

INTERFACE WEB PARA EXPERIMENTOS COM REGRESSÃO SIMBÓLICA



Guilherme Seidyo Imai Aldeia, Fabrício Olivetti de França

Centro de Matemática, Computação e Cognição - Universidade Federal do ABC - Heuristics, Analysis and Learning Laboratory (HAL)
Avenida dos Estados, 5001 - Bairro Santa Terezinha, Santo André - CEP: 09210-580
{guilherme.aldeia@aluno.ufabc.edu.br, folivetti@ufabc.edu.br}

Resumo. A regressão simbólica é uma técnica que busca por equações matemáticas que sejam capazes de se ajustar e descrever uma série de medidas. Este projeto teve como objetivo o estudo e implementação de métodos de regressão simbólica baseados na estrutura de dados *Interação-Transformação* (IT) em uma interface web, capaz de inferir equações baseadas em dados mensurados, podendo ser utilizada como assistente em laboratórios didáticos. Ainda, o projeto estudou a aplicação do uso da estrutura de dados em algoritmos de regressão simbólica em computação de baixo custo.

Palavras-chave. regressão simbólica, programação genética.

ESTRUTURA IT

Uma expressão IT é uma lista de tuplas (função, expoentes) que formam uma combinação linear, como por exemplo:

$$[(\sin, [2, 0]), (\sqrt{}, [3, -1])]$$

Que se traduz na expressão matemática:

$$f(x, y) = \sin(x^2) + \sqrt{\frac{x^3}{y}}$$

METODOLOGIA

Criamos uma ferramenta web desenvolvida com JavaScript implementando algoritmos de regressão simbólica relacionados com a estrutura IT. Essa plataforma encontra-se disponível em <https://galdeia.github.io/>, e funciona até mesmo em smartphones.

Entrada de dados

Os dados podem ser digitados manualmente, ou você pode fazer o upload de um arquivo .csv contendo seus dados. Caso queira, a primeira linha pode ser usada para nomear as variáveis. No site há uma página mais elaborada sobre os [dados de entrada](#).

Entrada manual de dados

Upload de arquivo local

Digite os dados aqui. Um ponto em cada linha, separados por vírgulas, espaços ou tabulações

Choose File No file chosen

☐ Usar a primeira linha da entrada como rótulo?

Usar valores digitados

Usar arquivo enviado

Escolhi um exemplo

FIGURA 1: Interface de entrada de dados.

Resultados

Com os dados carregados, basta escolher um algoritmo abaixo:

Algoritmo SymTree (Symbolic Tree)

Algoritmo IT-LS (Interaction Transformation Local Search)

Algoritmo IT-ES (Interaction Transformation Evolution Strategies)

Resultado:

Algoritmo: SymTree

Expressão: 6.28*x⁰*h + 6.28*x²

Score: 0.999999999996491

Tempo (em ms): 242

FIGURA 2: Seleção de algoritmos e resultados.

Foram implementados três algoritmos: *SymTree*, baseado em uma heurística em árvores; *IT-LS*, baseado em busca local; e *IT-ES*, baseado em estratégias evolutivas.

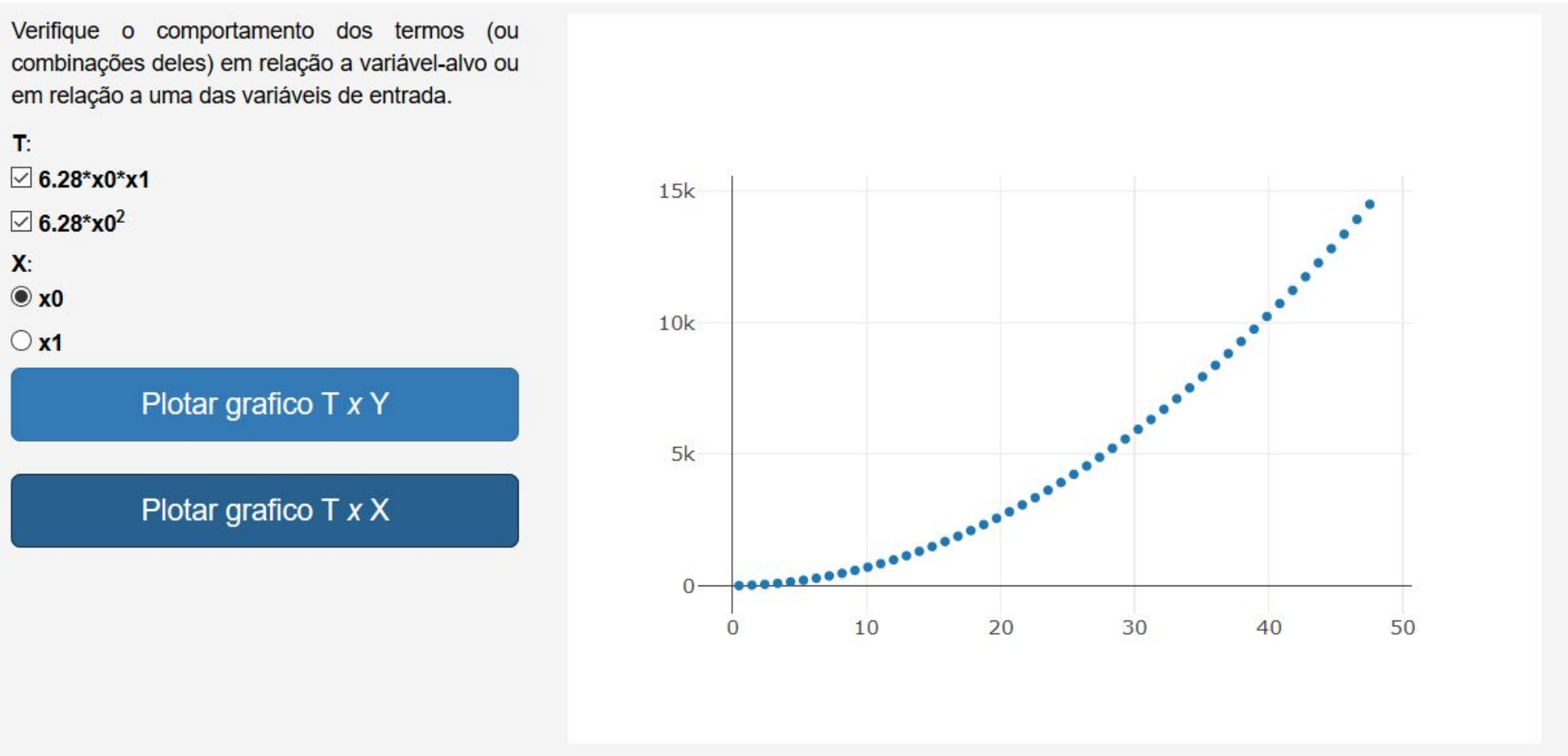


FIGURA 3: Gráficos para análise do resultado.

Para testar a ferramenta, utilizamos 14 equações da física e engenharia representáveis pela IT e 6 não representáveis. Foi criada uma base com 20 amostras num domínio experimentalmente plausível para cada expressão, e, então, cada algoritmo foi executado 30 vezes para cada equação, tendo seu desempenho médio reportado.

TABELA 1: Expressões utilizadas para validação.

Id	Função	Nome
1	$\Delta P = \rho g \Delta h$	Variação da Pressão Vertical
2	$E = mc^2$	Equivalência Massa-Energia
3	$I_x = \frac{bh^3}{12}$	Momento de Inércia de um Retângulo
4	$F = k\Delta l$	Lei de Hooke
5	$A = bh$	Área de um Retângulo
6	$W = \cos(\Theta) f \Delta d$	Trabalho
7	$Pv = nRT$	Lei Geral dos Gases
8	$T = 2\pi\sqrt{\frac{L}{g}}$	Equação do Pêndulo
9	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$	Média Aritmética
10	$A = 4\pi r^2$	Área Superficial de uma Esfera
11	$A = 2\pi r^2 + 2\pi rh$	Área Superficial de um Cilindro
12	$E = \frac{k\Delta l^2}{2}$	Energia Potencial Elástica
13	$\rho = \frac{m}{V}$	Densidade
14	$\beta = 10\log(\frac{I}{I_0})$	Intensidade Sonora

RESULTADOS

TABELA 2: *Score* médio dos algoritmos para cada equação representável.

Id	SymTree	IT-LS	IT-ES	Eureqa
1	1.00	0.93 ± 0.22	0.95 ± 0.18	1.00
2	1.00	1.00	1.00	1.00
3	1.00	0.03 ± 0.18	0.00	1.00
4	1.00	1.00	0.93 ± 0.21	1.00
5	1.00	0.97 ± 0.18	1.00	1.00
6	1.00	0.00	0.00	1.00
7	1.00	0.50 ± 1.00	0.83 ± 0.37	1.00
8	1.00	1.00	1.00	1.00
9	1.00	0.56 ± 0.25	0.63 ± 0.33	1.00
10	1.00	1.00	1.00	0.90 ± 0.30
11	1.00	0.07 ± 0.25	0.23 ± 0.42	1.00
12	1.00	0.97 ± 0.18	0.77 ± 0.42	1.00
13	1.00	1.00	1.00	1.00
14	1.00	0.89 ± 0.24	0.33 ± 0.09	0.75 ± 0.21

Observando os resultados reportados na tabela 2, é possível perceber que o algoritmo SymTree apresentou os melhores resultados na obtenção da expressão geradora. Um exemplo da saída do algoritmo é mostrado na Fig. 2.

Além disso, os resultados obtidos foram comparados com um software comercial (Eureqa) em relação à exatidão dos resultados restritos a um tempo máximo de execução [2].

CONCLUSÕES

Os resultados são uma prova de conceito da estrutura de dados IT, e mostram que a estrutura permite a realização de regressão simbólica com baixo custo computacional, mantendo um resultado interpretável e competitivo com técnicas de regressão comumente utilizadas.

REFERÊNCIAS BIBLIOGRÁFICAS

- FRANÇA, F. O. **A greedy search tree heuristic for symbolic regression**, Information Sciences, vol 442-443, p.18-32, 2018.
- ALDEIA, G. S. I & FRANÇA, F. O. **Lightweight symbolic regression with the interaction-transformation representation**, IEEE Congress on Evolutionary Computation, 2018.

AGRADECIMENTOS

Agradecemos ao Programa de Iniciação Científica da UFABC (PIC/UFABC) por financiar o trabalho.