Partial Sampling Operator and Structural Distance Ranking for Multi-Objective GP

Makoto OHKI
Tottori University, Japan
1. Introduction

Applications of Genetic Programming (GP)

- Program Synthesis
- Function Generation
- Rule Set Discovery

1. Introduction
2. MOGP with SD
3. Partial Sampling
4. Verification
5. Conclusion
1. Introduction

Program Synthesis [David2017]

```ocaml
type list = Nil | Cons of int * list
let rec even x =
  match x with
  | Nil -> Nil
  | Cons(u, Nil) -> Cons(u, Nil)
  | Cons(u, Cons(_, us))
    -> Cons(u, even us)
let rec sum x =
  match x with
  | Nil -> 0
  | Cons(u, us) -> u + sum us
let rec sum_even x = ??
let main x = assert (sum (even x) = sum_even x)
```

Let rec sum_even x =
match x with
  | Nil -> 0
  | Cons(u, Nil) -> u
  | Cons(u, Cons(_, us)) -> u + sum_even us
1. Introduction

- Function Generation [Jamali2017]

\[ U(x_p) = -Ce^{\frac{(x_p - x_g)^2}{r^2}} \]

Potential Function
1. Introduction

Rule Set Discovery [Ohmoto2013]

Observed Data

\[ x \text{ is } A' \]

Knowledge Base

if \( x \) is \( A_1 \) then \( y \) is \( B_1 \)
if \( x \) is \( A_2 \) then \( y \) is \( B_2 \)
\ldots \ldots

\[ \text{if } x \text{ is } A_n \text{ then } y \text{ is } B_n \]

\[ A_1, A_1, \ldots, A_n, B_1, B_2, \ldots, B_n : \text{Fuzzy Sets} \]

1. Introduction
2. MOGP with SD
3. Partial Sampling
4. Verification
5. Conclusion
1. Introduction

They can be expressed by a tree structure data.

Non-terminal node

Terminal node

Optimization

Genetic Programming: GP
1. Introduction

In this paper,

- A technique of Multi-Objective GP by applying NSGA-II
  - index of goodness of the tree
  - the Size of the tree
  - tree position in the population by Structural Distance (SD)

- Apply SD instead of Crowding Distance (CD) of NSGA-II

- Partial Sampling (PS) operator instead of Crossover and Mutation

- Double Spiral Problem for verification
2. Multi-Objective GP with Structural Distance

3 Objective Functions

① objective function according to Goodness of the tree structure

\[ h_1(\text{indiv}_i) = \text{performance}(\text{root}_i) \]

② objective function according to the size of the tree structure

\[ h_2(\text{indiv}_i) = \frac{1}{\text{size}(\text{root}_i)} \]

③ objective function according to average of SD in the population

\[ h_3(\text{indiv}_i) = \frac{1}{N_{\text{pop}}} \sum_{k=1}^{N_{\text{pop}}} \text{SD} \; \text{indiv}_i, \text{indiv}_k \]
2. Multi-Objective GP with Structural Distance

- Structural Distance (SD)

1. Introduction
2. MOGP with SD
3. Partial Sampling
4. Verification
5. Conclusion
2. Multi-Objective GP with Structural Distance

- Structural Distance (SD)

\[
\text{SD}(\text{root}_i, \text{root}_k) = \frac{1}{24} + \frac{1}{24} + \frac{1}{8} = \frac{5}{24}
\]

- : match
- : mismatch
- : do not care

- : conformity

1. Introduction
2. MOGP with SD
3. Partial Sampling
4. Verification
5. Conclusion
2. Multi-Objective GP with Structural Distance

- NSGA-II

![Diagram]

1. Introduction
2. MOGP with SD
3. Partial Sampling
4. Verification
5. Conclusion

$R_g^g$ : population
$P_g^g$ : parents
2. Multi-Objective GP with Structural Distance

NSGA-II with SD instead of CD

- non-dominated sorting
- parents
- children
- mating
- next generation

1. Introduction
2. MOGP with SD
3. Partial Sampling
4. Verification
5. Conclusion
3. Partial Sampling Operator for Mating

- Proliferation in Partial Sampling (PS) Operator

1. Introduction
2. MOGP with SD
3. Partial Sampling
4. Verification
5. Conclusion

---

1. $T_{\text{sub}}$
2. $N_{\text{cand}}$
3. $T_{\text{new}}$

---

terminate according to $p_t$
3. Partial Sampling Operator for Mating

- Proliferation Terminate Probability $p_t$

\[ p_t^0 = \frac{1}{\text{AverageSize} \ R^g} \]

\[ p_t^{g+1} = \frac{\text{Succ} \ R^g - p_t^0 \cdot \text{Succ} \ P^g}{\text{Succ} \ P^g - p_t^0 \cdot \text{Succ} \ R^g} \cdot p_t^g - p_t^0 + p_t^0 \]

$R^g$ : population

$P^g$ : parents
3. Partial Sampling Operator for Mating

- 2 kinds of metastasis

![Diagram showing partial sampling operator for mating with parent nodes and new generated tree](image)

- Initial proliferation
- Random metastasis
- Upper node depend metastasis
4. Verification by **Double Spiral Problem**

\[ f(x, y) > 0 \iff (x, y) \in D_1 \]
\[ f(x, y) < 0 \iff (x, y) \in D_2 \]
\[ f(x, y) = 0 \iff \text{FALSE} \]

difficult even by the neural network.

- non-terminal node \( \in +, -, *, \div, \sin, \cos, \tan, \text{ifltz} \)
- terminal node \( \in x, y, \text{constant} \)

\[ \text{ifltz}(a, b, c) \triangleq \begin{cases} 
  b & \text{if } a < 0 \\
  c & \text{(otherwise)}
\end{cases} \]
4. Verification by **Double Spiral Problem**

- Objective function $h_1$ according to the goodness of tree

\[
h_1(\text{indiv}_i) = \text{performance(root}_i) = \frac{1}{|D_1 \cup D_2|} \sum_{k=1}^{|D_1 \cup D_2|} g(x_k, y_k)
\]

\[
g(x, y) = \begin{cases} 
1 & f(x, y) > 0 \land (x, y) \in D_1, \\
0 & f(x, y) > 0 \land (x, y) \in D_2, \\
1 & f(x, y) < 0 \land (x, y) \in D_2, \\
0 & f(x, y) < 0 \land (x, y) \in D_1, \\
0 & f(x, y) = 0
\end{cases}
\]
4. Verification by **Double Spiral Problem**

- Final Solution Distribution

![Graph showing the relationship between inverse of the number of nodes and correct classification rate for different methods: CO+MU & CD, CO+MU & SD, PSM & CD, and PSM & SD.](image)

### Sections
1. Introduction
2. MOGP with SD
3. Partial Sampling
4. Verification
5. Conclusion
4. Verification by **Double Spiral Problem**

- Comparison among 3-Objective, 2-Objective, 1-Objective GPs

![Graph with SD](image1)

![Graph with CD](image2)

with SD

with CD
4. Verification by Double Spiral Problem

- Comparison of results on MS-Norm plane

![Graph showing comparison of results on MS-Norm plane]
5. Conclusion

In this paper,

- **Multi-Objective GP**
- In addition to goodness of the tree, 2 objective functions
  - tree size
  - Structural Distance (SD)
- **Partial Sampling (PS)** for mating

- Double Spiral Problem for verification.

- The proposed technique (PS + SD $\rightarrow$ NSGA-II) is effective.
5. Conclusion

In the future,

- Enhance the capability of numerical optimization
- Ranking Selection technique harmonizing CD and SD
- Mechanism to forcibly exit from PS
Thank you very much!

Ask me simply, if you have.
**MS and Norm**

**degree of spread of $FFS$**

$$MS = \sqrt{\sum_{i=1}^{m} \left( \frac{|FFS|}{\max f_i(x_j)} - \frac{|FFS|}{\min f_i(x_j)} \right)^2}$$

**degree of convergence to $POS$**

$$\text{Norm} = \frac{1}{|FFS|} \sum_{j=1}^{m} \sqrt{\sum_{i=1}^{m} f_i(x_j)^2}$$
3. Partial Sampling Operator for Mating

- 2 kinds of metastasis

\[ p_{met} = 0.5 \]

Initial proliferation
- random metastasis
- upper node depend metastasis