

MODELO HÍBRIDO PARA PREDIÇÃO DE ESTRUTURAS DE PROTEÍNAS USANDO ALPHAFOLD2 E BRKGA

Lucas Gonçalves Erdmann, Nicholas Wojeicchowski, Rafael Stubs Parpinelli

INTRODUÇÃO

Este trabalho propõe um modelo híbrido para predição de estrutura de proteínas que combina o Biased Random-Key Genetic Algorithm (BRKGA) com o AlphaFold2. O modelo utiliza o AlphaFold2 para gerar estruturas proteicas iniciais, que são então refinadas por meio do BRKGA.

DESENVOLVIMENTO

Os experimentos foram realizados em um computador com o sistema operacional Linux Mint 20.3 Cinnamon, equipado com um processador Intel® CoreTM i5-10, com 4 núcleos e 8 threads, operando a 1,6 GHz, e 16 GB de RAM.

M.A. Londe et al. (2025) demonstram, por meio da análise de mais de 150 artigos, que o Biased Random-Key Genetic Algorithm (BRKGA) tem se mostrado promissor na resolução de uma variedade de problemas de otimização relacionados a diferentes áreas, ao mesmo tempo em que destacam como uma abordagem híbrida do BRKGA pode gerar resultados relevantes, embora possa exigir maior poder computacional. Eles apresentam uma revisão detalhada da literatura existente sobre o BRKGA, a fim de identificar os problemas mais estudados, as modificações realizadas no framework e os métodos de hibridização mais frequentemente explorados.

Os experimentos consistiram na análise do gráfico de convergência da função Score3 para cada proteína utilizada, além do cálculo da média e do desvio padrão, e da construção de um boxplot dessa função para o método proposto, Random Search e AlphaFold. O conjunto de proteínas utilizado foi extraído do CASP 16 e consiste em 6 proteínas que variam de 32 a 107 aminoácidos. Deve-se observar que proteínas maiores não foram executadas devido a uma limitação do ambiente colabFold, que permite a execução de proteínas de até 150 aminoácidos em sua versão gratuita. Os parâmetros do experimento incluíram uma população de 30 indivíduos, 1.000 gerações do algoritmo, 5 indivíduos elite e 5 indivíduos aleatórios gerados em cada execução.

O primeiro passo na otimização das estruturas de proteínas geradas pelo AlphaFold foi a extração dos dados dos ângulos ϕ (phi) e ψ (psi) a partir dos arquivos pdb produzidos em cada execução. Esses ângulos representam graus de liberdade essenciais na cadeia polipeptídica e são fundamentais para modelar a conformação espacial da proteína. Cada conjunto de ângulos extraídos foi convertido em um indivíduo dentro da população inicial do algoritmo de otimização. Além disso, o algoritmo de otimização mais eficiente sem comprometer a qualidade da solução. Para aplicar o BRKGA no modelo proposto, foi necessário utilizar um codificador e um decodificador, responsáveis por mapear os ângulos ϕ (phi) e ψ (psi), que variam de -180° a 180° , para números reais no intervalo $[0, 1]$ com 14 casas decimais. Essa estratégia, baseada na abordagem proposta por Marchi, F e Parpinelli, utiliza a representação decimal dos indivíduos na população para codificar os ângulos. Cada número real com 14 dígitos após a vírgula é dividido em dois segmentos de 7 dígitos, onde cada segmento corresponde, respectivamente, a um dos ângulos ϕ e ψ . Em cada geração, cinco indivíduos de elite, aqueles com as melhores avaliações de aptidão, são automaticamente selecionados para compor a próxima população, garantindo a preservação de soluções promissoras. Além disso, cinco

novos indivíduos são gerados aleatoriamente a cada geração para introduzir diversidade genética na população e evitar a convergência prematura para soluções subótimas.

RESULTADOS

O método se concentrou em refinar as estruturas proteicas previstas, explorando o espaço conformacional por meio de operações evolutivas e avaliando soluções candidatas usando uma função de aptidão baseada em energia do conjunto Rosetta.

Com base nos resultados obtidos, o método proposto demonstrou um desempenho promissor em comparação com outras abordagens, apresentando uma vantagem em termos de resultados médios para as métricas avaliadas.

CONSIDERAÇÕES FINAIS

Apesar da melhora no desempenho, o modelo apresentou um alto desvio padrão, sugerindo maior variabilidade nos resultados.

Palavras-chave: Proteínas, Predição de estruturas proteicas, Alpha Fold, BRKGA, Modelo híbrido

REFERÊNCIAS BIBLIOGRÁFICAS

1. LONDE, M. A. et al. Biased random-key genetic algorithms: A review. *European Journal of Operational Research*, 1 mar. 2024.
2. MARCHI, F.; PARPINELLI, R. S. A Multi-objective Approach to the Protein Structure Prediction Problem using the Biased Random-Key Genetic Algorithm. *2022 IEEE Congress on Evolutionary Computation (CEC)*, p. 1070–1077, 28 jun. 2021.

DADOS CADASTRAIS

BOLSISTA: Lucas Gonçalves Erdmann

MODALIDADE DE BOLSA: PIBITI/CNPq (IT)

VIGÊNCIA: 09/2024 a 08/2025 – Total: 12 meses

ORIENTADOR(A): Rafael Stubs Parpinelli

CENTRO DE ENSINO: CCT

DEPARTAMENTO: Departamento de Ciência da Computação

ÁREAS DE CONHECIMENTO: Ciências Exatas e da Terra / Ciência da Computação

TÍTULO DO PROJETO DE PESQUISA: Reconhecimento de Padrões com Inteligência Computacional e Processamento de Alto-desempenho para Auxílio de Diagnóstico - Parte 2
Nº PROTOCOLO DO PROJETO DE PESQUISA: NPP3192-2020