

# Novel Application of Mutual Information in Transfer Learning for Genetic Programming

Yilin Liu, Gareth Taylor, Zhengwen Huang  
Department of Electrical and Electronic Engineering,  
Brunel University of London



## Introduction

### How to find Building Blocks early?

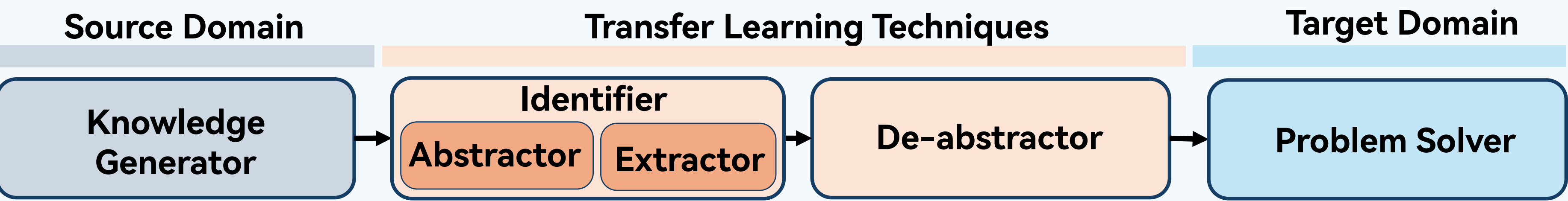
Identifying building blocks early can help guide evolution efficiently toward the optimum.

### Transfer learning is effective:



Knowledge from solved problems directs evolution in similar tasks.

## Transfer learning in GP



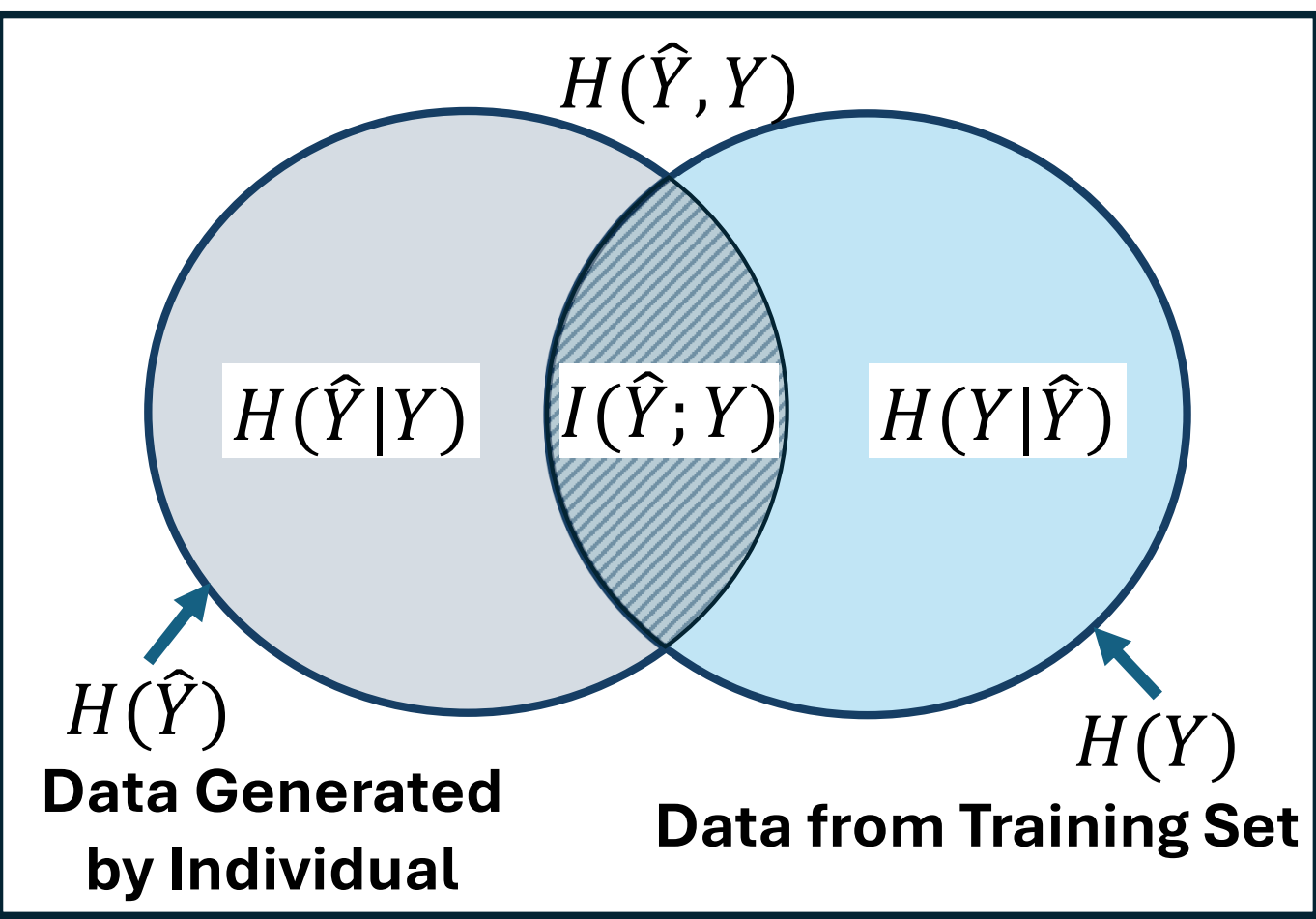
### Key Challenge

Identifying valuable knowledge for the target domain.

Existing methods often struggle with low significance, local optima, and strong dependence on specific source problems.

## Methodology

### Mutual Information – Indicator of Useful Knowledge



$$I(\hat{Y}; Y) = \sum_{\hat{y} \in \hat{Y}} \sum_{y \in Y} p(\hat{y}, y) \log \left( \frac{p(\hat{y}, y)}{p(\hat{y})p(y)} \right)$$

#### Normalization Methodology

target-norm MI:

By  $H(Y)$  – Emphasize redundancy

$$I_N(\hat{Y}; Y) = \frac{I(\hat{Y}; Y)}{H(Y)}$$

joint-norm MI:

By  $H(Y, \hat{Y})$  – Emphasize both

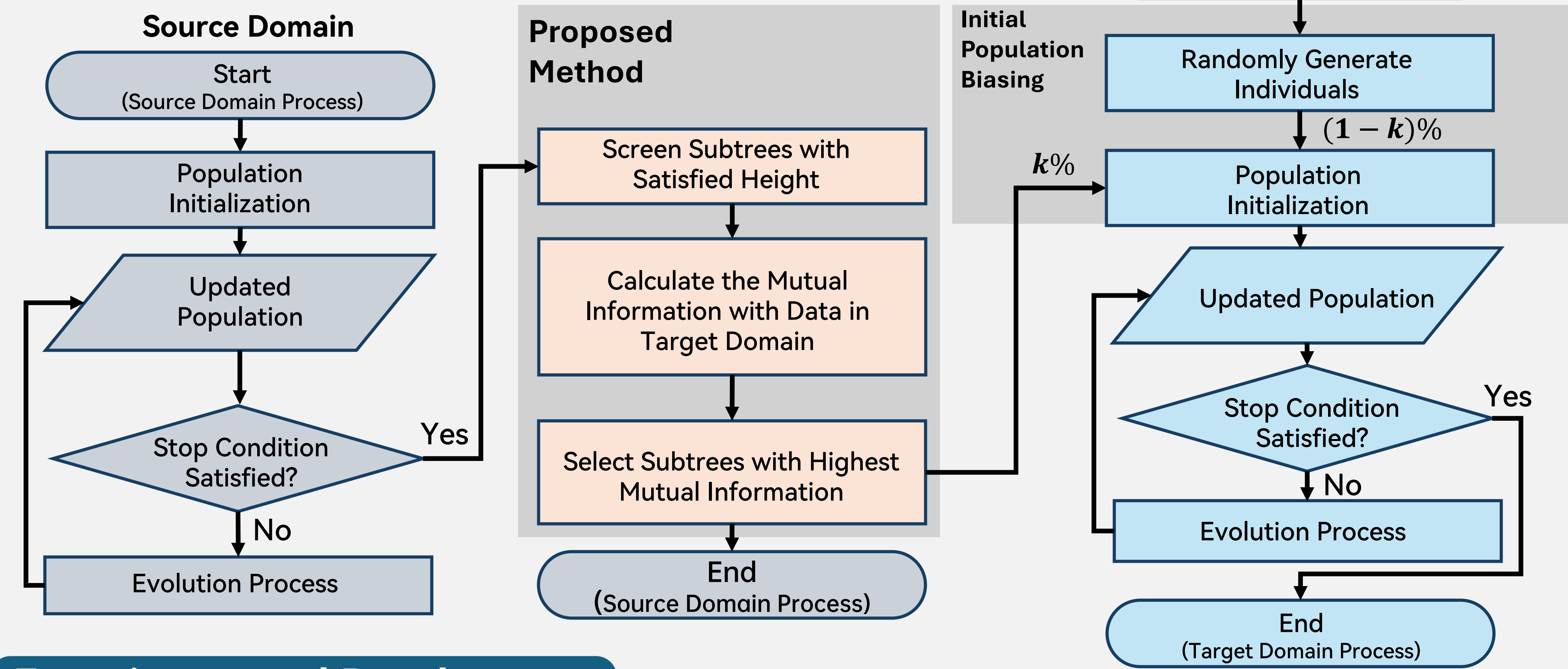
$$I_N(\hat{Y}; Y) = \frac{I(\hat{Y}; Y)}{H(Y)}$$

### Two important characteristics

- No assumption on the distribution of data before measuring.
- Remains unchanged after a smooth invertible operation.

### Proposed Method

### Mutual Information-based Transfer Learning GP



### What to Transfer?

#### Subtrees with High Mutual Information

- These subtrees approximate the global optimum and can evolve into it via smooth and investable structural changes.
- High overlap with target data makes them more relevant and reduces error.

### How to Transfer? & When to Transfer?

#### Initial Population Biasing

- Avoids disrupting the search space.
- Guides evolution to focus on small, high-information subspaces.
- Flexibility to further self-adaptation.

## Experiment and Results

### Trial Problems

10 symbolic regression problems from existing work<sup>[1]</sup>

### Baseline

Standard GP and SubTree50<sup>[1]</sup>

| Problems             | Poly-1         | Poly-2        | Poly-3        | Poly-4         | Poly-5              | Trig-1        | Trig-2        | Trig-3        | Trig-4           | Trig-5        |
|----------------------|----------------|---------------|---------------|----------------|---------------------|---------------|---------------|---------------|------------------|---------------|
| Mean of Best Fitness |                |               |               |                |                     |               |               |               |                  |               |
| Standard GP          | 0.0180         | 0.0204        | 0.0379        | 0.0506         | 0.0622              | 0.0365        | 0.0088        | 0.0159        | 0.0156           | 0.0947        |
| SubTree50            | 0.0139         | 0.0138        | 0.0258        | 0.0419         | 0.0563              | 0.0230        | 0.0057        | 0.0177        | 0.0105           | 0.0933        |
| no-norm MI           | 0.0139         | 0.0129        | 0.0238        | 0.0418         | 0.0516              | 0.0141        | <b>0.0052</b> | 0.0103        | <b>0.0043</b>    | <b>0.0470</b> |
|                      | (+,=)          | (+,+)         | (+,+)         | (+,+)          | (+,+)               | (+,+)         | (+,+)         | (+,+)         | (+,+)            | (+,+)         |
| target-norm MI       | 0.0127         | <b>0.0128</b> | 0.0254        | <b>0.0377</b>  | <b>0.0449</b>       | 0.0172        | 0.0066        | 0.0126        | 0.0075           | 0.0485        |
|                      | (+,+)          | (+,+)         | (+,+)         | (+,+)          | (+,+)               | (+,+)         | (+,-)         | (+,+)         | (+,+)            | (+,+)         |
| joint norm MI        | <b>0.0117</b>  | 0.0129        | <b>0.0222</b> | 0.0488         | 0.0540              | <b>0.0129</b> | 0.0057        | <b>0.0089</b> | 0.0070           | 0.0627        |
|                      | (+,+)          | (+,+)         | (+,+)         | (+,-)          | (+,+)               | (+,+)         | (+,=)         | (+,+)         | (+,+)            | (+,+)         |
| Median of Test Error |                |               |               |                |                     |               |               |               |                  |               |
| Standard GP          | 126.965        | 47.900        | 91.983        | 156.348        | 5955826.581         | 2.892         | 4.973         | 6.426         | 8.910E-17        | 1.200         |
| SubTree50            | <b>116.750</b> | 38.099        | 90.793        | 155.527        | <b>5169346.282</b>  | 2.854         | 7.912         | 6.172         | 8.351E-17        | 1.286         |
| no-norm MI           | 125.207        | <b>26.601</b> | 82.773        | <b>132.057</b> | <u>23010016.817</u> | <b>2.614</b>  | 2.766         | <u>4.077</u>  | <b>8.042E-17</b> | <u>1.137</u>  |
|                      | (+,-)          | (+,+)         | (+,+)         | (+,+)          | (-,-)               | (+,+)         | (+,+)         | (+,+)         | (+,+)            | (+,+)         |
| target-norm MI       | 126.428        | 43.425        | <b>67.320</b> | <u>139.002</u> | <u>23010016.840</u> | <u>2.732</u>  | 3.175         | 4.796         | <u>8.188E-17</u> | <b>1.000</b>  |
|                      | (+,-)          | (+,-)         | (+,+)         | (+,+)          | (-,-)               | (+,+)         | (+,+)         | (+,+)         | (+,+)            | (+,+)         |
| joint norm MI        | 126.529        | <u>34.971</u> | <u>76.556</u> | 157.080        | <u>23010016.801</u> | <u>2.662</u>  | <b>2.746</b>  | <b>4.062</b>  | <u>8.243E-17</u> | <u>1.080</u>  |
|                      | (+,-)          | (+,+)         | (+,+)         | (-,-)          | (-,-)               | (+,+)         | (+,+)         | (+,+)         | (+,+)            | (+,+)         |

“+” Outperforming | “-” Outperformed | “=” Similar Performance

Underlined: Passed Wilcoxon Significance Test Compared with Standard GP

Bolded: The Best in One Test Problem

Smaller figure indicates better results.

### Observations

For MI-based Methods:

- Improve training fitness and reduce test error in most tasks.
- More often passes significance tests, better preserves initial subtree advantages.

## Insights and Future Works

- Mutual information is key for selecting transferable subtrees and boosting GP performance.
- It helps identify relevant parts of past models for current tasks.
- Future work will explore its use in complex systems like power systems to guide model reuse.

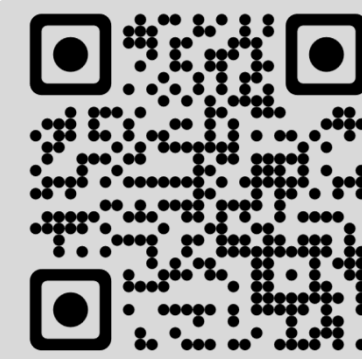
[1] T. T. Huong Dinh, T. H. Chu and Q. U. Nguyen, “Transfer learning in genetic programming,” 2015 IEEE Congress on Evolutionary Computation (CEC), Sendai, Japan, 2015.

[2] S. W. Card and C. K. Mohan, “Information theoretic indicators of fitness, relevant diversity & pairing potential in genetic programming,” 2005 IEEE Congress on, Applied Soft Computing, Evolutionary Computation, Edinburgh, UK, 2005.

[3] Zahra Zojaji, Mohammad Mehdi Ebadzadeh, Hamid Nasiri, “Semantic schema based genetic programming for symbolic regression”, Volume 122,2022.



Full Paper



Email to the Author