geração anterior, por meio da aplicação de operadores genéticos.

Os AGs são métodos de otimização eficientes por utilizarem duas técnicas denominadas *Exploration* e *Exploitation* para encontrar o ótimo global da função a ser maximizada [4]. Ambas as técnicas têm a mesma tradução para o português, exploração. Entretanto, *Exploitation* significa exploração no sentido de absorver informações presentes nas soluções encontradas e *Exploration* diz respeito à exploração de busca à procura por novas soluções.

Neste presente trabalho utilizou-se Algoritmos Genéticos Binários. Trata-se de um tipo especial de AG onde os pontos do espaço de solução são codificados como uma cadeia de bits dentro dos indivíduos. Cada bit seria uma alusão a um gene do cromossomo ou indivíduo.

As principais razões da escolha dos AGs Binários para resolver o problema de decisão de lotes a cortar foram:

- A facilidade em codificar os pontos do espaço de solução deste problema em cadeia de bits;
- Por ter um bom desempenho garantido pelo Teorema dos Esquemas [6] na otimização de funções.

3. Algoritmo Genético para a escolha de lotes de cana-de-açúcar a cortar

Os seguintes tópicos esquematizam a especificação do algoritmo genético proposto:

- Entrada do algoritmo;
- Codificação do problema;
- Espaço de soluções associados;
- Função objetivo;
- Método de Seleção e operadores genéticos;
- Critério de parada.

Cada um desses tópicos é apresentado nas subseções que se seguem.

3.1. Entrada do algoritmo

O algoritmo proposto necessita ter como entrada uma tabela onde se encontra especificado para cada lote da unidade agrícola: a área; e as estimativas de PCC, TCH e Fibra no dia em que será estabelecido o corte. Outra entrada para o algoritmo é a tonelagem de cana a ser cortada.

As estimativas dos indicadores PCC, TCH e Fibra para cada lote podem ser feitas através de modelos analíticos de maturação da cana ou modelos obtidos através de algoritmos de aprendizagem de máquina treinados com dados de produção de safras anteriores. Redes neurais artificiais [7] tem sido utilizadas com sucesso na criação automática de modelos que estimam esses parâmetros com base em dados históricos de produção [2] [8] [9].

Como os dados utilizados para treinar os algoritmos de aprendizado de máquina são ruidosos, isto acarretará na incorporação de ruídos (ou imprecisão) também nas estimativas dos indicadores. Assim, o nível de ruído presente na estimativa dos indicadores não pode ser muito alta, pois isto poderia induzir o algoritmo genético a convergir para soluções avaliadas como boas mas na prática ruins.

3.2. Codificação do Problema

No AG proposto, cada indivíduo representa uma provável solução do problema, contendo informações sobre quais lotes serão selecionados para colheita.

Cada cromossomo é composto por N genes, onde N é o número de lotes que compõem a unidade agrícola em questão. Cada gene pode assumir valores 0 ou 1, indicando respectivamente lote não-selecionado ou lote selecionado.

3.3. Espaço de Soluções Associados

A população inicial é composta de 30 indivíduos, este valor foi definido a partir de experimentos realizados com a ferramenta desenvolvida neste trabalho onde foi verificado que a utilização de valores maiores apenas aumentava o tempo de processamento do algoritmo, sem que houvesse ganhos significativos na qualidade da solução encontrada.

Os indivíduos da população inicial são gerados obedecendo ao seguinte processo: A primeira metade é gerada de forma aleatória e a segunda metade é gerada invertendo-se os bits de cada indivíduo da primeira metade, garantindo-se dessa forma, que boa parte do espaço de busca da solução seja representada.

Na avaliação dos indivíduos é utilizado o método de Ordenamento Linear, de modo a garantir que os valores da aptidão não assumam valores muito próximos ou que alguns indivíduos assumam valores muito elevados em relação ao restante da população, incorrendo em problemas como busca aleatória e convergência prematura [4].

3.4. Função Objetivo

Na Equação 1 é mostrada a função multiobjetivo que se deseja maximizar para a escolha dos lotes a colher dos N lotes de cana-de-açúcar disponíveis na usina.

$$h(x,t) = \alpha * c(x) + (1-\alpha) * f(x) - \beta * \Phi(x,t) - \gamma * \Phi(x) \quad (1)$$

As Funções $c(x)$ e $f(x)$, especificadas nas Equações 2 e 3, determinam respectivamente os valores do PCC total e Fibra total que se pretende maximizar, afim de maximizar a quantidade de açúcar obtida (produto almejado) e a quantidade de matéria seca após a moagem (combustível para as caldeiras das usinas) dos lotes selecionados para colheita.

$$c(x) = \frac{100 * \sum_1^N x_i * PCC_i}{\sum_1^N PCC_i} \quad (2)$$

$$f(x) = \frac{100 * \sum_1^N x_i * Fibra_i}{\sum_1^N Fibra_i} \quad (3)$$

As funções $c(x)$ e $f(x)$ obedecem a um critério de ponderação α que é fornecido pelo usuário determinando o grau de relevância de cada uma das funções. Este critério de ponderação fornecido fará com que o algoritmo faça uma busca pela melhor solução de acordo com o que o usuário achar mais importante maximizar: PCC ou Fibra.

A função $\Phi(x,t)$ é uma restrição para a tonelagem que se deseja cortar, e é especificada nas Equações 4 e 5.

$$\Phi(x,t) = \frac{100 * |Min(0, (t - Test))|}{|t - \sum_1^N Area_i * TCH_i|} \quad (4)$$

$$Test = \sum_1^N x_i * Area_i * TCH_i \quad (5)$$

$\Phi(x,t)$ corresponde ao mínimo do módulo da diferença entre a tonelagem desejada t (fornecida pelo usuário) e a tonelagem estimada $Test$, dividido por constante de normalização que corresponde ao máximo do módulo da diferença possível entre as variáveis t e $Test$. $Test$ é calculada como o somatório das estimativas de tonelagem dos lotes selecionados. Esta função fará com que o algoritmo encontre uma solução com a tonelagem total o mais próximo possível da tonelagem desejada.

A função $\Phi(x)$ é uma restrição quanto o número de lotes selecionados para o corte, veja a Equação 6. Esta função ajuda o algoritmo a cortar um número mínimo de lotes.

$$\Phi(x) = \frac{100 * \sum_1^N x_i}{N} \quad (6)$$

As funções $c(x)$, $f(x)$, $\Phi(x,t)$, $\Phi(x)$ estão normalizadas para uma escala de 0 a 100.

As variáveis β e γ são constantes de pênalti utilizadas para controlar a intensidade da penalidade. Foram utilizadas β igual a 2 e γ igual a 0,75. Estes valores foram encontrados experimentalmente em testes preliminares.

3.5. Método de Seleção e Operadores Genéticos

A seleção por amostragem estocástica universal é utilizada para determinar os indivíduos mais aptos a gerar novos descendentes na próxima geração.

Logo após a seleção e cruzamento, os indivíduos gerados passam pelo processo de mutação.

Foram utilizados 3 tipos de métodos de cruzamento, permitindo ao usuário trocá-los livremente:

- Um ponto;
- Dois pontos;
- Uniforme.

O operador de mutação trabalha alternando o valor de cada bit de um indivíduo sorteado pseudoaleatoriamente com uma determinada probabilidade.

Os operadores de cruzamento e mutação são dependentes de probabilidades predefinidas pelo usuário. Essas probabilidades foram definidas como padrão: 80% para cruzamento e 3% para mutação. A pressão de seleção utilizada foi definida como 0,5. Estes valores foram obtidos através de experimentos preliminares.

Além dos operadores básicos, um esquema de elitismo foi também incorporado ao AG. A função do elitismo é preservar as melhores soluções encontradas,

para que essas não se percam durante o processo de evolução.

Toda a população é substituída em cada geração, ou seja, são criados N filhos para substituir N pais. Os k melhores pais nunca são substituídos por filhos piores. Quanto maior for o valor de k , maior o risco de convergência prematura. No AG proposto o valor de k foi definido como 10.

Em testes preliminares o método de cruzamento de dois pontos mostrou-se mais efetivo, levando o AG a melhores soluções num menor número de gerações.

3.6. Critério de Parada

Foram adotados dois tipos possíveis de critérios de parada do AG, os quais também são parâmetros possíveis de serem determinados pelo usuário:

- Número de Gerações: o AG irá evoluir a população de soluções até o número de gerações fornecido;
- Convergência do AG: passada a quantidade de gerações fornecida pelo usuário sem que o melhor indivíduo sofra alguma modificação, o AG terá convergido e irá parar a evolução.

O critério de parada padrão foi determinado como o número de gerações igual a 1000.

4. Metodologia

A linguagem escolhida para o desenvolvimento do protótipo foi C#, uma linguagem de programação orientada a objetos elaborada pela Microsoft Corporation. Foi utilizado o ambiente de desenvolvimento integrado (IDE - *Integrated Development Environment*) Microsoft Visual Studio .NET 2003 para desenvolver este protótipo.

A entrada de dados do protótipo se dá através de uma planilha eletrônica MS-Excel compatível com versões superiores à versão 97. A planilha deve conter informações sobre identificador de cada lote, área de plantio do lote e respectivas estimativas de PCC, TCH e Fibra.

Foram realizados vários testes para verificar a efetividade do software desenvolvido neste trabalho. Foram utilizadas como entradas duas planilhas: uma com 15 lotes (Cana15) e outra com 380 lotes (Cana380).

As soluções encontradas pelo AG, dada a planilha, critérios e restrições estabelecidos, foram comparadas às melhores soluções previamente conhecidas para o mesmo cenário de teste.

Neste presente trabalho, as planilhas contêm estimativas de PCC, TCH e Fibra, para um conjunto de lotes candidatos, advindas de uma rede neural treinada

com dados de produção de safras anteriores de uma usina de açúcar situada no noroeste do estado de São Paulo [2].

5. Resultados

São apresentados a seguir a especificação e os resultados de três testes realizados para testar o software desenvolvido.

A entrada do software foi uma das duas planilhas especificadas (Cana15 ou Cana380), e um dos cenários de teste especificados na Tabela 1. Na Tabela 1 são exibidas informações sobre a Tonelagem desejada, e as ponderações de PCC e Fibra para cada teste.

Tabela 1. Testes

Teste	Tonelagem Desejada	Ponderação PCC	Ponderação Fibra
1	2000	50%	50%
2	4800	30%	70%
3	9000	0	100%

Em cada teste os parâmetros do AG utilizados foram os valores padrão especificados anteriormente.

A solução final obtida pelo AG (melhor indivíduo da última geração), foi comparada com as melhores soluções conhecidas para cada planilha sobre o conjunto de critérios e restrições determinados em cada cenário de teste.

Na Tabela 2 a melhor solução obtida pelo AG para a planilha Cana15 e os respectivos cenários de teste são ilustrados.

Tabela 2. Melhores soluções para os testes com a planilha Cana15.

Teste	Tonelagem	PCC Total	Fibra Total	Lotes a Cortar
1	2036,45	28,41	32,07	10, 14
2	4756,67	55,29	65,64	4,6, 10, 14
3	9052,05	85,75	96,66	4, 6, 10, 11, 14, 15

As soluções mostradas na Tabela 2 são as soluções ótimas para a planilha Cana15 e os respectivos cenários estabelecidos. A maioria das execuções dos AG convergiram para estas soluções em menos de 300 gerações, sendo a convergência para outras soluções mínimos locais pouco provável (1 a cada 10 execuções).

Na Tabela 3 mostra-se a melhor solução obtida pelo AG para a planilha Cana380 e os respectivos cenários de teste.

Com a planilha Cana380, as soluções obtidas não são ótimas, consistindo de mínimos locais que apenas

representam boas soluções. Não foi verificada nenhuma tendência de convergência do AG a nenhuma solução em múltiplas execuções de cada teste. Este fato se explica pela grande quantidade de mínimos locais presentes e a dificuldade da presente configuração do AG em convergir para os mínimos globais.

Tabela 3. Melhores soluções para os testes com a planilha Cana380.

Teste	Tonelagem	PCC Total	Fibra Total	Lotes a Cortar
1	2038,20	45,70	45,09	193, 211, 328
2	4630,60	232,31	241,32	112, 116, 131, 154, 158, 166, 233, 242, 270, 277, 279, 310, 325, 336, 341, 342
3	9035,70	380,80	378,24	57, 77, 92, 94, 112, 132, 150, 151, 165, 187, 202, 207 213, 214, 227, 235, 244, 248, 261, 290, 297, 298, 301, 336 340, 341

6. Conclusão

Neste trabalho evidenciou-se a possibilidade de se utilizar algoritmos genéticos para dar suporte à decisão de frentes de corte de cana-de-açúcar.

Para isso foi desenvolvido um protótipo de software que implementa o AG proposto e experimentados vários valores de parâmetros, onde foi observado um melhor desempenho para os valores padronizados nesse trabalho.

O protótipo desenvolvido mostrou-se capaz de auxiliar o gerente agrícola no processo de decisão do corte de lotes de cana-de-açúcar, já que ofereceu nos testes realizados soluções com valores máximos de PCC e Fibra, dado um valor de Tonelagem desejada para um teste envolvendo dezenas de lotes, e boas soluções para um teste envolvendo centenas de lotes. Assim, com a utilização do protótipo desenvolvido pelo gerente agrícola é possível vislumbrar um aumento do retorno da produção agrícola.

Como trabalhos futuros, pretendemos investigar:

- O uso de algoritmos híbridos que mesclam AGs com outros métodos de otimização, como *Simulated Annealing* e Subida de Encosta, afim de diminuir o custo computacional e permitir a obtenção de soluções mais refinadas e próximas das soluções ótimas, principalmente em unidades agrícolas com centenas de lotes;
- A incorporação na função objetivo de outras variáveis ou fatores que influenciem na produtividade ou retorno econômico da cultura, como por exemplo a distância entre lotes a cortar;
- A utilização uso de Algoritmos Genéticos multi-objetivo baseados na relação de Pareto, a fim de remover alguns parâmetros empíricos da função de avaliação do AG;
- Comparação com outras abordagens de solução do problema desenvolvidas paralelamente à apresentada neste trabalho [10];
- O teste do sistema por gerentes de uma usina.

7. Referências

[1] Castro, P. R. C. “Maturadores Químicos em cana-de-açúcar”. In: Semana da Cana-de-açúcar de Piracicaba, 4, Piracicaba, 1999. Anais Saccharum, v. 1, 1999, p.12-16.

[2] Neto, F. B. de L. *Suporte a Decisão Gerencial Baseado em Redes Neurais Artificiais – nDSS*. Dissertação de Mestrado apresentada ao Departamento de Informática da Universidade Federal de Pernambuco, Recife - Brasil, 1998.

[3] Papadimitriou, C. H., Steiglitz K. *Combinatorial Optimization: Algorithms and Complexity*. Prentice-Hall, Inc., 1982.

[4] Lacerda, E.G.M., Carvalho, A.C.P.L. “Introdução aos algoritmos genéticos”. In: Galvão, C.O., Valença, M.J.S. (orgs.) *Sistemas inteligentes: aplicações a recursos hídricos e ciências ambientais*. Porto Alegre: Editora da Universidade/UFRGS/ABRH, 1999, p. 87-150.

[5] Beasley, D., Bull, D. R., Martin, R. R. *An Overview of Genetic Algorithms: Part 1, Fundamentals. Part 2, Research Topics*. University Computing, 1993.

[6] Goldberg, D. E. *Genetic algorithms in search, optimization and machine learning*. Addison Wesley, 1989.

[7] Haykin, S. *Redes neurais - Princípios e prática*. 2a edição. Editora Bookman, 1999.

[8] Trigo, T. R., Júnior, P. C. B., Neto, F. B. de L. “Redes Neurais Artificiais em Colheita de Cana-de-Açúcar”. In: V Congresso Brasileiro de Agroinformática, Londrina - Brasil, Outubro, 2005.

[9] Pacheco, D. F., Regueira, F. S., Neto, F. B. de L. “Utilização de Redes Neurais Artificiais em Colheitas de Cana-de-açúcar para Predição de PCC, TCH e Fibra”. *ALCOOLBRÁS*, São Paulo - Brasil, v. 90, 2005, p. 60-63.

[10] Oliveira, R. S. F., et. al. “Intelligent Support Decision in Sugar Cane Harvest”. In: IV World Congress on Computers in Agriculture (WCCA 2006), Orlando – USA, Julho, 2006, p. 456-462.