

Paulo Augusto da Costa Filho e Ronei Jesus Poppi

Instituto de Química - Universidade Estadual de Campinas - CP 6154 - 13083-970 - Campinas - SP

Recebido em 30/4/98; aceito em 2/9/98

GENETIC ALGORITHM IN CHEMISTRY. Genetic algorithm is an optimization technique based on Darwin evolution theory. In last years its application in chemistry is increasing significantly due the special characteristics for optimization of complex systems. The basic principles and some further modifications implemented to improve its performance are presented, as well as a historical development. A numerical example of a function optimization is also shown to demonstrate how the algorithm works in an optimization process. Finally several chemistry applications realized until now is commented to serve as parameter to future applications in this field.

Keywords: chemometrics; optimization; genetic algorithm.

INTRODUÇÃO

Em química muitas vezes surgem situações onde se deve considerar uma série de fatores que exercem influência sobre um determinado processo. Devemos então decidir qual desses fatores, ou em quais níveis (quantidades), que irão produzira a melhor resposta baseada em algum parâmetro pré-estabelecido. Neste caso, deseja-se realizar uma otimização numérica do sistema, ou seja, pretende-se encontrar as condições que irão fornecer a melhor resposta desejada baseada nos parâmetros quantificáveis.

Por exemplo, em uma determinada síntese, deseja-se determinar quais valores de temperatura, pH e quantidade de catalisador que irão fazer com que numa determinada reação obtenha-se o maior rendimento possível. Outro exemplo seria encontrar quais as melhores condições operacionais de um equipamento para realizar-se uma análise. Como em cromatografia gasosa, onde seria determinada qual a melhor temperatura da coluna e a melhor vazão do gás de arraste que proporcione a separação dos picos.

A utilização de métodos para otimização numérica em química já é explorada há bastante tempo, podendo ser considerada uma das primeiras aplicações da disciplina que hoje se conhece como quimiometria¹. A quimiometria, na definição atualmente mais aceita, pode ser considerada como a disciplina que utiliza métodos matemáticos e estatísticos, assim como de lógica matemática, a problemas de origem química.

Os métodos normalmente utilizados para a otimização de sistemas químicos podem ser divididos em dois grandes grupos. Tem-se os métodos baseados em gradientes e os que não são. Nos métodos baseados em gradiente² explora-se o fato de que a partir de processos derivativos pode-se encontrar os mínimos e máximos. Os métodos que não são baseados em derivadas encontram-se assim no outro grupo, e dentre os mais largamente conhecidos e empregados em química destaca-se o método simplex³. Este se baseia no deslocamento do simplex, que é uma figura geométrica, sobre a superfície de resposta para encontrar os máximos e mínimos.

Mais recentemente tem surgido na literatura e causado grande impacto na comunidade científica o chamados métodos de inteligência artificial⁴. A inteligência artificial faz parte dos processos que simulam via "software" ou "hardware", o talento dos seres racionais para a representação e processamento de

informações. Na área de otimização uma técnica de inteligência artificial que vem sendo muito estudada atualmente é o algoritmo genético. Este é um processo de otimização baseado nos processos biológicos da evolução das espécies, simulando matematicamente a teoria de Darwin.

Este método de otimização tem recentemente recebido grande destaque na área de química, e neste artigo pretende-se introduzir os seus conceitos básicos, pois se percebe que ele ainda não está difundido dentro da comunidade química do país. Pretende-se também apresentar várias aplicações realizadas até o momento, para que seja possível ter uma visão do potencial da técnica e estimular a sua utilização.

ALGORITMO GENÉTICO

Histórico

A teoria da evolução das espécies originou-se das observações realizadas por Charles R. Darwin, durante uma expedição científica ao redor do mundo. Nesta ocasião Darwin assumia a tarefa de naturalista do navio *Eagle*⁵. Durante esta jornada que durou cinco anos, Darwin presenciou inúmeras situações que o levaram a pensar sobre a possível seleção natural das espécies. No arquipélago dos Galápagos, ele verificou que cada ilha possuía uma fauna característica, principalmente com relação a pássaros, tartarugas e lagartos. Em 1858, Darwin apresentou a teoria da seleção natural, numa sessão da Academia Britânica de Ciências. No ano seguinte publicou um resumo do livro "Natural Selection", onde introduzia a teoria da Evolução das espécies. Segundo Darwin, no mecanismo de seleção natural, os indivíduos mais adaptados têm maior probabilidade de sobrevivência do que os menos adaptados. Como consequência disso, os mais adaptados teriam maior número de descendentes, o que provocaria, na geração seguinte, um aumento no seu número, em relação aos não adaptados. Tais adaptações levam a evolução biológica, a qual ocasiona mudanças nas propriedades das populações dos organismos que transcendem o período de vida de um único indivíduo. As mudanças nas populações que são consideradas evolutivas são aquelas herdáveis via material genético, de uma geração para a outra. A evolução biológica pode ser pequena ou substancial, sendo estes efeitos responsáveis pelas adaptações dos organismos a diferentes ambientes⁵.

Com base na teoria evolutiva das espécies, John H. Holland um pesquisador da Universidade de Michigan, propôs nos anos 60 a construção de um algoritmo matemático para otimização

e-mail: ronei@iqm.unicamp.br

em sistemas complexos, sendo denominado de Algoritmo Genético⁶. Este algoritmo pretendia simular matematicamente todo o mecanismo da evolução biológica, com todas as características e vantagens desse processo.

A partir dos anos oitenta, o algoritmo genético recebeu um grande impulso em diversas áreas de aplicação científica, devido principalmente à versatilidade e excelentes resultados apresentados. A popularização dos computadores e o aparecimento de sistemas cada vez mais rápidos e potentes, também auxiliaram muito o seu desenvolvimento. O algoritmo genético tem sido aplicado nas mais variadas áreas, podendo-se destacar sua utilização em processamento de imagem; modelagem e identificação de sistemas; filtros para cancelamento de ruídos; robótica; e seleção de parâmetros de redes neurais⁷.

Mais recentemente, o algoritmo genético começou a ser utilizado em química, uma vez que nesta área problemas complexos de otimização também são normalmente encontrados. A primeira aplicação de algoritmo genético em química foi realizada no início dos anos noventa por Lucasius e Kateman⁸, num trabalho exploratório para seleção de comprimentos de onda em espectros no ultravioleta para análise de quatro nucleotídeos.

As aplicações deste algoritmo cresceram muito nos últimos anos, principalmente devido às vantagens que este apresenta quando comparado a outros métodos de otimização. Como principais vantagens pode-se destacar⁶:

- i - Não requer informações sobre o gradiente da superfície de resposta;
- ii - Eventuais descontinuidades da superfície de resposta não afetam a performance da otimização;
- iii - A presença de mínimos locais não reduzem a eficiência do algoritmo;
- iv - A performance do algoritmo tem apresentado excelentes resultados para problemas de otimização de grande escala.

Algoritmo Genético básico

O algoritmo genético foi estruturado de forma que as informações referentes a um determinado sistema pudessem ser codificadas de maneira análoga aos cromossomos biológicos. Desta forma o algoritmo proposto assimila-se muito ao processo evolutivo natural.

O algoritmo genético básico envolve cinco passos: codificação das variáveis, criação da população inicial, avaliação da resposta, cruzamento e mutação.

Codificação das variáveis

O cromossomo biológico é composto de genes, que são responsáveis por determinadas características do indivíduo, como por exemplo, cor dos olhos, altura, cor da pele, etc. Através de uma analogia é possível construir um cromossomo artificial. Num processo químico cada gene representará um determinado parâmetro que será otimizado como, por exemplo: temperatura, agitação, tempo, quantidade de catalisador (conforme figura 1).

Neste tipo de sistema, a forma mais simples de representar as informações relativas a cada parâmetro que será otimizado (como temperatura, catalisador, agitação e quantidade de catalisador), é codificando-as em números binários. Assim cada cromossomo artificial será representado por uma sequência de zeros e uns (forma binária para cada parâmetro). Esse tipo de codificação apresenta inúmeras vantagens no decorrer do processo de otimização como tornar as operações genéticas mais fáceis de serem executadas; tornar possível o trabalho com variáveis que assumam valores contínuos; facilitar a variação da faixa de trabalho dos parâmetros; além de apresentar relativa simplicidade na variação da precisão dos valores de cada parâmetro.

A Figura 1 mostra a codificação dos valores relativos aos quatro parâmetros que serão otimizados. Neste caso deve-se

Temperatura	Catalisador	Agitação	Tempo
35	8	43	56
1 0 0 0 1 1	0 0 1 0 0 0	1 0 1 0 1 1	1 1 1 0 0 0

Figura 1. Codificação dos parâmetros Temperatura, Catalisador, Agitação e Tempo de valores decimais para valores binários de 6 bits. Por exemplo, para a Temperatura que tem valor decimal igual a 35, após a conversão tem-se a seguinte sequência de zeros e uns : 100011.

codificar para binário o valor de temperatura igual a 35 graus, o valor do catalisador igual a 8 gramas, o valor de agitação igual a 43 rpm e o valor do tempo igual a 56 minutos. Tomando como 6 bits cada um dos parâmetros a serem otimizados, tem-se assim o cromossomo formado por 24 unidades binárias, sendo que os seis primeiros formam o gene correspondente à temperatura, e na sequência de 6 bits os genes para catalisador, agitação e tempo.

CRIAÇÃO DA POPULAÇÃO INICIAL

Através de um gerador randômico, cria-se uma população inicial de cromossomos, como mostrado na Figura 2 (neste caso uma população de sete indivíduos, ou cromossomos).

```

[1] 1000000000010010001010100000
[2] 00010010010010000000000100100
[3] 100001000000000000000100000001
[4] 000000000100000000000100000000
[5] 010001000000010000000000000000
[6] 000000000100000000010000000100
[7] 000001000000000100000001000100

```

Figura 2. Representação de 7 cromossomos gerados de maneira aleatória que serão utilizados para o início do processo de seleção, onde a sequência de bits está relacionada com as variáveis Temperatura, Catalisador, Agitação e Tempo conforme mostrado na figura 1.

Sendo a população inicial aleatória, garantimos que não haverá nenhum tipo de influência tendenciosa do meio externo. Desta forma tenta-se garantir a idoneidade da pesquisa. Entretanto há casos onde existem informações adicionais que podem ser introduzidas na geração da população inicial, minimizando o tempo de computação, ou seja, o tempo para encontrar o valor ótimo.

Avaliação da resposta (Aptidão)

Deve-se encontrar o valor associado à performance de cada cromossomo, relacionado ao sistema de interesse. A avaliação da resposta de cada cromossomo é o resultado mais importante no procedimento do algoritmo genético.

A resposta (aptidão) é uma característica intrínseca ao indivíduo. Esta indicará, a nível biológico, qual é a habilidade que um indivíduo possui para sobreviver a predadores, pestes e a outros obstáculos para a subsequente reprodução. Transportando para o

algoritmo matemático, seria sua habilidade de produzir a melhor resposta. Esta resposta, na maioria das vezes, é avaliada em termos de um erro calculado pela diferença entre o valor esperado e o valor previsto. O objetivo é encontrar o menor erro possível, e este será o responsável direto pela vida ou morte dos indivíduos.

Em alguns algoritmos, será esta resposta que irá determinar qual a probabilidade com que um determinado cromossomo irá cruzar. Neste caso, determinados indivíduos poderão realizar mais de um cruzamento durante uma mesma geração.

Cruzamento

Nesta etapa do algoritmo é realizado o cruzamento entre material genético de diferentes indivíduos. A partir da atual população, uma nova é formada pelo cruzamento aleatório entre os cromossomos. Portanto os filhos recebem informações de seus progenitores, através do material genético proveniente deste cruzamento.

Durante a permuta de material genético entre os indivíduos, haverá uma tendência da transmissão das características dominantes para as gerações futuras. Após algumas gerações pode-se observar que uma elevada taxa dos indivíduos possuem a presença de variáveis 'dominantes'. O cruzamento é responsável pela convergência para a situação de otimização desejada.

O processo de cruzamento dá-se pela escolha aleatória dos indivíduos e da posição onde ocorrerá a ruptura seguida pela troca de material genético. O número de rupturas geralmente varia entre 1 e 2. Isto se deve ao fato de que quanto maior for o número de rupturas menor será a semelhança entre os pais e os filhos, portanto dificultando a convergência.

A Figura 3 mostra a etapa do cruzamento entre dois indivíduos, para o gene (pedaço do cromossomo) relacionado à variável catalisador, a qual está sendo otimizada.

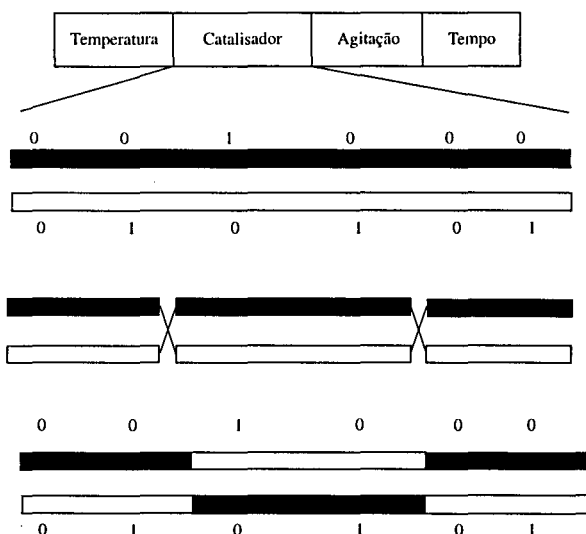


Figura 3. Etapa de cruzamento no gene referente à variável Catalisador. Neste caso ocorre : (a) o emparelhamento de dois cromossomos (um representado pela cor branca e outro pela cor negra) ; (b) rompimento e permuta do material genético ; (c) geração de dois novos cromossomos.

Mutação

Durante o processo de reprodução pode ocorrer uma perturbação das informações contidas dentro dos genes ocorrendo uma mutação. Esta promove alterações no código genético de uma pequena parcela da população, onde muitas vezes pode atuar de forma benigna. Se a alteração gerar indivíduos mais

adaptados, estas novas características serão transmitidas para os demais indivíduos ao longo das próximas gerações.

Este mecanismo ajuda a solucionar o problema do confinamento a mínimos locais na otimização, pois promove alterações que direcionam a pesquisa para outros locais da superfície de resposta. A Figura 4 representa dois cromossomos, onde ocorreu mutação no material genético. A mutação na representação binária é realizada pela troca de um por zero e vice-versa, conforme pode ser notado na Figura 4.

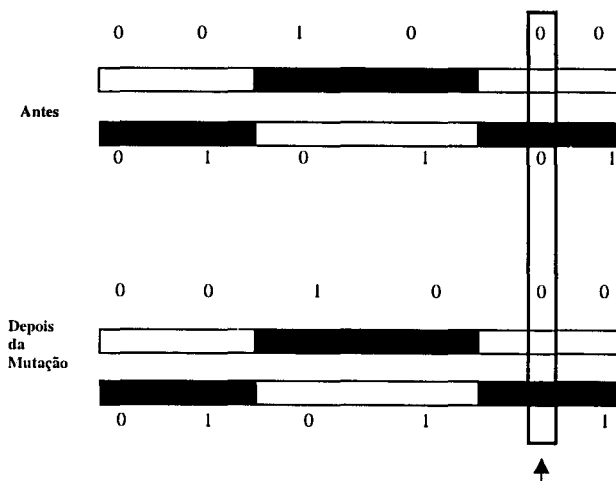


Figura 4. Operação de mutação nos cromossomos após o cruzamento. Neste caso, em uma das posições, tem-se a troca de um bit cujo valor é "0" pelo valor "1", gerando um novo indivíduo.

PRINCIPAIS ETAPAS DO ALGORITMO GENÉTICO BÁSICO

A Figura 5 mostra as principais etapas do algoritmo genético básico. Os passos 2 a 6 são continuamente repetidos até que um critério de parada seja satisfeito. Como critério de parada para o algoritmo genético, normalmente adota-se um número máximo de gerações ou até que um erro mínimo desejado seja alcançado.

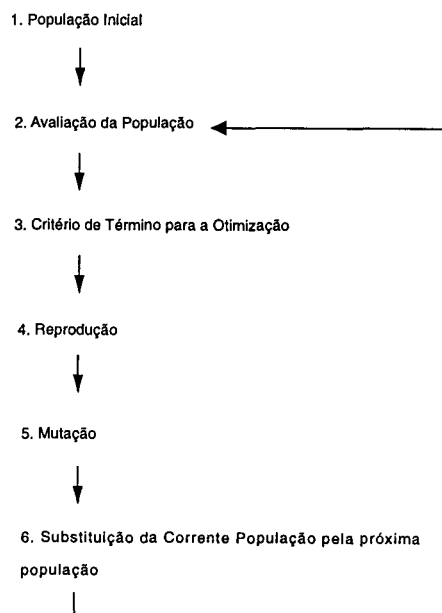


Figura 5. Fluxograma mostrando as principais etapas do algoritmo genético básico.

Basicamente existem quatro tipos principais de algoritmos genéticos: Algoritmo genético de Holland⁶, programa genético de Koza³, programa evolucionário de Fogel¹⁰ e estratégia evolucionária de Rechenberg¹¹. Eles se diferem pelo mecanismo de código genético, procedimento de seleção, e reprodução. Entretanto, o mais popular é o de Holland, que codifica as informações na forma de código binário (0,1), e é o algoritmo que está sendo tratado neste trabalho.

Para uma boa performance do algoritmo faz-se necessário otimizar parâmetros como: probabilidade de mutação, probabilidade de cruzamento, tamanho da população inicial e número de gerações. O tamanho da população inicial é escolhido normalmente entre 50-500 indivíduos^{6,7}. Utiliza-se probabilidade de mutação de aproximadamente 1%, e probabilidade de cruzamento por volta de 90%^{6,7}. O número de gerações irá depender do tamanho da população inicial, existindo um comprometimento entre o tamanho da população inicial e o número de gerações. Para um tamanho de população muito grande, muitas vezes torna-se impraticável, em termos de tempo de processamento, realizar-se um grande número de gerações. Por outro lado, com uma população maior pode-se realizar uma pesquisa mais completa em torno da superfície de resposta.

OPERAÇÕES E TÉCNICAS AVANÇADAS

Após o desenvolvimento do primeiro algoritmo genético inúmeras modificações foram propostas com o intuito de aperfeiçoar a performance e robustez do programa.

Os primeiros algoritmos eram compostos de basicamente quatro etapas: geração da população inicial, cruzamento, mutação e avaliação. Atualmente em programas mais sofisticados pode-se encontrar inúmeros outros operadores genéticos como: dominância, distinção de sexo e mecanismos para evitar incestos. Também para diminuir o tempo de processamento do programa, a computação paralela tem sido implementada nesse algoritmo.

DOMINÂNCIA

A dominância é estabelecida quando um determinado gene é responsável por uma característica que faz com que um indivíduo se destaque entre um determinado grupo. Desta forma, torna-se interessante a preservação desta informação, e a propagação para as gerações futuras. Isso pode fazer com que ocorra uma aceleração do processo de convergência para o ótimo.

O operador de dominância possui a função de localizar e preservar os genes responsáveis pelo destaque em sua performance. Isso é implementado no algoritmo matemático fazendo com que após um determinado número de gerações, inicie-se uma sub-rotina que emparelhe dois cromossomos e crie um novo indivíduo que possua as características dominantes dos progenitores. Por exemplo, pode-se considerar que os bits "zeros" conterão informações dominantes e os bits "uns" recessivas. Neste tipo de operação, conforme pode observado no esquema apresentado na Figura 6, nos cruzamentos entre "zeros" e "uns" prevalecerá sempre o bit "zero". Portanto além da população proveniente do cruzamento haverá também o acréscimo de mais estes novos indivíduos na população final desta geração.

DISTINÇÃO DE SEXOS

O cruzamento é a operação onde novos indivíduos são gerados a partir de cromossomos pais. O cruzamento pode apresentar-se de três formas distintas: homossexual, heterossexual e assexuado.

O cruzamento por distinção de sexo é realizado através de um bit no cromossomo reservado para distinguir o sexo. Por exemplo, pode-se representar um indivíduo do sexo masculino como tendo o bit "1", em uma determinada posição do cromossomo. A partir deste bit é possível controlar o cruzamento entre os indivíduos.

Nos cruzamentos homossexuais tem-se o cruzamento entre

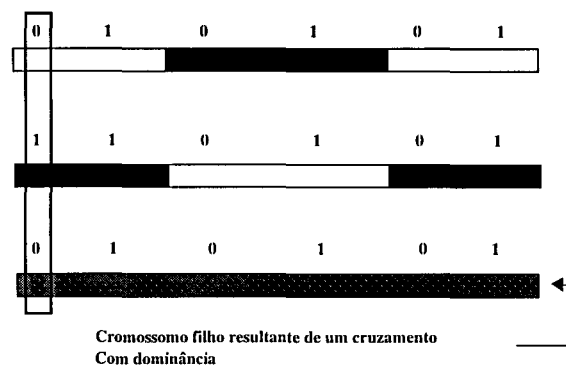


Figura 6. Operação de cruzamento com dominância. Esta operação genética não envolve ruptura dos cromossomos, apenas a comparação dos bits em cada posição dos cromossomos pais. Nessa comparação sempre irá prevalecer o bit dominante, conforme ilustrado na figura.

cromossomos de mesmo sexo, enquanto nos heterossexuais realiza-se o cruzamento entre cromossomos de sexo distintos. Existem também os assexuados, onde ocorre a troca de material genético dentro do próprio indivíduo.

O controle pela distinção de sexo faz com que a diversificação seja mais ampla. Determinadas características desejáveis muitas vezes somente podem ser encontradas num determinado sexo. Isso pode ser observado na natureza, já que apenas seres inferiores são hermafroditos.

MECANISMOS PARA EVITAR INCESTOS

No incesto tem-se o cruzamento entre indivíduos que possuem cromossomos muito semelhantes (pais, irmãos ou parentesco muito próximo), não levam a gerações futuras mais adaptadas. O cruzamento entre indivíduos muito semelhantes leva a formação de um cromossomo muito parecido ao dos pais. Desta forma não há nenhuma melhora significativa para as novas gerações. Muitas vezes, implementa-se nos procedimentos uma sub-rotina que evite o incesto durante a etapa de cruzamento. Isso é realizado evitando-se que cromossomos muito semelhantes se cruzem.

COMPUTAÇÃO PARALELA

Programas onde existe um grande número de operadores genéticos tornam-se demorados, uma vez que muitas operações devem ser realizadas. Com o objetivo de tornar a execução do programa mais rápida, tem-se proposto o uso da computação paralela¹² minimizando o tempo de processamento. A computação paralela é caracterizada por ser realizada em um ambiente onde há múltiplos processadores, e uma rede interligando os processadores através da distribuição de partes do programa a serem executadas em diferentes máquinas simultaneamente.

A construção de um programa em paralelo requer que o programador realize a decomposição do algoritmo, ou dos dados, em várias partes; distribua essas partes para os diferentes processadores que estão trabalhando simultaneamente; e coordene o trabalho e a comunicação entre esses processadores.

No algoritmo genético a computação paralela pode ser implementada na determinação da resposta de cada indivíduo, em cada geração. Esse cálculo, que muitas vezes é a etapa mais complexa dos cálculos, pode ser distribuído para ser realizada em diferentes processadores.

EXEMPLO DA OPERAÇÃO DE UM ALGORITMO GENÉTICO

Para melhor exemplificar o funcionamento do algoritmo genético, pode-se demonstrar sua aplicação na determinação do ponto máximo da seguinte função:

$$Z = (3 \times (1 - X)^2 \times \exp(-X^2 - (Y + 1)^2)) - (10 \times (\frac{X}{5} - X^3 - Y^5) \times \exp(-X^2 - Y^2)) - (\frac{1}{3} \times \exp(-(X + 1)^2 - Y^2) + \frac{3}{5} \times (\frac{17}{10} \times \sin(\pi \times X) + \frac{17}{10} \times \cos(2 \times Y)))$$

com o domínio de X e Y de -3 à +3, conforme pode ser observado na superfície de resposta da Figura 7. Este é um exemplo de uma função complexa, com máximos locais, além do máximo global.

Neste caso, pretende-se determinar quais os valores de X e Y que, dentro domínio especificado, irão produzir o maior valor de Z.

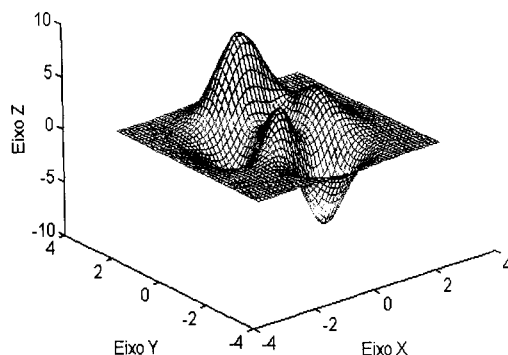


Figura 7. Superfície correspondente a função exemplificada acima.

Para a otimização da função, foram estabelecidos cromossomos com dois genes, sendo cada um destes sendo referente a uma variável independente da função (valores de X e Y). O próximo passo foi representar cada gene por um código binário com uma precisão de 0,00003 unidade. Para isto foram utilizados 17 bits, sendo 16 para representar cada valor numérico com a precisão requerida e 1 bit para representar o sinal do número (valores positivos ou negativos).

Desta maneira, cada cromossomo é representado por 34 unidades binárias, onde os 17 primeiros estão relacionados ao valor de X e os 17 últimos (outro gene) ao valor de Y. Assim a cada cromossomo (que representa os valores de X e Y), pode-se associar uma resposta que é representada pelo valor correspondente de Z da função. O objetivo do algoritmo genético é então encontrar qual cromossomo que irá produzir o maior valor de Z, dentro do domínio especificado.

Pelo fato dos cromossomos possuírem uma dimensão relativamente pequena, foram criados 8 cromossomos (ou seja, a população tem 8 indivíduos) e foram realizadas 30 gerações (reproduções).

Após a realização dos cruzamentos correspondentes às 30 gerações propostas, pode-se avaliar as respostas dos melhores indivíduos em cada geração como mostrada na Figura 8.

Analisando-se o gráfico mostrado na Figura 8 podem ser observados três patamares bastante distintos. O primeiro patamar, com valores de Z próximos a 1,0 encontra-se relacionado com um vale localizado entre 3 máximos e 2 mínimos, portanto trata-se de um local complexo para iniciar uma análise segundo os modelos tradicionais de otimização. O segundo patamar encontra-se relacionado à determinação de um máximo local, que perdurou como uma situação de pseudo-ótimo por 4 gerações. Somente após 6 gerações iniciais determina-se a existência do máximo global que se encontra caracterizado pelo terceiro patamar da Figura 8 (com valor de Z próximo a 8,0).

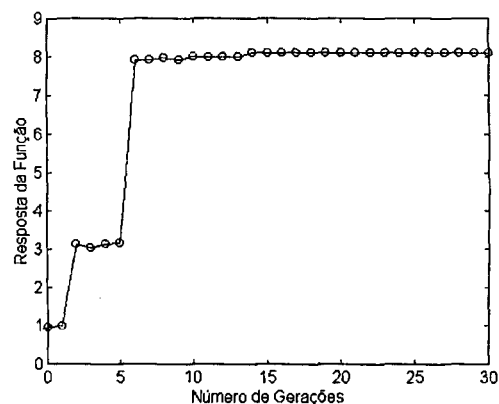


Figura 8. Resposta dos melhores indivíduos de cada geração.

Aplicação do algoritmo genético em química

Na química tem-se observado uma acentuada tendência na aplicação de algoritmo genético em inúmeros problemas de otimização de sistemas complexos essencialmente devido à robustez e versatilidade do algoritmo.

Tem-se utilizado o algoritmo genético para otimização de parâmetros internos (pesos) de rede neural¹³. Uma de suas aplicações é o treinamento da rede para análise quantitativa de espectros de fluorescência de raios-X¹⁴.

O algoritmo genético também tem sido utilizado no mapeamento de átomos, para determinar a distância mínima entre eles¹⁵. Outro campo que está sendo explorado é a pesquisa de planejamentos experimentais, que consiste no planejamento de "n" experimentos, onde há variação de alguns fatores para otimizar o sistema.¹⁶

Muitos sistemas químicos apresentam um alto nível de complexidade, onde muitos parâmetros necessitam ser otimizados num único processo. Para solucionar este tipo de problema utiliza-se uma poliotimização. Um exemplo de algoritmo genético em poliotimização em química analítica é a otimização de distâncias entre picos cromatográficos, como função das condições cromatográficas (composição da fase móvel, temperatura do injetor, valor de pH, etc.)¹⁷. Pode-se encontrar outro exemplo desse tipo de aplicação na determinação de glicose em sangue humano por fotometria. Essa poliotimização envolveu 12 parâmetros químicos e tecnológicos, com o intuito de monitorar seis propriedades ópticas. O resultado obtido foi comparado com os dos métodos clássicos regressão dos mínimos quadrados parciais não linear¹⁸.

Encontra-se ainda exemplos da aplicação nos seguintes casos : 1) determinação da constante de estabilidade de sistemas de cloreto de cádmio em meio perclorado com chumbo e éter coroa diclorohexil-18-coroa-6, a partir de dados calorimétricos e polarográficos¹⁸; 2) seleção e detecção de amostras anômalas dentre um conjunto de amostras, para construção do modelo de calibração.¹⁹⁻²⁰; 3) otimização da separação de bases fracas ou soluções ácidas por cromatografia líquida de fase reversa com variação de pH e modificação orgânica na fase móvel²¹; 4) otimização da estimativa dos parâmetros cinéticos de uma reação²²; 5) otimização da intensidade linhas de emissões atômicas de traços de elementos em pó de alumina como uma função da condição de excitação²³; 6) análise conformacional de biopolímero²⁴.

Provavelmente a aplicação mais difundida do algoritmo genético está na seleção de comprimentos de onda em determinações espectroscópicas, principalmente no infravermelho próximo, utilizando calibração multivariada²⁵. Geralmente nestes casos utiliza-se o método dos mínimos quadrados parciais²⁵ ou a regressão linear múltipla²⁶ para a construção de um modelo que possa correlacionar os espectros com as espécies de interesse. Este tem como objetivo determinar o número mínimo de

comprimentos de onda que possuam informação relevante para a análise que está sendo realizada.

Devido ao interesse despertado na seleção de variáveis, resolvemos dedicar um tópico deste trabalho a este tipo de aplicação, conforme apresentado a seguir.

Algoritmo genético na seleção de variáveis.

Tem-se observado nos últimos anos uma tendência na implementação de técnicas espectroscópicas em análises químicas rotineiras. Isto se deve essencialmente à simplicidade, rapidez, baixo custo, versatilidade, bem como outros atributos de diferenciação proporcionados pela técnica.

A calibração multivariada contribuiu de forma significativa para a consolidação desta técnica, viabilizando a determinação simultânea de mais de uma espécie mesmo em situações adversas, onde os métodos convencionais de análises seriam difíceis ou impossíveis de serem realizados (por exemplo, determinações "on-line" em processos). Tornou também análises qualitativas e quantitativas possíveis, bem como a detecção de amostras anômalas.

Em análises espectrofotométricas, muitas das aplicações de calibração multivariada deste método envolvem um elevado número de variáveis. Muitas vezes o uso de somente algumas das variáveis que contém mais informações, podem proporcionar maior segurança e facilidade na interpretação do modelo. Assim as variáveis (absorbância em diferentes comprimentos de onda) que possuem uma menor relação sinal/ruído ou/e não linearidade podem ser descartadas.

Atualmente o número de algoritmos que se propõe a selecionar variáveis é pouco expressivo. Dentre estes se destacam a regressão linear múltipla por passos (stepwise multiple linear regression)²⁷ e os que empregam o uso do algoritmo genético.

A implementação do algoritmo genético na seleção de variáveis difere-se das demais aplicações apresentadas no que tange à codificação do problema e a função de resposta, já que as outras etapas permanecem inalteradas.

Na codificação do problema em seleção de variáveis de espectros, considera-se que o cromossomo possui "p" genes, onde cada gene representa uma das variáveis do espectro (absorbância em cada comprimento de onda). Assim sendo, o cromossomo terá o mesmo número de variáveis contidas no espectro, conforme representada na Figura 9.

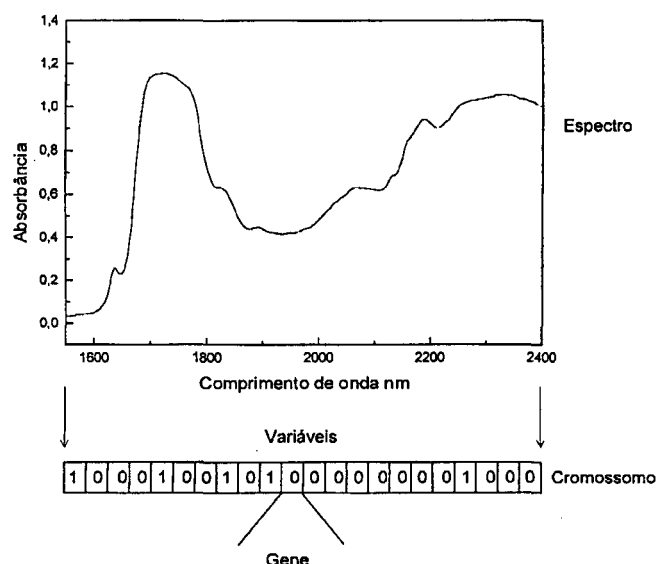


Figura 9. Codificação de um espectro na forma de um cromossomo. Neste caso cada valor de comprimento de onda corresponde a um gene do cromossomo.

Quando se utiliza o algoritmo genético para a seleção de variáveis, é necessário, dependendo do modelo de calibração, utilizar algumas "condições de contorno", como por exemplo, pré-estabelecer o número máximo de variáveis selecionadas, como no caso da regressão linear múltipla.

Na seleção de variáveis utiliza-se também o auxílio do código binário (0,1) para codificar o problema. Cada gene pode assumir o valor *um* ou *zero*, sendo que quando a posição referente a uma determinada variável for igual a *um*, isto implica na seleção desta variável. Por outro lado, se a posição conter o valor *zero*, isto indica que a variável não foi selecionada.

CONCLUSÃO

Observando as inúmeras vantagens apresentadas pelo algoritmo genético, chega-se a conclusão que suas aplicações tendem a crescer na área da química, principalmente onde métodos de otimização mais tradicionais não apresentam resultados satisfatórios.

As vantagens que devem ser ressaltadas são: versatilidade e robustez do modelo. Por ser um algoritmo versátil, torna-se fácil introduzir adaptações ao programa de forma a atender às diferentes características de cada sistema que está sendo otimizado.

Muitas vezes, para sistemas muito complexos, este tipo de otimização pode consumir um grande tempo de processamento. Este problema pode ser reduzido com o emprego da computação paralela, onde de partes do programa que requerem mais tempo podem ser executados simultaneamente.

Associando o algoritmo genético a técnicas quimiométricas já estabelecidas de otimização e análise multivariada, pode ser possível o desenvolvimento de modelos matemáticos mais simples, facilitando sua análise e entendimento.

É importante ressaltar que esta técnica de otimização como qualquer outra não garante que a melhor condição de uma otimização tenha sido determinada. Contudo tem-se observado que o algoritmo genético, em alguns casos apresenta resultados melhores que outras técnicas de otimização.

REFERÊNCIAS

1. Sharaf, M. A.; Illman, D. L.; Kowalski, B. R.; *Chemometrics*, John Wiley & Sons, New York 1986.
2. Jang, J. S. R.; Sun, C. T.; Mizutani, E.; *Neuro-Fuzzy and Soft Computing*, Prentice Hall, Upper Saddle River 1997.
3. Bruns, R. E.; Scarminio, I. S.; Neto, B. B.; *Planejamento e Otimização de Experimentos*, Editora da Universidade Estadual de Campinas 1995.
4. Cartwright, H. M.; *Applications of Artificial Intelligence in Chemistry*, Oxford Science Publications, New York 1995.
5. Ruse, M.; *Darwinism Defended. A Guide to the Evolution Controversies* John Wiley & Sons, New York 1982.
6. Goldberg, D. E.; *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading 1989.
7. Lucasius, C. B. and Kateman, G.; *Chemom. Intell. Lab. Syst.* **1993**, 19, 1.
8. Lucasius, C. B. and Kateman, G.; *Trends Anal. Chem.* **1991**, 10, 254.
9. Koza, J. R.; *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, Cambridge University Press 1992.
10. Fogel, L. J.; Owens, A. J. and Walsh, M. J.; *Artificial Intelligence Through Simulated Evolution*, Wiley 1996.
11. Rechenberg, I. *Evolutionstrategie: Optimierung Technischer Systeme nach Prinzipien der Biologischen Evolution*, Frommann-Holzboog 1973.
12. Snir, M.; Otto, S.; Huss-Lederman, S.; Walker, D.; Dongarra, J.; *MPI: The Complete Reference*, The MIT Press-Cambridge, Massachusetts 1996.

13. Zupan, J.; Gasteiger, J.; *Neural Networks for Chemists*, VCH, Weinheim 1993.
14. Bos, M. and Weber, H. T.; *Anal. Chim. Acta* **1991**, 247, 97.
15. Fontain, E.; *Anal. Chim. Acta* **1992**, 265, 227.
16. Broudiscou, A.; Leardi, R.; Phan-Tan-Luu, R. *Chemom. Intell. Lab. Syst* **1996**, 35, 105.
17. Wienke, D.; Lucasius, C.; Ehrlich, M. and Kateman, G.; *Anal. Chim. Acta* **1993**, 271, 253.
18. Hartnett, M. K.; Bos, M.; van-der-Linden, W. E.; and Diamond, D.; *Anal. Chim. Acta* **1995**, 316, 347.
19. Leardi, R.; *J. Chemom.* **1994**, 8, 65.
20. Walczak, B.; *Chemom. Intell. Lab. Syst.* **1995**, 28, 259.
21. Lopes-Marques, R. M.; Schoenmakers, P. J.; Lucasius, C. B. and Buyden, L.; *Chromatographia* **1993**, 36, 83.
22. Hibbert, D. B.; *Chemom. Intell. Lab. Syst.* **1993**, 19, 319.
23. Wienke, D.; Lucasius, C.; Kateman, G.; *Anal. Chim. Acta* **1992**, 265, 211.
24. Blommers, M. J. J.; Lucasius, C. B.; Kateman, G. and Kaptein, R.; *Biopolymers* **1992**, 32, 45.
25. Bangalore, A. S.; Shaffer, R. E.; Small, G. W. and Arnold, M. A.; *Anal. Chem.* **1996**, 68, 4200.
26. Jouan-Rimbaud, D.; Massart, D.L.; Leardi, R. and De-Noord, O. E.; *Anal. Chem.* **1995**, 67, 4295.
27. Martens, H. and Naes, T.; *Multivariate Calibration*, John Wiley & Sons, New York 1989.