

Analisi della neutralità degli spazi di ricerca booleani in Programmazione Genetica

Yuri Pirola

Anno Accademico 2004/05

La Programmazione Genetica

Introduzione

Obiettivo
Paesaggio di fitness
Analisi del paesaggio
Even-k parity
Misurare la neutralità
Campionamento

Risultati

Considerazioni finali

Funzionamento:

```
inizializza popolazione  
while non finito do  
    selezione  
    variazione  
end while
```

Obiettivo e motivazioni

Introduzione

Obiettivo

Paesaggio di fitness

Analisi del paesaggio

Even-k parity

Misurare la neutralità

Campionamento

Risultati

Considerazioni finali

Neutralità:

Soluzioni differenti poste allo stesso livello di qualità.

Obiettivo:

Studio del paesaggio per ricercare un collegamento fra *Neutralità* e *Difficoltà* del problema.

Paesaggio di fitness

Introduzione

Obiettivo

Paesaggio di fitness

Analisi del paesaggio

Even-k parity

Misurare la neutralità

Campionamento

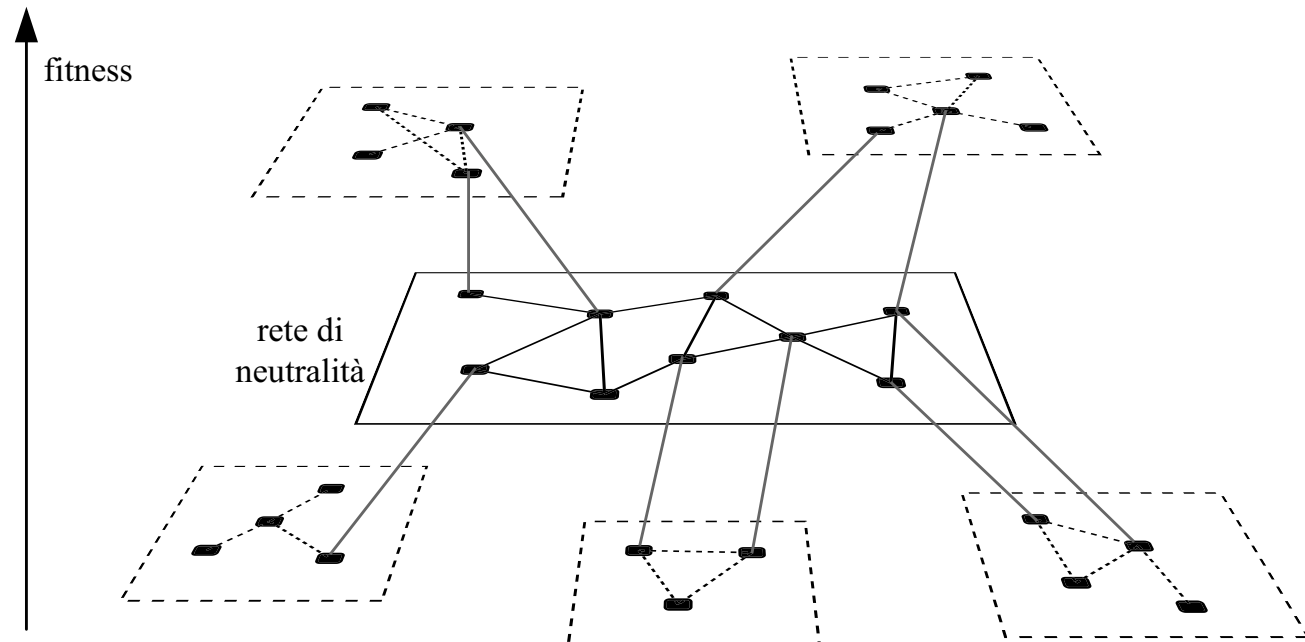
Risultati

Considerazioni finali

Paesaggio di fitness:

È un grafo $\mathcal{P} = (\mathcal{S}, \mathcal{V}, f)$ che modella il comportamento della PG

Rete di neutralità:



Analisi del paesaggio

Introduzione

Obiettivo
Paesaggio di fitness

Analisi del paesaggio

Even-k parity

Misurare la neutralità

Campionamento

Risultati

Considerazioni finali

Domanda:

Le caratteristiche di *neutralità* del paesaggio possono spiegare il diverso grado di *difficoltà* dei problemi?

Si deve analizzare *quantitativamente* la neutralità in diversi paesaggi.

Problemi:

- Quali paesaggi?
- Come *misurare* la neutralità?
- Come analizzare uno spazio di ricerca molto grande?

Il problema dell'even-parity

Introduzione

Obiettivo
Paesaggio di fitness
Analisi del paesaggio

Even-k parity

Misurare la neutralità

Campionamento

Risultati

Considerazioni finali

Even-k parity problem:

Determinare un'espressione di k variabili booleane che vale Vero se e solo se sono vere un numero *pari* di esse.

- La fitness di una soluzione è pari al numero di errori di approssimazione.
- Il 'grado di difficoltà' dipende
 1. dall'ordine k del problema
 2. dagli operatori booleani ammessi
 - con {XOR; NOT} il problema è 'facile'
 - con {NAND} il problema è 'difficile'

Misurare la neutralità

Introduzione

Obiettivo

Paesaggio di fitness

Analisi del paesaggio

Even-k parity

Misurare la neutralità

Campionamento

Risultati

Considerazioni finali

Misure 'di rete':

- a) Misure 'tradizionali'
 - Taglia e fitness di rete...
- b) Misure relative alla neutralità
 - *Tasso medio di neutralità*: percentuale di mutazioni neutre
 - *Δ -fitness media*: guadagno medio di fitness a seguito di una mutazione
 - *Tasso di subottimi e di subpessimi*: percentuale di individui, rispettivamente, non-migliorabili o non-peggiorabili

Metodo di campionamento

Introduzione

Obiettivo
Paesaggio di fitness
Analisi del paesaggio
Even-k parity
Misurare la neutralità

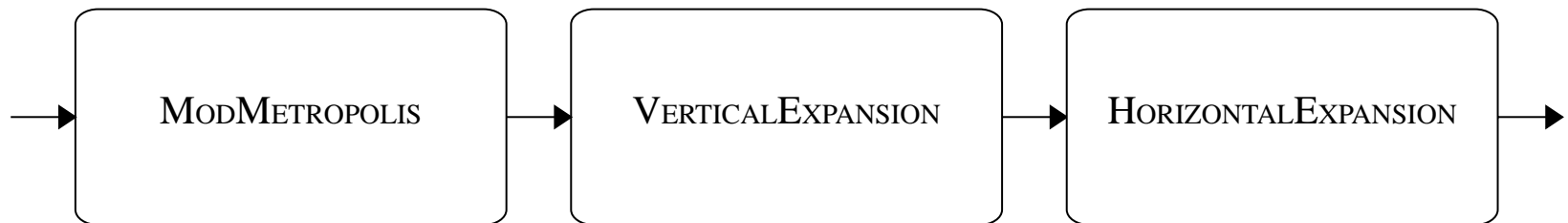
Campionamento

Risultati

Considerazioni finali

I metodi tradizionali sono insufficienti.

Nuovo metodo di campionamento:



ModMetropolis genera un campione di individui C con valori di fitness 'ben distribuiti'

VerticalExpansion arricchisce C con vicini possibilmente non-neutri

HorizontalExpansion aggiunge individui alle reti di neutralità *incomplete* troppo 'piccole'

Caratterizzazione di $\mathcal{P}^{\{\text{XOR}; \text{NOT}\}}$

Introduzione

Risultati

$\mathcal{P}^{\{\text{XOR}; \text{NOT}\}}$

Analisi sperim.

$\overline{T_N}$

$\overline{\Delta f}$

t_{so}

Considerazioni
finali

Obiettivo:

Date f e \mathcal{V} , determinare per via teorica alcune caratteristiche del paesaggio $\mathcal{P}^{\{\text{XOR}; \text{NOT}\}}$

- Solo tre livelli di fitness: 0, 0.5 e 1
- Un'unica 'grande' rete di neutralità a fitness 0.5 (→ rete *centrale*)
- Tutte le altre reti hanno taglia pari a 1 (→ reti *periferiche*)
- Tutte le reti periferiche sono collegate 'direttamente' a quella centrale

Analisi sperimentale

Introduzione

Risultati

$\mathcal{P}^{\{\text{XOR}; \text{NOT}\}}$

Analisi sperim.

$\overline{T_N}$

$\overline{\Delta f}$

t_{so}

Considerazioni
finali

Due insiemi degli operatori:

- $\{\text{XOR}; \text{NOT}\}$
- $\{\text{NAND}\}$

Due tipologie di analisi:

- *esaustiva* di paesaggi ‘limitati’
even-2 parity
- *campionaria* di paesaggi ‘realistici’
even-4 parity

Tasso medio di neutralità

Introduzione

Risultati

$\mathcal{P}^{\{XOR; NOT\}}$

Analisi sperim.

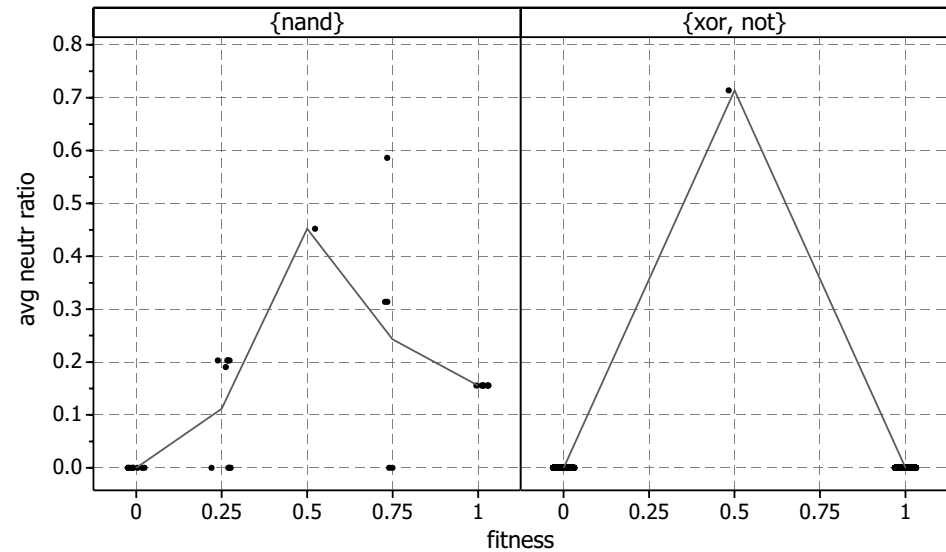
$\overline{T_N}$

Δf

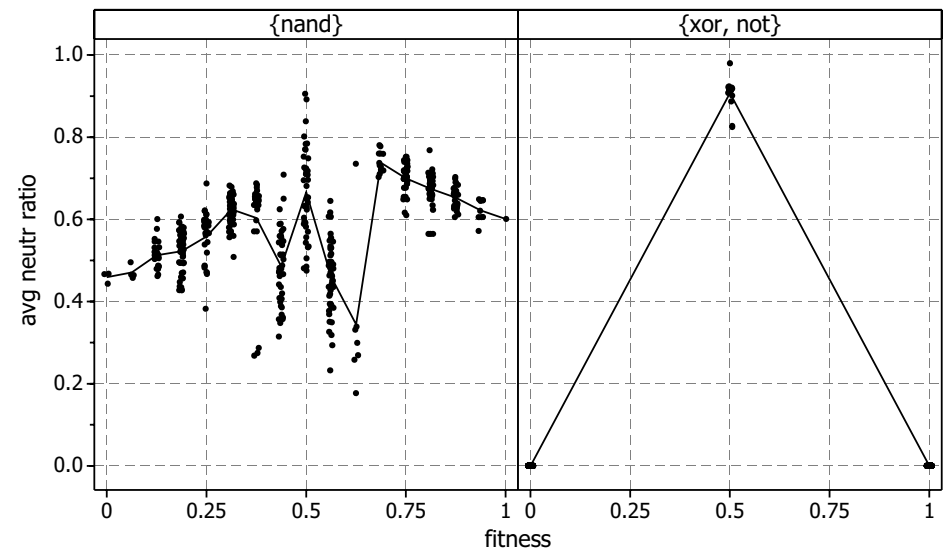
t_{so}

Considerazioni
finali

Even-2 parity



Even-4 parity



Δ -fitness media

Introduzione

Risultati

$\mathcal{P}^{\{XOR; NOT\}}$

Analisi sperim.

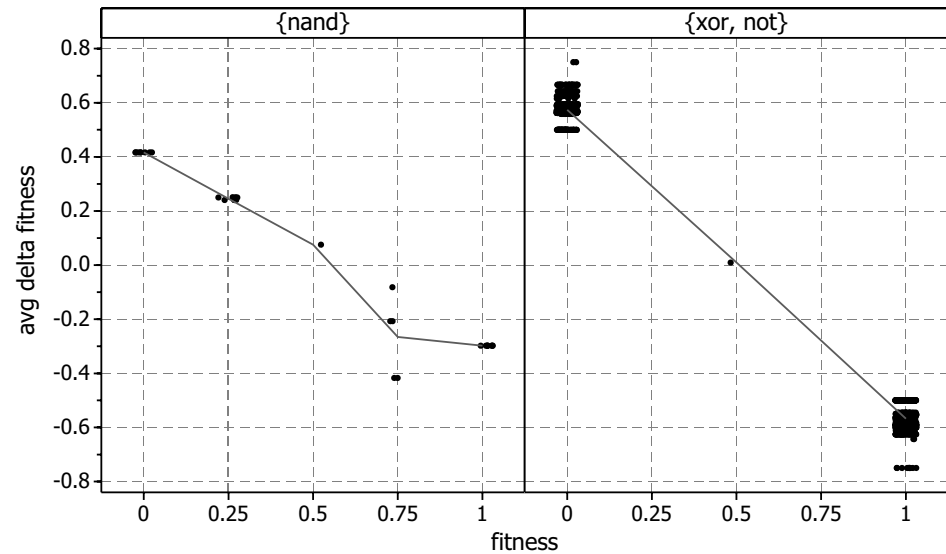
$\overline{T_N}$

$\overline{\Delta f}$

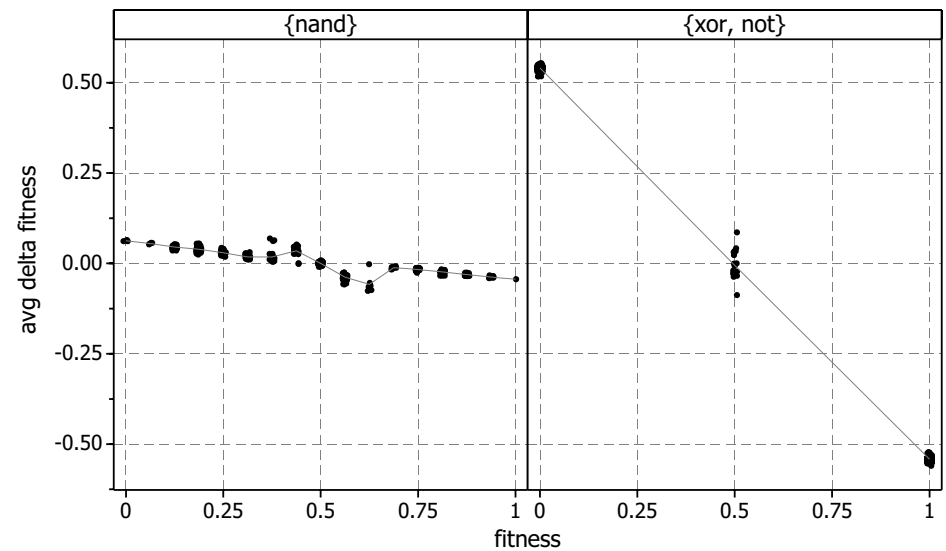
t_{so}

Considerazioni
finali

Even-2 parity



Even-4 parity



Tasso di subottimi

Introduzione

Risultati

$\mathcal{P}^{\{\text{XOR}; \text{NOT}\}}$

Analisi sperim.

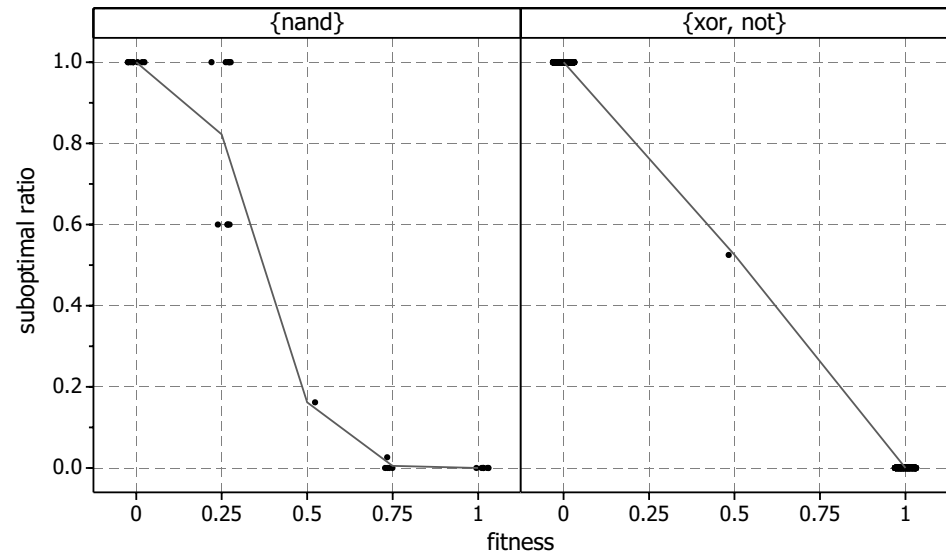
$\overline{T_N}$

$\overline{\Delta f}$

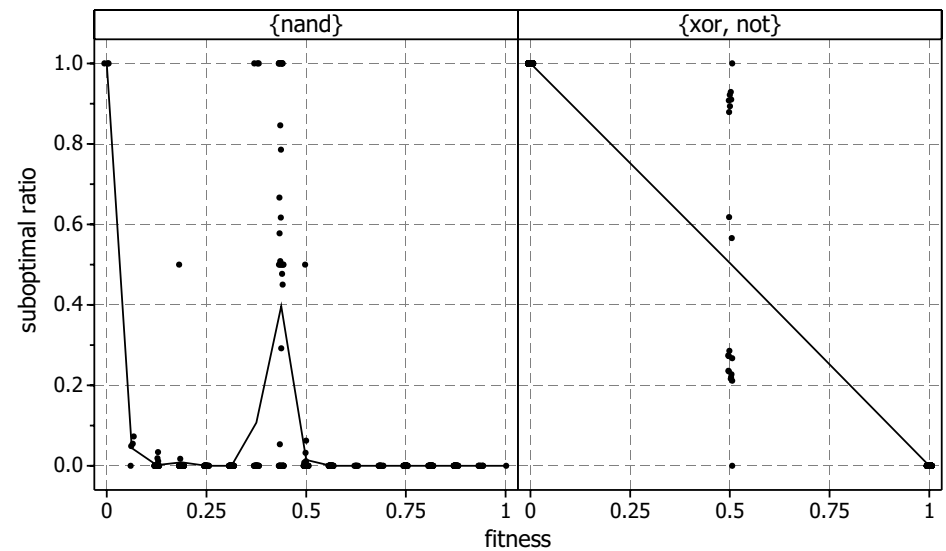
t_{so}

Considerazioni
finali

Even-2 parity



Even-4 parity



Conclusioni

Introduzione

Risultati

Considerazioni
finali

Conclusioni

Contributi

Sviluppi futuri

- Risultati delle analisi:
 - in $\mathcal{P}^{\{\text{NAND}\}}$, reti di fitness buona resistono al miglioramento
 - in $\mathcal{P}^{\{\text{XOR}; \text{NOT}\}}$, grande rete centrale facile da attraversare
- La diversa difficoltà dei problemi è parzialmente spiegata.
- Validità del metodo di campionamento:
 - genera campioni con caratteristiche simili allo spazio completo
 - riproduce la struttura teorica di $\mathcal{P}^{\{\text{XOR}; \text{NOT}\}}$

Contributi originali

Introduzione

Risultati

Considerazioni
finali

Conclusioni

Contributi

Sviluppi futuri

- Si distingue da lavori esistenti perché:
 - a) studia la PG 'standard'
 - b) studia la neutralità senza alterare artificialmente i paesaggi
- Contributi originali:
 - a) caratterizzazione teorica di $\mathcal{P}^{\{\text{XOR}; \text{NOT}\}}$
 - b) misure della neutralità e relativa analisi dei paesaggi
 - c) metodo di campionamento del paesaggio

Sviluppi futuri

Introduzione

Risultati

Considerazioni
finali

Conclusioni

Contributi

Sviluppi futuri

- Validità del metodo di campionamento
 - verifica sperimentale su nuovi paesaggi
 - indagine formale con strumenti statistico-matematici

- Studio della neutralità
 - analisi di altri tipi di paesaggio
 - misure della difficoltà del problema

Introduzione

Risultati

Considerazioni
finali

Extra

Paesaggio di
fitness

Vicinato e
mutazioni

Rappres. di
 $\mathcal{P}^{\{XOR; NOT\}}$

Misure

Campionamento

Dim. paesaggi

t_{sw}

t_{so} vs t_{sw}

Riferimenti

Extra

Paesaggio di fitness

Introduzione

Risultati

Considerazioni finali

Extra

Paesaggio di fitness

Vicinato e mutazioni

Rappres. di $\mathcal{P}^{\{\text{XOR}; \text{NOT}\}}$

Misure

Campionamento

Dim. paesaggi
 t_{sw}

t_{so} vs t_{sw}

Riferimenti

Paesaggio di fitness [Sta02]:

$$\mathcal{P} = (\mathcal{S}, \mathcal{V}, f)$$

dove

- \mathcal{S} , è l'insieme delle soluzioni ammissibili
- $\mathcal{V} : \mathcal{S} \rightarrow 2^{\mathcal{S}}$, è la funzione di vicinato
- $f : \mathcal{S} \rightarrow \mathbb{R}^+$, è la funzione di fitness

Vicinato neutro: $\mathcal{N}(s) = \{s' \in \mathcal{V}(s) \mid f(s) = f(s')\}$

Rete di neutralità: componente connessa del grafo (\mathcal{S}, E) dove $E = \{(s_1, s_2) \in \mathcal{S}^2 \mid s_2 \in \mathcal{N}(s_1)\}$

Vicinato e mutazioni

Introduzione

Risultati

Considerazioni finali

Extra

Paesaggio di fitness

Vicinato e mutazioni

Rappres. di $\mathcal{P}^{\{\text{XOR}; \text{NOT}\}}$

Misure

Campionamento

Dim. paesaggi

t_{sw}

t_{so} vs t_{sw}

Riferimenti

\mathcal{V} è definita in base agli operatori di mutazione

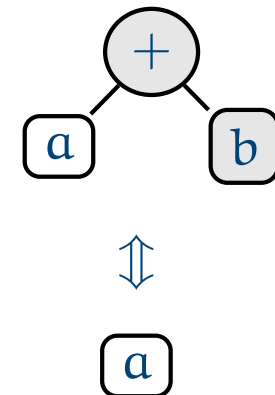
$$\mathcal{V}(s) = \{s' \in \mathcal{S} \mid s' \text{ può essere ottenuta variando } s\}$$

– Due criteri:

1. adeguata capacità esplorativa
2. semplicità del paesaggio risultante

– *Mutazioni strutturali strette*:

- versione semplificata degli operatori di mutazione strutturale
- mutazioni *deflate* e *inflate*: trasformano sotto-alberi di altezza 1 in foglie (e viceversa)
- mutazione di terminale



Rappresentazione di $\mathcal{P}^{\{\text{XOR}; \text{NOT}\}}$

Introduzione

Risultati

Considerazioni
finali

Extra

Paesaggio di
fitness

Vicinato e
mutazioni

**Rappres. di
 $\mathcal{P}^{\{\text{XOR}; \text{NOT}\}}$**

Misure

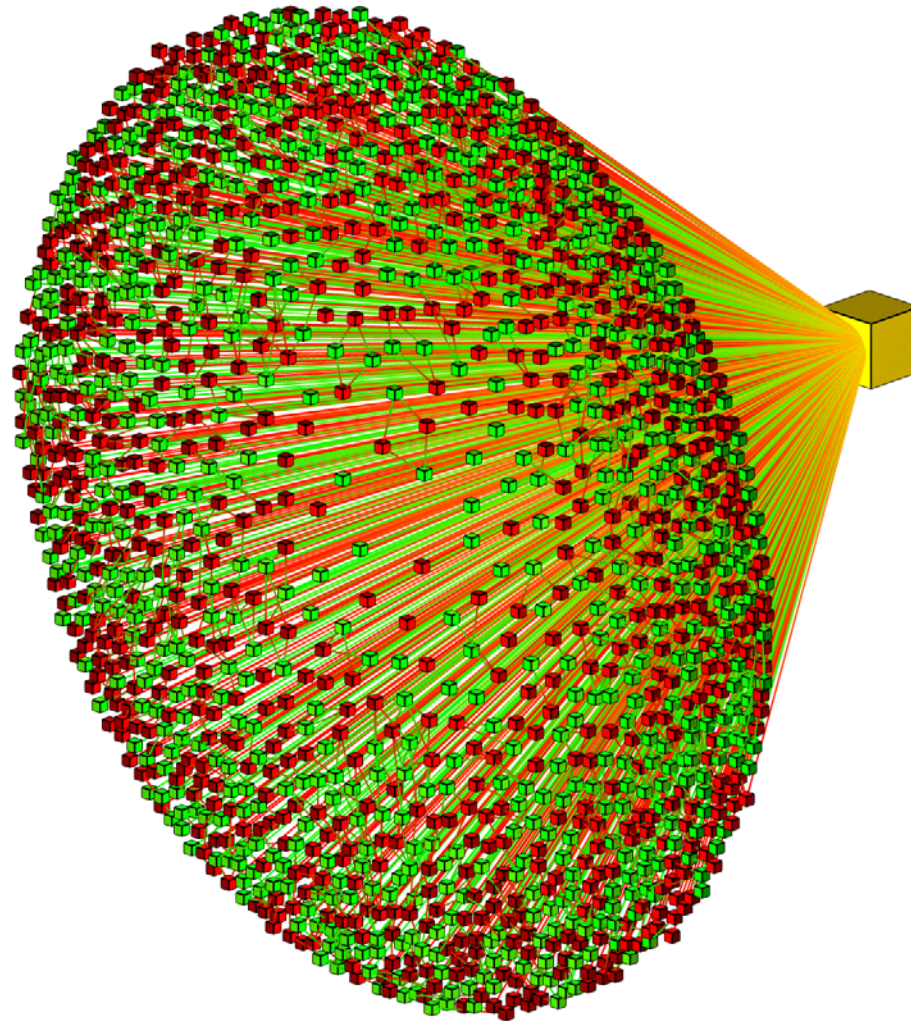
Campionamento

Dim. paesaggi

t_{sw}

t_{so} vs t_{sw}

Riferimenti



Misure di neutralità

Introduzione

Risultati

Considerazioni
finali

Extra

Paesaggio di
fitness

Vicinato e
mutazioni

Rappres. di
 $\mathcal{P}_{\{XOR; NOT\}}$

Misure

Campionamento

Dim. paesaggi

t_{sw}

t_{so} vs t_{sw}

Riferimenti

- *Tasso medio di neutralità:*
percentuale di mutazioni neutre
- *Δ -fitness media:*
guadagno medio di fitness a seguito di una
mutazione
- *Tasso di subottimi e di subpessimi:*
percentuale di individui, rispettivamente,
non-migliorabili o non-peggiorabili

Metodo di campionamento

Obiettivo:

Generare campioni di individui di molti livelli di fitness che formano reti di neutralità sufficientemente ‘grandi’ e interconnesse tra loro.

Ovvero i campioni devono contenere:

1. individui di molti livelli di fitness
2. reti di neutralità abbastanza grandi
3. reti di neutralità collegate tra loro

- Introduzione
- Risultati
- Considerazioni finali
- Extra
- Paesaggio di fitness
- Vicinato e mutazioni
- Rappres. di $\mathcal{P}_{\{XOR; NOT\}}$
- Misure
- Campionamento**
- Dim. paesaggi
- t_{sw}
- t_{so} vs t_{sw}
- Riferimenti

Dimensione paesaggi

Dimensione dei paesaggi considerati:

ordine k	2	4	
prof. max	3	6	8
{XOR; NOT}	10^3	10^{29}	10^{178}
{NAND}	10^3	10^{25}	10^{166}

- Introduzione
- Risultati
- Considerazioni finali
- Extra
- Paesaggio di fitness
- Vicinato e mutazioni
- Rappres. di $\mathcal{P}_{\{XOR; NOT\}}$
- Misure
- Campionamento
- Dim. paesaggi**
- t_{sw}
- t_{so} vs t_{sw}
- Riferimenti

Tasso di subpessimi

Introduzione

Risultati

Considerazioni finali

Extra

Paesaggio di fitness

Vicinato e mutazioni

Rappres. di $\mathcal{P}^{\{XOR; NOT\}}$

Misure

Campionamento

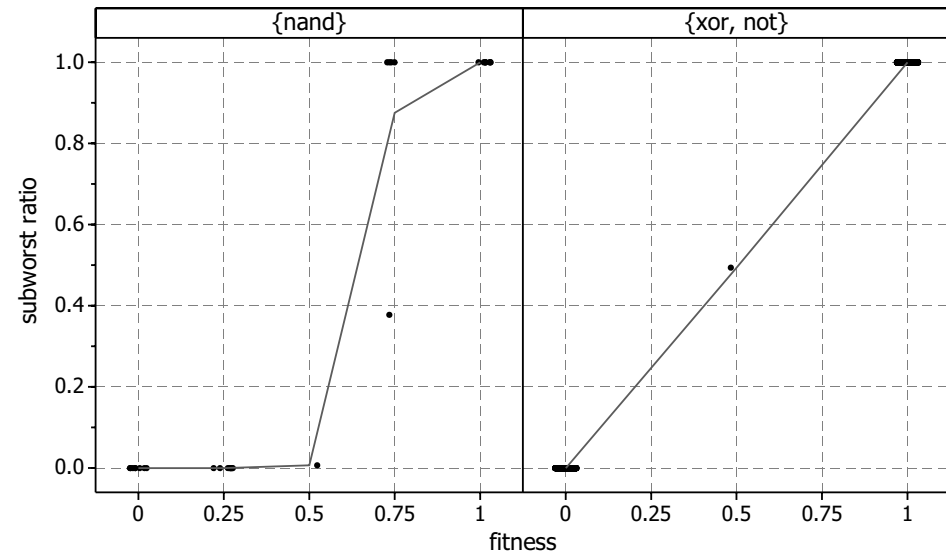
Dim. paesaggi

t_{sw}

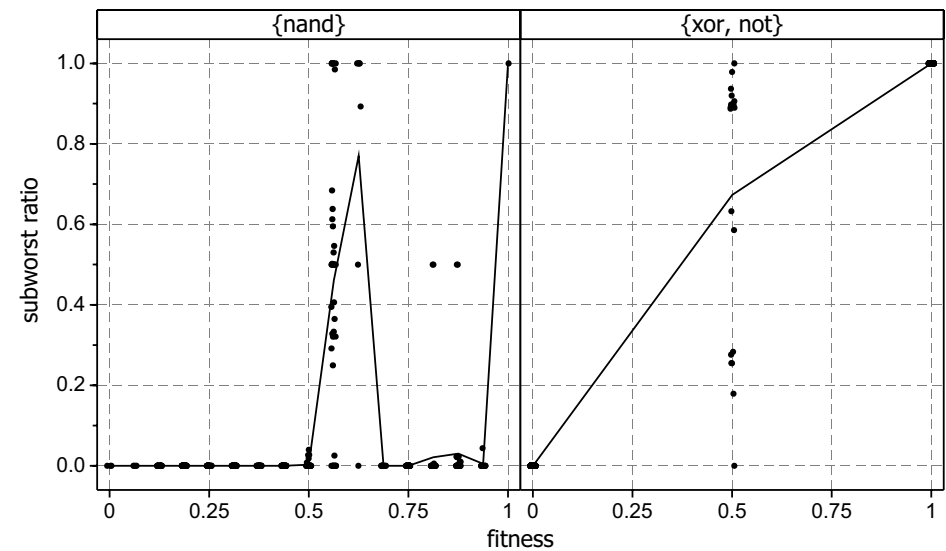
t_{so} vs t_{sw}

Riferimenti

Even-2 parity



Even-4 parity



t_{so} vs. t_{sw} - $\mathcal{P}\{\text{XOR}; \text{NOT}\}$

Introduzione

Risultati

Considerazioni finali

Extra

Paesaggio di fitness

Vicinato e mutazioni

Rappres. di $\mathcal{P}\{\text{XOR}; \text{NOT}\}$

Misure

Campionamento

Dim. paesaggi

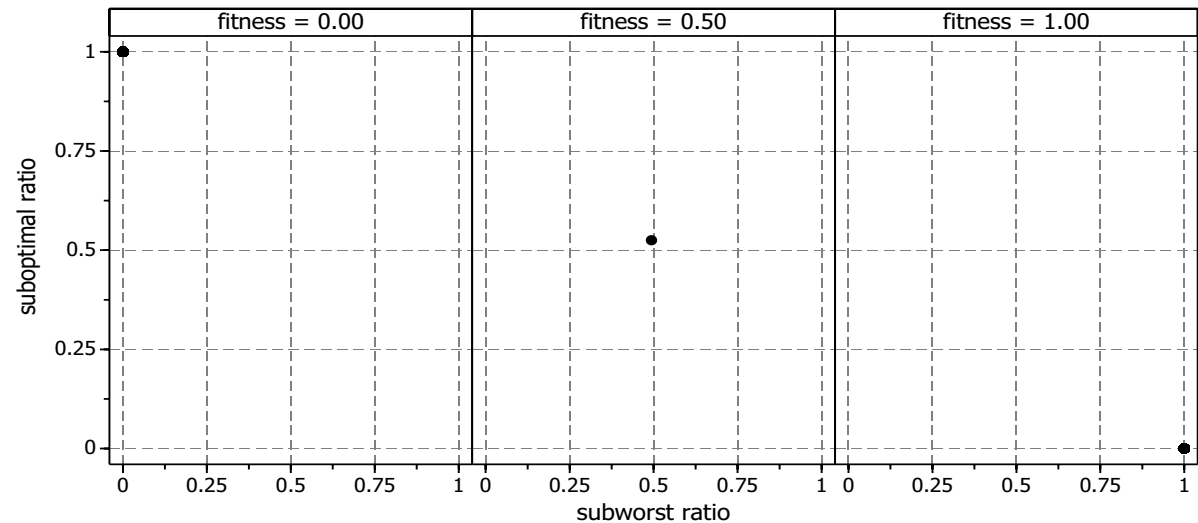
t_{sw}

t_{so} vs t_{sw}

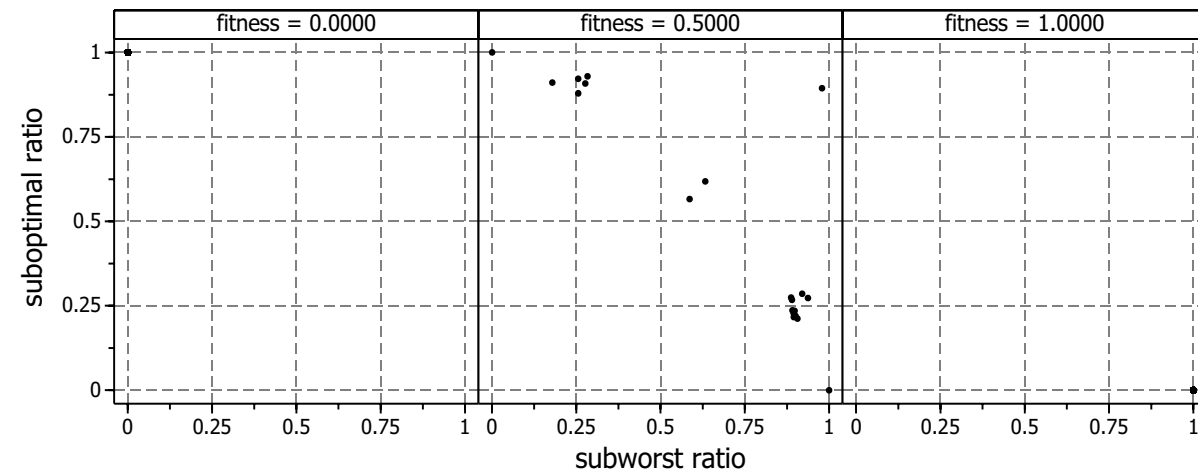
Riferimenti

{XOR; NOT}

Even-2 parity



Even-4 parity



t_{so} vs. t_{sw} - $\mathcal{P}^{\{NAND\}}$ - ordine 2

Introduzione

Risultati

Considerazioni finali

Extra

Paesaggio di fitness

Vicinato e mutazioni

Rappres. di $\mathcal{P}^{\{XOR; NOT\}}$

Misure

Campionamento

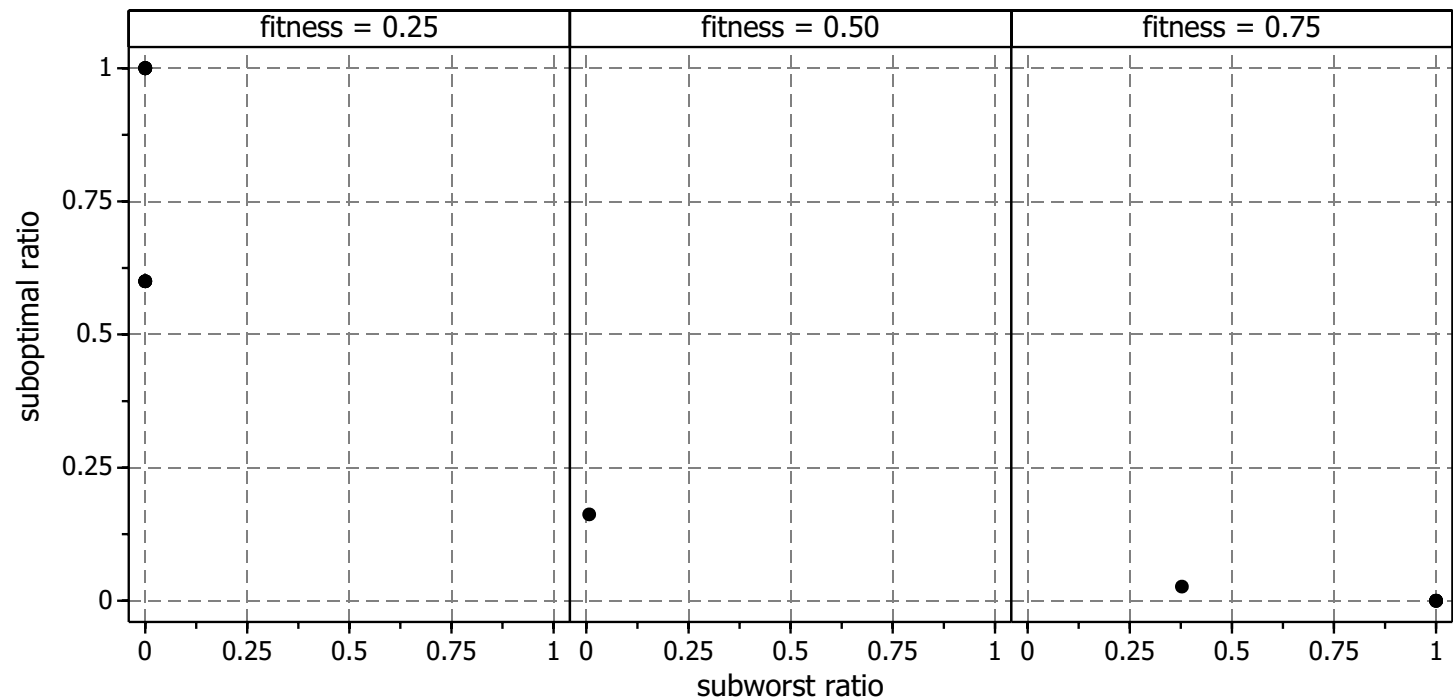
Dim. paesaggi

t_{sw}

t_{so} vs t_{sw}

Riferimenti

Even-2 parity, $\mathcal{P}^{\{NAND\}}$



t_{so} vs. t_{sw} - $\mathcal{P}^{\{NAND\}}$ - ordine 4

Introduzione

Risultati

Considerazioni finali

Extra

Paesaggio di fitness

Vicinato e mutazioni

Rappres. di $\mathcal{P}^{\{XOR; NOT\}}$

Misure

Campionamento

Dim. paesaggi

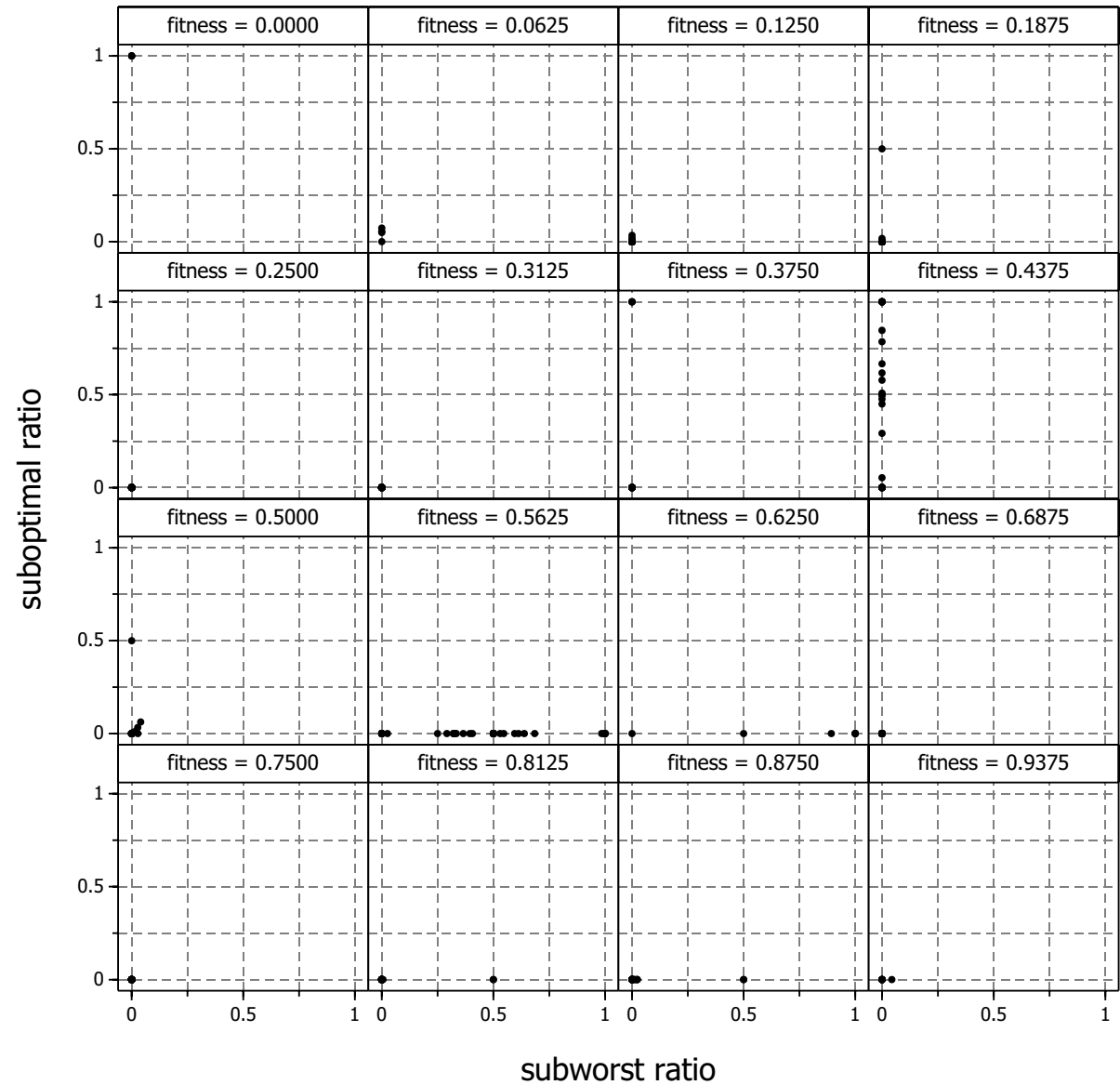
t_{sw}

t_{so} vs t_{sw}

Riferimenti

Even-4 parity,

$\mathcal{P}^{\{NAND\}}$



Riferimenti principali

Introduzione

Risultati

Considerazioni finali

Extra

Paesaggio di fitness

Vicinato e mutazioni

Rappres. di $\mathcal{P}^{\{XOR; NOT\}}$

Misure

Campionamento

Dim. paesaggi

t_{sw}

t_{so} vs t_{sw}

Riferimenti

- [Koz92] John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. 1992.
- [Sta02] P. F. Stadler. Fitness landscapes. In *Biological Evolution and Statistical Physics*, 2002.
- [VTCC03] L. Vanneschi *et al.*. Fitness distance correlation in structural mutation genetic programming. In *Genetic Programming, Proceedings of EuroGP'2003*.