

UMA HEURÍSTICA GRASP PARA SELEÇÃO DE CADEIAS DE CARACTERES

Válber Laux

Universidade Federal Rural do Rio de Janeiro (UFRRJ)
Av. Governador Roberto Silveira, Centro, Nova Iguaçu - RJ, 26020-740
valber@ufrj.br

Adria Lyra

Universidade Federal Rural do Rio de Janeiro (UFRRJ)
Av. Governador Roberto Silveira, Centro, Nova Iguaçu - RJ, 26020-740
adrialyra@ufrj.br

RESUMO

Diversos problemas na Biologia Molecular envolvem a comparação de sequências (cadeias de caracteres) e busca por regiões comuns ou próximas entre sequências ou conjunto de sequências de DNA, RNA e proteínas. Esses problemas têm várias aplicações, como busca de regiões conservadas em sequências não alinhadas, identificação de drogas genéticas, formulação de sondas (*probes*) genéticas, entre outras. No Problema da Sequência mais Próxima (PSMP) deseja-se encontrar, a partir de um conjunto S de sequências, todas de tamanho m formadas sobre um alfabeto Σ , uma nova sequência s_h que apresente a menor Distância de Hamming em relação a todas as outras do conjunto, ou seja, para cada $s_i \in S$, tenhamos $d_h(s_h, s_i) \leq d$, para $i=1\dots m$. O problema foi provado ser NP-difícil, o que motivou os autores a usarem de heurísticas na busca de soluções aproximadas para o problema. Diversos algoritmos aproximativos, exatos e heurísticos têm sido propostos. Neste trabalho é proposto um algoritmo para o PSMP baseado na meta-heurística GRASP (*Greedy Randomized Adaptive Search Procedure*), que tem se mostrado bastante competitiva em problemas combinatórios de grande complexidade. O GRASP é um processo iterativo “*multi-start*”, onde cada iteração do processo consiste basicamente em duas fases: construção e busca local. Ele se distingue de outras heurísticas por privilegiar a construção de uma solução inicial de melhor qualidade, que será apenas melhorada na busca local. Na fase de construção, cada solução inicial é construída de forma iterativa, onde uma função gulosa mede o benefício da adição de cada elemento candidato à solução. Os candidatos melhor avaliados são adicionados a uma Lista Restrita de Candidatos, da qual um elemento é selecionado de forma aleatória. Isso permite a geração de variadas soluções, uma vez que não se limita somente na escolha do melhor candidato. Na segunda fase, uma busca local é aplicada trocando-se uma base arbitrária em s por outra diferente. Se houver melhora na qualidade, a nova solução substitui a anterior. A melhor dentre todas soluções encontradas nas iterações do GRASP é retornada como resultado. Experimentos computacionais foram realizados com o algoritmo proposto e comparados com outros existentes na literatura. Conseguiu-se a execução de instâncias maiores do que as normalmente apresentadas na literatura com um tempo bastante competitivo, além de melhora na qualidade das soluções obtidas. Como trabalhos futuros, estão a otimização e paralelização do algoritmo, assim como melhorias na busca local.

PALAVRAS-CHAVE. Meta-heurísticas, Problema da Sequência mais Próxima, GRASP.