

Um Estudo Paramétrico do Algoritmo Evolutivo baseado na representação Interação-Transformação para Regressão Simbólica

Guilherme S. I. Aldeia¹, Fabrício O. de França¹

¹ Centro de Matemática, Computação e Cognição (CMCC)
Heuristics, Analysis and Learning Laboratory (HAL)
Universidade Federal do ABC (UFABC) – Santo André, SP – Brasil

{guilherme.aldeia, folivetti}@ufabc.edu.br

Dada uma base de dados com variáveis explanatórias \mathbf{x} e uma variável alvo y , onde exista uma função desconhecida $f(\mathbf{x}) = y$, a regressão simbólica busca por uma função $\hat{f}(\mathbf{x})$ que se aproxime suficientemente bem de $f(\mathbf{x})$. Ela costuma ser feita por meios da programação genética, um método de busca e otimização que simula conceitos da teoria da evolução para construir e ajustar $\hat{f}(\mathbf{x})$ através da manipulação de populações de soluções, mas a forma em que as soluções são expressas (por meio de árvores sintáticas) apresenta um espaço de busca amplo e resultados pouco interpretáveis. Nesse contexto, foi introduzida recentemente uma nova representação para a regressão simbólica — a Interação-Transformação (IT) — de forma a evitar esses problemas, limitando o espaço de busca em expressões simples, sendo capaz de retornar resultados com maior interpretabilidade que se mantém competitivos com a literatura. Algoritmos de regressão simbólica apresentam um importante compromisso entre aproximação dos dados e complexidade da função encontrada. Essa representação foi usada no contexto de um algoritmo evolutivo, com a finalidade de encontrar expressões concisas que tenham um erro de aproximação pequeno e resultados competitivos com a literatura. Em [Aldeia and de França 2020], os autores investigaram a influência dos hiper-parâmetros do algoritmo evolutivo proposto sobre os resultados, em termos de erro de aproximação e complexidade da função simbólica encontrada. Os resultados mostraram que o número máximo de termos é mais importante que o limite de expoentes permitido, e existe um limite onde aumentar o tamanho da expressão gera benefícios, podendo ser explicado pelo *overfit*. Como o limite dos expoentes tem menor influência, ele pode ter um valor definido como padrão sem impactar nos resultados, mas diminuindo a complexidade das funções em termos de sua interpretabilidade.

Referências

Aldeia, G. S. I. and de França, F. O. (2020). A parametric study of interaction-transformation evolutionary algorithm for symbolic regression. In *2020 IEEE World Congress on Computational Intelligence (WCCI)*. IEEE. O artigo foi aceito para publicação.