

**Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**

---

**Abordagens Heurísticas para Problemas de Seleção de Strings**Jean P. T. Torres<sup>1</sup>

Faculdade de Computação, UFMS, Campo Grande, MS

Edna A. Hoshino<sup>2</sup>

Faculdade de Computação, UFMS, Campo Grande, MS

Este trabalho tem como objetivo estudar três dos principais problemas de seleção de strings: *Closest String Problem* (CSP), *Farthest String Problem* (FSP) e o *Distinguishing String Selection Problem* (DSSP). De maneira geral, no CSP deseja-se criar, para um dado conjunto de strings, uma string alvo que seja a mais próxima de todas as strings do conjunto, enquanto no FSP a string alvo deve ser a mais distante possível. O DSSP, no entanto, busca criar uma string alvo que seja a mais próxima possível de um conjunto  $S^c$  e simultaneamente mais distante de outro conjunto  $S^f$ . Também foram estudadas as variantes dos problemas CSP, FSP e DSSP que buscam por substrings alvos e são denotadas por CSSP, FSSP e DSSSP, respectivamente.

Os problemas são definidos em [5], enquanto o modelo matemático utilizado é semelhante ao apresentado por Meneses [6] para o DSSP. Foram adicionadas ao modelo duas variáveis,  $d_c$  e  $d_f$ , para representar a maior diferença da string alvo em relação às strings em  $S^c$  e a menor diferença em relação às strings em  $S^f$ , respectivamente, as quais permitem definir uma função objetivo que minimiza a diferença  $d_c - d_f$ .

Dentre os problemas estudados, o CSP é aquele que mais recebeu atenção na literatura. A qualidade da sua relaxação linear é reportada em [1], onde os autores propuseram duas heurísticas para o problema, o RA (Rounding Algorithm) e o BCPA (Basic Core Problem Algorithm). A primeira aplica procedimentos de arredondamento na solução do problema relaxado enquanto a segunda explora as variáveis com valores inteiros na solução relaxada. A qualidade da relaxação linear também foi discutida em [7], onde o DSSP e o DSSSP foram estudados. Foram aplicados e comparados o RA e uma heurística VNS (Variable neighbourhood Search). Este VNS corresponde ao *Basic VNS* apresentado em [3] e é iniciado com a solução gerada pelo RA. Ele utiliza a solução relaxada das variáveis para definir suas vizinhanças e aplica a busca *first improvement* mencionada em [3].

Desta forma, este trabalho é uma expansão de [7], abordando os seis diferentes problemas de seleção de strings e tem como objetivo avaliar a qualidade da relaxação linear produzida pelos modelos e propor melhorias ao VNS. Na implementação de um algoritmo exato dos modelos propostos, foram utilizados o resolvidor *CPLEX* [4] e a biblioteca de otimização *SCIP* [2]. Durante os testes, as instâncias foram separadas em três grupos,

---

<sup>1</sup>tremeschin.jean@gmail.com<sup>2</sup>eah@facom.ufms.br

onde o tamanho das strings varia em relação à cardinalidade dos conjuntos. Observou-se que, conforme o tamanho da string torna-se muito maior que a cardinalidade dos conjuntos, a relaxação linear do problema tende a perder qualidade. Ainda, o DSSP e DSSSP se apresentaram muito mais difíceis que os outros problemas.

Nos experimentos realizados, o VNS apresentou bons resultados para todos os problemas, melhorando a solução produzida pelo RA. No entanto, para o CSP e FSP, a relaxação linear é muito boa, o que levou o RA a gerar resultados muito próximos do ótimo. De fato, ele produziu uma solução ótima em cerca de 51.24% das instâncias e, nos demais casos, obteve soluções a uma distância média de, no máximo, 3. Isto impacta no desempenho do VNS, que consegue produzir pouca melhoria na solução do RA. Desta forma, embora o VNS melhore as soluções do RA em até 94% para o DSSSP e 52% para o DSSP, nesses problemas mais fáceis esta taxa de melhoria não ultrapassa 34%.

O VNS, baseado no RA, embora consiga boas soluções para os problemas de seleção de substrings, é limitado quando o RA já gera uma boa solução. Isso se dá pois a primeira vizinhança é muito menor que as outras. Atualmente este trabalho está estudando novas vizinhanças, novas buscas locais e incorporando técnicas do BCPA no VNS.

## Referências

- [1] F. Della Croce e F. Salassa, Improved lp-based algorithms for the closest string problem, *Computers & Operations Research*, 39:746–749, 2014. DOI:10.1016/j.cor.2011.06.010.
- [2] G. Gamrath, T. Fischer, T. Gally, A. M. Gleixner, G. Hendel, T. Koch, S. J. Maher, M. Miltenberger, B. Müller, M. E. Pfetsch, C. Puchert, D. Rehfeldt, S. Schenker, R. Schwarz, F. Serrano, Y. Shinano, S. Vigerske, D. Weninger, M. Winkler, J. T. Witt e J. Witzig, The scip optimization suite 3.2, *ZIB Technical Report*, 15–60, 2016.
- [3] P. Hansen, N. Mladenović e J. A. Moreno Pérez, Variable neighbourhood search: methods and applications, *A Quarterly Journal of Operations Research*, 6:319–360, 2008. DOI: 10.1007/s10288-008-0089-1.
- [4] IBM, IBM ILOG CPLEX Optimization Studio CP Optimizer User’s Manual, Version 12 Release 6, 2015.
- [5] J. K. Lanctot, M. Li, B. Ma, S. Wang e L. Zhang, Distinguishing string selection problems, *Proceedings of the Tenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 185:633–642, 1999.
- [6] C. N. Meneses, P. Pardalos, M. Resende e A. Vazacopoulos, Modeling and solving string selection problems, In *Second international symposium on mathematical and computational biology*, Petrópolis, Rio de Janeiro, Brasil, 2005.
- [7] J. Torres, E. Silva e E. A. Hoshino, Heuristic approaches to the Distinguishing Substring Selection Problem, *Electronic Notes in Discrete Mathematics*, 66:151-158, 2018. DOI: 10.1016/j.endm.2018.03.020.