

Aplicação da Técnica de Arrefecimento Simulado na Solução do Problema de Distância Geométrica: Uso na Visualização de Moléculas a Partir de Dados RMN

Miscelânea Pedrosa de Araújo

Emerson Alexandre de Oliveira Lima

Programa de Pós-Graduação em Engenharia de Sistemas, PPGES-UPE

Escola Politécnica da Universidade de Pernambuco

Rua Benfica, 455 - Madalena - Recife/PE CEP: 50720-001

E-mail: misspedrosa@gmail.com, eal@poli.br,

RESUMO

As proteínas exercem diversas funções nos organismos e a maior parte do comportamento e propriedades das proteínas depende unicamente de sua estrutura geométrica tridimensional ([2]). Técnicas de análise tais como a ressonância magnética nuclear ou RMN permitem conhecer as distâncias entre *dipletos* - conjunto de dois átomos consecutivos de uma cadeia - as distâncias e ângulos entre *tripletos* - conjuntos de três átomos consecutivos da cadeia - e as distâncias, ângulos, e ângulos de torção entre *quadripletos* - conjuntos de quatro átomos consecutivos de uma cadeia - sempre que tais distâncias não ultrapassem o limite tecnológico de 5\AA .

O problema da conformação molecular é uma subclasse do problema mais geral chamado de problema da Distância Geométrica ([3],[4],[6] e [5]) que consiste, de forma simplificada, em obter uma representação de um grafo no plano de tal maneira que as distâncias - supostas previamente conhecidas - entre os vértices sejam obedecidas.

O problema geral da distância geométrica, no qual todas as distâncias entre quaisquer pares de vértices são conhecidas, é um problema de solução em tempo $\mathcal{O}(n)$ não sendo, portanto, complexo. Na prática, contudo, quando não se conhecem todas as distâncias envolvidas¹, mas apenas um subconjunto destas ou quando os sempre presentes erros de medição estão envolvidos de forma que temos não as distâncias mas um intervalo no qual cada distância efetiva deve estar situada, temos um problema NP-Difícil ([3]).

Neste trabalho aplicamos as instâncias deste problema específico, chamado de modelo molecular linear discreto ou modelo de Lavor, o algoritmo clássico de arrefecimento simulado ([6]) - *Simulated Annealing* - descrito no próximo tópico, correlacionando o tempo necessário, medido como o número de vezes que a função objetivo foi calculada antes de obter a convergência do método, para obtenção de uma solução do mesmo com o número de átomos presentes em dada molécula.

Para testar o algoritmo de arrefecimento simulado (SA) aplicado ao problema da conformação molecular, foram seguidos os seguintes passos metodológicos: 1. Geração de todas as moléculas (instâncias de Lavor) para $n = 10$ ($2^{10-3} = 128$ moléculas), $n = 11$ ($2^8 = 256$ moléculas) e $n = 12$ ($2^9 = 512$ moléculas); 2. Uso de 10 casos aleatoriamente escolhidos de cada instância (valor de n) acima para calibração do método; 3. Geração de 20 dos 2^{100} casos possíveis de moléculas com $100 + 3$ átomos para teste do método calibrado com os parâmetros dos casos de menor valor de n .

Ao gerar cada caso, a informação a ser recuperada pelo algoritmo são as posições (conhecidas) de cada átomo da molécula. Para tanto, o algoritmo foi alimentado com as seguintes informações: Distâncias inter-atômicas de todos os singletos, isto é, pares $(i, i + 1)$, $i = 1, 2, \dots, n - 1$; Distâncias inter-atômicas de todos os dipletos, isto é, pares $(i, i + 2)$, $i = 1, 2, \dots, n - 2$; Distâncias inter-atômicas

¹O caso típico das informações provenientes do RMN

de todos os tripletos, isto é, pares $(i, i + 3)$, $i = 1, 2, \dots, n - 3$ e Distâncias inter-atômicas de todos os pares de átomos (i, j) para os quais tal distância é igual ou menor a 5Å

O conjunto de todos os pares (i, j) cujas distâncias são fornecidas ao algoritmo SA é denotado por S . Os ângulos escolhidos como ângulos inter-atômicos foram todos iguais a 1.91 radianos ([1]). As distâncias entre as ligações foram todas de 1.5226Å .

A calibração do método forneceu como parâmetros que garantiam a convergência para todas as moléculas de 10, 11 e 12 átomos testados: 1. Dado n o número de átomos, o vetor binário inicial utilizado como chute inicial do métodos SA foi o vetor no qual todas as posições são iguais a 0. As primeiras 3 posições são ignoradas pois os primeiros 3 átomos do modelo discreto tem suas posições fixadas; 2. Número máximo de iterações fixadas em 1.000 iterações ou 1% de 2^n (número total de possibilidades), o que for menor; 3. Para lei de atualização de temperatura, $T \leftarrow \alpha T$ onde $\alpha = 0.01$; 4. Para geração de vizinhanças, dado um vetor binário, uma vizinhança é gerada escolhendo aleatoriamente um dos bits do vetor e o invertendo.

Houve, com estes parâmetros, convergência para o valor zero de energia (posição verdadeira encontrada pelo algoritmo) NA TOTALIDADE dos casos testados para $n = 10, 11$ e 12. Note que apenas 10 casos de cada valor de n foram utilizados na calibração do método e foram testados o restante dos $128 + 256 + 521 - 3 \times 10 = 886$ casos.

Para os 20 casos de tamanho $100 + 3$ átomos, o método falhou em convergir para energia zero (solução exata) em 3 das instâncias. Para estas instâncias, fixar o valor máximo de iterações em 5.000 garantiu a convergência do método.

O algoritmo SA mostrou-se bastante adequado a resolução do problema da conformação molecular ao menos em instâncias discretas de moléculas de proteína.

Os parâmetros de calibração sugerem que o algoritmo pode ser melhorado para convergência em casos nos quais o número de iterações torna-se proibitivamente elevado.

Observamos que o algoritmo, mesmo sem fazer uso de derivadas - o que é comum em algoritmos eficientes de otimização - é capaz de recuperar de forma rápida e precisa as posições dos átomos de uma molécula através de seus dados RMN. Isto indica que métodos mais robustos de otimização podem fornecer resultados mais rápidos de convergência.

Trabalhos futuros incluem, além de ampliar a quantidade de átomos da molécula até o limite de tempo de CPU aceitável, testar a robustez do método com moléculas reais, melhorando assim os parâmetros de calibração do algoritmo de arrefecimento simulado. Outras ações futuras incluem o teste e comparação de outros algoritmos de otimização, em particular, envolvendo o uso de derivadas. A elaboração e inclusão de modelos discretos mais complexos de moléculas também podem ser testados em trabalhos futuros assim como o uso de algoritmos de inteligência computacional como alternativa a este problema podem ser interessantes tópicos de pesquisa aplicada em monografias futuras.

Palavras-chave: *Otimização, Arrefecimento Simulado, Conformação Molecular, RMN*

Referências

- [1] Crippen, G.M. and Havel, T.F. "Distance geometry and molecular conformation" Computational Chemistry vol 11 (1988) pp265–266"
- [2] Florence, A.T. and Attwood, D. "Princípios Físico-químicos em farmácia" São Paulo EdUSP 3ªed. (2003)
- [3] Lavor, C.,Liberti, L. and Maculan, N. "Molecular distance geometry problem" Encyclopedia of optimization(2009) pp 2305–2311
- [4] Liberti, L., Lavor, C. and A. Mucherino "Molecular distance geometry methods: From continuous to discrete" International Transactions in Operational Research vol. 18 (2010)
- [5] Moré, J.J. and Wu, Z. "Distance geometry optimization for protein structures" Journal of Global Vol 15 (1999) pp219–234
- [6] William H. Press, Saul Teukolsky, William Vetterling and Brian Flannery "Numerical Recipes. The Art of Scientific Computing" 2sd. Cambridge University Press (2007)